Volume 11 • Number 5 • 2008
VALUE IN HEALTH

Bayesian Meta-Analysis of Multiple Treatment Comparisons: An Introduction to Mixed Treatment Comparisons

Jeroen P. Jansen, PhD,[1] Bruce Crawford, MA, MPH,[1] Gert Bergman, PhD,[2] Wiro Stam, PhD[2]

[1]Mapi Values Ltd, Boston, MA, USA; [2]Mapi Values, Houten, The Netherlands

ABSTRACT

Recently, mixed treatment comparisons (MTC) have been presented as an extension of traditional meta-analysis by including multiple different pairwise comparisons across a range of different interventions. MTC allow for indirect comparisons and can therefore provide very useful information for clinical and reimbursement decision-making in the absence of head-to-head data. In this article, we provide an introductory overview of MTC illustrated with example analyses of different drug treatments in rheumatoid arthritis using a continuous patient-reported end point. As a background, we start with an overview of the traditional meta-analyses for pairwise trials, and the difference between a traditional approach and a Bayesian approach. Next, the Bayesian MTC for continuous outcomes are presented. We finish with a discussion of how MTC can best be presented in order to maximize acceptance by target audiences, i.e., clinicians and market access decision-makers.

Keywords: Bayesian methods, evidence synthesis, meta-analysis, mixed treatment comparisons.

Introduction

Systematic reviews are considered the standard practice to inform evidence-based decision-making regarding the efficacy and safety of medical technology. As part of a systematic review, quantitative results of several similar studies can be combined by means of meta-analysis to summarize the available evidence into a pooled estimate of the outcome of interest. Recently, mixed treatment comparisons (MTC) have been presented as an extension of traditional meta-analysis (where all included studies compare the same intervention with the same comparator) by including multiple different pairwise comparisons across a range of different interventions [1–3]. With MTC, the relative efficacy (or safety) of a particular intervention versus competing interventions can be obtained in the absence of head-to-head comparisons; indirect comparison of two interventions is made through a common comparator. These types of evidence synthesis methods are often performed as a basis for cost-effectiveness decision-making [4]. Moreover, MTC provide very useful information regarding the relative clinical value of medical interventions and can therefore support medical decision-making; MTC allow us to rank-order available interventions regarding efficacy, safety, or patient-reported outcomes.

In this article, we provide an introductory overview of MTC illustrated with a example analysis of a continuous patient-reported outcome in rheumatoid arthritis. To provide a background, we start with an overview of traditional meta-analyses for pairwise trials, and the difference between a frequentist and a Bayesian approach. Next, Bayesian MTC for continuous outcomes are introduced with a fixed-effects model and a random-effects model. We conclude with a discussion of how MTC can best be presented in order to maximize acceptance by target audiences.

Address correspondence to: Jeroen P Jansen, Mapi Values, 15 Court Square, Suite 620, Boston, MA 02108, USA. E-mail: jeroen.jansen@mapivalues.com

10.1111/j.1524-4733.2008.00347.x

Standard Meta-Analysis for Pairwise Trials

This section briefly reviews the basic features of standard meta-analysis. For a more thorough description, see for example Sutton et al. [5]. Meta-analysis is the statistical tool for estimating a treatment effect obtained from several sources of evidence available. One of the aims of meta-analysis is to combine results of comparable studies in order to obtain an overall estimate of effect (e.g., odds ratio, relative risk, or difference in change from baseline) thereby reducing uncertainty. Meta-analysis uses the magnitude of the effect and its uncertainty from each study to produce a weighted mean. Roughly, meta-analyses can be differentiated into fixed-effect approaches and random-effects approaches, and meta-analysis can be performed with traditional frequentist statistics or with Bayesian statistics.

Fixed-Effect and Random-Effects Meta-Analysis

With the fixed-effect approach, the assumption is made that each of the individual studies aims to estimate the

  1098-3015/08/956 956–964

same true treatment effect (i.e., the underlying effect), and that differences between studies are solely due to chance. The observed effect of each study equals a fixed effect common to all studies plus sampling error. The true treatment effect (estimated with the pooled effect) is calculated as the weighted average of study-specific effects, with weights based on the inverse of the uncertainty (standard error squared) of each study. The uncertainty of the pooled estimate is calculated as the sum of inverse weights [5]. This method is often referred to as the inverse-variance-weighted method.

With the random-effects approach, it is assumed that, in addition to sampling error, differences between studies are caused by heterogeneity between studies. In other words, it is assumed that for each of the different studies, the true effect may be study specific and vary across studies. In addition, it is often assumed that these true effects are described by a normal distribution [6]. Hence, the variation in observed individual study results is caused not only by sampling error (as with the fixed-effect approach) but also by the variation in the true (underlying) effects of each study, which is called the *random-effects variance*. The use of the random-effects model has been advocated if there is heterogeneity between study results, caused by different study populations across studies, or methodological differences, for example. Again, the true treatment effect is calculated as the weighted average of study specific effects. Now, however, the weights are based on a combination of the uncertainty (standard error squared) of each study and the random-effects variance. The uncertainty of the pooled estimate is calculated as the sum of inverse weights [5]. Hence, the confidence interval of the pooled estimate obtained with the random-effects model is a function of both sampling error of the individual studies and between-study variance, and is therefore wider than the confidence interval obtained with a fixed-effect model.

Neither of the methods is ideal. Fixed-effect models ignore heterogeneity, but the random-effects model uses distributional assumptions for the heterogeneity that can be argued to be unrealistic or unjustified. In general, when there is evidence of heterogeneity that is not explained by covariates in a meta-regression analysis, random-effects methods are preferred [5].

Traditional Approach and Bayesian Approach

With a frequentist approach, the result of the meta-analysis is a point estimate along with a 95% confidence interval. It has to be noted that confidence intervals obtained with a frequentist approach cannot be interpreted in terms of probabilities (the 95% confidence interval does *not* mean there is 95% probability that "true" or population value is between the boundaries of the interval) [7].

Bayesian methods involve a formal combination of a prior probability distribution (that reflects a prior belief of the possible values of the pooled effect) with a (likelihood) distribution of the pooled effect based on the observed data to obtain a posterior probability distribution of the pooled effect [5]. (The likelihood informs us about the extent to which different values for the parameter of interest is supported by the data. Frequentists use the sampling distribution as the basis of statistical inference that is proportional to the likelihood function) [8]. The posterior distribution (as obtained with the Bayesian approach) can be interpreted in terms of probabilities, which allows for a more intuitive interpretation (e.g., "There is an x% probability that treatment A results in a larger cholesterol reduction than treatment B"). This is in contrast to findings with a conventional frequentist approach. Another major advantage of the Bayesian approach is that the method naturally leads into a decision framework to support decision-making [5,9,10]. Other advantages of a Bayesian meta-analysis include the straightforward way of making predictions, and the possibility of incorporating different sources of uncertainty [5,10]. In order not to influence the observed results by the prior distribution, an often heard critique of the Bayesian approach, a noninformative prior distribution can be used for the pooled treatment effect. With such a "flat" prior, it is assumed that before seeing the data, any value for the pooled effect is equally likely to occur. As a consequence, posterior results are not influenced by the prior distribution but totally driven by the data as with a conventional frequentist meta-analysis.

Within the Bayesian framework, analyses consist of data, likelihood, parameters, and a model [4]. The simple Bayesian random-effects meta-analysis for continuous outcomes [11] can be written according to

$$\begin{aligned} &diff_i \sim N(\delta_i, \sigma_i^2) && \text{for each study } i \text{ (Likelihood)} \\ &\delta_i \sim N(d, \sigma_\delta^2) && \text{(Random-effects model)} \\ &d \sim [-,-],\ \sigma_\delta \sim [-,-] && \text{(Prior distributions)} \end{aligned} \quad (1)$$

where $diff_i$ reflects the observed treatment effect (e.g., difference in a continuous end point) of treatment A versus B in study i (the *data*). The observed $diff_i$ is described by the study-specific true effect δ_i and sampling Error σ_i, according to a normal distribution. This is called the *likelihood*. (Often, the standard error of a study is used for σ_i). The *model* describes how the true study-specific effects δ_i are distributed. The basic *parameters* of the model are the pooled treatment effect d and between-study variance σ_δ^2 reflecting the heterogeneity of the true study-specific effects, which both need to be estimated. (Actually, the objective of the meta-analysis is to estimate the pooled treatment effect d and heterogeneity σ_δ^2.) In essence, the model describes how the parameters d and σ_δ^2 relate to the data. (If σ_δ^2 is set to 0, this implies there is no between-study heterogeneity, and the model reflects a fixed-

effect approach.) Because the analysis is performed within a Bayesian framework, *prior distributions* need to be defined for d and σ_δ^2 reflecting prior belief of the likely values for these parameters. [–,–] symbolizes an arbitrarily prior distribution for the pooled effect and heterogeneity. (For example, a normal distribution with a mean of 0 and a variance of 1000 can be used, and an inverse gamma prior distribution for the heterogeneity). The prior distributions are updated according to the data to result in a posterior distribution for d and σ_δ^2. It is not an easy task to assign uninformative priors, in particular to variance parameters reflecting the heterogeneity, and therefore the model should be tested for sensitivity to alternative specification of uninformative priors. (Remark: with random-effects meta-analysis according to a traditional frequentist approach, the between-study variance is calculated and the uncertainty in this estimate is ignored in the confidence interval of the pooled estimate. Therefore, in contrast to a Bayesian random-effects meta-analysis, a frequentist random effects meta-analysis results in underestimates of uncertainty of the pooled effect estimate) [12].

Indirect Comparisons and Mixed Treatment Comparisons

There is often an interest among physicians, decision-makers, and drug manufacturers to identify the most effective treatment or to rank-order the treatments among a range of alternatives. A randomized controlled trial (RCT) comparing all different interventions provide such information. However, RCTs are often designed for registration purposes and therefore do not include all available comparator interventions. The comparator arms of such trials are often limited to a placebo intervention as well as one common or best practice intervention. In order to obtain insight into the relative efficacy (or safety) versus excluded interventions, one has to rely on indirect comparisons. Of key importance in indirect comparisons is not to "break randomization" [13]: For example, if one trial compares drug A versus placebo regarding cholesterol reduction and a second trial compares drug B versus placebo, it is incorrect to simply compare the absolute cholesterol reduction observed with drug A with the absolute reduction observed with drug B. One reason is that part of the absolute reduction can be attributed to the efficacy of the drug, whereas another part is due to a placebo effect. (RCTs are designed to separate drug effects from other effects.) Another reason to avoid "breaking randomization" is that differences may be observed in absolute treatment effects as a result of different baseline risks, even where the relative risk is consistent between trials [5]. In order not to "break randomization," one can only compare the relative effect of drug A versus placebo from one trial with the relative effect from other trials (this principle also applies to *combining* results of similar trials in a meta-analysis; only placebo-subtracted or relative effects are pooled, otherwise we "create" an observational study).

Meta-analysis of placebo controlled drug-A trials (PA trials) provides a direct estimate of the *true* relative effect of A versus placebo (d_{PA}). Meta-analysis of PB trials provides a direct estimate of the true relative effect d_{PB}. If the included PA and PB trials are alike and the fixed-effects assumption applies (i.e., the true comparison-specific relative effects are fixed and differences between PA and PB trials are only caused by the different treatment and no other factors) or the random-effects assumption applies (i.e., the true comparison-specific relative effects are drawn from a population of effects and exchangeable), then the true relative efficacy of the different types of comparisons are mathematically related, as illustrated in Fig. 1a. In the absence of "head-to-head" evidence comparing drugs A and B, an indirect estimate for the relative *true* effect of B versus A (d_{AB}) can be obtained from the *true* effect d_{PA} and from the *true* effect d_{PB}. In essence, this implies that the same true d_{AB} is obtained as would have been estimated in a meta-analysis focusing on drug A versus B using three-arm PAB trials, if available. Fundamentally, the only difference between combining results of two PA trials and an indirect comparison of one PA and one PB trial is that we are not taking the (weighted) average of two treatment effects, but are now subtracting two treatment effects. Both are calculations with two treatment effects.

MTC is a generalization of standard meta-analysis for pairwise trials to a simultaneous analysis of multiple pair-wise comparisons [2]. As for the example illustrated in Fig. 1a, the meta-analyses for PA and PB trials are performed simultaneously. Figure 1b represents the situation when, in addition to interventions A and B, the intervention C is of interest as well. For this latter intervention, direct estimates from AC trials are available. Given the network of direct comparisons across the range of interventions, indirect estimates can be obtained for d_{PA}, d_{PB}, d_{AB}, d_{BC}, and d_{PC}. Given the mathematical relations between the true underlying estimates of the different comparisons in the network, we have both direct and indirect evidence available for all the pairwise comparisons, except for the BC comparisons (only indirect evidence) and the AC comparisons (only direct evidence; see Fig. 1b). Hence, the advantages of the simultaneous analysis with MTC are that 1) estimates for indirect comparisons are obtained; and 2) indirect comparisons can support evidence for direct estimates [4].

In general, with MTC, the same assumptions apply as with traditional meta-analysis for one type of comparison. If it is assumed appropriate to combine results of different studies with standard meta-analysis for one type of comparison with a fixed- or random-effects

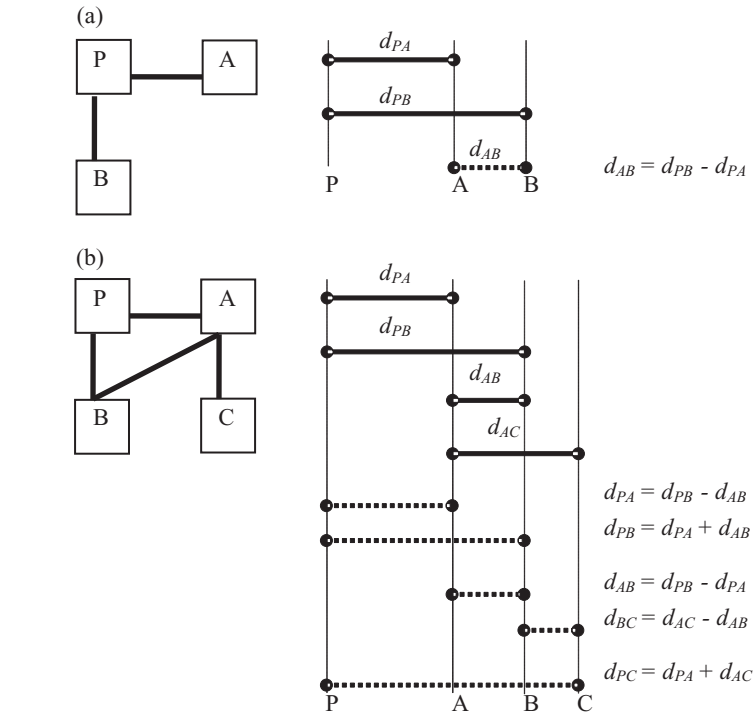

Figure 1 Network of studies reflecting indirect comparison of treatment A and B given PA trials with PB trials (a), and network of studies reflecting mixed treatment comparisons of PA trials, PB trials, AB trials, and AC trials (b). See text for explanation of direct and indirect estimates.

assumption, then MTC have to be considered an appropriate methodology when studies across the range of interventions can be considered identical or exchangeable (other than differences in treatments compared). With MTC with a fixed-effects model, it is assumed that differences in true relative treatment effects (whether estimated directly or indirectly) are only caused by the difference in treatment and no other factors. There is no heterogeneity in true relative treatment effects beyond differences in treatment effect caused by the differences in the types of interventions compared. With a random-effects assumption, differences in study-specific treatment effects (beyond the differences attributable to the interventions compared) are exchangeable (and heterogeneity is constant between the different comparisons).

As outlined earlier, a meta-analysis within a Bayesian framework has advantages over frequentist analysis. For MTC, an additional advantage of a Bayesian approach is that it allows calculation of the probability which of the treatments is best and which directly appeals to clinicians and reimbursement decision-makers, among other probability statements and predictions [7]. The random-effects model according to Equation 1 can be extended with terms reflecting indirect or mixed treatment comparisons (as presented in Fig. 1). With [–,–] representing arbitrarily prior distributions, the Bayesian random-effects model for MTC can be written according to

$$\begin{aligned} &diff_i \sim N(\delta_i, \sigma_i^2) && \text{for each study } i \text{ (Likelihood)} \\ &\delta_{i,b,k} \sim N(d_{b,k}, \sigma_\delta^2) \sim N(d_{P,k} - d_{P,b}, \sigma_\delta^2) && \text{(Random-effects model)} \\ &d_{P,k} \sim [-,-],\ \sigma_\delta \sim [-,-] && \text{(Prior distributions)} \end{aligned} \quad (2)$$

Now $diff_i$ reflects the observed relative treatment effects of the different comparisons (for example: the observed difference of treatment A versus placebo or A versus C in study *i*). Again, the observed relative effects are described by the study-specific true relative effects according to the likelihood. The *model* defines how the true study and comparison-specific relative effects $\delta_{i,b,k}$ are distributed and related according to basic parameters as illustrated in Fig. 1 ($d_{b,k} = d_{P,k} - d_{P,b}$). *P* refers to the placebo intervention, *k* refers to the active intervention, and *b* is a comparator intervention. Because we use a Bayesian approach we need to define prior

Table 1 Change in HAQ score by intervention as observed in different clinical trials

	Placebo (P)			Methotrexate (A)			Anti-TNFα (B)			Anti-TNFα + methotrexate (C)		
	Mean	SD	N	Mean	SD	N	Mean	SD	n	Mean	SD	n
Strand et al. 1999 [14]	0.07	0.50	101	–0.30	0.47	169						
Van de Putte et al. 2004 [15]	–0.07	0.49	110				–0.39	0.60	434			
Genovese et al. 2002 [16]				–0.19*	0.58	217	–0.24*	0.60	415			
Keystone et al. 2004 [17]				–0.24	0.52	200				–0.58	0.53	419
Weinblatt et al. 2003 [18]				–0.27	0.57	62				–0.58	0.58	209
Maini et al. 1999 [19]				–0.30	0.75	88				–0.40	0.85	340

*Change in HAQ estimated from proportion of subjects showing reduction of at least 0.5 point.
HAQ, Health Assessment Questionnaire; SD, standard deviation.

distributions for these true direct and indirect estimates as well as for the heterogeneity (σ_δ^2).

Example: Health Assessment Questionnaire (HAQ) in Rheumatoid Arthritis

In Table 1 six RCTs are presented that report change in HAQ in rheumatoid arthritis by different interventions that form a network according to Fig. 1b. In Table 2, the study-specific comparisons are presented as differences in HAQ. (For the sake of this example, we assume that all treatments are an option for a given level of disease severity, and not that one treatment is considered first-line and another second-line, for example. Also, we did not differentiate between different anti-TNFα compounds. The selected studies were not identified by means of a thorough systematic review, but were primarily selected to create an example that, in our opinion, helps understanding of MTC using continuous outcomes. Hence, the presented output of the MTC needs to be interpreted with caution because additional relevant evidence might be available) [14–19]. The different studies can be combined with a fixed-effects model or a random-effects model. Bayesian analyses were performed with WinBUGS v1.4, a Bayesian software package using Markov Chain Monte Carlo (MCMC) simulation.

Meta-Analysis of Pairwise Comparison

In Table 2, the pooled results of the three studies comparing anti-TNFα-methotrexate versus methotrexate using a frequentist and a Bayesian meta-analysis are presented. The confidence interval of the frequentist random-effects model is wider than obtained with the frequentist fixed-effect model because with the random-effects model it is assumed that differences between studies are caused by both sampling error and heterogeneity. The results of the Bayesian fixed-effect model are the same as with the frequentist approach given the use of a noninformative prior distribution. Regarding the random-effects models, the interval obtained with the Bayesian approach is wider than with the frequentist approach because uncertainty in the between-study heterogeneity is taken into consideration with the Bayesian approach. (To differentiate the uncertainty regarding the point estimate with a Bayesian approach, 95% credible intervals were used instead of 95% confidence intervals.)

Mixed Treatment Comparisons; Fixed-Effects Model

Figure 1 shows the network of direct and indirect comparisons possible given the six studies analyzed simultaneously. With a fixed-effects mixed treatment comparison, differences in the observed effect across studies focusing on the same comparisons are solely

Table 2 Change in HAQ score expressed in differences for specific pairwise head-to-head comparisons

		Difference		
Comparison	Study	Mean	SE/SD	95%CI/95%CrI
Methotrexate vs. placebo	Strand et al. 1999 [14]	–0.37	0.06	–0.49; –0.25
Anti-TNFα vs. placebo	van de Putte et al. 2004 [15]	–0.32	0.05	–0.42; –0.22
Anti-TNFα vs. methotrexate	Genovese et al. 2002 [16]	–0.06	0.05	–0.16; 0.04
Anti-TNFα + methotrexate vs. methotrexate	Keystone et al. 2004 [17]	–0.34	0.04	–0.42; –0.26
	Weinblatt et al. 2003 [18]	–0.31	0.08	–0.47; –0.15
	Maini et al. 1999 [19]	–0.10	0.09	–0.28; 0.08
Pooled results, anti-TNFα + methotrexate vs. methotrexate	Frequentist fixed-effects model	–0.30	0.03	–0.37; –0.24
	Frequentist random-effects model	–0.27	0.07	–0.40; –0.13
	Bayesian fixed-effects model*	–0.30	0.03	–0.37; –0.24
	Bayesian random-effects model*	–0.27	0.16	–0.51; 0.03

*Several analyses were performed with different normal prior distributions for the treatment effect, all distributions with a mean of 0 and different values for the variance (>10, given the boundaries of the HAQ [0–3 points]). Results were similar for all prior distributions. For the heterogeneity of the random-effects model, an inverse gamma prior distribution was used.
95%CI, 95% confidence interval of individual studies and frequentist meta-analysis; 95%CrI, 95% credible interval for the Bayesian analysis; HAQ, Health Assessment Questionnaire; SE, standard error of individual studies and frequentist meta-analysis; SD, standard deviation of posterior distribution of Bayesian meta-analysis.

Table 3 Results of MTC with a fixed-effects model and a random-effects model

	Fixed effects*			Random effects*		
Comparison	Mean	95%CrI		Mean	95%CrI	
Methotrexate vs. placebo	–0.32	–0.41	–0.23	–0.33	–0.73	0.06
Anti-TNFα vs. placebo	–0.35	–0.43	–0.27	–0.35	–0.75	0.04
Anti-TNFα + methotrexate vs. placebo	–0.63	–0.74	–0.52	–0.60	–1.06	–0.10
Anti-TNFα vs. methotrexate	–0.03	–0.11	0.06	–0.02	–0.42	0.37
Anti-TNFα + methotrexate vs. methotrexate	–0.30	–0.37	–0.24	–0.27	–0.53	0.03
Anti-TNFα + methotrexate vs. anti-TNFα	–0.27	–0.38	–0.17	–0.24	–0.70	0.26
P (placebo is best)	0%			1%		
P (methotrexate is best)	0%			1%		
P (anti-TNFα is best)	0%			8%		
P (anti-TNFα + methotrexate is best)	100%			90%		

*Several analyses were performed with different normal prior distributions for the treatment effects, all distributions with mean 0 and different values for the variance (>10, given the boundaries of the HAQ [0–3 points]). Results were similar for all prior distributions. For the heterogeneity of the random-effects model, a uniform distribution for the standard deviation was used.
95%CrI, 95% credible interval.

caused by sampling error, and σ^2_δ for these treatment-specific comparisons are assumed to be zero. There is no heterogeneity across comparisons other than due to comparison-specific effects. Hence, with a fixed-effects MTC, the $\delta_{i,b,k} \sim N(d_{P,k} - d_{P,b}, \sigma^2_\delta)$ of the random-effects model (Equation 2) translates into $\delta_{i,b,k} = d_{P,k} - d_{P,b}$, where d is true effect, P refers to the placebo intervention, k refers to the active intervention, and b is a comparator intervention. In order not to influence the observed results by the prior distribution, noninformative ("flat") normal distributions were defined for $d_{P,k}$. In Table 3, the results of direct and indirect estimates are presented.

It is interesting to compare the results of the fixed-effects model presented in Table 3 with the results presented in Table 2. The point estimates of methotrexate versus placebo, anti-TNFα versus placebo, and anti-TNFα versus methotrexate are slightly different with the MTC given the closed network of the available head-to-head comparisons [14–16]. (In Fig. 1b, the interventions correspond to nodes P, A, and B.) More important, the uncertainty intervals obtained with the fixed-effects MTC are smaller because more evidence (i.e., both direct and indirect evidence) is available. The result for anti-TNFα + methotrexate versus methotrexate is similar with the MTC as with the pairwise meta-analysis because there is no additional indirect evidence for this comparison (See Fig. 1b, node C and A). Furthermore, the MTC provides indirect estimates of anti-TNFα with methotrexate versus placebo and anti-TNFα + methotrexate versus anti-TNFα.

Given the Bayesian nature of the analysis by which the posterior distributions can be interpreted in terms of probabilities (and the use of MCMC simulation), it is also possible to calculate which intervention is best, i.e., determine for each of the intervention the probability that it results in the greatest improvement in HAQ given the level of uncertainty (see Table 3).

Mixed Treatment Comparisons; Random-Effects Model

The uncertainty intervals of the comparisons with the random-effects model are larger than those observed with the fixed-effects approach because between-study heterogeneity is also taken into account (Table 3). The model is adjusted accordingly by including a second-level statistical model that describes the distribution of true effects, as well as a prior distribution for the variance reflecting heterogeneity. Heterogeneity is estimated using all available data, i.e., all the different studies, and accordingly the heterogeneity is assumed constant between the different comparisons (see Equation 2). As a consequence, the estimated heterogeneity is also assumed applicable to comparisons where only one study is available (and for which no heterogeneity could have been assessed without other studies). It can be argued that an advantage of MTC is that in the absence of multiple studies focusing on one particular comparison, it is still possible to take heterogeneity into consideration for this comparison (given the availability of other pairwise comparisons). In essence, it is assumed that factors responsible for differences in true effects for one type of comparison apply to and have the same impact in other comparisons. Given the assumed constant heterogeneity between comparisons with random-effects MTC, the results are arguably more persuasive than results from indirect comparisons where heterogeneity is not taken into consideration.

Acceptance and Value of Indirect Comparisons or Mixed Treatment Comparisons to Support Medical Decision-Making

Both for national reimbursement and local market access (i.e., local payers, budget holders, and formulary committees) it is important to provide evidence that demonstrates the added value of a new drug versus other treatment options. RCTs comparing the alternative interventions of interest provide the most convincing evidence. However, randomized direct or

head-to-head comparisons of the comparisons of interest are often not available. As an alternative, one has to rely on indirect comparisons or MTC. In order to convince target audiences of the findings of indirect comparisons or MTC, they do need to understand the methodology, the method needs to be presented transparently, and assumptions regarding the comparability of trials need to be tested or evaluated.

In essence, MTC is a combination of two or more meta-analyses. Hence, the same methodological difficulties apply to ordinary meta-analysis where results of several studies are combined (e.g., differences in methodological quality, length of follow-up, patient characteristics, etc.) [13]. Now with MTC, the comparability of the different types of comparisons must also be considered. With an indirect comparison of trials, the value of randomization of patients assigned to interventions indirectly compared does not hold across trials. (The value of randomization only applies to the treatments compared within a study. Please note that the same applies to traditional meta-analysis.) As a result, there is the risk that patients assigned to the different trials are not comparable and the indirect estimates can be biased in a comparable way to selection bias in observational studies. If, in addition, there is also an association between these patient characteristics and treatment effect (i.e., the patient characteristics are known or expected to be effect modifiers), these patient characteristics will act as confounding variables on the observed indirect effect estimate in a comparable fashion to confounding bias in an observational study. Hence, the similarity of all the trials involved should always be inspected and inclusion of different studies or comparisons in the indirect comparison or MTC should be based on clinical grounds and preferably be evaluated with multiple analyses.

We want to make several suggestions that, in our experience, improve transparency and helps nonstatistical targeted audiences understand the methodology and findings. First, present the placebo-subtracted or relative effects of the individual studies (as illustrated in Table 2) rather then presenting findings per intervention (as in Table 1). This allows for the straightforward comparison of the "input data" with the results of the MTC analysis, thereby greatly improving transparency (e.g., compare Table 2 with Table 3). Second, it is very helpful presenting results of Bayesian meta-analyses for each type of direct comparison available before presenting results obtained with the simultaneous analysis of the MTC. Third, it is suggested to perform a series of MTC starting with two types of pairwise comparisons (i.e., one indirect comparison) in the first analysis and build upon this by including the other head-to-head comparisons in the analysis. This provides insight into whether and how the inclusion of additional pairwise comparisons influences the relative (pooled) estimates of the initial comparisons. This might be indicative of discrepancy and bias between direct and indirect estimates caused by heterogeneity across comparisons. Sensitivity analysis by excluding or including different studies can be helpful as well.

It is tempting to incorporate a covariate in the analysis in an attempt to make studies more exchangeable thereby adjusting for "meta-confounding" [20,21]. This can be considered a key issue in assessing rheumatoid arthritis treatments because patient characteristics such as disease duration and severity can have a significant impact on relative treatment effect. Nixon et al. have recently presented an analysis that simultaneously compared several rheumatoid arthritis treatments with adjustment for study-level covariates [21]. The number of studies, however, is often relatively low and the observed impact of patient characteristics can therefore be questionable [13,22,23].

An interesting question is whether an indirect comparison of RCTs provides greater evidence of the relative effect of one intervention relative to another than a nonrandomized study comparing these two interventions directly. With nonrandomized studies, there is the risk of confounding bias caused by unknown or immeasurable covariates that are associated with the outcome of interest and that are not comparably distributed over the two intervention groups. With indirect comparisons of RCTs (or any meta-analysis), differences in covariates or patient characteristics can exist between the trials in a similar fashion to between intervention groups in nonrandomized studies. Glenny et al. [13] and Deeks [24] have outlined that it can be expected that differences in patient characteristics across trials in indirect comparisons of RCTs with binary outcomes (i.e., success or failure) result in less bias than observational direct comparisons (i.e., nonrandomized controlled clinical trials or cohort studies). In contrast to an observational study, the impact of covariates on the observed number of events would affect both intervention groups in an RCT in a proportionate manner because the distribution of covariates is comparable. Hence, the effect measure of interest (i.e., odds ratio or relative risk) is not influenced by differences in patient characteristics; despite differences in patient characteristics across trials, the observed effect measures are comparable. Indirect comparisons are only biased if patient characteristics or baseline risk differences across trials are associated with differences in the observed treatment effect. In other words, indirect comparisons are biased if the factors that differ across trials are known to be effect modifiers of the (direct) treatment effect (and this heterogeneity cannot be taken satisfactory into consideration with a random-effects approach).

In this article, we focused on a Bayesian approach. Indirect comparisons can also be performed with a non-Bayesian approach. The adoption a Bayesian

approach, however, has both methodological advantages and advantages regarding interpretation and presentation, as outlined earlier [5]. In our experience, the ability to calculate the probability which intervention provides greatest outcomes resonates well with target audiences. An additional advantage is the ability to predict the observed treatment effect when a new head-to-head study will be performed. This can help answer the question of whether it is worthwhile to perform such a study. To avoid the often mentioned prejudice toward a Bayesian approach (which relates to the use of prior belief), it is advised to use a flat prior distribution by which the findings are only driven by the data.

In this introductory article, we have illustrated MTC with an example with a continuous end point. Of course dichotomous outcomes (i.e., frequency measures) can be used as well, and the same principles apply. Other published illustrations of the flexibility of a Bayesian framework include MTC with outcomes measured repeatedly over time [25], incorporation of study weights based on the internal validity of individual studies [20], the use of continuous effect sizes (e.g., standardized treatment effects) [26], and combination of RCTs and real-world studies [27]. In last example the difference in treatment effect between RCTs and real-world studies was taken into consideration in addition to the (indirect) treatment effect. This predictive distribution of the difference was used to predict the expected real-world findings based on RCT evidence for a new drug not yet available on the market.

In conclusion, MTC has been presented as an extension of traditional meta-analysis by including multiple different pairwise comparisons across a range of different interventions. The advantages of Bayesian MTC include comparison of drugs in the absence of head-to-head data; probability statements that one drug is better (e.g., more efficacious, safer) than another; and probability calculations that your drug is best (rank-order the interventions). Hence, MTC can provide very useful information for (medical) decision-making.

We would like to thank Professor Tony Ades of the University of Bristol for his useful comments on earlier versions of this article. The suggestions of anonymous reviewers are also greatly appreciated.

References

1 Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. Stat Med 2003;22:2995–3016.
2 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105–24.
3 Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897–900.
4 Ades AE, Sculpher M, Sutton AJ, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006;24:1–19.
5 Sutton AJ, Abrams KR, Jones DR, et al. Methods for Meta-Analysis in Medical Research. London: Wiley, 2000.
6 DerSimomian R, Laird N. Meta-analysis of clinical trials. Contr Clin Trials 1986;7:177–88.
7 Goodman SN. Towards evidence based medical statistics: 1. The P value fallacy. Ann Intern Med 1999;120:995–1004.
8 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluations, pp. 14–46. Chichester: John Wiley & Sons, 2004.
9 Luce BR, Claxton K. Redefining the analytical approach to pharmacoeconomics. Health Econ 1999; 8:187–9.
10 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluations, pp. 80–5. Chichester: John Wiley & Sons, 2004.
11 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluations, p. 286. Chichester: John Wiley & Sons, 2004.
12 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluations, pp. 167–74. Chichester: John Wiley & Sons, 2004.
13 Glenny AM, Altman DG, Song F, et al. International Stroke Trial Collaborative Group. Indirect comparisons of competing interventions. Health Technol Assess 2005;9:1–134.
14 Strand V, Tugwell P, Bombardier C, et al. Function and health-related quality of life: results from a randomized controlled trial of leflunomide versus methotrexate or placebo in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Invest Group Arthritis Rheum 1999;42:1870–8.
15 van de Putte LB, Atkins C, Malaise M, et al. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. Ann Rheum Dis 2004;63:508–16.
16 Genovese MC, Bathon JM, Martin RW, et al. Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. Arthr Rheum 2002;46:1443–50.
17 Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum 2004;50:1400–11.
18 Weinblatt ME, Keystone EC. Furst DE, et al. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. Arthritis Rheum 2003; 48:35–45. Erratum in: Arthritis Rheum 2003;48: 855.
19 Maini R, St Clair EW, Breedveld F, et al. Infliximab (chimeric anti-tumour necrosis factor alpha mono-

clonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet 1999;354:1932–9.
20 Jansen JP. Self-monitoring glucose in type 2 diabetes mellitus: a Bayesian meta-analysis of direct and indirect comparisons. Curr Med Res Opin 2006;22:671–81.
21 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. Stat Med 2007;26:1237–54.
22 Lambert PC, Sutton AJ, Jones DR. A comparison of summary patient-level covariates in metaregression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86–94.
23 Berlin JA, Santanna J, Schmid CH, et al. Individual patient- versus group-level data meta-regression for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002;21: 371–87.
24 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med 2002;21:1575–600.
25 Lu G, Ades AE, Sutton AJ, et al. Meta-analysis of mixed treatment comparisons at multiple follow-up times. Stat Med 2007;26:3681–99.
26 Stam WB, Gaugris S, Jansen JP. Efficacy of etoricoxib in the treatment of osteoarthritis as compared to acetaminophen, non-selective NSAIDs and cox-2 selective NSAIDs: a Bayesian meta-analysis of direct and indirect comparisons. 71st Annual Meeting of the American College of Rheumatology and 42nd annual meeting of the Association of Rheumatology, Boston, 2007.
27 Jansen JP, Crawford B. A Bayesian Approach to Predict Effectiveness of Newly Introduced Drugs in Daily Practice Based on the Relation between Efficacy and Effectiveness of Competing Interventions. Philadelphia, PA: ISPOR, 2006.