METHODOLOGY

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

Edward J Mills[1]
Nick Bansback[2,8]
Isabella Ghement[3]
Kristian Thorlund[4]
Steven Kelly[5]
Milo A Puhan[6]
James Wright[7]

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

Correspondence: Edward Mills
Faculty of Health Sciences, University of Ottawa, 43 Templeton St, Ottawa, Ontario K1H 8M5, Canada
Tel +1 778 317 8530
Fax +1 613 562 5149
Email edward.mills@uottawa.ca

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
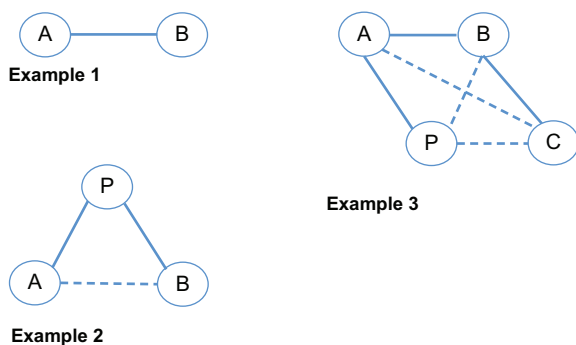
Introduction

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 1980s[3,4] 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,[6] 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,[7–9] 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 al[6] 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,[13] 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 MTC methods 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.

Figure 1 Direct and indirect comparisons. Circled letters represent trial arms of drug A (A), drug B (B), drug C (C), and placebo (P). Flat lines represent direct trials, dotted lines represent indirect comparisons. Example 1: Direct comparison of drug A and drug B. Example 2: Adjusted indirect comparison where drug A and drug B have not been evaluated directly. Example 3: A multiple treatment comparison where drug A and drug C, drug B and placebo, and drug C and placebo have not been evaluated directly.

Table 1 Characteristics of published multiple treatment comparison (MTC) analysis in oncology

Author, year	Condition	Period	Number of trials*	Number of interventions	Number of trials used in MTC	Median RCT sample size (IQR)	Minimum/maximum number of RCTs in arms (minimum/maximum patients in arms)	Interventions collapsed into class effects	Outcomes	Comments
Kyrgiou et al,[18] 2006	Ovarian cancer	1971–2006	198	120	60	103 (53–234)	2/11 (NA)	Yes	Overall survival	Combination of both first- and second-line treatments
Golfinopoulos et al,[57] 2007	Colorectal cancer	1967–2007	242	137	40	152 (81–283)	1/9 (28/4566)	Yes	Overall survival, disease progression	Combination of first-, second-, and third-line treatments. Patient status improved by 6.2% per decade
Golfinopoulos et al,[17] 2009	Cancers of unknown site	1980–2009	10	10	10	73 (49–87)	1/5 (17/170)	Yes	Overall survival	Eight trials of untreated patients, two trials unknown
Mauri et al,[19] 2008	Advanced breast cancer	1971–2007	370	22	172	141 (87–262)	1/153 (NA)	Yes	Overall survival	Assessed interventions to classification of "older combinations"
Hawkins et al,[58] 2009	Nonsmall cell lung cancer	2000–2007	6	4	6	731 (651–1257)	1/4 (104/1692)	No	Overall survival	Only second-line treatments
Griffin et al,[59] 2006	Ovarian cancer	2004 alone	3	3	3	NA	2/2 (NA)	No	Overall survival, progression-free survival	

Abbreviations: IQR, interquartile range; NA, not applicable; RCT, randomized controlled trial.

An assumption of fixed effects is that the effect size is the same across studies and the observed variability results from chance alone. This is commonly referred to as the statistical homogeneity assumption.[14] Although several common interpretations exist,[15] usually, an assumption of random effects is that there may be genuine diversity in the results of various trials owing to differences between these trials in study and patient characteristics, so a between-study variance component is incorporated into the calculations to capture this diversity (commonly referred to as statistical heterogeneity). When there is no observed between-study heterogeneity, the fixed- and random-effects approaches coincide. Otherwise, the random-effects approach provides wider confidence intervals (CIs) for the relative intervention effect and is thus considered more clinically conservative.[16] Although a random-effects approach explicitly models between-study heterogeneity, it does not explain it. Attempts to explain the between-study heterogeneity would have to rely on meta-regression, a technique that allows one to study whether or not relevant trial-level covariates act as modifiers of the relative intervention effect.

In an MTC analysis, the assumptions made about statistical heterogeneity are of prime importance, as assessments of heterogeneity are not yet established and conventional measurements of heterogeneity do not exist (ie, τ^2 or I^2). Note that in this setting, one would need to consider the issue of statistical heterogeneity in relation to each possible pair-wise comparison of interventions.

Clinical heterogeneity may induce statistical heterogeneity. In a recent MTC analysis involving 60 RCTs of cancers of unknown sites published between 1971 and 2006,[17] the populations range from poor-risk patients who had received previous therapy to favorable- and intermediate-risk patients as time progressed (a 6% performance status improvement per decade). Therefore, clinicians will need to determine for themselves whether the underlying risk of events is sufficiently similar across time. This appears to be an issue across differing diseases, as an MTC examining breast cancer,[19] including trials from 1971 to 2007, demonstrated changing disease risks over time. This possibly reflects the cointerventions that improve outcomes for patients and that have been used for breast cancer since 1971.[20]

Methodological heterogeneity may also induce statistical heterogeneity. Therefore, in addition to intervention and clinical similarities, the MTC analysis requires an assumption of similarity on methodological grounds. In particular, are trials measuring a similar estimate of effect? Is the length of follow-up sufficiently similar? Are adjuvant therapies considered? Were any trials stopped early?[21–23] 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.[24] However, without consideration of these issues, it may be impossible to determine whether and where biases are affecting results. Song et al[25] 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.[7,9] In a meta-analysis addressing smoking cessation,[26] 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.[27,28] Given that this type of hypothesis testing can suffer from low power as well from a lack of balance,[27] Dias et al[29] 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.[29]

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).[29] Although 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.[29] 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).[6] 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.[7] 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[11] 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.[30] 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.[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[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 unbalanced (eg, three trials compare A with B, but nine trials compare A with C), four trials are likely to be an underestimate.[27] Hence, a spurious result due to imprecision within one treatment comparison is likely to contaminate the overall inferences drawn from the MTC.[34] 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.[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 adjunct 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.[36,38] As a result, most RCTs are underpowered to assess overall survival.[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.[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.[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.[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.[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.[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.[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.[45] Some study designs and opinions do not permit a crossover of patients to the intervention arm if a patient fails first-line therapy.[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.[25] Song et al[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.[10] However, the pitfalls of subgroup analysis, meta-regression, and meta-analysis acknowledged in the conventional meta-analysis literature[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.[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.[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.[8]

Using Bayesian priors to deal with challenging situations

Although indirect and MTC comparisons have been conducted from the frequentist approach,[9,51] the Bayesian approach has
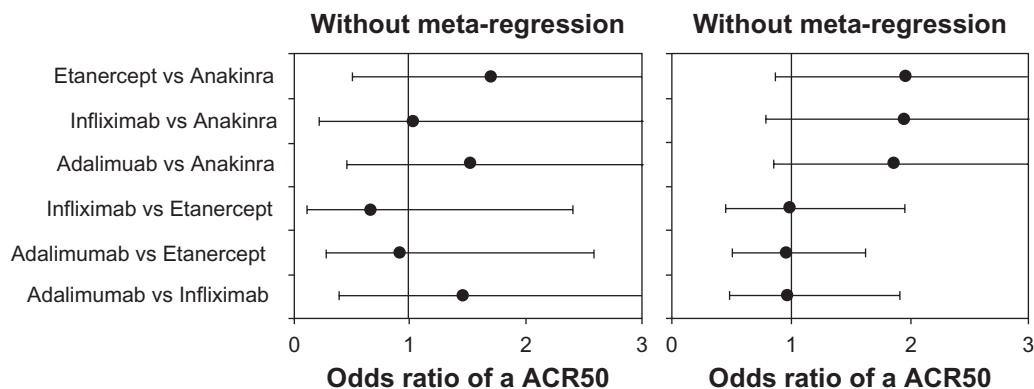
Figure 2 Use of baseline adjustments versus crude analysis in rheumatoid arthritis multiple treatment comparison meta-analysis.[49] Legend: American College of Rheumatology 50th percentile of response (ACR50).

been increasingly promoted as a more flexible method for incorporating "knowledge" into the results.[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.[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.[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.[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

References

1. Sutton AJ, Higgins JP. Recent developments in meta-analysis. *Stat Med*. 2008;27:625–650.
2. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005;9:1–134, iii–iv.
3. Stampfer MJ, Goldhaber SZ, Yusuf S, et al. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. *N Engl J Med*. 1982;307:1180–1182.
4. Yusuf S, Peto R, Lewis J, et al. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*. 1985;27:335–371.
5. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci*. 1993;703:156–163; discussion 63–65.
6. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50:683–691.
7. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23:3105–3124.
8. Cooper NJ, Sutton AJ, Morris D, et al. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28:1861–1881.

9. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21:2313–2324.

10. Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ*. 2009;338:b1147.

11. O'Regan C, Ghement I, Eyawo O, et al. Incorporating multiple interventions in meta-analysis: an evaluation of the mixed treatment comparison with the adjusted indirect comparison. *Trials*. 2009;10:86.

12. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15:2733–2749.

13. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. *Ann Intern Med*. 2008;148:544–553.

14. Borenstein M, Hedges L, Higgins JP, Rothstein H. *Introduction to Meta-analysis*. Chichester: John Wiley & Sons, 2009.

15. Sanchez-Meca J, Marin-Martinez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychol Methods*. 2008;13:31–48.

16. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–188.

17. Golfinopoulos V, Pentheroudakis G, Salanti G, et al. Comparative survival with diverse chemotherapy regimens for cancer of unknown primary site: multiple-treatments meta-analysis. *Cancer Treat Rev*. 2009;35:570–573.

18. Kyrgiou M, Salanti G, Pavlidis N, et al. Survival benefits with diverse chemotherapy regimens for ovarian cancer: meta-analysis of multiple treatments. *J Natl Cancer Inst*. 2006;98:1655–1663.

19. Mauri D, Polyzos NP, Salanti G, et al. Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *J Natl Cancer Inst*. 2008;100:1780–1791.

20. Cardoso F, Bedard PL, Winer EP, et al. International guidelines for management of metastatic breast cancer: combination vs sequential single-agent chemotherapy. *J Natl Cancer Inst*. 2009;101:1174–1181.

21. Wilcox RA, Djulbegovic B, Moffitt HL, et al. Randomized trials in oncology stopped early for benefit. *J Clin Oncol*. 2008;26:18–19.

22. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294:2203–2209.

23. Bassler D, Ferreira-Gonzalez I, Briel M, et al. Systematic reviewers neglect bias that results from trials stopped early for benefit. *J Clin Epidemiol*. 2007;60:869–873.

24. Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. *Lancet*. 2006;368:1470–1472.

25. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol*. 2008;61:455–463.

26. Mills EJ, Wu P, Spurden D, et al. Efficacy of pharmacotherapies for short-term smoking abstinence: a systematic review and meta-analysis. *Harm Reduct J*. 2009;6:25.

27. Mills EJ, Ghement I, O'Regan C, Thorlund K. Estimating the power of indirect comparisons: a simulation study. *PLOS One*. 2011;6:e16237.

28. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ*. 2009;181:488–493.

29. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29: 932–944.

30. Thorlund K, Anema A, Mills E. Interpreting meta-analysis according to the adequacy of sample size. An example using isoniazid chemoprophylaxis for tuberculosis in purified protein derivative negative HIV-infected individuals. *Clinical Epidemiology*. 2010;2:1–10.

31. Thorlund K, Devereaux PJ, Wetterslev J, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *Int J Epidemiol*. 2009;38:276–286.

32. Trikalinos TA, Churchill R, Ferri M, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol*. 2004;57:1124–1130.

33. Berkey CS, Mosteller F, Lau J, Antman EM. Uncertainty of the time of first significance in random effects cumulative meta-analysis. *Control Clin Trials*. 1996;17:357–371.

34. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive: trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *Int J Epidemiol*. 2009;38:287–298.

35. Saad ED, Katz A, Hoff PM, Buyse M. Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. *Ann Oncol*. 2009;21:7–12.

36. Di Leo A, Bleiberg H, Buyse M. Overall survival is not a realistic end point for clinical trials of new drugs in advanced solid tumors: a critical assessment based on recently reported phase III trials in colorectal and breast cancer. *J Clin Oncol*. 2003;21:2045–2047.

37. Di Leo A, Buyse M, Bleiberg H. Is overall survival a realistic primary end point in advanced colorectal cancer studies? A critical assessment based on four clinical trials comparing fluorouracil plus leucovorin with the same treatment combined either with oxaliplatin or with CPT-11. *Ann Oncol*. 2004;15:545–549.

38. Thomsen IM, Crawford ED. Adjuvant radiotherapy for pathological T3N0M0 prostate cancer. *J Urol*. 2009;181:956.

39. Mouridsen H, Gershanovich M, Sun Y, et al. Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a phase III study of the International Letrozole Breast Cancer Group. *J Clin Oncol*. 2001;19:2596–2606.

40. Mouridsen H, Gershanovich M, Sun Y, et al. Phase III study of letrozole versus tamoxifen as first-line therapy of advanced breast cancer in postmenopausal women: analysis of survival and update of efficacy from the International Letrozole Breast Cancer Group. *J Clin Oncol*. 2003;21:2101–2109.

41. Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol*. 2009;27:2874–2880.

42. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325:652–654.

43. Mills EJ, Chan AW, Wu P, et al. Design, analysis, and presentation of crossover trials. *Trials*. 2009;10:27.

44. Carroll KJ. Analysis of progression-free survival in oncology trials: some common statistical issues. *Pharm Stat*. 2007;6:99–113.

45. Akl EA, Briel M, You JJ, et al. LOST to follow-up Information in Trials (LOST-IT): a protocol on the potential impact. *Trials*. 2009;10:40.

46. Tuma R. Progression-free survival remains debatable endpoint in cancer trials. *J Natl Cancer Inst*. 2009;101:1439–1441.

47. Davey-Smith G, Egger M, Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials. *Systematic Reviews in Health Care: Meta-analysis in Context*. 2nd ed. London, England: BMJ Publishing Group. 2001:143–156.

48. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21:1559–1573.

49. 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;26:1237–1254.

50. Cristofanilli M, Hayes DF, Budd GT, et al. Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. *J Clin Oncol*. 2005;23:1420–1430.

51. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50:683–691.

52. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technol Assess*. 2000;4:1–130.

53. Ades AE, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics*. 2006;24:1–19.

54. Miksad RA, Gonen M, Lynch TJ, Roberts TG Jr. Interpreting trial results in light of conflicting evidence: a Bayesian analysis of adjuvant chemotherapy for non-small-cell lung cancer. *J Clin Oncol*. 2009;27: 2245–2252.

55. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490.

56. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924–926.

57. Golfinopoulos V, Salanti G, Pavlidis N, Ioannidis JP. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: a meta-analysis. *Lancet Oncol*. 2007;8:898–911.

58. Hawkins N, Scott DA, Woods BS, Thatcher N. No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. *Value Health*. 2009;12: 996–1003.

59. Griffin S, Bojke L, Main C, Palmer S. Incorporating direct and indirect evidence using Bayesian methods: an applied case study in ovarian cancer. *Value Health*. 2006;9:123–131.