NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Briefing paper for methods review working party on mixed treatment comparisons

The briefing paper is intended to provide a brief summary of the issues that are proposed for discussion by the Methods Review Working Party to inform an update to the Institute's Guide to Methods of Technology Appraisal. It is not intended to reflect a comprehensive or systematic review of the literature. The views presented in this paper are those of the authors and do not reflect the views of the Institute.

1 Review of the 'Guide to Methods of Technology Appraisal'

The Institute is reviewing the 'Guide to the methods of technology appraisal', which underpins the technology appraisal programme.

The original Methods Guide was published in February 2001, and revised versions were published in 2004 and 2008. The Methods Guide provides an overview of the principles and methods used by the Institute in assessing health technologies. It is a guide for all organisations considering submitting evidence to the technology appraisal programme and describes appraisal methodology.

The revised draft of the Methods Guide will be available for a 3-month public consultation, expected to begin in xxx. We encourage all interested parties to take part in this consultation.

2 Background

2.1 Relevance of topic to NICE technology appraisals

The quantity and nature of clinical evidence submitted for technology appraisals varies considerably. Commonly there may be one or two directly relevant head-to-head trials which compare an intervention of interest with a comparator of interest, but evidence to draw comparisons across the full range of treatment options specified as comparators in the scope is lacking. In such situations, it is also common for there to exist a number of indirectly relevant trials in which the intervention(s) of interest, or the comparator(s) of interest, are compared with other treatments which may or may not be within the appraisal scope. The use of mixed treatment comparisons to synthesise such evidence is becoming increasingly used for NICE technology appraisals. This may be the consequence of a number of factors including a lack of direct head-to-head trials of all relevant decision alternatives, increased awareness of indirect methods, developing methodology as well as the direction of the 2008 Methods Guide. Where such approaches are employed, it is essential that the scope and methods of evidence synthesis are appropriate, robust and transparent for NICE's Appraisal Committees.

As with any pooling of studies, it is crucial that there can be confidence that the trial populations and methods are comparable and that decisions about trial inclusion into the network are both unbiased and transparent. However, it is very rare for manufacturers' submissions to present a full critical appraisal of the mixed treatment comparison which includes full details of how the mixed treatment comparison has been constructed and full details of the trials and participants included in the mixed treatment comparison. In addition, the network of trials can often be very large, which from a practical viewpoint, can result in problems for the Evidence Review Groups and Assessment Groups in systematically reviewing and appraising mixed treatment comparisons. In addition, manufacturers' submissions rarely present a full examination of the effects of individual trials on the results of the mixed treatment comparison

and sensitivity analyses exploring the inclusion and exclusion of key trials are rarely submitted.

A critical appraisal checklist has recently been developed as part of the DSU series of Technical Support Documents on evidence synthesis methods (see Appendix 1).[1-7] This checklist[7] covers a number of pertinent synthesis issues including the scope of the analysis, the search strategy used to identify relevant trials for inclusion in the analysis, the definition of the interventions, the choice of outcome measure(s), the presentation of data, statistical methods employed, software considerations, issues surrounding inconsistency, and the use of the analysis within economic decision models.

2.2 Introduction to mixed treatment comparisons

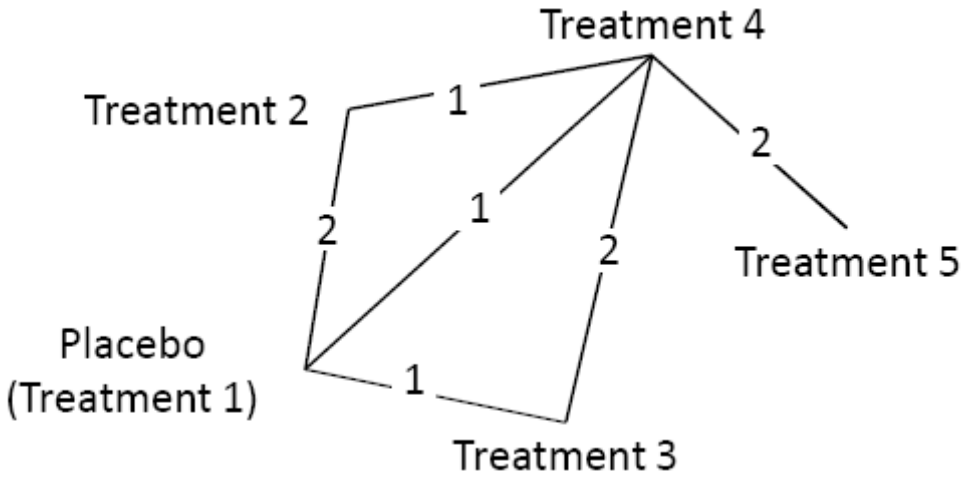
Frequently, and particularly in the case of newly licensed technologies, there are very few head-to-head randomised controlled trials that directly compare the intervention of interest (that is, the new technology) with the comparator of interest (that is, routine and standard practice in the NHS). It is therefore common in practice to see indirect comparisons or mixed treatment comparisons conducted in order to provide sufficient evidence on which to ascertain the relative effectiveness of the new technology compared with the comparator(s) of interest.

In order for a robust mixed treatment comparison to be possible, a number of conditions must be satisfied. Firstly, the trial populations must be truly comparable at baseline and secondly there must be comparable treatment circumstances. For example, if the trial populations or trial methodologies differ greatly, then it may be inappropriate to pool these studies. In addition, the whole trial network needs to be constructed in an unbiased manner (that is, the same inclusion and exclusion criteria are applied to all of the trials considered). These conditions are consistent with those that would be expected if conducting a more conventional head-to-head or ‘classical’ piecewise meta-analysis.

In order to explain the concept of indirect and mixed treatment comparisons, consider a scenario in which there are three technologies of interest, A, B and

C. Define the effect in a trial which compares A to B by d_{AB}. This is a *direct* estimate of AB. An *indirect* estimate of AB can also be obtained if there are AC an BC trials since $d_{BC} - d_{AC} = d_{AB}$. A mixed treatment comparison analysis (also known as network meta-analysis or mixed treatment meta-analysis) allows the synthesis of AB, AC, BC and ABC (i.e. three-arm) trials and estimates each pairwise treatment effect from both the direct and indirect evidence *without breaking randomisation*. Mixed treatment comparisons, are essentially extension of a traditional meta-analysis; these comparisons synthesise data from a series of trials allowing different comparisons to be made among the technologies of interest. Mixed treatment comparisons require a connected network; that is, for each treatment, there is a chain of pairwise comparisons that connects it to every other treatment. The construction of network diagrams can clearly describe the different possible evidence structures (see Figure 1). Within this form of network diagram, each edge represents a treatment; connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison.

Figure 1 Example network diagram[1]



The methodology can be particularly useful when no, or little, direct head-to-head evidence exists on comparisons of interest. Also, when conducting a series of pairwise meta-analyses, it is difficult or impossible to rank all technology options in terms of effectiveness or cost-effectiveness. In contrast,

this is straightforward within a mixed treatment comparison and allows the estimation of the probability each technology is optimal across individual or multiple clinical endpoints.

It is important to recognise that mixed treatment comparison networks can become relatively large and complicated. This happens when there are a number of relevant interventions and comparators. The benefit of combining all of the direct and indirect evidence is that a decision is being made on all of the available evidence. However, there are a number of issues with mixed treatment comparisons that frequently arise within NICE technology appraisals. Firstly, the size of the network, that is, the number of trials and additional comparators included within the mixed treatment network, can become very large. In this situation, undertaking a comprehensive review and appraisal of the mixed treatment comparison can become cumbersome and time consuming. This can cause problems for Evidence Review Groups and Assessment Groups, especially in terms of checking that all the relevant studies have been included, that no inappropriate studies have been included and that the results of the analysis are both robust and reliable.[1]

In practice, previous mixed treatment comparisons from similar appraisals are often used as a basis for the network, into which additional trials are added. This may mitigate the intensity of the checking activity required if this ‘base network’ is considered reliable. However, checking of the original network, and any amendments made to it, must always be conducted. In addition, conducting an appraisal of a mixed treatment comparison can also include checking with experts in the field and comparator manufacturers, relevant stakeholders and conducting additional systematic reviews. From a practical point of view, these activities can be very resource intensive especially when the scope of network is large and complex and if the mixed treatment comparison has been submitted later on in the appraisal process, for example in response to an Appraisal Consultation Document.

A second issue is that the presentation of the mixed treatment comparisons usually does not facilitate understanding of the individual trials and of the trial participants and characteristics. Often, the descriptions of the individual trials

are limited and key differences between trials are not exposed. This means that it is often difficult to assess the face validity of the results from a mixed treatment comparison. The same criticism can also however be made with respect to classical piecewise meta-analysis which can also contain a large number of trials. As with any pooling of studies (such as in a conventional meta-analysis or indirect comparison), it is essential that the studies included are comparable in terms of design, participants, and other key factors. However, in mixed treatment comparisons, especially those with large networks, there are more trials and more decisions being made when constructing the network and therefore an increased possibility for trials that are not completely comparable to enter the network. In instances where the network (and hence individual trials) is poorly described, it can also be difficult to exactly ascertain how comparable the trials are and what the effect of this may be.

One example of when a mixed treatment comparison can be strongly affected by trial inclusion is when trials with non-comparable control groups are included in a network. For example, consider a Technology A that in trials appears slightly better than a placebo, but that the placebo arm in that trial also performed relatively well. Technology B in trials (the comparator to Technology A) appears to be much more effective than placebo, but the placebo arm in that trial has performed relatively poorly. In this situation, the relative effectiveness of Technology A compared with placebo is small and the relative effectiveness of Technology B and placebo is large. If these were combined in a mixed treatment comparison together with a number of other trials within the evidence network, it is possible that the results of the mixed treatment comparison would be misleading and the reasons for this inconsistency would be difficult to tease out if the network and the individual trials are poorly presented. In this situation one can, for example, question whether the scale of measurement (log ORs) is correct, or one can adjust for baseline risk if one believes that this has an impact of the *relative* effects.[5]

A further point, however, applies to pairwise comparisons or to cases where there is just one trial in the evidence base. Suppose that we only had the

Technology A vs placebo trial, and the target population was in fact the one that appears in the Technology B vs placebo trial, or vice versa. In both cases we would completely misjudge the efficacy of the active treatment. Or consider we had three A vs placebo trials, all with different baseline efficacies. Whilst these are clearly difficult situations for the interpretation of evidence, it would be a mistake to blame indirect comparisons as the root cause of the problem.[3]

A third issue is that often mixed treatment comparisons are presented as the reference case and little, or no, exploration of the suitability of the mixed treatment comparison is presented. In particular, the results of some mixed treatment comparison networks may be heavily influenced by one or two key trials and the inclusion and exclusion of these trials and the subsequent effect of this on the overall result is rarely presented clearly as sensitivity or scenario analyses to the Appraisal Committee. Particularly in the cases described above, whereby the inclusion of some trials may be open to question, it is important that the effect of these studies on the overall results is clearly presented.

In summary, there remains a need to undertake coherent analyses and further clarity about when this should include a mixed treatment comparison could be beneficial. In addition, there is an outstanding need for further direction on the reporting standards and appropriate sensitivity and scenario analyses that should be conducted when undertaking a mixed treatment comparison.

The decision support unit (DSU) have been commissioned to write a number of technical support documents (TSDs) that address many oif the points raised in this briefing paper. In particular TSD7 is a checklist for reviewers of mixed treatment comparisons and many consider that this would be of great value going forward and could have a role within the methods guide itself.

2.3 What the current Methods Guide advises with respect to mixed treatment comparisons

During the last review of the methods guide, the subject of mixed treatment comparisons was a central discussion point. As a result, the methods guide

includes a number of paragraphs (5.3.13 to 5.3.22) on mixed treatment comparisons. The methods guide states the following:

5.3.13 Data from head-to-head RCTs should be presented in the reference-case analysis, if available. When head-to-head RCTs exist, evidence from mixed treatment comparison analyses may be presented if it is considered to add information that is not available from the head-to-head comparison. This mixed treatment comparison must be fully described and presented as additional to the reference-case analysis (a 'mixed treatment comparison' includes trials that compare the interventions head-to-head and indirectly). When multiple technologies are being appraised that have not been compared within a single RCT, data from a series of pairwise head-to-head RCTs should be presented. Consideration should also be given to presenting a combined analysis using a mixed treatment comparison framework if it is considered to add information that is not available from the head-to-head comparison. If data from head-to-head RCTs are not available, indirect treatment comparison methods should be used (an 'indirect comparison' is a synthesis of data from a network of trials). The principles of good practice for standard meta-analyses should also be followed in mixed and indirect treatment comparisons.

5.3.14 The Institute has a preference for data from head-to-head RCTs and these should be presented in the reference-case analysis when available.

5.3.15 An 'indirect comparison' refers to the synthesis of data from trials in which the technologies of interest have not been compared in head-to-head trials, but have been compared indirectly using data from a network of trials that compare the technologies with other interventions. A 'mixed treatment comparison' refers to an analysis that includes trials that compare the interventions of interest head-to-head and trials that compare them indirectly. The principles of good practice for systematic reviews and meta-analyses should be

carefully followed when conducting mixed and indirect treatment comparisons. The rationale for the identification and selection of the RCTs should be explained, including the rationale for the selection of treatment comparisons that have been included. A clear description of the methods of synthesis is required. The methods and results of the individual trials should be documented. If there is doubt about the relevance of a particular trial, sensitivity analysis should also be presented in which these trials are excluded. The heterogeneity between results of pairwise comparisons and inconsistencies between the direct and indirect evidence on the technologies should be reported.

5.3.16 *There may be circumstances in which data from head-to-head RCTs are less than ideal (for example, the sample size may be small or there may be concerns about the external validity). In such cases additional evidence from mixed treatment comparisons can be considered. In these cases, mixed treatment comparisons should be presented separately from the reference-case analysis and a rationale for their inclusion provided. Again, the principles of good practice apply.*

5.3.17 *When multiple technologies are being appraised, data from RCTs (when available) that compare each of the technologies head-to-head should be presented in a series of pairwise comparisons. Consideration may be given to presenting an additional analysis using a mixed treatment comparison framework. In these situations, the Appraisal Committee will consider the results of both analyses with particular reference to the methods of synthesis and the appropriateness of the inclusion or exclusion of studies.*

5.3.18 *There may be situations when data from head-to-head RCTs of the technologies (and/or comparators) are not available. In these circumstances, indirect treatment comparison analyses should be considered.*

5.3.19 When evidence is combined using indirect or mixed treatment comparison frameworks, trial randomisation must be preserved. A comparison of the results from single treatment arms from different randomised trials is not acceptable unless the data are treated as observational and appropriate steps taken to adjust for possible bias and increased uncertainty.

5.3.20 Analyses using indirect or mixed treatment comparison frameworks may include comparator interventions (including placebo) that have not been defined in the scope of the appraisal if they are relevant to the development of the network of evidence. The rationale for the inclusion and exclusion of comparator interventions should be clearly reported. Again, the principles of good practice apply.

5.3.21 Evidence from a mixed treatment comparison may be presented in a variety of ways. The network of evidence may be presented in tabular form. It may also be presented diagrammatically as long as the direct and indirect treatment comparisons are clearly identified and the number of trials in each comparison is stated.

5.3.22 When sufficient relevant and valid data are not available for including in meta-analyses of head-to-head trials, or mixed or indirect comparisons, the analysis may have to be restricted to a qualitative overview that critically appraises individual studies and presents their results. In these circumstances, the Appraisal Committee will be particularly cautious when reviewing the results of analysis.

3 Proposed issues for discussion

After consideration of the developments in this methodological area, the current Methods Guide and the requirements of the Institute's Technology Appraisal Programme, it is proposed that the following key areas are discussed by the Methods Guide Review Working Party.

Currently indirect and mixed treatment comparisons are described in great detail in the methods guide. However, the consistency in submissions varies widely:

- Should further direction be given of the use of mixed treatment comparisons?
    - Is the current content in the methods guide regarding mixed treatment comparisons excessive?

What would be the impact of reducing the level of content on mixed treatment comparisons in the next methods guide?

- Should components of 'best practice' in conducting mixed treatment comparisons be more clearly specified?
    - As a minimum, should a full list of all trials included in the mixed treatment comparison, with baseline participant characteristics and key outcomes be provided?
    - Can any guidance on the size of networks by provided?

How should the technical support documents created by the decision support unit be incorporated into the Methods Guide?

- Should TSD 7 (checklist for reviewers) be recommended as a standard reference within the methods guide?

What are the potential consequences of requiring a full list of all trials (with participant characteristics and key outcomes)? Should the methods guide state how the information should be presented?

- Should guidance be provided on checking the face validity of a mixed treatment network (for example, contacting experts in the field, checking other appraisals in the same disease area)?

What would be the impact of providing instruction within the next methods guide on checking face validity?

- Should sensitivity and scenario analyses involving the mixed treatment comparison networks always be requested?

What would be the impact of always requesting sensitivity analyses? Should sensitivity analyses only be requested if there are inconsistencies within the mixed treatment comparison?

- Should potential inconsistencies within a mixed treatment comparison network always be formally explored?

What are the consequences of requesting formal exploration of inconsistencies within the method guide? Could specific methodology be referred to if this was included in the methods guide?

4 References

1. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. (2011) NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. Available from http://www.nicedsu.org.uk
2. Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated August 2011. Available from http://www.nicedsu.org.uk
3. Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. (2011) NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Available from http://www.nicedsu.org.uk
4. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Guobing, L. & Ades, A.E. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. 2011. Available from http://www.nicedsu.org.uk
5. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. (2011) NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. Available from http://www.nicedsu.org.uk

6. Dias, S., Sutton, A.J., Welton, N.J. & Ades, A.E. NICE DSU Technical Support Document 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. 2011. Available from http://www.nicedsu.org.uk

7. Ades T, Caldwell TM, Reken S, Welton NJ, Sutton AJ, Dias S. NICE DSU Technical Support Document 7: Evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. Available from http://www.nicedsu.org.uk

5 Author/s

This document was prepared by Rebecca Trowman, Andrew Stevens and Paul Tappenden. Thanks for Tony Ades and Alex Sutton for their helpful comments.

Appendix 1 Checklist Table. Abbreviations: Y/N/na Yes, No, Not Applicable; SA Sensitivity Analysis.

		Y/N/na	Comments, SA needed ?
A. DEFINITION OF THE DECISION PROBLEM ***A1. Target population for decision***			
A1.1	*Has the target patient population for decision been clearly defined?*		
A2. Comparators			
A2.1	*Decision Comparator Set: Have all the appropriate treatments in the decision been identified?*		
A2.2	*Synthesis Comparator Set: Are there additional treatments in the Synthesis Comparator Set, which are not in the Decision Comparator Set?*		
A3 Trial inclusion / exclusion			
A3.1	*Is the search strategy technically adequate?*		
A3.2	*Have all trials involving at least two of the treatments in the Synthesis Comparator Set been included?*		
A3.3	*Have all trials reporting relevant outcomes been included?*		
A3.4	*Have additional trials been included?*		
A4 Treatment Definition			
A4.1	*Are all the treatment options restricted to specific doses and co-treatments, or have different doses and co-treatments been "lumped" together?*		
A4.2	*Is a dose-response model fitted, or are the sub-components of the treatment modelled?*		
A5 Trial outcomes and scale of measurement chosen for the synthesis			
A5.1	*Where alternative outcomes are available, has the choice of outcome measure used in the synthesis been justified?*		
A5.2	*Have the assumptions behind the choice of scale been justified?*		
A6 Patient population: trials with patients outside the target population			
A6.1	*Do some trials include patients outside the target population?*		
A6.2	*What assumptions are made about the impact, or lack of impact this may have on the relative treatment effects?*		
A6.3	*Has an adjustment been made to account for these differences? If so, comment on the adequacy of the evidence presented in support of this adjustment, and on the need for a sensitivity analysis.*		
A7 Patient population: heterogeneity within the target population			
A7.1	*Has there been a review of the literature concerning potential modifiers of treatment effect?*		
A7.2	*Are there apparent or potential differences between trials in their patient populations, albeit within the*		

	target population?		
A8 Risk of Bias			
A8.1	*Is there a discussion of the biases to which these trials, or this ensemble of trials, are vulnerable?*		
A9. Presentation of the data			
A9.1	*Is there a clear table or diagram showing which data have been included in the base-case analysis?*		
A9.2	*Is there a clear table or diagram showing which data have been excluded and why?*		
B. METHODS OF ANALYSIS AND PRESENTATION OF RESULTS ***B1 Meta-analytic methods***			
B1.1	*Is the statistical model clearly described?*		
B1.2	*Has the software implementation been documented?*		
B2. Heterogeneity in the relative treatment effects			
B2.1	*Have numerical estimates been provided of the degree of heterogeneity in the relative treatment effects?*		
B2.2	*Has a justification been given for choice of random or fixed effect models? Should sensitivity analyses be considered?*		
B2.3	*Has there been adequate response to heterogeneity?*		
B2.4	*Does the extent of unexplained variation in relative treatment effects threaten the robustness of conclusions?*		
B3 Baseline model for trial outcomes			
B3.1	*Are baseline effects and relative effects estimated in the same model? If so, has this been justified?*		
B3.2	*Has the choice of studies to inform the baseline model been explained?*		
B3.3	*Has the statistical heterogeneity between baseline arms been discussed?*		
B4 Presentation of results of analyses of trial data			
B4.1	*Are the relative treatment effects (relative to a placebo or "standard" comparator) tabulated, alongside measures of between-study heterogeneity if a RE model is used?*		
B4.2	*Are the absolute effects on each treatment, as they are used in the CEA, reported?*		
B5 Synthesis in other parts of the natural history model			
B5.1	*Is the choice of data sources to inform the other parameters in the natural history model adequately described and justified?*		
B5.2	*In the natural history model, can all the differences between treatments be explained by their differences on randomised trial outcomes?*		

C. ISSUES SPECIFIC TO NETWORK SYNTHESIS ***C1 Adequacy of information on model specification and software implementation***			
C1.1	*Is the statistical model described, or was a citation for the statistical model given?*		
C1.2	*Is the source of the computer code used in the synthesis cited?*		
C1.3	*Is programming code for the synthesis provided?*		
C2. Multi-arm trials			
C2.1	*If there are multi-arm trials, have the correlations between the relative treatment effects been taken into account?*		
C3 Connected and disconnected networks			
C3.1	*Is the network of evidence based on randomised trials connected?*		
C4 Inconsistency			
C4.1	*How many inconsistencies could there be in the network?*		
C4. 2	*Are there any* a priori *reasons for concern that inconsistency might exist, due to systematic clinical differences between the patients in AB, AC, etc trials?*		
C4.3	*Have adequate checks for inconsistency been made?*		
C4.4	*If inconsistency was detected, what adjustments were made to the analysis, and how was this justified?*		
D EMBEDDING THE SYNTHESIS IN A PROBABILISTIC COST EFFECTIVENESS ANALYSIS ***D1. Uncertainty Propagation***			
D1.1	*Has the uncertainty in parameter estimates been propagated through the model?*		
D2 Correlations			
D2.1	*Are there correlations between parameters? If so, have the correlations been propagated through the model?*		