Statistics
in Medicine

Research Article

Received 11 October 2011, Accepted 11 May 2012 Published online 11 July 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5470

Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates‡

Sarah Donegan,[a]*† Paula Williamson,[a] Umberto D'Alessandro[b] and Catrin Tudur Smith[a]

Mixed treatment comparison (MTC) meta-analysis allows several treatments to be compared in a single analysis while utilising direct and indirect evidence. Treatment by covariate interactions can be included in MTC models to explore how the covariate modifies the treatment effects. If interactions exist, the assumptions underlying MTCs may be invalidated. For conventional pair-wise meta-analysis, important benefits regarding the investigation of such interactions, gained from using individual patient data (IPD) rather than aggregate data (AD), have been described. We aim to compare IPD MTC models including patient-level covariates with AD MTC models including study-level covariates. IPD and AD random-effects MTC models for dichotomous outcomes are specified. Three assumptions are made regarding the interactions (i.e. independent, exchangeable and common interactions). The models are applied to a dataset to compare four drugs for treating malaria (i.e. amodiaquine-artesunate, dihydroartemisinin-piperaquine (DHAPQ), artemether-lumefantrine and chlorproguanil-dapsone plus artesunate) using the outcome unadjusted treatment success at day 28. The treatment effects and regression coefficients for interactions from the IPD models were more precise than those from AD models. Using IPD, assuming independent or exchangeable interactions, the regression coefficient for chlorproguanil-dapsone plus artesunate versus DHAPQ was statistically significant and assuming common interactions, the common coefficient was significant; whereas using AD, no coefficients were significant. Using IPD, DHAPQ was the best drug; whereas using AD, the best drug varied. Using AD models, there was no evidence that the consistency assumption was invalid; whereas, the assumption was questionable based on the IPD models. The AD analyses were misleading. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: mixed treatment comparison; multiple treatments meta-analysis; network meta-analysis; similarity; consistency; individual patient data

1. Introduction

When numerous treatments exist for the same clinical indication (e.g. A, B, C, D), often a systematic review will aim to compare the relative effect of all the available treatments (e.g. A vs B vs C vs D). For each pair of treatments, a relative treatment effect based on direct evidence may be estimated from a single randomised controlled trial or from several similar trials using conventional pair-wise meta-analysis. For a particular treatment comparison (e.g. C vs B), a treatment effect estimate may also be

[a]*Department of Biostatistics, Faculty of Health and Life Sciences, University of Liverpool, Shelley's Cottage, Brownlow Street, Liverpool, L69 3GS, U.K.*
[b]*Prince Leopold Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerp, Belgium*
**Correspondence to: Sarah Donegan, Department of Biostatistics, Faculty of Health and Life Sciences, University of Liverpool, Shelley's Cottage, Brownlow Street, Liverpool, L69 3GS, U.K..*
†*E-mail: sarah.donegan@liverpool.ac.uk*
‡*Supporting information may be found in the online version of this article.*

obtained from indirect evidence, when, for example, each of the treatments (e.g. B and C) have been compared with a third common treatment (e.g. A) in separate trials [1, 2]. Moreover, when the treatments form a connected network, such that each treatment has been directly compared with one or more of the other treatments, relative effects for all treatment pairings (e.g. B vs A, C vs A, D vs A, C vs B, D vs B, D vs C) can be estimated based on a combination of direct and indirect evidence using a single mixed treatment comparison (MTC) meta-analysis model [3–7].

The core assumption underlying MTCs is consistency, that is, for each treatment pairing, the true treatment effect is assumed to be similar for every trial in the analysis irrespective of whether the trial allocated patients to those two treatments. The consistency assumption can be thought of as an extension of the homogeneity assumption (i.e. similarity of treatment effects from trials that directly compare the same two treatments) that also underlies conventional pair-wise meta-analysis. More formally, the consistency assumption is satisfied if the so called 'consistency equations' hold [3]. For instance, for a network of three treatments A, B, C, one consistency equation must hold: $d_{BC} = d_{AC} - d_{AB}$ where d_{BC}, d_{AC} and d_{AB} denote the treatment effect of C relative to B, C relative to A, and B relative to A, respectively.

Assessment of the consistency and homogeneity assumptions is vital to ensure the results of the MTC are valid and interpreted appropriately. The consistency assumption is difficult to assess because each treatment effect may not actually be estimated by each trial. Even so, methods to assess consistency have been proposed and include comparing characteristics across the trials, or investigating the effect of patient or trial characteristics on the MTC results using subgroup analysis, sensitivity analysis, or meta-regression [2, 3, 5, 6, 8–14]. Methods to assess homogeneity are well established [15].

To assess consistency, an approach that warrants further research is the inclusion of treatment by covariate interactions in MTC models. The consistency assumption holds when, for each treatment comparison, the true treatment effect is expected to be similar across all trials. If one or more true treatment effect is modified by a particular characteristic (commonly referred to as a treatment by covariate interaction) and included trials differ with respect to the characteristic, then the consistency assumption may be violated. If the results of an MTC model including interactions reveal interactions exist, the consistency assumption may not be satisfied; whereas if no interactions are identified, the assumption may be satisfied, providing there are no unknown covariates that could modify one or more treatment effects.

Furthermore, including the interactions in the MTC model provides useful information regarding how the treatment effects are modified by the characteristic, for instance it can answer the following questions: Is the same treatment effect applicable to all patients? Is the treatment effect for patients in population X greater than for patients in population Y? Is drug A more effective for population X and drug B more effective for population Y?

Treatment by covariate interactions can be explored in meta-analysis models using individual patient data (IPD) or aggregate data (AD). With IPD, a patient-specific covariate value, (e.g. age), is available for each individual; whereas, with an AD approach, a study-specific covariate value (e.g. average age), is available for each trial. It has been well documented that interactions identified from across-trial information in an AD meta-analysis, may not be the same as the associations that would be detected when investigating within-trial information in an IPD meta-analysis. When exploring interactions in conventional pair-wise meta-analysis models, the use of IPD often provides greater benefits as compared with AD. In a simulation study, Lambert *et al.* [16] demonstrated that the power of models including trial-level covariates was dramatically lower than for models including patient-level covariates, the estimates of model parameters rarely agreed, and a large number of trials was required to detect interactions using study-level covariates. By comparing the ability of patient-level and trial-level covariates to detect interactions, Simmonds and Higgins [17] found that study-level covariates were more powerful when the distributions of the covariate were very different across trials; the patient-level approach was more powerful when the covariate distributions were similar across trials, but patients within the same trial differed with respect to their covariate values. Berlin *et al.* [18] recommended that IPD should be acquired to study patient characteristics to avoid ecological bias that could be introduced by investigating aggregated, study-level summaries of patient-level covariates. Through application of various approaches to explore interactions using IPD or AD, Tudur Smith *et al.* [19] found that a more thorough explanation of heterogeneity was obtained using patient-level covariates. Using real data, Schmid *et al.* [20] concluded that treatment by study-level covariate interactions can be detected when the treatment effects of trials are heterogeneous, but patient-level covariates are needed to detect interactions when treatment effects are homogeneous.

A number of published methodological articles have described MTC models including treatment by study-level covariate interactions. Nixon *et al.* [14] applied MTC models for dichotomous outcomes that included two study-level covariates using a case study from rheumatoid arthritis. Cooper *et al.* [12] presented MTC models for dichotomous outcomes with a trial-level covariate and applied the models to a network of stroke prevention treatments. Salanti *et al.* [8] applied meta-regression in an MTC setting to compare topical fluoride treatments for preventing dental caries. Currently, to our knowledge, only one methodological article has illustrated the inclusion of patient-level covariates in an MTC model; Tudur Smith *et al.* [13] applied a Cox proportional hazards model to investigate differences in the effectiveness of antiepileptic drugs for patients with different seizure types.

As yet, any advantages gained from using IPD, rather than AD, to explore treatment by covariate interactions in an MTC model are unknown. Certainly, the potential for interactions to exist is greater in an MTC, as compared with a conventional pair-wise meta-analysis, because the trials included in an MTC differ with respect to the allocated treatments and may therefore be more likely to differ with respect to other characteristics (e.g. patient inclusion criteria, location, or recruitment date), as compared with trials in a standard pair-wise meta-analysis. Furthermore, even when no direct evidence is available for a particular treatment comparison, interactions can be explored using indirect evidence in an MTC model. Consequently, IPD may be even more valuable for MTC meta-analysis than for standard pair-wise meta-analysis.

The aims of this article are to extend the existing AD MTC models for dichotomous outcomes described by Cooper *et al.* [12] to allow for patient-level outcomes and covariates; and to compare the models to evaluate the potential benefits of IPD. In the next section, AD MTC models are described then modified to allow for IPD and treatment by patient-level covariate interactions. The IPD models are further developed to contrast within-trial and across-trial interactions. In Section 3, the methodology is illustrated through application to real data, with differences between IPD and AD models discussed in Section 4.

2. Methods

2.1. Notation

Using notation that is often used in the MTC literature, denote j to be the trial where $j = 1 \ldots\ldots NS$ where NS is the number of independent trials included in the MTC. Denote k to be the treatment where $k \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$ and A is a referent treatment. Any single treatment k can be chosen to be the referent treatment A [21]. Let NT be the number of treatments. The methodology can be applied to any number of treatments by extending the notation. A prerequisite of the MTC models is that all treatments form a connected network [3, 22].

Let d_{XY} denote the treatment effect of treatment Y relative to treatment X. Under an MTC frame-work, because A is the referent treatment, the treatment effects relative to A (e.g. d_{AB}, d_{AC}, d_{AD}), are denoted as basic parameters and are estimated by the MTC model. The remaining treatment effects (e.g. d_{BC}, d_{BD}, d_{CD}) are functional parameters, linear combinations of the basic parameters (e.g. $d_{\mathrm{BC}} = d_{\mathrm{AC}} - d_{\mathrm{AB}}$; $d_{\mathrm{BD}} = d_{\mathrm{AD}} - d_{\mathrm{AB}}$; and $d_{\mathrm{CD}} = d_{\mathrm{AD}} - d_{\mathrm{AC}}$) and can be defined this way as proven by Bucher *et al.* [1].

The notation of Cooper *et al.* [12] is used to allow for the inclusion of treatment by covariate interactions in the MTC model. Suppose β_{XY} is the regression coefficient for the interaction for the comparison of treatment Y relative to treatment X. The (NT-1) regression coefficients corresponding to the basic parameters (e.g. β_{AB}, β_{AC}, β_{AD}), are estimated by the model and are used to estimate the remaining coefficients (e.g. $\beta_{\mathrm{BC}} = \beta_{\mathrm{AC}} - \beta_{\mathrm{AB}}$; $\beta_{\mathrm{BD}} = \beta_{\mathrm{AD}} - \beta_{\mathrm{AB}}$; and $\beta_{\mathrm{CD}} = \beta_{\mathrm{AD}} - \beta_{\mathrm{AC}}$).

2.2. Models

Two types of MTC models for dichotomous outcomes will be described: AD models based on event rates (i.e. number of patients with an event and the total number of patients, in each treatment group of each study) and trial-level covariate values (as described previously by Cooper *et al.* [12]), and models that have been modified to accommodate IPD for the outcomes and patient-level covariate values.

For AD and IPD in turn, four MTC model specifications will be specified: the model without interactions and a further three models that include treatment by covariate interactions with differing assumptions regarding the interactions. One model specification assumes that the (NT – 1) regression coefficients corresponding to the basic parameters (e.g. β_{AB}, β_{AC}, β_{AD}) are independent, that is, different

and unrelated to each other. This model makes the weakest assumption and is suitable for all treatment networks providing data are not limited. Data limitation issues would arise, for example, if for one or more treatment comparisons, the regression coefficient(s) could not be reliably estimated because of the small number of contributing trials, or because covariate values were alike. Another model specification assumes that the $(NT-1)$ regression coefficients are exchangeable with each other, that is, the coefficients are different but related, and the magnitude of the coefficients cannot be predicted. For example, this model may be appropriate if the treatments (i.e. A, B, C, D) are drugs within the same treatment class. The last model specification assumes that the $(NT-1)$ regression coefficients are identical. Clearly, this model makes the strongest assumption and may only be suitable in some contexts, when treatments are very similar.

All models assume random effects, that is, for each treatment comparison, there is no single underlying treatment effect, but that, different trials are estimating unique treatment effects that are realisations from a normal distribution. Also, the models initially described in this section assume that each study compares two treatments b and k, where $b,k \in \{\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D}\}$ and k is after b alphabetically. In this article, $k \succ b$ indicates that k is after b in the alphabet. The models are extended to allow for multi-arm trials in Section 2.4.

2.2.1. Aggregate data mixed treatment comparison meta-analysis. As described by Cooper *et al.* [12], let r_{jk} be the number of events on a treatment k in trial j and let n_{jk} be the number of patients on treatment k in trial j. The probability of an event on treatment k in study j is denoted by p_{jk}. Assume that the outcomes of patients within the same treatment group, r_{jk}, are independent and are distributed as $r_{jk} \sim \text{binomial}(p_{jk}, n_{jk})$.

No interactions (model 1). The model without interactions is specified as follows:

$$\text{logit}(p_{jk}) = \begin{cases} \mu_{jb} & \text{if } k = b, b \in \{\mathrm{A},\mathrm{B},\mathrm{C}\} \\ \mu_{jb} + \delta_{jbk} & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$$

where the log odds of an event in treatment group b in the jth trial is given by μ_{jb}; δ_{jbk} represents the log odds ratio of k versus b in the jth trial; and the trial-specific log odds ratios, δ_{jbk}, are assumed to be realisations from a normal distribution with mean d_{bk} and variance τ^2_{bk}. By writing the functional parameters in terms of basic parameters (i.e. setting $d_{bk} = d_{\mathrm{A}k} - d_{\mathrm{A}b}$), $\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$ is replaced by $\delta_{jbk} \sim N\left(d_{\mathrm{A}k} - d_{\mathrm{A}b}, \tau^2_{bk}\right)$ where $d_{\mathrm{AA}} = 0$.

Independent treatment by covariate interactions (model 2). Independent interactions can be included in the model as follows:

$$\text{logit}(p_{jk}) = \begin{cases} \mu_{jb} & \text{if } k = b, b \in \{\mathrm{A},\mathrm{B},\mathrm{C}\} \\ \mu_{jb} + \delta_{jbk} + \beta_{bk} z_j & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$$

where z_j is a trial-level summary of the patient-level covariate values (such as, the mean of continuous covariate values or the proportion of patients with a particular dichotomous covariate value) of the jth trial; β_{bk} represents the difference in the log odds ratio of k versus b per unit increase in the covariate z_j; and d_{bk} represents the mean log odds ratio of k versus b when the covariate value is zero (i.e. $z_j = 0$). The model can be written in terms of basic parameters by setting $d_{bk} = d_{\mathrm{A}k} - d_{\mathrm{A}b}$ and $\beta_{bk} = \beta_{\mathrm{A}k} - \beta_{\mathrm{A}b}$:

$$\text{logit}(p_{jk}) = \begin{cases} \mu_{jb} & \text{if } k = b, b \in \{\mathrm{A},\mathrm{B},\mathrm{C}\} \\ \mu_{jb} + \delta_{jbk} + (\beta_{\mathrm{A}k} - \beta_{\mathrm{A}b})\, z_j & k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{\mathrm{A}k} - d_{\mathrm{A}b}, \tau^2_{bk}\right)$$

where $d_{\mathrm{AA}} = 0$ and $\beta_{\mathrm{AA}} = 0$.

Exchangeable treatment by covariate interactions (model 3). Under the assumption of exchangeable interactions, the model specification is identical to model 2, but with a distribution placed on the regression coefficients $\beta_{\mathrm{A}k} \sim N\left(m_{\mathrm{B}}, \tau^2_{\mathrm{B}}\right)$ where m_{B} is the mean and τ^2_{B} is the variance.

Common treatment by covariate interactions (model 4). Common interactions can be included in the model as follows:

$$\text{logit}(p_{jk}) = \begin{cases} \mu_{jb} & \text{if } k = b, b \in \{\text{A}, \text{B}, \text{C}\} \\ \mu_{jb} + \delta_{jbk} & \text{if } k \succ b \text{ and } b \neq \text{A} \\ \mu_{jb} + \delta_{jbk} + \beta z_j & \text{if } k \succ b \text{ and } b = \text{A} \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$$

where β represents the difference in the log odds ratio per unit increase in the covariate z_j. When the functional parameters are written in terms of basic parameters (i.e. $d_{bk} = d_{\text{A}k} - d_{\text{A}b}$), $\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$ is replaced by $\delta_{jbk} \sim N\left(d_{\text{A}k} - d_{\text{A}b}, \tau^2_{bk}\right)$ where $d_{\text{AA}} = 0$.

Notice that no interaction term is included for each comparison of k versus b when $b \neq \text{A}$, because the common regression coefficient cancels out (i.e. $\beta_{bk} = \beta_{\text{A}k} - \beta_{\text{A}b} = \beta - \beta = 0$); but an interaction term is included for each comparison of k versus A because $\beta_{\text{AA}} = 0$ (i.e. $\beta_{bk} = \beta_{\text{A}k} - \beta_{\text{AA}} = \beta$) [12]. Also, note that the mean log odds ratio of k versus A when the covariate value is zero, is given by $d_{\text{A}k}$, and the mean log odds ratio of k versus A when the covariate value is one is $d_{\text{A}k} + \beta$. Therefore, for each comparison of k versus A, the log odds ratio increases by the same constant with an increase in the covariate value. Yet, for each comparison of k versus b when $b \neq \text{A}$, the log odds ratio is d_{bk} for all covariate values.

2.2.2. Individual patient data mixed treatment comparison meta-analysis. Suppose $y_{ijk} = 1$ if the ith patient in the jth trial on treatment k experiences the event and $y_{ijk} = 0$ if the ith patient in the jth trial on treatment k does not experience the event, where $i = 1 \ldots . N_j$ such that N_j is the number of patients in the jth trial. Assume that the outcomes of patients, y_{ijk}, are independent and distributed as $y_{ijk} \sim \text{bernoulli}(p_{ijk})$ where p_{ijk} is the probability of an event for the ith patient in the jth trial on treatment k.

No interactions (model 5). The model without interactions is specified as follows:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} & \text{if } k = b, b \in \{\text{A}, \text{B}, \text{C}\} \\ \mu_{jb} + \delta_{jbk} & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$$

where the log odds of an event in treatment group b in the jth trial is given by μ_{jb}; δ_{jbk} represents the log odds ratio of k versus b in the jth trial; and the trial-specific log odds ratios, δ_{jbk}, are assumed to be realisations from a normal distribution with mean d_{bk} and variance τ^2_{bk}. By writing the functional parameters in terms of basic parameters (i.e. setting $d_{bk} = d_{\text{A}k} - d_{\text{A}b}$), $\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$ is replaced by $\delta_{jbk} \sim N\left(d_{\text{A}k} - d_{\text{A}b}, \tau^2_{bk}\right)$ where $d_{\text{AA}} = 0$

Independent treatment by covariate interactions (model 6). Independent interactions can be included in the model as follows:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} + \beta_{0jb} x_{ijk} & \text{if } k = b, b \in \{\text{A}, \text{B}, \text{C}\} \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} + \beta_{bk} x_{ijk} & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$$

where x_{ijk} is a patient-level covariate for the ith patient in the jth trial on treatment k (such as, a continuous covariate value or an indicator variable for a dichotomous covariate);β_{0jb} is a study-specific regression parameter that represents the difference in the log odds of an event in treatment group b per unit increase in the covariate x_{ijk}; β_{bk} represents the difference in the log odds ratio of k versus b per unit increase in the covariate x_{ijk}; and d_{bk} represents the mean log odds ratio of k versus b when the covariate value is zero (i.e. $x_{ijk} = 0$). The model can be written in terms of basic parameters by setting $d_{bk} = d_{\text{A}k} - d_{\text{A}b}$ and $\beta_{bk} = \beta_{\text{A}k} - \beta_{\text{A}b}$:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} + \beta_{0jb} x_{ijk} & \text{if } k = b, b \in \{\text{A}, \text{B}, \text{C}\} \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} + \left(\beta_{\text{A}k} - \beta_{\text{A}b}\right) x_{ijk} & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{\text{A}k} - d_{\text{A}b}, \tau^2_{bk}\right)$$

where $d_{\text{AA}} = 0$ and $\beta_{\text{AA}} = 0$.

Exchangeable treatment by covariate interactions (model 7). Assuming exchangeable interactions, the model specification is identical to model 6, but with a distribution placed on the regression coefficients $\beta_{Ak} \sim N\left(m_B, \tau_B^2\right)$ where m_B is the mean and τ_B^2 is the variance.

Common treatment by covariate interactions (model 8). Common interactions can be included in the model as follows:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} + \beta_{0jb} x_{ijk} & \text{if } k = b, b \in \{A, B, C\} \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} & \text{if } k \succ b \text{ and } b \neq A \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} + \beta x_{ijk} & \text{if } k \succ b \text{ and } b = A \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau_{bk}^2\right)$$

where β represents the difference in the log odds ratio per unit increase in the covariate x_{ijk}. When the functional parameters are written in terms of basic parameters (i.e. $d_{bk} = d_{Ak} - d_{Ab}$), $\delta_{jbk} \sim N\left(d_{bk}, \tau_{bk}^2\right)$ is replaced by $\delta_{jbk} \sim N\left(d_{Ak} - d_{Ab}, \tau_{bk}^2\right)$ where $d_{AA} = 0$.

2.2.3. Exploring within-trial and across-trial treatment by covariate interactions using individual patient data mixed treatment comparison meta-analysis. The IPD models (models 6–8) described in Section 2.2.2 assume that within-trial interactions and across-trial interactions are identical, with the caveat that the trial-level covariate is considered to be the mean of the patient-level covariates in the trial [23, 24]. However, the interactions based on across-trial information in an AD meta-analysis, may not be equivalent to the interactions detected when investigating within-trial IPD. Across-trial interactions can be affected by ecological biases and confounding; whereas within-trial interactions are not influenced [23, 24]. If ecological biases or confounding are at play, across-trial interactions may be detected when within-trial interactions are not observed, because the interactions across trials are caused by other study-level differences (e.g. study location); in such cases the genuine effect of the covariate is observed within-study. Furthermore, the within-trial and across-trial interactions may differ because of the covariate distributions of the specific dataset. If the study-specific covariate values vary across trials, but the patient-level covariates within each trial are alike, AD meta-analysis may identify interactions that would remain undetected using within-trial IPD. Conversely, if the covariate values of patients within the same trial differ, but study-level summaries of the covariate are similar across trials, IPD analyses could detect within-trial interactions that would be impossible to explore using AD.

It is possible to explore how the across-trial and within-trial interactions compare by fitting models that separate the two types of interaction, as described by Riley *et al.* [23, 24]. To highlight any differences between the two types of interaction, Riley *et al.* [23] recommend that results from IPD models that separate the interaction types should be reported whenever results from IPD models that assume identical within-trial and across-trial interactions are presented. If the two types of interaction are similar, inferences can be based on models that assume that within-trial and across-trial interactions are identical. However, if the within-trial and across-trial interactions differ, ecological bias or confounding may be at play and therefore the results from such models may be biased because across-trial interactions contribute to the parameter estimates. In such cases, a meta-analyst may base inferences on the within-trial interactions estimated by models that separate the two types of interaction. We describe how this can be accomplished in the IPD MTC context. Here we assume the within-trial interaction is common across all trials and the trial-level covariate, z_j, is the mean of the patient-level covariate values of the jth trial.

Independent treatment by covariate interactions (model 9). With independent interactions, model 6 can be extended as follows:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} + \beta_{0jb} x_{ijk} & \text{if } k = b, b \in \{A, B, C\} \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} + \beta_{bk}^w (x_{ijk} - z_j) + \beta_{bk}^a z_j & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau_{bk}^2\right)$$

where z_j is the mean of the patient-level covariate values of the jth trial; β_{bk}^w represents the difference in the log odds ratio of k versus b per unit increase in the covariate x_{ijk}; β_{bk}^a represents the difference in the log odds ratio of k versus b per unit increase in the covariate mean z_j; and d_{bk} represents the

mean log odds ratio of k versus b when the covariate values are zero ($z_j = x_{ijk} = 0$). When the functional parameters are written in terms of basic parameters (i.e. $d_{bk} = d_{Ak} - d_{Ab}$, $\beta^w_{bk} = \beta^w_{Ak} - \beta^w_{Ab}$, $\beta^a_{bk} = \beta^a_{Ak} - \beta^a_{Ab}$) the model becomes:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} + \beta_{0jb} x_{ijk} & \text{if } k = b, b \in \{A, B, C\} \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} + \left(\beta^w_{Ak} - \beta^w_{Ab}\right)\left(x_{ijk} - z_j\right) + \left(\beta^a_{Ak} - \beta^a_{Ab}\right) z_j & \text{if } k \succ b \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{Ak} - d_{Ab}, \tau^2_{bk}\right)$$

where $d_{AA} = 0$, $\beta^w_{AA} = 0$, and $\beta^a_{AA} = 0$.

Exchangeable treatment by covariate interactions (model 10). Under the assumption of exchangeable interactions, the model specification is identical to model 9, but with a distribution placed on the within-trial regression coefficients $\beta^w_{Ak} \sim N\left(m^w_B, \left(\tau^w_B\right)^2\right)$ where m^w_B is the mean and $\left(\tau^w_B\right)^2$ is the variance; and the across-trial regression coefficients $\beta^a_{Ak} \sim N\left(m^a_B, \left(\tau^a_B\right)^2\right)$ where m^a_B is the mean and $\left(\tau^a_B\right)^2$ is the variance.

Common treatment by covariate interactions (model 11). Common interactions can be included in the model as follows:

$$\text{logit}(p_{ijk}) = \begin{cases} \mu_{jb} + \beta_{0jb} x_{ijk} & \text{if } k = b, b \in \{A, B, C\} \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} & \text{if } k \succ b \text{ and } b \neq A \\ \mu_{jb} + \beta_{0jb} x_{ijk} + \delta_{jbk} + \beta^{\text{w}}\left(x_{ijk} - z_j\right) + \beta^{\text{a}} z_j & \text{if } k \succ b \text{ and } b = A \end{cases}$$
$$\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$$

where β^w represents the difference in the log odds ratio per unit increase in the covariate x_{ijk}; and β^a represents the difference in the log odds ratio per unit increase in the covariate mean z_j. By writing the functional parameters in terms of basic parameters (i.e. setting $d_{bk} = d_{Ak} - d_{Ab}$), $\delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)$ is replaced by $\delta_{jbk} \sim N\left(d_{Ak} - d_{Ab}, \tau^2_{bk}\right)$ where $d_{AA} = 0$.

2.3. *Between trial variances*

Many different specifications of the variance structure have been described [3, 4, 25]. Lu and Ades [25] showed that the consistency equations also constrain the variance parameters. Therefore, if the between trial variances $\left(\text{i.e. } \tau^2_{bk}\right)$ are allowed to differ across comparisons, such constraints must be considered. However, more often, it is assumed that between trial variances for the treatment effects are identical across treatment comparisons $\left(\text{i.e. } \tau^2_{bk} = \tau^2\right)$. The advantage of making this assumption is that there is only one variance–covariance parameter to estimate, even when multi-arm trials contribute, because the covariance between any two treatment effects estimated from the same trial becomes $\tau^2/2$ under the assumption [3, 4, 7].

2.4. *Models including multi arm trials*

When a trial compared more than two treatments, for example A, C and D, the trial-specific treatment effects $\delta_{jAC}, \delta_{jAD}$, and δ_{jCD} are correlated [3, 6, 7, 12]. The models described in the previous sections would treat multi-arm trials as two-arm trials with a common treatment b and therefore would not take account of the correlation. The models can be adapted to take the correlation structure into account by allowing the trial-specific treatment effects to follow a multivariate normal distribution [21]. Such models have been described extensively [3, 4, 6–8, 11, 12, 21, 26–32].

3. Application: artemisinin-based combinations for treating uncomplicated malaria

The IPD models and AD models described in the previous section are applied to real data in this section.

3.1. The dataset

Individual patient data from a single multicentre randomised trial (4ABC trial) carried out at study sites across seven African countries were made available [33]. The trial's primary objective was to compare four artemisinin-based combination therapies (ACTs) for the treatment of uncomplicated *Plasmodial falciparum* malaria in children. Each site compared up to three of the four ACTs: amodiaquine-artesunate (AQ+AS), dihydroartemisinin-piperaquine (DHAPQ), artemether-lumefantrine (AL), or chlorproguanil-dapsone plus artesunate (CD+A) (Table I). A primary outcome was unadjusted treatment success at day 28.

The randomisation sequence was stratified by site. The treatments allocated at each site were chosen based on the local current first line treatments, antimalarial resistance and malaria endemicity. Therefore, as part of the trial's analysis, meta-analysis was applied to provide a pooled estimate of treatment efficacy across all sites, pooling across sites rather than trials, as is conventional in meta-analysis. Further details are given in the original trial report [33].

A total of 3874 patients were included in the primary analysis of unadjusted treatment success at day 28, that is, 94.1% of the 4116 children that were randomized. Reasons for exclusion from the analysis were: lost-to-follow up (2.6%) and withdrawal from the study for reasons definitely or probably not related to malaria or treatment (3.3%).

The effect of one potentially treatment effect-modifying covariate, patient age, is considered here and was listed in the statistical analysis plan because it was believed that in endemic areas, older patients are more likely to achieve treatment success because they have comparatively greater protective immunity.

Table I displays the number of patients that achieved unadjusted treatment success at day 28 and the total number of patients in each treatment group of each site along with covariate information. The 17 sites each included two or three of the following treatments: DHAPQ, AQ+AS, AL, and CD+A. There were six possible pair-wise comparisons, all of which were supported by direct evidence (Figure 1).

The AD was generated from the IPD such that both datasets were based on the same patients. This ensured that any differences in the results would be due to the data type (i.e. IPD or AD).

Table I. Event rate of each treatment group of each site for unadjusted treatment success at day 28 and the mean age of patients at each site.

Site	ACT (number of patients that achieved treatment success/number of patients)				Total number of patients that achieved treatment success/total number of patients	Mean age (SD), years
	DHAPQ	AQ+AS	AL	CD+A		
Manhica (after CD+A)	94/100	78/97	—	—	172/197	2.88 (1.30)
Mbarara (after CD+A)	63/65	59/70	—	—	122/135	2.43 (1.07)
Nanoro	187/219	199/290	115/292	—	501/801	2.24 (1.18)
Gabon	62/63	67/76	65/70	—	194/209	2.83 (1.28)
Afokang	67/72	78/83	84/87	—	229/242	2.94 (1.28)
Pamol	60/65	73/79	73/80	—	206/224	2.66 (1.36)
Ndola	67/67	63/69	63/75	—	193/211	2.45 (1.20)
Manhica (before CD+A)	78/82	70/86	—	42/84	190/252	2.82 (1,00)
Mbarara (before CD+A)	72/80	64/79	—	53/80	189/239	2.60 (1.10)
Rukara (after CD+A)	46/47	—	46/50	—	92/97	3.08 (0.92)
Jinja (after CD+A)	160/167	—	157/168	—	317/335	2.33 (1.17)
Tororo (after CD+A)	54/75	—	33/77	—	87/152	1.99 (0.99)
Mashesha (after CD+A)	49/52	—	51/52	—	100/104	2.90 (1.05)
Rukara (before CD+A)	22/23	—	18/21	4/23	44/67	2.71 (1,00)
Jinja (before CD+A)	37/39	—	35/38	34/40	106/117	2.62 (1.19)
Tororo (before CD+A)	109/141	—	88/138	71/142	268/421	2.11 (0.85)
Mashesha (before CD+A)	23/24	—	23/23	18/24	64/71	2.92 (1.09)
Total number of patients that achieved treatment success/total number of patients	1250/1381	751/929	851/1171	222/393	3074/3874	—

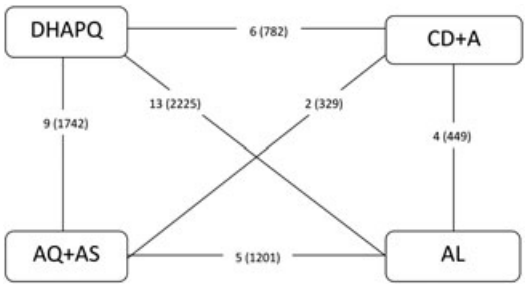

Figure 1. Network diagram of ACTs. Number of sites (number of patients) displayed.

3.2. Implementation

Models were applied using WinBUGS (MRC Biostatistics Unit, Cambridge) 1.4.3 and the R2WinBUGS package in R (R Foundation for Statistical Computing, Vienna, Austria). The between site variances were assumed to be equal across treatment comparisons (i.e. $\tau^2_{bk} = \tau^2$) and the correlation between site-specific treatment effects from the same site was taken into account in the modelling. The continuous covariate was centred at its mean to aid convergence of the chains and parameter estimates were back transformed so that the interpretation of the model parameters is as described in Section 2. Noninformative normal prior distributions were chosen for μ_{jb}, $d_{\mathrm{A}k}$, $\beta_{\mathrm{A}k}$, β_{0jb}, m_{B}, β, $\beta^w_{\mathrm{A}k}$, $\beta^a_{\mathrm{A}k}$, m^w_{B}, m^a_{B}, β^w and β^a, that is, normal (0,100000). A uniform prior distribution was selected for the between-site standard deviation, that is, $\tau \sim$ uniform$(0, 10)$. Weakly informative prior distributions were specified for the standard deviations of the exchangeable regression coefficients (i.e. τ_{B}, τ^w_{B} and τ^a_{B}), that is, uniform $(0, 2)$.

Initial values for IPD models were identical to those of the AD models, with the exception of the site-specific treatment effects because the models required specification of the initial values in different formats. Initial values were chosen arbitrarily. For each model, three MCMC chains with different initial values were run for 300,000 iterations including a burn-in period of 100,000 draws, therefore producing an overall sample of 600,000 draws. Every 10th iteration in the overall sample was retained to provide parameter estimates. Convergence of the chains was assessed using Gelman Rubin convergence methods and by inspecting plots of the draws. A potential scale reduction factor of less than 1.1 was indicative of convergence [34].

For each data type (i.e. IPD and AD), models were compared by evaluating the statistical significance of the regression coefficients for treatment by age interactions and by assessing the reduction in the between site variance following the inclusion of interactions in the model. The deviance information criterion was not used to compare models with and without interactions because it was anticipated that the random-effects model without interactions would fit any data reasonably well and including an interaction would not substantially improve the model fit [35].

Log odds ratios, regression coefficients, and odds ratios for all six possible treatment comparisons (i.e. AQ+AS vs DHAPQ, AL vs DHAPQ, CD+A vs DHAPQ, AL vs AQ+AS, CD+A vs AQ+AS, CD+A vs AL) were obtained. However, only results for AQ+AS versus DHAPQ, AL versus DHAPQ, and CD+A versus DHAPQ, are presented because DHAPQ was used as the referent treatment in the modelling and therefore the results for these three comparisons were estimated in the MTC model and used to calculate the results for the remaining three comparisons (i.e. AL vs AQ+AS, CD+A vs AQ+AS, CD+A vs AL). Odds ratios and drug rankings for children aged one year and aged five years were presented for illustration because describing the difference between these age groups is the most interesting clinically. For patients aged one, the probability that each drug was best was estimated by ranking the treatment effects corresponding to the basic parameters (i.e. $d_{\mathrm{AB}} + \beta_{\mathrm{AB}}, d_{\mathrm{AC}} + \beta_{\mathrm{AC}}, d_{\mathrm{AD}} + \beta_{\mathrm{AD}}$) at each iteration of the chain, and counting the number of iterations for which each drug ranked first. Similarly, for patients aged five, we ranked the treatment effects (i.e. $d_{\mathrm{AB}} + 5\beta_{\mathrm{AB}}, d_{\mathrm{AC}} + 5\beta_{\mathrm{AC}}, d_{\mathrm{AD}} + 5\beta_{\mathrm{AD}}$) at each iteration.

As exploratory analyses, within-site interactions were further examined by fitting logistic regression models to the IPD from each site in turn.

3.3. Results

3.3.1. Aggregate data mixed treatment comparison meta-analysis. The Brooks–Gelman–Rubin convergence plots and the trace plots of the generated draws are convincing in terms of convergence.

However, the trace plot for the variance parameter for the exchangeable regression coefficients of model 3 shows that the estimated variance is greatly influenced by the vague prior distribution. The number of exchangeable regression coefficients is small (i.e. three coefficients), therefore limited data contribute to the estimation of this variance parameter. For each AD model (models 1–4), the number of parameters that require estimation is less than the number of data points, therefore the models are not over parameterised (Table II).

Table II shows results from AD models with and without treatment by age interactions (models 1–4). Using AD, the regression coefficients for the interactions are not statistically significant (Table II). Note that the mean regression coefficient of model 3 (–0.10; 95% credibility interval (CrI) (–2.32, 2.04)) differs from the common regression coefficient of model 4 (0.35; 95% CrI (–1.29, 2.05)) (Table II). The estimate of the between site variance decreases from 0.71 in the model without interactions (model 1) to 0.55 with independent or exchangeable interactions (models 2–3) but increases to 0.78 with common interactions (model 4), suggesting that some variability is explained by mean age in the models with independent or exchangeable interactions (Table II). The plots of the log odds ratio of ACTs versus DHAPQ versus age indicate that interactions are present, although the plots do not take precision into account, therefore should not be over-interpreted (Figure S4). The models with independent or exchangeable interactions (models 2–3) show similar results to each other in terms of the magnitude of effect. Both models (models 2–3) show that the magnitude of the interactions differ across comparisons (i.e. AQ+AS vs DHAPQ, AL vs DHAPQ and CD+A vs DHAPQ) suggesting that forcing the interactions to be common across comparisons may not be appropriate. The results for the model with common interactions (model 4) differ from those for the models with independent or exchangeable interactions (models 2–3) because of the different underlying assumption.

Without interactions (model 1), DHAPQ is found to be significantly more effective than AQ+AS, AL, and CD+A (Figure 2). With interactions (models 2–4), the odds ratios are estimated with considerably low precision (Figure 2).

Without interactions (model 1), DHAPQ is the best drug with 100% probability and the remaining drugs (i.e. AQ+AS, AL and CD+A) each have 0% probability of being the best. However, after allowing for interactions (models 2–4), very different rankings are obtained using each of the three model specifications (Figure 3). With independent interactions (model 2), the drug rankings for sites with a mean age of one are: CD+A, AQ+AS, DHAPQ, AL; and with a mean age of five: AL, DHAPQ, AQ+AS, CD+A. The model with exchangeable interactions (model 3) gives the same ranking as the model with independent interactions (model 2) for sites with a mean age of five; and for sites with a mean age of one, the drug ranking is: DHAPQ, AQ+AS, CD+A, AL. With common interactions (model 4), the order of the drugs in terms of effectiveness is the same for all patients; the ranking is: DHAPQ, AL, AQ+AS, CD+A.

3.3.2. Individual patient data mixed treatment comparison meta-analysis. The Brooks–Gelman–Rubin convergence plots and the trace plots reveal that convergence is achieved for all model parameters. The trace plot for the variance parameter for the exchangeable regression coefficients of model 7 shows that the parameter is less influenced by the vague prior distribution, relative to the AD model (model 3). For each IPD model (models 5–8), the number of parameters that require estimation is far fewer than the number of data points, therefore the models are not over parameterised.

Table II shows results from IPD models with and without treatment by age interactions (models 5–8). Using IPD, the regression coefficient for the interaction for CD+A versus DHAPQ is statistically significant in the models with independent or exchangeable interactions (models 6–7); with common interactions (model 8), the common regression coefficient is statistically significant (Table II). The estimates of the regression coefficients obtained from the IPD models (models 6–8) are closer to the null and more precise than the corresponding estimates from the AD models (models 2–4) (Table II). The mean regression coefficient of model 7 (–0.28; 95% CrI (–0.92, 0.33)) is comparable with the common regression coefficient of model 8 (–0.26; 95% CrI (–0.47, –0.06)) (Table II). The estimate of the between site variance increases from 0.71 in the model without interactions (model 5) to 0.77 with independent interactions (model 6); 0.79 with exchangeable interactions (model 7); and 0.81 with common interactions (model 8) (Table II). The plots of the log odds ratio of ACTs versus DHAPQ versus age suggest that there are weak interactions (Figure S4). The three different model specifications that include interactions (models 6–8) seem reasonably similar in terms of the magnitude and direction of the interactions.

Table II. Results of MTC models including treatment by age interactions for unadjusted treatment success.

Description		No interactions		Independent interactions		Exchangeable interactions		Common interactions	
		AD (model 1)	IPD (model 5)	AD (model 2)	IPD (model 6)	AD (model 3)	IPD (model 7)	AD (model 4)	IPD (model 8)
Number of data points		45	3874	45	3874	45	3874	45	3874
Number of model parameters*		21	21	24	41	23	40	22	39
Log odds ratios (uncentred)	AQ+AS vs DHAPQ	−1.00 (−1.69, −0.35)	−1.00 (−1.69, −0.36)	0.58 (−6.33, 7.03)	−0.51 (−1.42, 0.39)	−0.13 (−5.86, 5.32)	−0.46 (−1.36, 0.41)	−1.91 (−6.40, 2.35)	−0.40 (−1.26, 0.45)
	AL vs DHAPQ	−0.95 (−1.55, −0.32)	−0.95 (−1.54, −0.32)	−4.00 (−8.44, 0.24)	−0.46 (−1.29, 0.40)	−3.20 (−7.38, 1.00)	−0.41 (−1.23, 0.45)	−1.83 (−6.16, 2.36)	−0.34 (−1.13, 0.50)
	CD+A vs DHAPQ	−2.32 (−3.17, −1.54)	−2.32 (−3.18, −1.55)	2.69 (−4.05, 10.12)	−1.16 (−2.37, 0.00)	0.13 (−6.06, 6.14)	−1.42 (−2.58, −0.24)	−3.23 (−7.71, 0.99)	−1.74 (−2.74, −0.78)
Regression coefficients for interactions	AQ+AS vs DHAPQ	—	—	−0.59 (−3.05, 2.01)	−0.22 (−0.47, 0.03)	−0.33 (−2.40, 1.85)	−0.23 (−0.47, 0.01)	0.35 (−1.29, 2.05)	−0.26 (−0.47, −0.06)†
	AL vs DHAPQ	—	—	1.24 (−0.44, 2.99)	−0.21 (−0.46, 0.03)	0.90 (−0.76, 2.56)	−0.23 (−0.46, 0.00)	0.35 (−1.29, 2.05)	−0.26 (−0.47, −0.06)†
	CD+A vs DHAPQ	—	—	−1.94 (−4.84, 0.66)	−0.49 (−0.83, −0.15)†	−0.94 (−3.29, 1.42)	−0.38 (−0.74, −0.09)†	0.35 (−1.29, 2.05)	−0.26 (−0.47, −0.06)†
Mean for distribution of regression coefficients		—	—	—	—	−0.10 (−2.32, 2.04)	−0.28 (−0.92, 0.33)	—	—
Variance for distribution of regression coefficients		—	—	—	—	1.52 (0.02, 3.83)	0.04 (0.00, 2.19)	—	—
Between site variance		0.71 (0.26, 1.84)	0.71 (0.26, 1.86)	0.55 (0.16, 1.66)†	0.77 (0.27, 2.06)	0.55 (0.16, 1.64)†	0.79 (0.28, 2.09)	0.78 (0.29, 2.04)	0.81 (0.30, 2.15)

Posterior median (95% CrI, i.e. 2.5th and 97.5th percentiles of the posterior distribution) presented.

* Model parameters include: 17 log odds for treatment group b (μ_{jb}); 17 differences in the log odds for group treatment b per unit increase in the covariate (β_{0jb}) for IPD models; three log odds ratios (d_{Ak}); one between site variance (τ^2); three regression coefficients (β_{Ak}) for the model with independent interactions; the mean (m_B) and variance (τ_B^2) for the model with exchangeable interactions; the common regression coefficient (β) for the model with common interactions.

† Statistically significant regression coefficient (i.e. zero excluded from the CrI) or between site variance decreases after including the interaction.

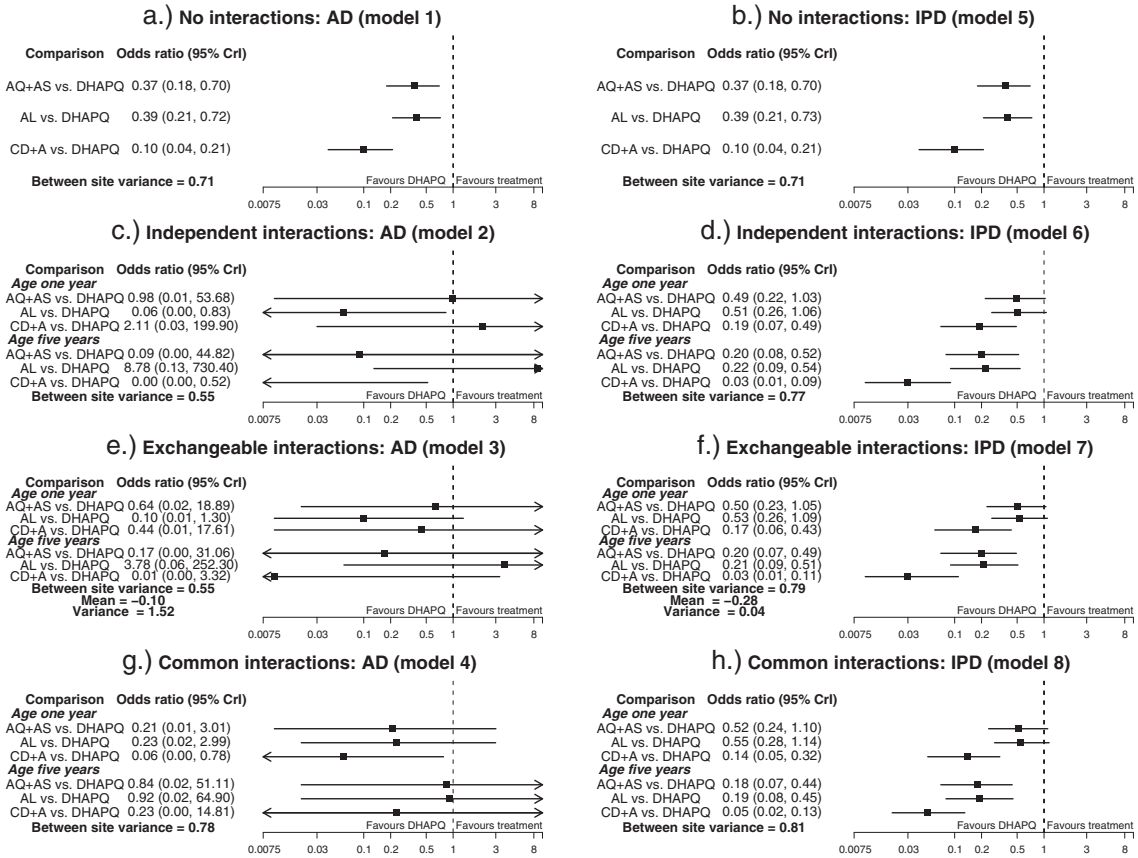

Figure 2. Odds ratios from MTC models including treatment by age interactions for unadjusted treatment success. Posterior median (95% CrI, i.e. 2.5th and 97.5th percentiles of the posterior distribution) presented. 'Mean' and 'variance' are the mean and variance of the distribution of the regression coefficients for the interactions.

Because the IPD model without interactions (model 5) is equivalent to the same AD model (model 1), almost identical results are obtained. With interactions (models 6–8), DHAPQ is consistently more efficacious than AQ+AS, AL and CD+A, with stronger effects for children aged 5 (Figure 2). The 95% CrIs of the odds ratios estimated by IPD (models 6–8) are much narrower than those from the AD models (models 2–4).

Without interactions (model 5), DHAPQ is the best drug with 100% probability. When interactions are included in the model (models 6–8), DHAPQ remains the best drug for patients of any age (Figure 3).

3.3.3. Exploring within-site and across-site treatment by age interactions. When exploring across-site and within-site treatment by age interactions using IPD MTC models (models 9–11), the trace plots for the log odds ratios and regression coefficients for the across-site interactions reveal poor mixing and autocorrelation. The problem is due to the dependence between the log odds ratio (i.e. treatment effect when covariate values are zero) and the across-site regression coefficient (representing the change in the treatment effect with each unit increase in the mean covariate value). Again, the trace plots reveal that the within-site and across-site variance parameters for the exchangeable regression coefficients of model 10 are influenced by the vague prior distributions, particularly the variance associated with the across-site relationships.

Table III displays results obtained when exploring the across-site and within-site interactions using models 9–11. The within-site regression coefficient for the interaction for CD+A versus DHAPQ is statistically significant in the models with independent or exchangeable interactions (models 9–10); and the within-site coefficient for the interaction for AL versus DHAPQ is statistically significant in the model with exchangeable interactions (model 10). With common interactions (model 11), the common within-site regression coefficient is statistically significant. No regression coefficients for across-site

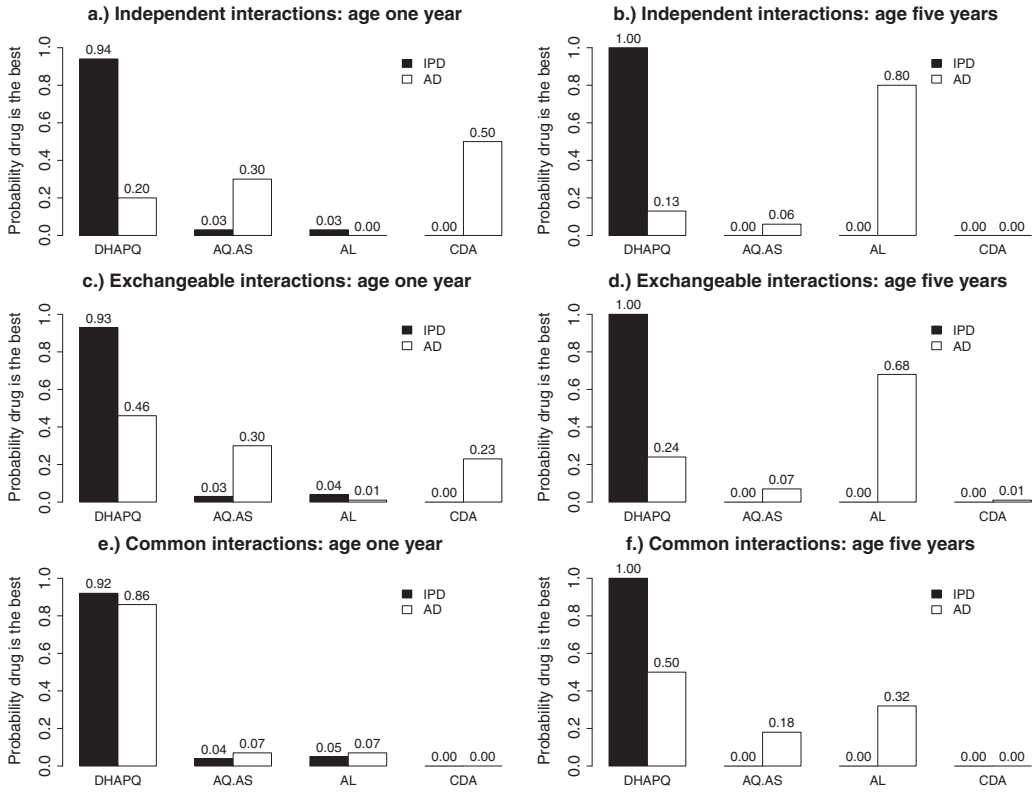

Figure 3. Probability each drug is the best from MTC models including treatment by age interactions for unadjusted treatment success.

interactions are statistically significant. These results support the findings of the original AD models (models 2–4) and IPD models (models 6–8), with the exception of the significant interaction detected for AL versus DHAPQ (Table II). The Bayesian P value indicates that there is a difference between the within-site and across-site regression coefficients for AL versus DHAPQ in the model with independent interactions (model 9) ($P = 0.08$). The estimate of the between site variance is lower than in the original IPD model with independent interactions (i.e. 0.64 (model 9) rather than 0.77 (model 6)) and the original model with exchangeable interactions (i.e. 0.64 (model 10) rather than 0.79 (model 7)); whereas with common interactions, the estimate of the between site variance is higher than in the original model (i.e. 0.87 (model 11) rather than 0.81 (model 8)).

As exploratory analyses, within-site treatment by age interactions were looked at. At each site, age varies between 6 months to 5 years; whereas, across sites, the mean ages are similar (1.99 to 3.08 years) (Figure S5). Therefore, treatment by age interactions are more likely to be detected within sites than across sites if the interactions exist. When a model that includes a treatment by age interaction was fitted to the IPD from each site in turn, more informative prior distributions were required on many occasions, but convergence was achieved. Bearing in mind that multiple testing issues apply because a total of 28 models with interactions were fitted, site-specific regression coefficients indicate an interaction may exist for AQ+AS versus DHAPQ in Pamol and Mbarara (before CD+A is discontinued), and for CD+A versus DHAPQ in Mbarara (before CD+A is discontinued) (Table SIV).

4. Discussion

In this article, the inclusion of treatment by covariate interactions in MTC models to assess the consistency assumption was described. For each treatment comparison, when the true treatment effect is similar across all trials, the consistency assumption is satisfied. Therefore, when no interactions are detected, the assumption may be reasonable provided there are no unknown treatment effect modifying covariates. The consistency assumption would not hold when one or more of the true treatment effects is modified by a particular characteristic and the included trials differ with respect to

Table III. Results of investigation of across-site and within-site treatment by age interactions using IPD MTC models for unadjusted treatment success.

Description		Independent interactions (model 9)	Exchangeable interactions (model 10)	Common interactions (model 11)
Number of data points		3874	3874	3874
Number of model parameters*		44	42	40
Log odds ratios (uncentred)	AQ+AS vs DHAPQ	0.26 (−7.82, 6.67); 0.16	−0.01 (−5.47, 5.63); 0.12	−1.59 (−6.44, 2.25);0.09
	AL vs DHAPQ	−4.17 (−9.30, 0.04); 0.09	−2.99 (−6.96, 1.43); 0.08	−1.49 (−6.20, 2.25); 0.09
	CD+A vs DHAPQ	2.66 (−4.57, 9.95); 0.15	0.26 (−6.16, 6.79); 0.14	−2.91 (−7.81, 0.87); 0.09
Within-site regression coefficients for interactions	AQ+AS vs DHAPQ	−0.22 (−0.47, 0.03); 0.00	−0.24 (−0.48, 0.00); 0.00	−0.27 (−0.48, −0.06)†; 0.00
	AL vs DHAPQ	−0.23 (−0.48, 0.01), 0.00	−0.25 (−0.48, −0.02)†; 0.00	−0.27 (−0.48, −0.06)†; 0.00
	CD+A vs DHAPQ	−0.47 (−0.82, −0.13)†; 0.00	−0.37 (−0.72, −0.08)†; 0.00	−0.27 (−0.48, −0.06)†; 0.00
Across-site regression coefficients for interactions	AQ+AS vs DHAPQ	−0.50 (−2.94, 2.54); 0.06	−0.40 (−2.54, 1.67); 0.04	0.20 (−1.25, 2.04); 0.04
	AL vs DHAPQ	1.28 (−0.39, 3.31); 0.04	0.80 (−0.93, 2.37); 0.03	0.20 (−1.25, 2.04); 0.04
	CD+A vs DHAPQ	−1.97 (−4.82, 0.83); 0.06	−1.03 (−3.57, 1.43); 0.05	0.20 (−1.25, 2.04); 0.04
Bayesian P value (difference between within-site and across-site regression coefficients)‡	AQ+AS vs DHAPQ	0.82	0.88	0.56
	AL vs DHAPQ	0.08	0.22	0.56
	CD+A vs DHAPQ	0.28	0.64	0.56
Mean for distribution of within-site regression coefficients		—	−0.28 (−0.89, 0.28); 0.00	—
Variance for distribution of within-site regression coefficients		—	0.03 (0.00, 2.03); 0.00	—
Mean for distribution of across-site regression coefficients		—	−0.20 (−2.45, 1.91); 0.04	—
Variance for distribution of across-site regression coefficients		—	1.50 (0.01, 3.84); 0.03	—
Between site variance		0.64 (0.19, 1.95)†; 0.01	0.64 (0.18, 1.92)†; 0.01	0.87 (0.32, 2.30); 0.01

Posterior median (95% CrI, i.e. 2.5th and 97.5th percentiles of the posterior distribution); Monte Carlo error presented.

* Model parameters include: 17 log odds for treatment group b (μ_{jb}); 17 differences in the log odds for baseline treatment b group per unit increase in the covariate (β_{0jb}); three log odds ratios (d_{Ak}); one between site variance (τ^2); six regression coefficients $\left(\beta^a_{Ak}, \beta^w_{Ak}\right)$ for the model with independent interactions; two means $\left(m^a_B, m^w_B\right)$ and two variances $\left(\tau^a_B\right)^2, \left(\tau^w_B\right)^2\Big)$ for the model with exchangeable interactions; two common regression coefficients (β^a, β^w) for the model with common interactions.

† Statistically significant regression coefficient (i.e. zero excluded from the CrI) or between site variance decreases after including the interaction.

‡ For each treatment pairing, a Bayesian P value was estimated by calculating the difference (i.e. *diff*) between the within-site regression coefficients and the across-site regression coefficient, at each iteration of the chain, and counting the number of iterations for which $diff \geqslant 0$. It was then possible to calculate the probability (i.e. *prob*) that the within-site regression coefficient exceeded the across-site regression coefficient, by dividing the number of counted iterations by the total number of iterations of the chain. Lastly, assuming that the posterior distribution of the differences (i.e. *diff*) was symmetric and unimodal, the P value was obtained by $P = 2 \times \text{minimum}(\text{prob}, 1 - \text{prob})$ [11, 36]. Results were considered to be significantly different when $P \leqslant 0.1$.

the characteristic. Consequently, if interactions are detected, the assumption may be unrealistic and the results from the original MTC would be questionable. However, valid inferences may be drawn from the model including interactions provided the interactions have explained any pre-existing inconsistency. Methods to assess whether interactions have explained inconsistency have been developed (Donegan *et al.*, unpublished research).

The benefits of using IPD, rather than AD, to explore interactions in an MTC were studied using random-effects Bayesian modelling. The AD models have previously been described by Cooper *et al.* [12] and were extended in this article to allow for IPD. For each data type (i.e. IPD and AD), three model specifications including interactions were described, each making different assumptions (i.e. independent, exchangeable, or common interactions). The IPD and AD approaches were compared by applying the models to a real dataset from the 4ABC trial.

For the 4ABC trial, differences in results were found between the IPD and AD approaches and between different model specifications. Using IPD, several regression coefficients for treatment by age interactions were statistically significant; whereas using AD, no coefficients were significant. On the basis of these results, when using AD, there was no evidence that the consistency assumption was invalid; conversely, the assumption seemed questionable when exploring IPD.

The estimates of between site variability differed by data type. Following the inclusion of independent or exchangeable interactions in the model, the variance decreased using AD (models 2–3); and the variance increased using IPD (models 6–7). The variance increased using the IPD models (models 6–7) because age did not explain the differences in the treatment effects across sites. The decreasing variance using AD (models 2–3) suggested that mean age partly explained the between site variability. However, because the within-site interactions and across-site interactions obviously differed, the relationships between treatment effects and average age were artificial and certainly influenced by biases and confounding.

For both data types, the variance increased following the inclusion of common interactions (models 4 and 8). The increase may have been due to the underlying assumption of the regression coefficients. If the regression coefficients for the interactions were different for each treatment comparison and the model has forced the coefficients to be identical, the variability between the interactions for different comparisons may have contributed to the between site variance and therefore caused the variation to increase relative to the model without interactions.

Importantly, the 95% CrIs for the treatment effects and regression coefficients for interactions from the IPD models (models 6–8) were substantially narrower than those from AD models (models 2–4) because the effect of the covariate both within-sites and across-sites contributed to the parameter estimates of the IPD models, whereas results from AD models were solely based on across-site associations. Using AD, the 95% CrIs for treatment effects and regression coefficients for the interactions were wide and it was impossible to distinguish whether interactions existed, but more data were needed to detect the interactions, or whether there were no underlying interactions. Therefore, we would have been unable to detect the treatment by age interactions if IPD were unavailable. Also, generally, across-trial interactions are prone to ecological bias, whereas the within-trial associations are not affected. For these reasons, when assessing the consistency assumption, acquiring and using patient-level covariates rather than study-level covariate summaries seems beneficial.

However, the distribution of the covariate within studies and across studies should be considered to establish if the available data permits investigation of such interactions. The 4ABC trial included children aged 6 months to 5 years and the distribution of age at each site generally spanned this age range, therefore the IPD models should have detected interactions if they were present. In contrast, the site-specific mean age of patients did not show substantial differences across the sites (1.99–3.08 years); therefore, the AD models may have struggled to detect interactions if they existed.

The probability that each drug was the best also differed between the IPD and AD approaches. Each of the IPD models showed that DHAPQ was consistently by far the best drug for patients of any age; whereas the AD models showed that the best drug varied depending on mean age and the model specification, and in some cases the drugs were quite evenly matched.

The most important limitation of this research is that the results presented were restricted to one specific dataset and one covariate. In this particular example, a continuous covariate (i.e. age) with quite a narrow age range was studied (i.e. 0.5–5 years) and the covariate distribution was similar for each site. If adults were included in the trial, or the covariate distributions were highly variable across sites, more dramatic, or different results, may have been observed. Further research is needed to establish the benefits of IPD in various settings. Additional investigations using diverse, simulated, or real datasets and

other potential treatment effect-modifying covariates would provide further evidence regarding when IPD are more valuable than AD in an MTC setting.

Here, models for dichotomous outcomes were described and compared, but they could easily be altered to enable application to other types of outcomes, such as continuous outcome measurements. The AD models described in this article were based on event rates and study-level covariates; however, MTC models based on study-specific treatment effect estimates and their standard errors have previously been described [6]. MTC models with treatment by age interactions based on log odds ratios and standard errors were applied (results not presented) and the results were reasonably consistent with the AD models based on event rates.

A further restriction is that the models applied assume the same between site variance for each treatment comparison, which may not be realistic. Exploring different variance structures would improve the feasibility of the model; however, more complex variance structures may be inappropriate if the number of studies contributing to some treatment pairings is low, as was the case for the 4ABC trial [3, 4, 25]. When data are limited, more informative prior distributions for the parameters could be applied, for example, based on evidence from nonrandomised studies, or the beliefs of clinical experts. To explore the influence of the prior distributions on the MTC results, sensitivity analyses could also be carried out by applying various informative or vague prior distributions for variance parameters.

The specified models assumed random treatment effects $\left(\text{i.e. } \delta_{jbk} \sim N\left(d_{bk}, \tau^2_{bk}\right)\right)$ and fixed effects for the log odds of the baseline treatment b group (μ_{jb}), the study-specific regression parameter (β_{0jb}), and the regression parameter for the treatment by covariate interaction (β_{bk}). Alternative models could specify random effects for the log odds of the baseline treatment b group $\left(\text{i.e. } \mu_{jb} \sim N\left(0, \tau^2_{\mu}\right)\right)$, the study-specific regression parameter $\left(\text{i.e. } \beta_{0jb} \sim N\left(0, \tau^2_{\beta_0}\right)\right)$, or the regression parameter for the interaction $\left(\text{i.e. } \gamma_{jbk} \sim N\left(\beta_{bk}, \omega^2_{bk}\right)\right)$ [37]. In such models, the correlation between the random effects (e.g. δ_{jbk} and γ_{jbk}) must also be taken into account. The adapted models would only be useful when there is enough data to accurately estimate all additional model parameters.

The robustness of the results presented in this article was not tested by reapplying the same models using different referent treatments. When no interactions are included in the MTC model, the selection of the referent treatment should not affect the results. However, the results could be influenced when treatment by covariate interactions are examined [32]. DHAPQ was chosen as the referent drug because this was described in the statistical analysis plan of the 4ABC trial prior to the trial commencing. Further research could question how to choose the referent treatment and how to proceed when conflicting results are obtained from different model parameterisations.

Three model specifications were applied that made different assumptions regarding the treatment by covariate interactions (i.e. independent, exchangeable or common interactions). The three interaction types have previously been described by Cooper *et al.* [12] for AD models. The results from the different specifications were not compared in detail because this was not the purpose of this research. However, it seems clear that common interactions would rarely be a reasonable assumption. Moreover, the results from models with common interactions would differ with the choice of referent treatment. Future investigations could further examine the different specifications to discover when each assumption is realistic.

Also, although acquiring patient-level covariates and IPD may enhance application of meta-analysis generally, review authors often do not do so, either because of time or resource limitations, or because the IPD cannot be accessed [38, 39]. For MTC meta-analysis, the number of included trials may be greater than in a standard pair-wise meta-analysis and therefore acquiring, cleaning, and analysing IPD, for all studies is an even larger undertaking. When IPD can only be obtained for a subset of trials and AD is available for the remaining trials, the IPD and AD can be combined in a single MTC model (Donegan *et al.*, unpublished research).

In conclusion, treatment by covariate interactions can be included in MTC models to assess the consistency assumption. In this example, using AD models, there was no evidence that the consistency assumption was invalid; whereas, the assumption was questionable based on the IPD models. The AD analyses were misleading.

For this empirical example, the treatment effects and drug rankings from IPD MTC models that included patient-level covariates differed from those of AD models that included site-level covariates. The inclusion of patient-level, rather than site-level, covariates produced more precise treatment effects. Therefore, including treatment by patient-level covariate interactions was preferable.

Acknowledgements

Special thanks to the collaborators of the 4ABC trial for collecting and supplying the IPD for use in this research. Also, thanks to Dr Nicola Cooper and Prof Alex Sutton for providing the WinBUGS code applied in Cooper *et al.* [12] and for useful discussions. We are also grateful to the two anonymous reviewers for their helpful comments on the manuscript. This research contributed to Sarah Donegan's Ph.D., which was funded by the Liverpool Reviews and Implementation Group, Liverpool (JXG 10265). The 4ABC trial was sponsored by Prince Leopold Institute of Tropical Medicine (Antwerp) and funded by The European and Developing Countries Clinical Trials Partnership.

Catrin Tudur Smith and Paula Williamson proposed the idea of comparing individual patient data with aggregate data when exploring treatment by covariate interactions in mixed treatment comparison models. Sarah Donegan extended existing aggregate data methodology to allow for patient-level outcomes and covariates, carried out the analysis, interpreted the results, and wrote the first draft of the manuscript. Tudur Smith contributed to the interpretation of results and writing of the manuscript. Umberto D'Alessandro provided valuable discussions regarding potential treatment-effect modifying covariates and other clinical aspects with respect to the dataset. All authors commented on the manuscript.

References

1. Bucher H, Guyatt G, Griffith L, Walter S. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology* 1997; **50**(6):683–691. DOI: 10. 1016/S0895-4356(97)00049-8.
2. Glenny A, Altman D, Song F, Sakarovitch C, Deeks J, D'Amico R, Bradburn M, Eastwood A. Indirect comparisons of competing interventions. *Health Technology Assessment* 2005; **9**(26).
3. Lu G, Ades A. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; **101**(474):447–459. DOI: 10.1198/016214505000001302.
4. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004; **23**(20):3105–3124. DOI: 10.1002/sim.1875.
5. Lumley T. Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* 2002; **21**(16):2313–2324. DOI: 10.1002/sim.1201.
6. Salanti G, Higgins J, Ades A, Ioannidis J. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008; **17**(3):279–301. DOI: 10.1177/0962280207080643.
7. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**(24):2733–2749. DOI: 10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.
8. Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of Clinical Epidemiology* 2009; **62**(8):857–864. DOI: 10.1016/j.jclinepi.2008.10.001.
9. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Controlled Clinical Trials* 2000; **21**(5):488–497. DOI: 10.1016/S0197-2456(00)00055-6.
10. Song F, Altman DG, Glenny A-M, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *British Medical Journal* 2003; **326**(7387):472. DOI: 10.1136/bmj.326.7387.472.
11. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* 2010; **29**(7-8):932–944. DOI: 10.1002/sim.3767.
12. Cooper N, Sutton A, Morris D, Ades A, Welton N. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine* 2009; **28**(14):1861–1881. DOI: 10.1002/sim.3594.
13. Tudur Smith C, Marson A, Chadwick D, Williamson P. Multiple treatment comparisons in epilepsy monotherapy trials. *Trials* 2007; **8**:34. DOI: 10.1186/1745-6215-8-34.
14. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statistics in Medicine* 2007; **26**(6):1237–1254. DOI: 10.1002/sim.2624.
15. Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons Ltd: Chichester, England, 2008.
16. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 2002; **55**(1):86–94. DOI: 10.1016/S0895-4356(01)00414-0.
17. Simmonds MC, Higgins JPT. Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* 2007; **26**(15):2982–2999. DOI: 10.1002/sim.2768.
18. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* 2002; **21**(3):371–387. DOI: 10.1002/sim.1023.
19. Tudur Smith C, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. *Journal of Evaluation in Clinical Practice* 2005; **11**(5):468–478. DOI: 10.1111/j.1365-2753.2005.00559.x.

20. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *Journal of Clinical Epidemiology* 2004; **57**(7):683–697. DOI: 10.1016/j.jclinepi.2003.12.001.
21. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials, 2011. Available from: http://www.nicedsu.org.uk. Accessed: 9 June 2011.
22. Salanti G, Kavvoura F, Ioannidis J. Exploring the geometry of treatment networks. *Annals of Internal Medicine* 2008; **148**(7):544–553.
23. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Boutitie F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* 2008; **27**(11):1870–1893. DOI: 10.1002/sim.3165.
24. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods* 2010; **1**(1):2–19. DOI: 10.1002/jrsm.4.
25. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; **10**(4):792–805. DOI: 10.1093/biostatistics/kxp032.
26. Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2010; **173**(3):613–629. DOI: 10.1111/j.1467-985X.2010.00639.x.
27. Woods B, Hawkins N, Scott D. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: A tutorial. *BMC Medical Research Methodology* 2010; **10**:54. DOI: 10.1186/1471-2288-10-54.
28. Govan L, Ades AE, Weir CJ, Welton NJ, Langhorne P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. *Statistics in Medicine* 2010; **29**(12):1340–1356. DOI: 10.1002/sim.3869.
29. Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. *Value in Health* 2010; **13**(8):976–983. DOI: 10.1111/j.1524-4733.2010.00784.x.
30. Whitehead A. *Meta-Analysis of Controlled Clinical Trials*. Wiley: Chichester, 2002.
31. Jansen J. Network meta-analysis of survival data with fractional polynomials. *BMC Medical Research Methodology* 2011; **11**(1):61. DOI: 10.1186/1471-2288-11-61.
32. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Guobing L, Ades AE. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials, 2011. Available from: http://www.nicedsu.org.uk. Accessed: 9 June 2011.
33. The Four Artemisinin-Based Combinations Study Group. A head-to-head comparison of four artemisinin-based combinations for treating uncomplicated malaria in african children: a randomized trial. *PLoS Medicine* 2011; **8**(11):e1001119. DOI: 10.1371/journal.pmed.1001119.
34. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC: London, UK, 1996.
35. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639. DOI: 10.1111/1467-9868.00353.
36. Marshall E, Spiegelhalter D. Identifying outliers in Bayesian hierarchical models: a simulation-based approach. *Bayesian Analysis* 2007; **2**(2):409–444. DOI: 10.1214/07-BA218.
37. Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; **20**(15):2219–2241. DOI: 10.1002/sim.918.
38. Clarke MJ, Stewart LA. Meta-analyses using individual patient data. *Journal of Evaluation in Clinical Practice* 1997; **3**(3):207–212. DOI: 10.1046/j.1365-2753.1997.00005.x.
39. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal* 2010; **340**:c221. DOI: 10.1136/bmj.c221.