GUIDELINE

COMPARATORS & COMPARISONS:

Direct and indirect comparisons

Adapted version (2015) based on

“COMPARATORS & COMPARISONS: Direct and indirect comparisons” - February 2013
The primary objective of EUnetHTA JA1 WP5 methodology guidelines was to focus on methodological challenges that are encountered by HTA assessors while performing a rapid relative effectiveness assessment (REA) of pharmaceuticals.

The guideline “COMPARATORS & COMPARISONS: Direct and indirect comparisons” has been elaborated during JA1 by experts from HIQA, reviewed and validated by HAS and all members of WP5 of the EUnetHTA network, the whole process was coordinated by HAS.

During Joint Action 2 the wording in this document has been revised by WP7 in order to extend the scope of the text and recommendations from pharmaceuticals only to the assessment of all health technologies. Content and recommendations remained unchanged.

This guideline represents a consolidated view of non-binding recommendations of EUnetHTA network members and in no case an official opinion of the participating institutions or individuals.
Table of contents

1. Acronyms – Abbreviations ................................................................. 4
2. Summary and recommendations .................................................... 5
3. Summary .............................................................................................. 5
4. Recommendations .................................................................................. 5
5. 1. Introduction .................................................................................... 7
   1.1. Definitions and general information .............................................. 7
   1.2. Context ............................................................................................ 7
   1.3. Scope/Objective(s) of the guideline ................................................ 7
   1.4. Relevant EunetHTA documents ....................................................... 7
6. 2. Summary of the analysed literature ................................................ 8
   2.1. Networks of evidence ................................................................. 8
   2.2. Direct comparisons ........................................................................ 9
   2.3. Indirect comparisons .................................................................... 10
   2.4. Mixed treatment comparisons ..................................................... 11
   2.5. Critical comparisons of techniques .............................................. 12
   2.6. Additional considerations ............................................................ 13
7. 3. Discussion ......................................................................................... 16
8. 4. Conclusion ......................................................................................... 18
9. Annexe 1. Methods of documentation and selection criteria ............... 19
10. 3.1. Sources of information .......................................................... 19
11. 3.2. Bibliographic search strategy ..................................................... 20
12. 3.3. Selection criteria ................................................................. 20
13. Annexe 2. Bibliography ....................................................................... 21
Acronyms – Abbreviations

1. IPD – individual patient data
2. MTC – mixed treatment comparison
3. NICE – National Institute for Health and Care Excellence
4. RCT – randomised controlled trial
5. REA – relative effectiveness assessment
Summary and recommendations

Summary

To make the best use of available evidence on the efficacy of a treatment, it is common to combine results from several randomised controlled trials (RCTs) in a meta-analysis. This guideline focuses on the methods available for treatment comparisons. Their strengths and limitations are discussed and recommendations are provided in order to support Relative Effectiveness Assessors in their activity. The planning stages of a systematic review are not covered here.

Direct comparisons combine the results of multiple head-to-head trials to generate a pooled estimate of the relative effectiveness of the two treatments using well described meta-analysis methods. These methods can be broadly split into frequentist and Bayesian approaches. When there is no prior expectation about what distributions parameters will take, Bayesian methods generate results that are approximately equivalent to frequentist methods. Bayesian methods can offer more sophisticated techniques than frequentist approaches to incorporate data from other sources such as observational studies. Sometimes insufficient data are available to reliably estimate the relative effectiveness of the two treatments or there may be a need to compare more than two treatments simultaneously in which case multiple treatment comparison methods are required.

Multiple treatment comparisons can be used to infer the relative effectiveness of two treatments in the absence of direct head-to-head evidence or through the combination direct and indirect evidence. A number of methods of multiple treatment comparison are available including Bucher’s method of adjusted indirect comparison, Lumley’s method of network meta-analysis, and Bayesian mixed treatment comparison (MTC).

The comparisons available for analysis form a network of evidence. Depending on the method of multiple treatment comparison used, there may be restrictions on the type of evidence networks that can be analysed. Some methods of multiple treatment comparison produce a measure of inconsistency that can highlight where direct and indirect evidence is divergent. The application of multiple treatment comparison has become increasingly common since first being adopted in the last 1990’s. Although Bucher’s method of adjusted indirect comparison was initially the most popular methodology, its use has been surpassed by Bayesian MTC in recent years.

There are many issues that must be taken into account when conducting treatment comparisons. It is assumed that the relative effectiveness of a treatment is the same across all studies included in a meta-analysis. Heterogeneity across studies should be measured, reported and, if possible, explained. Study-level covariates can be used to partially account for heterogeneity using a meta-regression approach. If the observed relative effectiveness varies substantially across studies, then combining the results may not be appropriate. If there is substantial heterogeneity, then a random effects model may be preferable to a fixed effect model. Sources of bias such as publication bias may impact on results and attempts should be made to reduce the potential for bias to impact on the analysis. It is useful to assess whether there are influential or outlying observations and to test the impact of these studies using sensitivity analysis. The results of a meta-analysis using Bayesian methods can be sensitive to the choice of priors and sensitivity analysis should be considered to test alternative formulations of priors.

Direct comparison is generally thought to provide the best evidence of relative effectiveness and is in general recommended. However, the use of multiple treatment comparisons allows the consideration of a larger evidence base. When both direct and indirect evidence are available, it may be pragmatic to investigate both first separately and then pool the results. Although Bucher’s method of adjusted indirect comparison is the most computationally straightforward of the multiple treatment methods, Bayesian MTC can be used to analyse very complex networks and can incorporate meta-regression to include study-level covariates. The choice of methodology is ultimately context specific and should be appropriate to the data available.

Recommendations

1. A systematic literature search is a pre-requisite to conducting a direct or indirect comparison. This must be documented according to existing guidelines. A comprehensive review will maximise the evidence base.
2. The application of direct or indirect comparisons relies on the assumption that only comparable studies should be combined. Studies that differ substantially in one or more key characteristics (e.g. participants, interventions, outcomes measured) should not be combined. Methods such as meta-regression that account for study level covariates may be used, although the power to detect effect differences is reduced.

3. The choice between a fixed and random effects model should be made based on the characteristics of the studies being analysed. Heterogeneity should be assessed and a clear justification for the choice of model must be provided. Where a random effects model is preferred, results from a fixed effect model can still be presented in special situations (e.g. few studies and where sample sizes vary considerably).

4. Potential sources of bias should be investigated, if identified (e.g. funnel plots for publication bias).

5. Attention should be given to determining the presence of outliers or influential observations that may have an undue impact on results. Sensitivity analysis should be used to determine the impact of those influential or outlying studies.

6. The choice between direct and indirect comparison is context specific and dependent on the question posed as well as the different evidence available. Where sufficient good quality head-to-head studies are available, direct comparisons are preferred as the level of evidence is high. Should substantial indirect evidence be available, then it can act to validate the direct evidence. When there is limited head-to-head evidence or more than two treatments are being considered simultaneously, the use of indirect methods may be helpful.

7. If both direct and indirect evidence are available, they can be evaluated separately. Attempts should be made to explain any discrepancies between the results obtained in terms of study characteristics. In the event of indirect results differing substantially from the direct evidence, there must be close scrutiny of the data, although there is no consensus in the literature on how to deal with these discrepancies.

8. Only adjusted methods of indirect comparison that maintain randomisation should be used. Unadjusted indirect comparisons are not recommended.

9. The choice of indirect comparison method relies on the network of available evidence. Preference should be given to the most transparent method available (i.e. one should favour Bucher’s method of adjusted indirect comparison over MTC if the data permit its usage and the appropriate assumptions are satisfied).

10. An indirect comparison should only be carried out if underlying data from comparable studies are homogeneous and consistent, otherwise the results will not be reliable.

11. The assumptions made for indirect comparisons must be explicitly stated. Particular attention should be given to the homogeneity, similarity and consistency assumptions. A general assumption of indirect comparisons is that the relative effectiveness of a treatment is the same across all studies included in a meta-analysis.

12. When Bayesian methods are applied, the choice of the prior distributions should be justified and documented. In the case of non-informative priors, where alternative specifications exist they should be applied as part of a sensitivity analysis. When informative priors are used, the source of that data must be clearly documented and consideration given to testing the impact of using a non-informative prior in place of the informative prior.

13. The complexity of a model is not a measure of its accuracy or utility and preference is for the most parsimonious model whose assumptions can be justified.
1. Introduction

1.1. Definitions and general information

Direct comparison – the combination of multiple head-to-head trials to generate a pooled estimate of the relative effectiveness of the two treatments

Indirect comparison – the estimation of the relative effectiveness of two or more treatments in the absence of any head-to-head trials

Multiple treatment comparison – the estimation of the relative effectiveness of three or more treatments

Mixed treatment comparison – the simultaneous estimation of the relative effectiveness of three or more treatments using a combination of direct and indirect evidence

To compare two or more treatments, meta-analytic techniques are generally used to combine the results of multiple studies in an attempt to provide the best evidence base. A meta-analysis is the formal evaluation of the quantitative evidence from two or more studies addressing the same question. This most commonly involves the statistical combination of summary statistics from the various studies, but the term is sometimes also used to refer to the combination of raw data. Direct comparisons enable evidence synthesis based on multiple head-to-head trials. Where direct head-to-head evidence is lacking, indirect evidence can be used to supplement the relative effectiveness data from the direct comparisons available.

1.2. Context

1.2.1. Problem statement

“What methods for direct and indirect comparisons are used; are more advanced methods like Bayesian mixed treatment comparison used?”

1.2.2. Discussion (on the problem statement)

A variety of methods are available to conduct direct and indirect comparisons. Each method rests on a number of assumptions about the data used. The choice of method and associated assumptions may impact on the findings of a relative effectiveness assessment (REA).

1.3. Scope/Objective(s) of the guideline

This guideline is intended to describe the main methods of direct, indirect and mixed treatment comparison available in terms of the types of relationship they can model and the assumptions inherent in them. The guidelines are not intended to give a detailed understanding of the meta-analytic techniques described, but rather to explain the main strengths and weaknesses of the methodologies. The guideline discusses some common issues in meta-analysis that must be considered when interpreting results. Finally, the guideline provides a set of recommendations regarding the use of direct and indirect comparisons in a relative effectiveness assessment (REA).

1.4. Relevant EunetHTA documents

Not applicable
2. Summary of the analysed literature

The relative effectiveness of two treatments is generally measured in a randomised controlled trial (RCT) setting. To make the best use of the available evidence base it is common to combine results from several trials in a meta-analysis. The process of combining RCT results involves some form of weighted average usually related to precision of the treatment effect estimation. In cases where two treatments are being compared, there is sometimes insufficient data available to reliably estimate the relative effectiveness of the two treatments in which case it may be possible to estimate the relative effectiveness using an indirect comparison. (1) For the purposes of this document it is presumed that sufficient data of acceptable quality are available to justify a meta-analysis. It is also assumed that the collection of the data contributing to the comparisons involved an exhaustive search of published and unpublished trials and a rigorous selection process based on the methodological quality of the trials.

2.1. Networks of evidence

The studies available for analysis form a network of evidence. Depending on the method of comparison used, there may be restrictions on the type of networks that can be analysed. For direct comparisons only a standard pair-wise meta-analysis can be used. Networks can include: a star pattern in which two or more treatments have a common comparator (e.g. A-B, C-B, D-B); a ladder where treatment comparisons appear in a sequence (e.g. A-B, B-C, C-D); a closed loop in which there is both direct and indirect evidence (e.g. A-B, A-C, C-B); or a complex combination of patterns such as a closed loop with a star (see Figure 1 for examples). (5)

Figure 1. Types of evidence network

Direct, or standard pair-wise, meta-analysis

Indirect: simple star

Indirect: ladder

Indirect: star with pair-wise contrasts

Indirect: closed loop

Direct and indirect: extended network

Direct and indirect: multi-loop

Direct and indirect: network with a closed loop
2.2. Direct comparisons

Direct comparisons involve a meta-analysis combining the results of multiple trials that all compare the treatment of interest to the same comparator (e.g. placebo). Meta-analysis involves the computation of a summary statistic for each trial followed by the combination of these studies into a weighted average.\(^2\)

Outcomes can be expressed as binary or continuous data. The summary statistic can be odds ratios, risk ratios, risk differences, hazard ratios, differences in means or effect sizes (standardised mean differences).

The same summary statistic must be computed for each study. Standard meta-analytic techniques are applied for direct comparisons. The primary decision in direct comparisons relates to the choice between fixed and random effects meta-analysis (see 3.6.2 below). Approaches to direct comparisons meta-analysis can be sub-divided into two methodologies: frequentist and Bayesian. The former are standard for direct comparisons primarily due to the ease of application and the variety of software packages available to apply them.

2.2.1. Frequentist approach

The frequentist methods available for direct comparison meta-analysis are divided into fixed and random effects methods. Fixed effect models include inverse variance, Mantel-Haenszel and Peto methods. Inverse variance methods can be used to pool binary or continuous data and weights are proportional to the inverse squared standard errors of the studies. Inverse variance methods are less reliable when data are sparse.

The Mantel-Haenszel method provides more robust weighting when data are sparse and gives similar weights to inverse variance methods when data are not sparse, but may lack credibility when there are low numbers of events. The Mantel-Haenszel method can be applied to both binary and continuous data. The Peto method is used for odds ratios and can be extended for pooling time-to-event data. Peto’s method has been shown to fail when treatment effects are very large and when the sizes of the trial arms are very unbalanced. The Peto method performs well when event rates are very low. Fixed effect methods tend to give small weights to small studies and large weights to large studies.

The most common random effects model is DerSimonian and Laird. A heterogeneity statistic is incorporated into the computation. When the heterogeneity statistic is small then the weights are equivalent to those given by the inverse variance method. With increasing heterogeneity the weights for each studies become increasingly similar to each other. The consequences of this property are that small studies get weights more similar to larger studies and the confidence intervals around the pooled estimate become wider than would be observed in the fixed effect meta-analysis. The differences between fixed and random effects are discussed further in section 2.6.2.

For certain outcomes, such as rate ratios, a study with zero cases can be problematic for some weighting approaches such as the inverse-variance method. The common solution to this problem is to apply a continuity correction by adding a constant (typically 0.5) to the number of cases. The constant may be added to all studies, only to all studies with zero cases or only to all study arms with zero cases. The addition of a constant can be done irrespective of whether any of the studies have zero cases. The use of a continuity correction can impact on the significance and interpretation of results.\(^3\)

While meta-analyses typically combine study level effect estimates, it is also possible to pool individual patient data (IPD) from studies. Use of individual data can improve the ability to include comparable subgroups or common outcomes which may not be reported in published studies. Analysis of patient data...
also enables more detailed time-to-event to be combined. The methods of IPD meta-analysis can be broadly
classified into two groups: a one-step analysis, in which all patients are analysed simultaneously as though
in a mega-trial, but with patients clustered by trial; or a two-step analysis in which the studies are analysed
separately, but then summary statistics are combined using standard meta-analysis techniques. Hybrid
methods are also available to combine individual patient data and aggregate study data. By modelling the
individual risk across hundreds or thousands of patients, IPD meta-analyses generally have much higher
power to detect differences in treatment effect than the equivalent aggregate data analyses that may have
few studies. The main disadvantage of IPD meta-analysis is that it may not be possible to acquire individual
level data from all relevant studies.

2.2.2. Bayesian approach

Bayesian methods for meta-analysis of direct comparisons are analogous to frequentist methods with the
primary distinction being the use of prior distributions for the mean of the overall estimate, the means of the
individual estimates of each study, and the between-study variance (for random effects models).(4) Priors
provide an expectation of the distributions that parameters will take. Priors can be defined as non-informative
or informative distributions. The former are used to make inferences that are not greatly affected by external
information. The latter are based on some prior knowledge and have a stronger influence on the posterior
distribution and hence on the estimate of relative effectiveness. The use of non-informative priors will
generally result in effect estimates that are comparable to those in a frequentist approach. However, in some
instances (e.g. expert opinion available in absence of collected data) it may be appropriate to form
informative priors by way of other data. The use of informative priors is likely to generate results that depend
on the choice of prior distribution and may differ to those from a frequentist approach.

The strength of Bayesian approaches in this context is that they can incorporate data from a wide variety of
sources and are not restricted to using data from RCTs. For example, expert opinion can be used to elicit
useful information that can then be incorporated into a Bayesian meta-analysis. While model flexibility is a
strength of Bayesian methods, it is also a potential weakness as additional information may be incorporated
in a biased or non-transparent manner. The use of informative priors must be accompanied by a justification
and a clear description of how they were generated to maintain transparency.

2.3. Indirect comparisons

The need for indirect comparisons arises when comparing treatments A and B, but the only available
evidence comes from studies comparing A to C and B to C. By using a common comparator, in this case
treatment C, it is possible to generate an indirect comparison of treatments A and B. For a variety of reasons,
placebo-controlled trials are commonly conducted in preference to head-to-head trials giving rise to the need
for indirect comparisons.(5,15)

Methods for indirect comparisons have only been readily applied since the late 1990s and have quickly
gained popularity. In unadjusted indirect comparisons, the results from individual arms of different trials are
compared naïvely as if they were from a single controlled trial.(6) Adjusted indirect comparisons preserve
randomisation and should always be used in preference to unadjusted methods.

2.3.1. Unadjusted indirect comparison

Unadjusted indirect comparisons combine study data as though they had come from a single large trial.(1) A
weighted summary effect is computed for all study arms involving treatment A and is compared to a weighted
summary effect for all study arms including treatment B. The two summary effects are then compared to
determine the relative effectiveness of treatment A compared to treatment B.

The primary flaw of this approach is that it ignores the randomised nature of individual trials. Glenny et al.
showed that when compared to direct estimates, unadjusted direct comparisons resulted in a large number
of discrepancies in the significance and direction of relative effectiveness.(14) Although theoretically
unbiased, this method clearly results in unpredictable results and is flawed by not acknowledging
randomisation. As such this method of indirect comparison should not be used.
2.4. Mixed treatment comparisons

A mixed treatment comparison combines direct and indirect evidence to determine the relative effectiveness of a treatment compared to two or more other treatments.

2.4.1. Lumley’s method of network meta-analysis

The method of network meta-analysis proposed by Lumley allows the combination of both direct and indirect evidence. In the subsequent text, ‘network meta-analysis’ refers specifically to the Lumley method. This methodology requires the data to contain a closed loop structure. Depending on the complexity of the closed loop design, it is generally possible to compute relative effectiveness by a number of routes. It is possible to compute the amount of agreement between the results obtained when different linking treatments are used. This agreement forms the basis of an incoherence measure which is used to estimate the consistency of the network paths. Incoherence is used to compute the 95% confidence interval for the indirect comparison.

As with Bucher’s adjusted indirect comparison method, it is assumed for network meta-analysis that the relative effectiveness of a treatment is the same across all trials used in the comparison. Network meta-analysis also does not automatically account for correlations that may exist between different effect estimates when they are obtained from a single multi-armed trial. In trials with more than two treatment arms, estimates of relative treatment effects will be correlated due to the structure of the network of evidence. For example, in a multi-arm placebo-controlled trial the comparison between any two treatments will be correlated with the comparison of each of those treatments with placebo. Accounting for this correlation is possible and can impact on the confidence interval around the relative treatment effect.

2.4.2. Bayesian mixed treatment comparison

Bayesian mixed treatment comparison (MTC) meta-analysis is a generalisation of standard pair-wise meta-analysis for A vs. B trials to more complex networks of evidence. Any combination of studies can be combined as long as all are connected in some way. Both direct and indirect evidence can be combined and there is no restriction to the number of arms in any given trial. Bayesian MTC facilitates simultaneous inference about all of the treatments included in the analysis, allowing estimation of effect estimates for all pair-wise comparisons and for treatments to be ranked according to relative effectiveness. Bayesian MTC is implemented in a Bayesian framework which provides more sophisticated methods for incorporating other sources of data such as observational studies or expert opinion. Methods have also been proposed and applied that combine Bayesian MTC and meta-regression which enables the incorporation of study-level covariates (e.g. setting, average patient characteristics, year of publication) in the Bayesian MTC framework as a means to reduce inconsistency although this adaptation has implications for compromised power.

As with the other approaches, a Bayesian MTC assumes that the true effect of a given treatment is the same in all trials included in the indirect comparison. Bayesian MTC models assume that the correlation terms in the variance-covariance matrix have the same value and that the event rates, on the logit scale, follow a
multivariate normal distribution. Being a Bayesian approach, there is scope for defining informative priors. While priors may be legitimately generated, it is critical that they are credible and clearly justified.

### 2.4.3. Other approaches

The confidence profile method provides a general framework for carrying out multi-parameter meta-analysis. As well as incorporating trials with different treatment comparisons, it can encompass different designs, outcomes and measures of effect. The confidence profile method also allows explicit modelling of biases. Although this method is typically implemented as a fully Bayesian model, it can be formulated without prior distributions and fitted using maximum likelihood methods. Where there is direct and indirect evidence available, cross-validatory predictive checking can be used to determine evidence consistency. In the case of drug assessments, if different doses of the same pharmaceutical were studied, looking at dose-response relationships can also generate cross-validatory information, provided the trial populations are comparable. The models available for this methodology are based on fixed-treatment effects although both fixed and random study-effects are possible. It is presumed that it is valid to combine the studies – that is the studies are measuring the same treatment effect.

Regression techniques can be used to combine trial data to evaluate relative effectiveness. Where the primary outcome is binary and data are available in the form of 2×2 frequency tables for each study, logistic regression can be used. Generalised linear mixed models (GLMMs) have also been proposed as a method of combining trial data in a regression framework. GLMMs can be applied to individual level patient data which can be difficult to obtain. The advantage of regression techniques is the potential for including study level covariates that may be used to explain heterogeneity in the measured effects.

### 2.5. Critical comparisons of techniques

A number of reviews have compared the methods of direct, indirect and mixed treatment comparison outlined above. Comparisons have generally revolved around discussion of the principles and assumptions of each method followed by application to a sample dataset to compare the estimates generated by each method.

In terms of direct comparisons, frequentist and Bayesian approaches have been compared by Vandermeer et al. The authors investigated six outcomes relating to treatment for chronic insomnia. The two approaches produced very similar results, although the Bayesian approach yielded wider credible intervals (analogous to frequentist confidence intervals). The two approaches are essentially equivalent when non-informative priors are used. However, in a given context there may be multiple ways to define non-informative priors and sensitivity to the choice of priors was noted for certain parameters.

Glenny et al. used a sample data set of 15 trials to compare a variety of methods for indirect comparison. They compared Bucher’s method of adjusted indirect comparison (inverse variance, fixed and random effects), meta-regression (weighted linear and random effects), logistic regression and mixed models (that combined fixed and random effects). With the exception of the mixed models, the other seven models gave comparable effect estimates and confidence bounds. The mixed models, however, produced slightly lower estimates and larger confidence intervals.

Methods of indirect treatment comparison were compared by Edwards et al. Using a single common comparator for indirect comparison is relatively inefficient in mathematical terms. Four times the amount of data are required to provide the same precision around an indirect comparison as would be required for a direct comparison. Furthermore, the use of a single comparator limits the number of RCTs available for inclusion in the analysis as only three treatments can be compared in any single analysis. Bias may also be introduced by ‘lumping’ individual treatments into convenient categories (e.g. drug classes) to achieve maximum statistical power with the available RCTs.

Bayesian MTC and Bucher’s adjusted indirect comparison method were compared by O’Regan et al. across a range of network types. They found that in most cases the two methods produced similar results although there were a limited number of instances where they produced estimates that differed in direction. For single and multi loop networks the results were very similar although it was remarked that results could be expected to differ for general single loop networks on the grounds that the Bayesian MTC approach is based on all available information. For star shaped networks the results were equivalent and, given the ease of applying Bucher’s method of adjusted indirect comparison, it may be preferable over the more complex Bayesian MTC.
Wells et al. compared Bucher’s method of adjusted indirect comparison, network meta-analysis, the confidence profile method and Bayesian MTC. Each method was applied to sample data followed by a discussion of the relative merits of each approach. Bucher’s method of adjusted indirect comparison was extended to apply to relative risks, hazard ratios, risk differences and mean differences. The method was also adapted to be applicable to ladder networks. The authors conclude that although Bucher’s adjusted indirect comparison method cannot always be applied, the simplicity of its application and interpretation is considered highly advantageous.

2.5.1. Usage in practice

As part of the originally conducted literature search during JA1, on the basis of abstracts it was possible to classify papers that applied Bucher’s adjusted indirect comparisons, Lumley’s network meta-analysis and Bayesian MTC. There were also a number of studies that applied Bayesian methodology that could not be classified as either network meta-analysis or MTC. There were 40 studies identified that applied Bayesian MTC, making it the single most popular methodology. There were 27 and 24 studies applying network meta-analysis and Bucher’s method of adjusted indirect comparison, respectively. Other Bayesian methods were used in 18 studies. One paper was identified that applied a naïve approach to indirect comparison. Lumley’s network meta-analysis and Bayesian MTC appear to be gaining popularity, while Bucher’s adjusted indirect comparison method appears to no longer be commonly used. The total number of published multiple comparisons is increasing year on year peaking at 34 in 2009, the last full year included.

Song et al. conducted a survey of systematic reviews published between 2000 and 2007 that used multiple comparisons. They observed a year on year increase in the number of published papers that contained indirect comparisons. By 2007 the most common method was Bucher’s method of adjusted indirect comparison. The results of the survey included an assessment of methodological problems found in published indirect comparisons. These problems include: poor understanding of underlying assumptions; use of flawed or inappropriate methods; inadequate comparison and inappropriate combination of direct and indirect evidence; incomplete literature search; and poor assessment of trial similarity. While methods for indirect comparison are widely available and frequently applied, they are often used without understanding and in situations where it may be inappropriate. In some instances, insufficient detail is provided in the published study to fully evaluate whether the findings are either accurate or valid. Further work by this group has found that significant inconsistency between direct and indirect comparisons may be more prevalent than previously observed.

A systematic review by Schöttker et al. investigated multiple comparisons published between 1999 and 2008. They found Bucher’s method of adjusted indirect comparisons to be the most common methodology but noted the increasing popularity of Bayesian MTC. They also looked at the number of discrepancies in results where both direct and indirect comparisons were used. Discrepancies were highest when unadjusted indirect comparisons were used, followed by meta-regression, Bucher’s adjusted indirect comparison and Bayesian MTC. Discrepancies were most commonly found where the assumption of between-study homogeneity did not hold.

2.6. Additional considerations

The application of meta-analysis is multi-faceted and requires giving consideration to a wide range of issues that are not necessarily related to the methodology being used. Some of the key issues are outlined below.

2.6.1. Heterogeneity

It is assumed that the relative effectiveness of a treatment is the same across all studies included in a meta-analysis – that is, we assume similarity of studies. If the results of the studies are very different then we observe heterogeneity and combining the results may not be appropriate. Three broad forms of heterogeneity have been identified: statistical, where effect estimates vary more than expected by chance alone; clinical, due to differences in patient populations or study settings; and methodological, arising from differences in study design and analysis. It is possible to test for heterogeneity to provide evidence for whether or not the study results differ greatly. This is not an optimal way to assess heterogeneity and, where significant heterogeneity is observed, it is critical to closely examine the studies being combined. There can be many causes of heterogeneity such as variations in study design, study subjects, setting, geographic location, and outcome measures. In some instances it will be possible to partially explain heterogeneity.
between studies by differences such as those listed above. Even if the variability can be explained, there
must still be a decision whether or not to proceed with the meta-analysis and whether to consider subgroup
analyses. A subgroup analysis can involve including studies that are considered equivalent according to a
more narrowly defined set of criteria (e.g., age range of study participants).

Inconsistency is a measure of how closely a network of evidence fits together. A measure of inconsistency
can be calculated within Bayesian MTC while Lumley’s method of network analysis estimates incoherence,
which is analogous to inconsistency, but is calculated in a different manner. In mixed treatment comparisons,
consistency between direct and indirect evidence is assumed. That is, if direct evidence suggests that
treatment A is better than treatment B, then that evidence should not be contradicted by the indirect
evidence. Depending on the amount of evidence available, indirect comparisons can sometimes generate
estimates of relative effectiveness via two or more different paths. In comparing treatments A and B, the
relative effectiveness should be similar whether derived via common comparator C or D. A statistically
significant difference in the estimate of relative effectiveness would indicate inconsistency. A difference in
direction of relative effectiveness, even if not statistically significant, would also raise concerns about
consistency. In a multiple treatment comparison involving both direct and indirect evidence, the evidence
network can become very complex with many comparisons based on only one or two studies. With
increasing complexity and greater numbers of treatments the prospect of inconsistency increases. There is
also a power trade-off between the number of pair-wise comparisons and the number of studies included in
the analysis – too many comparisons with too few studies and the analysis may be underpowered to detect
true differences. Understanding the cause or source of inconsistency in a complex network can be
difficult to determine, which raises questions about how elaborate an evidence network should be in order to
be accepted for analysis. In the context of a multiple treatment comparison, the presence of heterogeneity
may mask inconsistency. If relevant heterogeneity is present, it is not advisable to proceed with a multiple
treatment comparison.

2.6.2. Fixed and random effects

In fixed effect meta-analyses the true effect of treatment is assumed to be the same in each study. Use of a
fixed effect model therefore follows from the assumption that variability between studies is entirely due to
chance. In a random effects meta-analysis the treatment effect in each study is assumed to vary around an
overall average treatment effect. As the random effects model assumes a different underlying effect for
each study it tends to result in wider confidence intervals than the fixed effect model. When the reported
effect sizes are homogeneous the fixed and random effects approaches yield very similar results. The choice
between random and fixed effect models is context specific and the decision is often made following an
assessment of heterogeneity. Examples of common heterogeneity measures include I^2 and Q. Substantial
heterogeneity suggests use of a random effects model but also raises the question of whether the studies
are actually comparable. The use of random effects has implications for the interpretation of results and the
distribution of effect estimates should be discussed. A measure of heterogeneity should be reported to
support the choice between a fixed and random effects model. If relevant or explicable heterogeneity is
present then using a fixed effect model would not be correct. In such instances a fixed effect model should
only be presented in special situations (e.g., few studies and where sample sizes vary considerably).

2.6.3. Publication bias

The issue of publication bias arises due to journals being more likely to publish studies showing beneficial
effects of treatments while equally valid studies showing no significant effect remain unpublished. The
consequence of this bias is that a meta-analysis will show a spurious significant effect. Publication bias may
be detectable using funnel plots or regression techniques but these are not particularly powerful
techniques. Asymmetry in a funnel plot may indicate publication bias or it may be a reflection of how
comprehensive the search strategy has been. English language bias and citation bias are forms of
publication bias in which studies with negative findings are more likely to appear in non-English language
publications and are less likely to be cited, respectively. It is of critical importance that the search strategy
element of the systematic review is as comprehensive as possible and that clinical trial registers are
searched, where possible. The presence of publication bias can impact on any meta-analysis irrespective of
the methodology used (i.e., direct, indirect or mixed treatment comparison).

2.6.4. Outlier analysis and influential studies

The results of a meta-analysis may be overly influenced by one or a small number of studies. Similarly, some
studies may be outliers in a statistical sense. Outliers and influential studies are not synonymous: an outlier
need not necessarily be influential and an influential study need not be an outlier. A first step is to visually inspect a forest plot to identify any unusual data points or where the pooled estimate appears to be driven by a single or small number of studies. In conventional direct comparisons, metrics such as Cook’s distance can be used to identify influential studies and standardised study-level residuals can be used to identify outliers. In Bayesian MTC, visual inspection of QQ-plots has been used to identify outliers. Sensitivity analysis techniques can be used to determine the impact of influential studies and outliers on the results of a meta-analysis. For example, an analysis can be conducted with and without a particular study to determine the impact of that study on the results. It is also useful to characterise outliers and gain an understanding of why they might be different from other studies.

2.6.5. Sensitivity to priors in a Bayesian approach

A common criticism of Bayesian techniques is that they rely on the use of priors for key parameters. A point of criticism is that priors are subjectively chosen. In reality, most Bayesian analyses employ vague or non-informative priors. However, even with a non-informative prior, assumptions are made about the distribution of that prior and often there are alternative formulations available so the use of sensitivity analysis is important. For a number of example analyses, Wells et al. found that the Bayesian methods were insensitive to the prior chosen for the mean, but were sensitive to the prior chosen for the between study variance. They also found that the between study prior sensitivity was directly related to the number of studies included in the analysis. A sensitivity analysis in this context is not trivial as there will be a number of parameters with specified priors and there may be several potential priors for each parameter. In the event that conclusions on the effects size are affected by the choice of prior then additional evidence will have to be gathered to justify the choice of priors.

2.6.6. Dose-response issues in drug assessments

For addressing the issue of REA of pharmaceuticals, it is also important to define which doses of pharmaceuticals have been compared and to address the question of whether the dose-response relationships of the individual pharmaceuticals allow extrapolation to other doses. In sponsored trials there is also a risk that the comparator treatment may be administered at a sub-optimal dose or frequency to show the active treatment in a more favourable light. For trials to be considered comparable it is therefore important to assess the comparability of treatment regimens across the included trials.
3. Discussion

Direct comparisons require a number of head-to-head studies to estimate the relative effectiveness of two treatments. Should a third treatment be considered, unless sufficient studies were three-arm trials, a direct comparison is no longer possible and a multiple treatment comparison would be required. (11) The ability to consider more complex networks of evidence is an advantage of indirect methods, but this benefit may increase the potential of introducing inconsistencies. It is recommended that comparisons should be based on a focused research question with careful attention to ensuring that only comparable studies are included.

In presenting the indirect comparison method, Bucher contended that direct methods should be used whenever possible in preference to indirect methods. (7) However, more recent developments suggest that the combination of direct and indirect evidence may achieve a more accurate and comprehensive analysis. (27) The use of indirect and mixed treatment comparisons highlights the importance of a thorough literature search and identification of all relevant studies. Multiple treatment comparisons can be sensitive to data, particularly when there may be only one study for a given pair-wise comparison. (28) Furthermore, there is a power trade-off between the number of pair-wise comparisons and the number of studies included in the analysis – too many comparisons with too few studies and the analysis may be underpowered to detect true differences. (10)

Irrespective of methodology, a key principle of meta-analysis is that only comparable studies should be combined. (29) For both direct and indirect comparisons, it is critical that only comparable studies are selected for inclusion in the analysis. As an adjunct to studies being comparable, the results of the studies should be consistent. That is to say, if A is more effective than B and B is more effective than C, then it is anticipated that A will be more effective than C. Such a relationship is not a given and there is the potential for inconsistencies to arise in indirect comparisons. Within Bayesian MTC there is the possibility of assessing the level of inconsistency between direct and indirect evidence. (30) Lumley’s network meta-analysis enables the estimation of incoherence in the evidence network which is analogous to inconsistency. (8) When direct and indirect evidence differ or inconsistencies arise then the first step is to assess the similarity of participants and interventions across studies. (1) In terms of study population or interventions, the studies providing direct evidence may be sufficiently different from those providing indirect evidence to account for inconsistency in the results. That is to say, there may be systematic heterogeneity between direct and indirect evidence which may explain the conflicting results. In the event of such identifiable differences, presenting both combined and individual results of direct and indirect comparisons is the most appropriate approach. It should be borne in mind that the results of indirect comparisons are not valid if relevant inconsistency is present.

In contrast to the frequentist approach, the output of a Bayesian approach can be interpreted in terms of probabilities. (26) Thus Bayesian approaches facilitate the computation of useful measures such as the probability of treatment superiority, the probability that a clinically meaningful difference exists, or the probability of clinical equivalence. (31) It is possible to compute the probability that a given treatment is the best amongst those analysed and how the other treatments are ranked, which constitutes information useful clinically and to decision makers. (32) Bayesian methods also facilitate predicting the utility of conducting additional head-to-head studies. (26) The benefits to the decision maker of Bayesian outputs need to be weighed against the increased complexity of the modelling and expertise required in applying these methods.

Indirect comparisons can be affected by the same forms of bias as direct comparisons but these forms of bias can arise in subtle ways. For example, studies of new treatments often find them to be more effective than existing treatments even though the benefit may be spurious. (33) In an indirect comparison the chronology of studies will impact on the relative effectiveness of each treatment and it may be difficult to tease out what is genuine and what may be spurious even with the aid of measures of inconsistency. With a direct comparison a cumulative meta-analysis, for example, could be used to reveal changes in relative effectiveness over time. In applying indirect comparisons, a thorough assessment of inconsistency or incoherence must be carried out and reported. Song et al. suggest that Bucher’s adjusted indirect comparison method could be used as a means to cross-examine evidence from direct comparisons as it may be subject to different forms of bias. (34) Hence the two approaches, direct and indirect comparison, may act to validate each other’s results. The use of indirect and mixed treatment comparisons highlights the importance of a thorough literature search and identification of all relevant studies, including observational studies and unpublished data.

The mixed treatment comparison incorporating direct evidence could be considered the ‘gold standard’ approach of carrying out indirect comparisons. (15) Of the indirect comparison methods, where the available
Evidence supports the use of either Bucher’s method of adjusted indirect comparison or Bayesian MTC, the former approach offers the most in terms of transparency and relative ease of application. (5,35) Bayesian MTC has gained popularity and has greater flexibility with regard to the network of evidence available but these benefits must be counterbalanced by the difficulty in applying this methodology and the scope for incorrect specification of the model. The application of Bayesian MTC involves many choices that may alter the conclusions of the analysis but the justifications for those choices may seem somewhat arbitrary. (36) The complexity of the Bayesian MTC models renders them less advantageous than Bucher’s method of adjusted indirect comparison or network meta-analysis when the latter methods can be applied.
4. Conclusion

To maximise the evidence on the relative effectiveness of an intervention, it is common to combine results from several randomised controlled trials (RCTs) in a meta-analysis. Direct comparisons combine the results of multiple head-to-head trials whereas indirect comparisons can infer the relative effectiveness in the absence of direct head-to-head evidence. Irrespective of the methodology used, a meta-analysis must be preceded by a properly conducted and transparent systematic literature review.

Direct methods use well-known meta-analysis techniques which may be either frequentist or Bayesian although the former are more common. Methods of indirect comparison are a more recent development and still evolving. Bucher’s adjusted indirect comparison method is being supplanted by the Bayesian MTC approach. Both of these methods generate similar results when applied to the same data but Bayesian MTC can be used to analyse far more complex networks of evidence and can include study-level covariates. It has been suggested by a number of commentators that the least complex model that can be applied in a given analysis is generally preferable.

There is no consensus on the best approach to comparisons, particularly when both direct and indirect evidence are available. In the case of a head-to-head comparison where sufficient good-quality studies are available for a direct comparison, an indirect comparison may not be necessary. However, where the direct evidence is limited, then combining the direct with indirect evidence is advantageous. Alternatively, if a number of treatments are being considered simultaneously, then indirect or multiple treatment comparison may be helpful. It may also be desirable to carry out direct and mixed treatment comparison separately to determine the impact of incorporating indirect evidence. Alternatively, if a number of treatments are being considered simultaneously, then a multiple treatment comparison is preferable. When both direct and indirect evidence exists, it is important to investigate inconsistencies between the direct and indirect evidence and to attempt to account for the discrepancies.

Bucher’s method of adjusted indirect comparison, Lumley’s method of network meta-analysis and Bayesian MTC are acceptable methods for indirect comparisons in the context of REA.

Bayesian methods are advantageous for a number of reasons (e.g. they can incorporate additional sources of evidence and they produce output that is inherently interpretable by decision makers). However, these advantages must be weighed against the sometimes increased complexity of the Bayesian models and a potential lack of transparency.

When conducting a treatment comparison, irrespective of methodology used there are a number of additional issues to consider: heterogeneity; fixed and random effects models; the presence of bias; influential observations and outliers; and, in a Bayesian context, to investigate the sensitivity of results to the choice of priors.

In making a comparison between two or more treatments, it is critical that the reported relative effectiveness is an accurate reflection of reality. That accuracy can only be achieved by using the best evidence available synthesised using the most appropriate methods. The choice of methodology is context-specific and should be based on an objective assessment of the quality and quantity of the direct and indirect evidence, the comparability of the selected studies, and of the fundamental assumptions in the different models. The complexity of a model is not a measure of its accuracy or utility and preference is for the most parsimonious model.
Annexe 1. Methods of documentation and selection criteria (related to original guideline elaboration in JA1)

Keywords that will be used for the bibliographic research:

- Meta-analysis
- Meta-regression
- Direct comparison
- Direct treatment comparison
- Indirect comparison
- Unadjusted indirect comparison
- Adjusted indirect comparison
- Mixed treatment comparison
- Indirect treatment comparison
- Multiple treatment comparison
- Network meta-analysis
- Bayesian

3.1. Sources of information

3.1.1. Databases
- MEDLINE
- MEDLINE via OVID
- DARE
- Cochrane Database of Systematic Reviews
- CADTH/CEDAC
- EBSCOhost

3.1.2. Websites
- National Guideline Clearinghouse
- National Institute for Health and Clinical Excellence
- ISPOR
- Pharmaceutical Benefits Advisory Committee (PBAC)
- Centre for Reviews and Dissemination, University of York
- University of Bristol
- Haute Autorité de Santé (HAS)

3.1.3. Guidelines, reports, recommendations already available
- Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

3.1.4. Books
- The topic is relatively new; no relevant books were identified.

3.1.5. Other
- Google and Google Scholar
- ScienceDirect
- Wiley-InterScience
- Hand searching of references cited in relevant documents
- The Cochrane Collaboration
- The Cochrane Methodology Register
3.2. Bibliographic search strategy

Direct comparisons are generally dealt with by way of standard meta-analytic techniques that are well-documented. Indirect comparisons are a more recent development and most applications have occurred since the paper of Bucher et al. in 1997. For PubMed, the search was limited to the period "last ten years". In EBSCO the search was limited to 2000 to 2010 (inclusive). In both cases the search was limited to English language publications and human subjects.

Database searches used the following search strategy:

("network meta-analysis" OR "mixed treatment" OR "multiple treatment" OR "multiple comparison" OR "cross-trial") AND (meta-analysis OR systematic)) OR ((indirect comparison OR "direct comparison") AND (meta-analysis OR systematic))

3.3. Selection criteria

Reports, papers and other guidance documents were assessed on the basis of whether they described, applied or assessed methods of direct and indirect treatment comparisons. Documents that only mentioned methods, but did not describe, apply or assess them were discarded after being checked for useful references. Documents describing new methods provided keywords and were also used to identify subsequent citing documents that then either applied or assessed those new methods. Documents that applied methods were used to determine the scope of application, utility and possible limitations of those methods. Finally, documents that assessed methods were used to compare methods directly and to elicit recommendations.

The full bibliography is provided in Annexe 2 and is split into two – methodological/review articles and application articles. The former articles were used to form opinion and recommendations. The latter articles were not reviewed in depth, but were used to determine the frequency with which the different methods are applied as a marker of acceptability and diffusion into common practice.
Annexe 2. Bibliography


(17) O'Regan C, Ghement I, Eyawo O, Guyatt GH, Mills EJ. Incorporating multiple interventions in meta-analysis: an evaluation of the mixed treatment comparison with the adjusted indirect comparison. Trials 2009;10(66).


(34) Song F, Harvey I, Lifton R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. Journal of Clinical Epidemiology 2008 May;61(5):455-63.


Annexe 3. Literature search results (related to original guideline elaboration in JA1)

Methodological and review articles:


Application of direct and indirect methods:

(2) Department of Error. Lancet 369(9572), 1518. 5-5-2007.

EUnetHTA – European network for Health Technology Assessment


Bolon MK, Wright SB, Gold HS, Carmelli Y. The magnitude of the association between fluorquinolone use and quinolone-resistant Escherichia coli and Klebsiella pneumoniae may be lower than previously reported. Antimicrob Agents Chemother 2004 Jun;48(6):1934-40.


(66) Centre for Reviews and Dissemination. Drug eluting stents: an updated meta-analysis of randomised controlled trials (Structured abstract).


Ref Type: General


Holodniy EM, Craig JC, Strippoli GF, Webster AC. Antiviral medications for preventing cytomegalovirus disease in solid organ transplant recipients. Cochrane Database of Systematic Reviews 2008 Jun;(2).


Lam SKH, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: a Bayesian network meta-analysis of randomised controlled trials. BMJ (Clinical Research Ed ) 2007 Nov 3;335(7626):925.


EUnetHTA – European network for Health Technology Assessment 33


