Multiple treatment comparison meta-analyses: a step forward into complexity

Edward J Mills1
Nick Bansback2,8
Isabella Ghiement3
Kristian Thorlund4
Steven Kelly5
Milo A Puhan6
James Wright7

1Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada; 2Centre for Health Evaluation and Outcomes Sciences (CHEOS), University of British Columbia, Vancouver, BC, Canada; 3Ghiement Statistical Consulting Company, Richmond, BC, Canada; 4Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; 5Department of Outcomes Research and Evidence Based Medicine, Pfizer Ltd, Walton Oaks, UK; 6Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 7Department of Oncology and Medicine, McMaster University, Hamilton, ON, Canada; 8School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

Abstract: The use of meta-analysis has become increasingly useful for clinical and policy decision making. A recent development in meta-analysis, multiple treatment comparison (MTC) meta-analysis, provides inferences on the comparative effectiveness of interventions that may have never been directly evaluated in clinical trials. This new approach may be confusing for clinicians and methodologists and raises specific challenges relevant to certain areas of medicine. This article addresses the methodological concepts of MTC meta-analysis, including issues of heterogeneity, choice of model, and adequacy of sample sizes. We address domain-specific challenges relevant to disciplines of medicine, including baseline risks of patient populations. We conclude that MTC meta-analysis is a useful tool in the context of comparative effectiveness and requires further study, as its utility and transparency will likely predict its uptake by the research and clinical community.

Keywords: network, multiple treatment comparison, mixed treatment comparison, meta-analysis

Introduction

New methods of evaluating the relative effectiveness of competing interventions may provide unique opportunities for comparative effectiveness research. As the utility of meta-analysis grows in popularity, so too does it grow in the complexity of methods and questions that it aims to answer.1 An increasingly common challenge to decision makers is to infer which of several competing interventions is likely to be most effective. This is particularly challenging when the interventions have not been directly evaluated in well-conducted randomized clinical trials (RCTs). This is referred to as an indirect comparison.2

Although meta-analysis has been used in clinical medicine since the 1980s1,4 and became commonly used in the 1990s, possibly due to the establishment of the Cochrane Collaboration,5 the methods to refine, reduce bias, and improve meta-analysis have developed slowly.1 Standard meta-analyses have typically investigated the effect of an intervention against a control, typically a placebo or another active intervention. However, such an analysis provides no inference into the relative effect of one intervention over another intervention that has not been compared directly in an RCT. The adjusted indirect comparison, first reported by Bucher et al,4 developed initial methods to make indirect comparisons and has since been extended to the multiple treatment comparison (MTC) meta-analysis, to provide more sophisticated methods for quantitatively addressing indirect comparisons of several competing interventions.
The MTC approach, based on developing methods by several investigators, is a generalization of standard pair-wise meta-analysis for drug A versus drug B trials, to data structures that include, for example, A versus B, B versus C, but no A versus C evaluation (Figure 1). The MTC requires that there is a network of pair-wise comparisons that connects each intervention to every other treatment. This approach can only be applied to connected networks of RCTs and has two important roles: i) strengthening inference concerning the relative efficacy of two treatments by including both direct and indirect comparisons of these treatments, and ii) facilitating simultaneous inference regarding all treatments, in order to simultaneously compare, and potentially rank, these treatments.7 The MTC approach yields several advantages over other indirect comparison approaches, such as those proposed by Bucher et al6 and Song et al,10 as it can deal with large numbers of indirect comparisons during a single analysis and can improve statistical power by combining both direct and indirect evidence.11,12

However, despite the sophistication and desirability of a network of compared trials,11 the MTC approach is hampered by several important concerns. First, it is a relatively new approach that is most commonly conducted in a Bayesian framework and will necessitate familiarity with Bayesian software (eg, WinBUGS [WinBUGS Project, Cambridge, UK] and R2BUGS [R2BUGS project, Columbia University, NY]). Second, the basic assumptions underlying the MTC approach are more complex than the assumptions concerning the standard pair-wise meta-analysis approach, and these are typically not well defined. Finally, interpreting MTC outputs may be misleading, as assessments of heterogeneity and statistical power are not commonly employed, resulting in a “black box” effect of the analysis. Assuming that these concerns can be overcome, MTCs are a powerful tool for decision making in medicine.

The aim of this article is to describe some of the current challenges of MTC for readers who are familiar with meta-analysis. We have chosen to illustrate the novelty and challenges of this approach in oncology medicine, although its use is not limited to any specific field of medicine. We chose oncology because it is a very well-funded area of medicine that frequently reports clinical trials and regularly has major clinical advances. We then describe in more detail some of the assumptions underlying clinical trials and interpretations. We finally discuss specific challenges that readers and the methodological community may consider if MTC is to be widely understood.

### Multiple treatment comparison meta-analyses in oncology

Despite the high profile and large number of clinical trials in oncology, there have been relatively few MTC meta-analyses conducted within this field. This is likely to be for two reasons: first, MTC is a new and sophisticated approach to meta-analysis that has yet to gain much popularity in the general academic community, most likely due to its statistical complexity; and, second, conducting MTC in cancer identifies unique, disease-specific challenges, due to both a rapidly changing therapeutic armory and progressive understanding of the disease and underlying risks. Using a systematic search of the medical literature with the search terms “(network OR multiple treatment comparison OR mixed treatment comparison)” and “meta-analysis”, up to January 2010, we identified six published MTC meta-analyses conducted in the field of oncology (Table 1). As the table displays, these analyses range from simple to very complex. In this article we discuss the challenges and some solutions to interpreting and conducting clinically relevant MTC analysis.

### Issues of methods

#### Assumptions of an MTC analysis

When conducting a standard pair-wise meta-analysis of RCTs comparing two interventions, we assume that included trials are broadly similar in terms of interventions tested and the expected direction of intervention effects across included patient populations. This similarity assumption is also required when conducting an MTC analysis aiming to compare more than two interventions.

In addition to trial similarity, effect size similarity is also of concern in a standard pair-wise meta-analysis of RCTs. The most common methods for pooling studies in a meta-analysis are the fixed- and random-effects models.
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Condition</th>
<th>Period</th>
<th>Number of trials $^*$</th>
<th>Number of trials used in MTC</th>
<th>Number of interventions</th>
<th>Median RCT sample size (IQR)</th>
<th>Minimum/maximum number of RCTs in arms (minimum/maximum patients in arms)</th>
<th>Interventions collapsed into class effects</th>
<th>Outcomes</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyrgiou et al, 2006</td>
<td>Ovarian cancer</td>
<td>1971–2006</td>
<td>198</td>
<td>120</td>
<td>60</td>
<td>103 (53–234)</td>
<td>2/1.1 (NA)</td>
<td>Yes</td>
<td>Overall survival</td>
<td>Combination of both first- and second-line treatments</td>
</tr>
<tr>
<td>Golfinopoulos et al, 2007</td>
<td>Colorectal cancer</td>
<td>1967–2007</td>
<td>242</td>
<td>137</td>
<td>40</td>
<td>152 (81–283)</td>
<td>1/9 (28/4566)</td>
<td>Yes</td>
<td>Overall survival, disease progression</td>
<td>Combination of first-, second-, and third-line treatments. Patient status improved by 6.2% per decade</td>
</tr>
<tr>
<td>Golfinopoulos et al, 2009</td>
<td>Cancers of unknown site</td>
<td>1980–2009</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>73 (49–87)</td>
<td>1/5 (17/170)</td>
<td>Yes</td>
<td>Overall survival</td>
<td>Eight trials of untreated patients, two trials unknown</td>
</tr>
<tr>
<td>Mauri et al, 2008</td>
<td>Advanced breast cancer</td>
<td>1971–2007</td>
<td>370</td>
<td>22</td>
<td>172</td>
<td>141 (87–262)</td>
<td>1/1.53 (NA)</td>
<td>Yes</td>
<td>Overall survival</td>
<td>Assessed interventions to classification of “older combinations”</td>
</tr>
<tr>
<td>Griffin et al, 2006</td>
<td>Ovarian cancer alone</td>
<td>2004</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>NA</td>
<td>2/2 (NA)</td>
<td>No</td>
<td>Overall survival, progression-free survival</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** IQR, interquartile range; NA, not applicable; RCT, randomized controlled trial.
of follow-up sufficiently similar? Are adjuvant therapies considered? Were any trials stopped early? Are doses of the intervention sufficiently similar? In many cases, differences across trials do not result in meaningful discrepancies in pooled results, and an MTC should not be any more conservative in terms of inclusion criteria than any other meta-analysis. However, without consideration of these issues, it may be impossible to determine whether and where biases are affecting results. Song et al have demonstrated that pooled indirect comparisons may, in some circumstances, provide less biased estimates of treatment effects than pooled direct (head-to-head) comparisons.

In the circumstance that both direct and indirect evidence is available in an MTC analysis (when the network of treatments in the MTC analysis includes at least one “closed” loop, where all interventions are connected in a network), it is possible to assess the consistency in treatment effects observed through the two comparisons by assessing the coherence (or incoherence) of the treatment estimates. Incoherence tells us whether the effect estimated from indirect comparisons differs from that estimated from direct comparisons. In a meta-analysis addressing smoking cessation, two head-to-head trials resulted in superiority of varenicline over bupropion (odds ratio [OR] 1.49, 95% CI 1.49–2.33, \( P < 0.001 \)), with similar effects in the indirect comparison of nine trials of varenicline and 31 trials of bupropion (OR 1.40, 95% CI 1.08–1.85, \( P < 0.01 \)), displaying reasonable coherence. Although coherence can be explored through formal hypothesis testing, resulting in a significant or nonsignificant result (yes/no), a critical assessment of why incoherence may exist may represent a more reasonable and useful strategy for assessing the impact of incoherence. Given that this type of hypothesis testing can suffer from low power as well from a lack of balance, Dias et al have proposed two sophisticated methods for checking consistency in an MTC analysis, which rely on back-calculation and node-splitting, respectively. The back-calculation method can be used for summary-level data, and the node-splitting method can be used for trial-level data. Trials identified on the basis of these methods as contributing the most to inconsistency should be excluded from the MTC analysis in order to achieve consistency.

To summarize, at least three issues of combinability need to be considered: a homogeneity assumption for each meta-analysis, a similarity assumption for individual comparisons, and a consistency assumption for the combination of evidence from different sources.
Assessing the heterogeneity of included trials in an MTC analysis

As with standard pair-wise meta-analysis, the assumption that trials include similar populations, methodological approaches, and interventions should be assessed using both visual assessment and, where possible, an assessment of statistical homogeneity. As no formal statistical tools exist for evaluating statistical heterogeneity in an MTC analysis, we suggest several possible steps here.

The first step involves assessing the statistical heterogeneity for each direct pair-wise comparison before conducting the MTC analysis. Specifically, for pair-wise comparisons where sufficient direct evidence is available, one can compute measures of between-study (statistical) heterogeneity in the context of standard pair-wise meta-analyses (eg, $I^2$). Because MTC typically assumes that statistical heterogeneity is constant between different pair-wise treatment comparisons, one could contrast these measures of heterogeneity across the relevant treatment comparisons to get a sense of whether or not this assumption is tenable. This approach is of limited use when the measures of heterogeneity are computed from a small number of studies, as these measures would likely be unreliable.

A second strategy for gauging whether or not to take into account between-study (statistical) heterogeneity when performing an MTC analysis is to fit both a fixed-effect and a random-effects MTC model to the data and then compare the resulting model fits using a measure of model fit adjusted for model complexity (eg, deviance information criterion). The fixed-effect MTC model assumes that there is no between-study heterogeneity, the random-effects MTC model would allow for between-study heterogeneity but would assume that this heterogeneity is constant across the different pair-wise treatment comparisons. If no substantial difference can be detected between the two model fits (ie, if the difference in the deviance information criterion for the two models would not exceed three points), heterogeneity may be low.

In some situations, it may be possible to relax the assumption of constant between-study heterogeneity across distinct pair-wise intervention comparisons and consider a random-effects MTC model that allows this heterogeneity to be different across these comparisons. The latter type of model could be compared against the random-effects MTC model introduced previously via the deviance information criterion to determine which assumption is more sensible for the data: constant or nonconstant between-study heterogeneity across pair-wise intervention comparisons. This would constitute the third strategy for evaluating heterogeneity in an MTC analysis.

The influence of methodological approach

With indirect and MTC comparisons in their infancy, it is not surprising that there has been little comparison between the influence of the different approaches on the estimation of relative intervention effects. The adjusted indirect comparison enables the construction of an indirect estimate of the relative effect of two interventions A and B by using information from RCTs comparing each of these interventions against a common comparator C (eg, A vs C and B vs C). The MTC, in contrast, enables the incorporation of direct comparisons of A versus B with indirect comparisons (A vs C and B vs C) to strengthen the inference of results. The MTC is a statistically more flexible approach and allows incorporation of various analyses at the same time. It is, however, often more complicated to implement and validate than adjusted indirect comparisons.

The influence of using each approach will be different depending on the data available, particularly in situations where both direct and indirect evidence is available. O’Regan et al have recently provided a comprehensive review of the influence of the different approaches in seven scenarios, each pertaining to a different number of trials with direct and indirect comparisons. Their findings demonstrate that depending on the evidence, the indirect and MTC comparisons can provide different results. In the scenario where all direct pair-wise comparisons involve a common comparator (corresponding to star-shaped networks of interventions), they found the results of these two approaches to be similar. However, where the network of trial evidence becomes more complicated, the MTC is more appropriate, as it enables the incorporation of more evidence, often reducing the variance in results. As a result, we have a starting point for selecting the most appropriate approach.

Information size

Precision and adequacy of sample size

An increasingly recognized weakness of pair-wise meta-analysis is the inadequate power or precision to confirm or refute some important intervention effects when only a few studies with a small number of events are available. A growing body of work has provided evidence that about 15%–30% of such meta-analyses are prone to yield spurious
inferences in the form of false-positive results or important overestimates of treatment effects.\textsuperscript{31-34} This is particularly problematic for MTC analyses, where several different interventions are being assessed and where authors may choose to rank the effectiveness of interventions according to probability values. For example, if three treatments, A, B, and C, are being compared in an MTC and all treatments have been compared head-to-head in a few RCTs, there is a considerable risk that one of the three pooled head-to-head treatment comparisons will yield an over- or underestimate of the comparative treatment effect due to the play of chance (imprecision). The scenario is less problematic if the indirect evidence adds sufficient precision to “correct” the spurious estimate. Unfortunately, indirect evidence is typically very imprecise. Glenny et al\textsuperscript{2} have recommended a rule of thumb that four RCTs contributing to an indirect comparison are required to approximately match the precision that a single direct (head-to-head) trial would contribute. When the number of trials included for the different treatment comparisons is imbalanced (eg, three trials compare A with B, but nine trials compare A with C), four trials are likely to be an underestimate.\textsuperscript{27} Hence, a spurious result due to imprecision within one treatment comparison is likely to contaminate the overall inferences drawn from the MTC.\textsuperscript{24} Given this circumstance, the precision of estimates may be affected by a few imprecise comparative treatment effect estimates.

**Trial-level challenges**

**The issue of crossing over**

In any clinical study, we should aim to evaluate patient-important outcomes. In many cases of disease, patients and their families are most concerned about quality of life and mortality. Oncology is one of the most funded fields of medicine, and there are many large trials published. However, using overall survival as an endpoint in cancer clinical trials may be an elusive goal. Although arguably objective and easy to measure, it is limited by requiring extended patient follow-up and being confounded by disease progression that may be unrelated to site-specific cancer.\textsuperscript{35-37} Further, as new therapies may provide effectiveness along a continuum of the disease, patient survival may be influenced as to whether they received adjacent therapies after randomization. It is common in cancer clinical trials that a failing patient crosses over (also called crossing in) to either the intervention under investigation or another salvage therapy, thus obfuscating the effect of the study drug on overall survival.\textsuperscript{36,38} As a result, most RCTs are underpowered to assess overall survival.\textsuperscript{36}

The largest challenge to employing overall survival as a study primary endpoint is that clinicians typically want to exhaust treatments in order to sustain a patient’s life, regardless of participation in a clinical trial. A patient who does not respond to the trial intervention may seek or be provided with the alternative study drug or an alternative existing or experimental treatment. However, when a patient crosses over to receive the intervention treatment, the extent of carryover effect or the contribution of a new treatment to hastening mortality cannot be known. In addition, the patient’s mortality status is frequently removed from the group they were assigned to. The following example exemplifies this concern. In a trial of tamoxifen or letrozole for first-line treatment in postmenopausal patients with endocrine-responsive advanced breast cancer, letrozole was significantly more effective than tamoxifen for response rates and time to tumor progression.\textsuperscript{39} However, no important differences in overall survival existed between the study arms. But when a sensitivity analysis censored patients who crossed over to the other study drug, the results indicated that the use of letrozole was associated with a survival benefit.\textsuperscript{40}

Although crossing over in cancer clinical trials is common, methods with which to deal with overall survival in the analysis of individual RCTs and meta-analysis are not well established.\textsuperscript{35,41} A clinical trial may consider a crossover as a failure of treatment (included in a progression of disease analysis) or may exclude a patient from analysis, as a result losing the benefits of the intent-to-treat principle.\textsuperscript{42} Despite the fact that these are RCTs, such crossing over may not occur randomly across arms, as one treatment may be genuinely more beneficial than the other. In meta-analysis, the inclusion of crossed-over patients represents an important challenge. Principally, the benefits of randomization are lost on the patients who crossed over. If this is a small number of patients in a moderate to large trial, the effects may be small. However, if this is a large number of patients, the overall survival analysis could be seriously biased in a meta-analysis. Further, the effects of the patient’s previous treatment (period A) may have a differing effect on response to the second treatment (period B), akin to the carryover concept in a crossover trial.\textsuperscript{43} It is possible that the survival endpoint remains valid if the trialists aim to determine what regimen or strategy (ie, what drug to start with) of first-line therapies provides overall survival benefits. However, this becomes increasingly confusing as patients switch from drug A to drug B or drug C/D. A common approach is to censor these patients from the survival analysis. However, this will bias a survival assessment, as the true time
point for progression of that patient is replaced by the time points of patients who were free of disease when that patient was censored, favoring the less effective intervention.\textsuperscript{41,44} A meta-analyst aiming to calculate a relative risk based on event rates would typically not have this information from a publication, and pooling hazard ratios may improperly ignore this issue. It seems appropriate that meta-analysis reports whether crossed-over patients could create bias, just as large loss to follow-up may.\textsuperscript{45} Some study designs and opinions do not permit a crossover of patients to the intervention arm if a patient fails first-line therapy.\textsuperscript{41,46} However, this creates dilemmas for clinicians and trialists, especially regarding patients with a poor survival prognosis or those for whom alternative experimental drugs may exist.

Issues pertaining to the use of patients’ characteristics

The requirement of evidence to inform decisions regarding health technologies has been a leading motivation in the development of indirect and MTC approaches. It is therefore not surprising that pragmatic decision makers have sought to identify characteristics or subgroups of patients in whom technologies may have greater benefit of improved safety.

As described previously, conventional meta-analysis requires included trials to be sufficiently homogeneous, but MTC approaches have an additional requirement that trials are similar for moderators of relative treatment effect.\textsuperscript{25} Song et al\textsuperscript{10} state that the average relative effect estimated in placebo-controlled trials of one therapy should be generalizable to patients in placebo-controlled trials of an alternative therapy, and vice versa. The role of meta-regression and subgroup analysis in circumstances where heterogeneity between sources of evidence is present can play a greater role than in conventional meta-analysis by enabling comparisons between “similar” groups.\textsuperscript{40} However, the pitfalls of subgroup analysis, meta-regression, and meta-analysis acknowledged in the conventional meta-analysis literature\textsuperscript{47,48} will carry through to indirect and MTC approaches, and, as a consequence, authors are recommended to use predefined characteristics and interpret the results cautiously.

We did not find examples of meta-regression in MTC for oncology. However, an example of meta-regression from one of the authors examined the effectiveness of biologic agents in rheumatoid arthritis.\textsuperscript{49} In this example, baseline disease duration, a characteristic known to be related to the effectiveness of biologics versus standard treatments, was included in the MTC in the form of a meta-regression. Although the inclusion of the meta-regression did not change the significance of the results, it did modify the odds to suggest that the three tumor necrosis factor antagonists etanercept, infliximab, and adalimumab were very similar, a result that has long been suspected in the wider literature (Figure 2).

Adjustments for baseline characteristics known to be predictive of survival and progression-free survival, such as the level of circulating tumor cells in metastatic breast cancer, are obvious candidates for the use of meta-regression in MTC.\textsuperscript{50} Another use is the adjustment of doses where different trials have used various dosages of the same treatment. Where a dose-response relationship exists, meta-regression within an MTC can potentially overcome issues where RCTs have used potentially inadequate dosages.\textsuperscript{8}

Using Bayesian priors to deal with challenging situations

Although indirect and MTC comparisons have been conducted from the frequentist approach,\textsuperscript{9,51} the Bayesian approach has
been increasingly promoted as a more flexible method for incorporating “knowledge” into the results.\textsuperscript{52} In the literature, it is common to see the use of “noninformative” priors (essentially assuming complete uncertainty), and the Bayesian approach is used for other purposes, such as enabling treatments to be ranked so that the probability of any one treatment being “best” can be obtained and used in economic evaluation.\textsuperscript{53} However, the use of informed priors provides a basis for dealing with situations where there is inconsistency in results from different trials, trials included have not reported results, and where a sparse network makes estimating heterogeneity challenging.

Bayesian statistics are poorly understood by clinicians and many statisticians. The use of prior knowledge can change the interpretation of a study or can provide increasing confidence in the results. For example, a recent analysis of pivotal trials of adjuvant chemotherapy in nonsmall cell lung cancer produced positive results that were at odds with previous studies.\textsuperscript{54} To deal with the problem of how to interpret the results of the new studies, the authors examined different priors: assume adjuvant chemotherapy has no benefit, use skeptical priors (which essentially add pessimism to the positive results of the new findings), and use expert beliefs. The results consistently found that by combining the evidence, adjuvant chemotherapy was effective in all scenarios, giving greater confidence in the results.

**Discussion**

This paper aims to inform readers who may be exposed to reading MTC analyses as manuscripts or contributing to them as investigators. Although we have found only a few examples of MTC in the oncology literature at present, with further consideration of some of the issues presented, it is anticipated that there will be more widespread use of MTC in the future. Although we believe that MTC has an important role to play, we are cautious that readers and users of this methodology are clear of both the assumptions and the benefits over conventional meta-analysis.

Two issues related to the way MTCs are interpreted are worthy of further consideration. The first regards the level of evidence an MTC can provide. The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group, that proposes methods for evaluating the strength of inference of observational studies, RCTs, and meta-analysis, has considered MTC.\textsuperscript{55,56} In an upcoming paper, the working group considers an MTC analysis as a weaker comparison than head-to-head meta-analysis, as differences in study methods may contribute to differing estimates of treatment effect and result in probably larger treatment effects. The basis for this consideration is that because the similarity assumption required for MTC is always in some doubt, the level of quality of the evidence is downgraded. The working group recommends grading down further if the similarity assumption is unconvincing.

A second issue relates to the way results are translated. It is increasingly common for authors of MTC to augment their results based on estimates of risk to the use of probabilities, enabling the ranking of treatments for a particular outcome. This approach has obvious appeal for lay readers of an analysis, providing simple knowledge translation rather than statistical outcomes. The concern is that such rankings overinterpret results that authors acknowledge to contain errors and biases occurring from the statistical approach and synthesis of evidence.

**Conclusion**

In conclusion, the appropriate use and interpretation of MTC will provide the methodological community and wider medical community with information to guide decisions. Further methodological research in MTC is required to overcome concerns related to their transparency and interpretability.

**Disclosure**

This study was supported by Pfizer Canada Inc. and Pfizer UK Ltd. The company had no role in the conduct, writing or interpretation of this paper, or choice to publish. SK is an employee of Pfizer UK Ltd and participated as an author.

**References**


