

**NICE DSU TECHNICAL SUPPORT DOCUMENT 1:
INTRODUCTION TO EVIDENCE SYNTHESIS
FOR DECISION MAKING**

REPORT BY THE DECISION SUPPORT UNIT

April 2011
(last updated April 2012)

Sofia Dias¹, Nicky J Welton¹, Alex J Sutton², AE Ades¹

¹School of Social and Community Medicine, University of Bristol, Canynge Hall, 39
Whatley Road, Bristol BS8 2PS, UK

²Department of Health Sciences, University of Leicester, 2nd Floor Adrian Building,
University Road, Leicester LE1 7RH, UK

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734

E-mail dsuadmin@sheffield.ac.uk

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University. The DSU is commissioned by The National Institute for Health and Clinical Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme.

Please see our website for further information www.nicedsu.org.uk

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

The NICE Guide to the Methods of Technology Appraisalⁱ is a regularly updated document that provides an overview of the key principles and methods of health technology assessment and appraisal for use in NICE appraisals. The Methods Guide does not provide detailed advice on how to implement and apply the methods it describes. This DSU series of Technical Support Documents (TSDs) is intended to complement the Methods Guide by providing detailed information on how to implement specific methods.

The TSDs provide a review of the current state of the art in each topic area, and make clear recommendations on the implementation of methods and reporting standards where it is appropriate to do so. They aim to provide assistance to all those involved in submitting or critiquing evidence as part of NICE Technology Appraisals, whether manufacturers, assessment groups or any other stakeholder type.

We recognise that there are areas of uncertainty, controversy and rapid development. It is our intention that such areas are indicated in the TSDs. All TSDs are extensively peer reviewed prior to publication (the names of peer reviewers appear in the acknowledgements for each document). Nevertheless, the responsibility for each TSD lies with the authors and we welcome any constructive feedback on the content or suggestions for further guides.

Please be aware that whilst the DSU is funded by NICE, these documents do not constitute formal NICE guidance or policy.

Dr Allan Wailoo

Director of DSU and TSD series editor.

ⁱ National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, 2008 (updated June 2008), London.

Acknowledgements

The DSU thanks Julian Higgins, Alec Miners, Georgia Salanti, Mike Spencer and the team at NICE led by Janet Robertson for reviewing this document. The editor for the TSD series is Allan Wailoo.

The production of this document was funded by the National Institute for Health and Clinical Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

This report should be referenced as follows:

Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>

EXECUTIVE SUMMARY

This document serves as a brief introduction and orientation to the series of Technical Support Documents on Evidence Synthesis. It describes (Section 2) the overall analytic approach to evidence synthesis, in terms of separate models for the “baseline”, that represents the natural history in the specified target population under a standard comparator, and the model for the treatment effects relative to that standard. It then clarifies the impact of the decision making context on synthesis methodology. It concludes that, in order for the estimates of treatment effects to be appropriate for cost-effectiveness analysis and comparative effectiveness research alike (and indeed for any scientific enquiry into the relative efficacy of medical interventions), the decision maker should restrict the study inclusion criteria to a target patient population that has a relatively narrow definition. In the specific context of cost-effectiveness analysis, the document then refers to the need for synthesis methods that are compatible with the use of probabilistic methods.

Criteria for inclusion of treatments are described (Section 3), distinguishing between the comparator set of treatments in the decision analysis, and the comparator set of treatments used in synthesis. Once a target patient population has been defined, a suggested trial inclusion rule that avoids potential ambiguity regarding the relevance of evidence is to include any trial that compares at least two treatments in the synthesis comparator set. We suggest that any exceptions to this should require special justification. Trials with small sample sizes or trials with sample size motivated for particular reasons (such as “non-inferiority trials”) should not be excluded unless specific reasons can be given. Attention should be drawn to known or potential effect modifiers, biases in the trial evidence, or markers associated with risk of bias, as consideration should be given to the potential role of methods for addressing heterogeneity and bias. Some guidance is suggested on approaches that can be taken to disconnected evidence networks.

Section 4 sets out some “good practice” in presenting the evidence base for relative treatment effects, transparency and reproducibility of methods, presentation of results, and presentation of model critique and model selection. We encourage the use of network diagrams and tables to clarify exactly what comparative evidence is used in the relative efficacy synthesis, and the production of tables that show both the relative and the absolute effects of treatments that are taken forward for use in the cost-effectiveness analysis. Justification and explanation should

be given for choice of statistical model, for example fixed or random effects synthesis. Sensitivity analyses should be used to explore robustness of conclusions to variations in assumptions, when the evidence is compatible with alternative interpretations or analyses. The document ends with a brief description of the other documents in the series.

CONTENTS

1. INTRODUCTION	8
2. OVERALL ANALYTIC APPROACH TO SYNTHESIS FOR COST EFFECTIVENESS ANALYSIS	9
2.1. SEPARATE MODELS FOR BASELINE NATURAL HISTORY AND FOR RELATIVE TREATMENT EFFECTS	9
2.2. THE DECISION-MAKING CONTEXT AND METHODOLOGY FOR EVIDENCE SYNTHESIS.....	9
2.3. ANALYTIC METHODS COMPATIBLE WITH PROBABILISTIC COST-EFFECTIVENESS ANALYSIS.....	11
3. DATASET FOR EVIDENCE SYNTHESIS OF RELATIVE TREATMENT EFFECTS	11
3.1. COMPARATOR SETS FOR DECISION AND FOR SYNTHESIS	12
3.2. TRIAL INCLUSION AND EXCLUSION	12
3.3. EXTENSION OF THE SYNTHESIS COMPARATOR SET	14
3.4. RECORDING OF TREATMENT EFFECT MODIFIERS	15
3.5. DEALING WITH DISCONNECTED NETWORKS.....	15
4. PRESENTATION OF THE EVIDENCE AND THE RESULTS OF EVIDENCE SYNTHESIS	16
4.1. THE EVIDENCE BASE	16
4.2. DESCRIPTION OF SYNTHESIS METHODOLOGY	17
4.3. PRESENTATION OF SYNTHESIS RESULTS	18
4.4. MODEL DIAGNOSIS AND SELECTION	20
5. USING THE TECHNICAL SUPPORT DOCUMENTS	20
6. REFERENCES	22

TABLES AND FIGURES

Table 1 Certolizumab Pegol (CZP) for Rheumatoid Arthritis: ²⁸ number of patients achieving ACR50 at 6 months, out of the total number of patients, in 12 trials comparing 6 treatments with Placebo, and mean disease duration (in years) for patients in each trial. Blank cells indicate that the treatment was not compared in that trial. All trial arms had Methotrexate in addition to the placebo or active treatment.	17
Table 2 Parkinson example: posterior mean, standard deviation (sd) and 95% Credible interval (CrI) of the mean off-time reduction for the fixed effects models for the treatment effects of Treatments 2 to 5 relative to Placebo, and absolute mean off-time reduction for Placebo and treatments 2 to 5.....	18
Table 3 Parkinson example: posterior mean, standard deviation (sd) and 95% Credible interval (CrI) of the relative effect of treatment Y compared to X for all possible treatment comparisons, for the network meta-analysis and separate pairwise meta-analyses with fixed effects.....	19
Figure 1 Treatment network where the decision and synthesis comparator sets differ. Lines represent a comparison of the connected treatments in at least one trial. Treatments relevant to the decision are in bold. Treatment X has been added to the synthesis set because it links treatment C to the rest of the network (dashed lines).....	12
Figure 2 Disconnected treatment network. Lines represent a comparison of the connected treatments in at least one trial. The network formed by treatments A, B and C is not connected to the network formed by treatments X and Y.....	16
Figure 3 Parkinson network: 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.	19

Abbreviations and definitions

CEA	Cost-effectiveness analysis
CrI	Credible interval
DMARD	Disease Modifying Anti-Rheumatics Drugs
IPD	Individual patient data
RCT	Randomised controlled trial
TSD	Technical Support Document

1. INTRODUCTION

This Technical Support Document (TSD) is the first in a series of seven documents on evidence synthesis methods in decision making. These are intended to support those presenting submissions to the National Institute for Health and Clinical Excellence (NICE). While the 2008 revision of the NICE Methods Guide¹ sets out what properties submissions should have, the evidence synthesis TSDs show how the required analyses can be implemented. There is no attempt to “prescribe” the form that analyses must take or the methods that must be used. Any methods can be used, as long as they have the required properties. The purpose of the TSDs is to further explain these properties, to set out the implications for what kinds of analyses meet the requirements, and to describe some methods that can be used.

The series is intended primarily to address the issues of the relative and absolute efficacy of interventions. Although oriented to the health technology appraisal process, they would be equally relevant to clinical guideline development, or to comparisons between medical devices, or between public health interventions. The series does not, however, cover questions arising in the synthesis of data on diagnostic test performance, or general synthesis of epidemiological data.

This document sets the scene for the remaining TSDs, providing a general outline of the overall evidence synthesis process, and referring forward to subsequent TSDs for more detailed information. We begin (Section 2) by describing the overall analytic approach, the distinction between a “baseline” model, and the relative treatment effect usually informed by randomised controlled trial (RCT) evidence. The distinction between evidence synthesis undertaken specifically for decision making and more general ideas of synthesis as a summary of what appears in the literature is also explored. Section 3 is about inclusion criteria for treatments and for trials. Here we suggest some simple rules that provide answers to questions such as: how far should an evidence network be extended? We also note the need to address potential effect modifiers or biases. Section 4 deals with presentation of the evidence base, the synthesis methodology, the synthesis results, and the attention that must be given to model diagnosis and model choice. Finally, Section 5 provides a brief description of the other documents in the series, guiding users to the most appropriate document for different analyses.

2. OVERALL ANALYTIC APPROACH TO SYNTHESIS FOR COST EFFECTIVENESS ANALYSIS

2.1. SEPARATE MODELS FOR BASELINE NATURAL HISTORY AND FOR RELATIVE TREATMENT EFFECTS

The most common formula that is followed in constructing cost-effectiveness analyses (CEAs) is to divide the model into two distinct components: a “baseline” natural history model, which in effect represents the natural history of patients with the target condition under a “standard” comparator treatment, and the relative treatment effects. The natural history for the “treatment arms” is based on a second model where the relative treatment effects are appropriately combined with the baseline model to generate predicted absolute effects for the non-standard treatments. The point in the natural history model at which the relative effects are considered to act depends on the model. In models for Rheumatoid Arthritis, for example, the treatment effect generally acts on the probability that patients will stay on treatment beyond the initial 3 to 6 months trial period.^{2,3} In models of smoking cessation, treatments are assumed to act on the probability of relapse within say 12 months.^{4,5} A common, though not invariable feature of such natural history models is that, conditional on the initial response, the subsequent “downstream” outcomes do not depend on treatment. Where Markov models are used, treatment effects are assumed to act on one or more transition probabilities.

The 2008 Methods Guide¹ requires that systematic literature search should be used to identify the data sources for *both* baseline, and relative treatment effects. TSD5⁶ gives further details on methods for constructing the baseline model, while synthesis of relative effects is covered in TSD2.⁷

2.2. THE DECISION-MAKING CONTEXT AND METHODOLOGY FOR EVIDENCE SYNTHESIS

The methods proposed throughout this series of documents are designed for decision making and, indeed, for probabilistic modelling (see below). It is important, however, to be clear that the decision making context impresses itself on evidence synthesis methodology in a way that goes far beyond the choice of analytic methods and software. A decision maker has to have in

mind a well-defined *decision problem*. This will consist of a *target population* of individuals not only with a specific condition, but probably at a specific point in a disease progression pathway. Accompanying this there will be a set of comparator interventions. Each candidate intervention has to consist of a pre-specified dosing regimen and/or a particular mode of delivery, possibly with concomitant treatments that must also be defined. There is latitude, of course, for dosing regimens that start at one level and may then be increased or decreased according to a pre-specified monitoring system. However, it has to be possible, at the end of the CEA to ascribe a single cost and a single effect to each intervention alternative. Costs may be variable, and if so they should be appropriately averaged. Effects may vary as well, and may need to be modelled. But, eventually, expected lifetime costs and effects must be calculable, if necessary under some set of clearly stated assumptions.

As a result it has long been recognised that the trial inclusion criteria for a synthesis undertaken for decision making, whether based on CEA or on an efficacy analysis alone,⁸ is likely to be drawn more narrowly than for syntheses whose objective can be described, perhaps, as a summary of literature. Likewise, previous authors have distinguished between evidence synthesis as “science” and evidence synthesis as “summary”,⁹ and the common use of random effect models to average over important heterogeneity has attracted particular criticism.¹⁰ Expressing a similar view, the Cochrane Handbook¹¹ states: “meta-analysis should only be considered when a group of studies is sufficiently homogeneous in terms of participants, interventions and outcomes to provide a meaningful summary”.

Nevertheless, this advice is not always followed. For example, an overview of reviews of treatments for enuresis presented trials that involved young children who were being treated for the first time alongside trials of older children and adolescents who had failed on the first-line treatments.¹² Similarly, in treatments for Rheumatoid Arthritis, one cannot assess the relative efficacy of biologics by combining trials of biologics in patients who have failed on standard therapies, with trials of patients who have failed on biologics.¹³ Summaries of effect sizes in these cases lack validity because young children and adolescents with enuresis, or patients who are naïve to biologics and those who have failed on them, represent different clinical populations and require different decisions, informed by different summaries of treatment effects. The effect of the decision-making context, therefore, is often to markedly reduce the *clinical heterogeneity* of the trials assembled for synthesis.

Synthesis of evidence from clinically heterogeneous populations, not only increases the risk of statistical heterogeneity and inconsistency, but often requires highly implausible assumptions, such as assuming that interventions are equally effective in a naïve population

or in a population that has already failed on that intervention or has contra-indications to its use.

The focus of this series of TSDs is, therefore, to provide a formal underpinning of studies of relative efficacy in specific decision contexts. These synthesis methods can be applied equally well within CEA, Comparative Effectiveness Research,¹⁴ or within intermediate forms of analysis that attempt to weigh the benefits of treatments against harms.¹⁵ What distinguishes the decision context from a literature summary rationale for synthesis appears to be the clinical heterogeneity and the degree to which it needs to be addressed. The methods we propose can be used outside what we have referred to as a decision making context, but certain key assumptions (TSD2⁷) may not hold. Furthermore, without a clearly defined decision problem it is not clear how the results should be interpreted.

2.3. ANALYTIC METHODS COMPATIBLE WITH PROBABILISTIC COST-EFFECTIVENESS ANALYSIS.

There are a wide range of methods for synthesising information on relative treatment effects. The NICE methods guide¹ states that probabilistic methods^{16,17} are preferred for cost-effectiveness analyses. This puts some constraints, if not on the kinds of synthesis methods that must be used, then certainly on the way outputs from the synthesis are fed into the CEA. The synthesis methods described in TSDs 2, 3 and 5 are specifically designed to be compatible with probabilistic CEA. A set of alternative approaches to incorporate the results from the synthesis into the CEA are described in TSD6.¹⁸

3. DATASET FOR EVIDENCE SYNTHESIS OF RELATIVE TREATMENT EFFECTS

Within the context of NICE Technology Appraisals or Guidelines, an initial scoping exercise will have defined the target population(s) for whom one or more decisions are to be made. The specification may refer, for example, to a stage of disease progression, a level of disease severity, previous treatments that have failed or that are contra-indicated for whatever reason, or to various combinations of these. The wider the definition of the target population, the greater the potential for clinical and statistical heterogeneity, as discussed in Section 2, raising the possibility of different decisions in different groups of patients.

3.1. COMPARATOR SETS FOR DECISION AND FOR SYNTHESIS

An initial set of treatments that are to be compared will have been identified in the scoping exercise. We will call this the *decision comparator set*. Ideally, this should include all the candidate treatments for the target population in question. If it is not possible to form a connected network of comparisons of these treatments based on randomised data, it may be possible to introduce further treatments so that a connected network can be formed. For example, in Figure 1 the treatments A and B have been compared in RCTs, but treatment C has not been compared to either A or B. An additional treatment X has been introduced because there are AX and CX trials, forming a connected network. Thus, the *synthesis comparator set* consists of A, B, C, and X and is different from the decision comparator set A, B, C.

Other ways to deal with disconnected networks are discussed in Section 3.5.

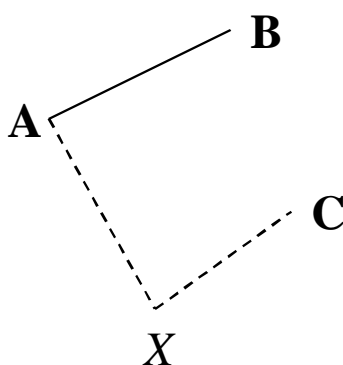


Figure 1 Treatment network where the decision and synthesis comparator sets differ. Lines represent a comparison of the connected treatments in at least one trial. Treatments relevant to the decision are in bold. Treatment X has been added to the synthesis set because it links treatment C to the rest of the network (dashed lines).

3.2. TRIAL INCLUSION AND EXCLUSION

Once the target population and the synthesis comparator set have been defined, the trial inclusion criteria are, in principle, relatively easy to implement. All trials on the target patient population *that compare two or more of the treatments from the synthesis comparator set* should be included. This wording is constructed with the intent of producing an unambiguous definition that eliminates the risk of what could be perceived as “gaming” with trial inclusion and exclusion criteria, that is, deliberately leaving out or including trials in order to obtain a particular result. It also clarifies that, even though the key comparisons might be considered

to be those involving a particular active treatment and one or more standard comparators, “head-to-head” trials comparing similar active treatments *must* also be included. Note that if there was a three arm trial comparing A vs B vs Z, and treatment Z is not in the synthesis comparator set, the Z arm would be dropped as it adds no further information, but the A vs B comparison would be used, as long as it was on the same patient population. It is important to include all the trials that compare any of the treatments in the synthesis comparator set. This is because decisions are usually based on an incremental CEA, which means that, for any treatments A, B, C, trials comparing AB, AC, and BC must all be considered equally informative. Indeed, in this context, it is technically impossible to distinguish “direct” and “indirect” evidence – direct evidence for one model parameter is indirect evidence for another.

It is sometime argued that trials which are not “double-blind”, or trials which are small should be excluded. In general, no trials that would be included if the above rules were followed should be excluded without a clear reason being given. For example, one conceivable reason for excluding small trials might be on the basis that they were more likely to be biased, often termed ‘small study bias’, where trial size is used as a proxy for factors associating trial quality with increased effect size. However, in this case, it would be prudent to consider how sensitive the CEA would be to inclusion or exclusion of “small” trials, and also what effect bias adjustment procedures might have (see TSD3¹⁹). Another reason that is sometimes given for excluding a trial is on the basis that it was designed as a “non-inferiority trial”. This is not, however, a valid reason for exclusion. The rationale for the sample size calculations has no obvious connection to the size or direction of the effect estimate.

Another reason given for not including data is that outcomes are reported in a different way. We show in later documents^{6,7,19} how different outcomes can in many cases be synthesized in a single coherent analysis, although this may not always be possible. However one frames the “rules” for trial inclusion and exclusion, subjective judgement is required about whether a trial’s target population does or does not match the target population for the present decision, or there may be doubts on whether the interventions compared are sufficiently similar to those defined in the synthesis comparator set. Investigators may wish to argue that the relative treatment effects will be the same in this group and that such trials should be included, or that they are not, in which case they would be excluded. In such cases it may be possible to use bias or covariate adjustment methods (TSD3¹⁹). Alternatively, some of the trial evidence could involve a mixture of patients, for example children and adults, while the target population for decision is adults alone. Whether or not such evidence should be

included depends on whether it can be used to inform the parameter of interest. Regression adjustment for the proportion of adults, for example, may be possible (see TSD3¹⁹), but the comments on ecological and aggregation biases in TSD3¹⁹ (Section 1) must be taken into account. Whenever trial inclusion or exclusion is debatable or depends on judgement, sensitivity analyses should be used to assess the impact on the decision.

3.3. EXTENSION OF THE SYNTHESIS COMPARATOR SET

It is sometimes asked: how far should a network be extended? There is no specific need to extend the synthesis comparator set beyond the comparator set for decision making, unless such an extension is required to produce a connected network. Extending the network further does have potential advantages such as, for example, strengthening inference,²⁰ producing a more robust analysis which is less sensitive to specific sources of data, the ability to check consistency more thoroughly (see TSD4²¹), and the ability to identify effect modifiers or carry out bias adjustment (see TSD3¹⁹). Thus, while extension of the network is not ruled out, in the spirit of the 2008 Methods Guide,¹ it would not be considered as the “base-case” analysis. Further, counter-balancing the potential advantages outlined above, there are a number of potential *disadvantages*. These would include an increased danger of introducing effect modifiers as the more “remotely” connected treatments are more likely to have been trialled on somewhat different patient populations. There may be a particular danger in extending the network to include treatments that were trialled earlier, particularly if – as is so often the case – date is associated with severity of condition.

The situation in Figure 1, where a new treatment is introduced to connect the network, suggests an exception to the above. In this case, if the network could have been connected either by the addition of BY and CY trials, *or* by the addition of AX and CX trials, then it would seem logical that *both* X and Y should be added to the synthesis comparison set. Then, in accordance with the rules given for trial inclusion, X vs Y trials would need to be included as well.

Finally, another reason for extending the set of trials beyond the set of immediate comparators is to estimate relationships between outcomes and use this to fill gaps in the evidence base. For example, a cost-effectiveness model for cancers may require information on percent responders, on time to tumour progression, and time to death. Some of this information may not be reported within the synthesis comparator set. Under these circumstances it may be possible to estimate the relationships from a wider set of trials, either

in the same statistical analysis (preferable) or separately, and then apply the estimated relationship to the treatments where it is missing. This approach has been adopted in advanced breast cancer²² but it can be applied in any situation where there are related multiple outcomes. Another example would be in treatments for influenza where trials may report one or more of time to end of fever, time to end of symptoms, time to return to work, and so on.^{23,24} The advantage of these more complex forms of synthesis is that they provide more stable estimates where data are sparse, and allow investigators to produce estimates for less commonly reported outcomes that are appropriately “calibrated” to the available data. However, they require a range of assumptions, particularly regarding the constancy of certain relationships between outcomes, that need to be checked against expert clinical opinion, and wherever possible, empirically.

3.4. RECORDING OF TREATMENT EFFECT MODIFIERS

It is important that investigators thoroughly research previous literature and consult clinical experts to assess whether there are known or likely treatment effect modifiers for any of the interventions in the comparator set for synthesis. The presence of effect modifiers in the trials should be recorded, and taken into consideration during the synthesis, as they may have a bearing on heterogeneity and bias adjustment (TSD3¹⁹) or inconsistency in networks (TSD4²¹).

3.5. DEALING WITH DISCONNECTED NETWORKS

Although the focus of the TSDs is for situations where a connected evidence network can be constructed from randomised evidence, there may be situations like the one depicted in Figure 2 where the evidence forms two, or even more, disconnected elements. In these cases the methods described in TSD2⁷ cannot be applied without further assumptions being made.

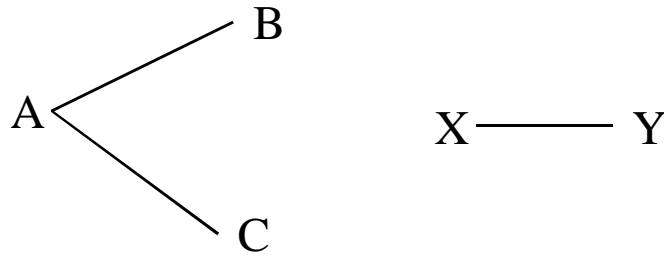


Figure 2 Disconnected treatment network. Lines represent a comparison of the connected treatments in at least one trial. The network formed by treatments A, B and C is not connected to the network formed by treatments X and Y.

These assumptions must take the form of a relative treatment effect linking the two networks. For example an assumption may be made about the relative efficacy of treatment X relative to A, of treatment Y relative to A, *or* of treatment X relative to B, and so on. From a technical point of view a distribution for such a treatment effect can readily be inserted into a synthesis (see methods for shared parameter models in TSD2⁷) and expert opinion or observational data can be used to inform the treatment effect. The difficulty, of course, lies in making a convincing case for the central estimate of the “missing” treatment effect and assigning it a variance that appropriately reflects the increased uncertainty attaching to a non-randomised comparison. The increased uncertainty will affect comparisons between, but not within, the two networks.

4. PRESENTATION OF THE EVIDENCE AND THE RESULTS OF EVIDENCE SYNTHESIS

4.1. THE EVIDENCE BASE

Where there are more than two treatments in the comparator set for synthesis, it is helpful to show a network diagram (such as Figure 1). Software for automatically drawing such diagrams is available as stand-alone programs with multiple capabilities (e.g. Pajek²⁵) or as packages and routines developed for R²⁶ among other programs.

Further refinements in network diagrams which can be implemented in various software include: adding the number of studies making that comparison to each connecting line (see Figure 3); having the thickness of the connecting lines reflect the number of trials on that

contrast and the size of the vertices reflect the number of patients included in that comparison.²⁷ A second useful presentation is a table of the sort shown as Table 1. This has a separate row for each trial, with columns for each treatment where the presence of data indicates which treatments have been compared in each trial arm. It is also convenient to add to the table trial level covariate values, or to indicate whether individual participant data (IPD) are available. Another useful representation of the same data is given in TSD3.¹⁹

Table 1 Certolizumab Pegol (CZP) for Rheumatoid Arthritis:²⁸ number of patients achieving ACR50 at 6 months, out of the total number of patients, in 12 trials comparing 6 treatments with Placebo, and mean disease duration (in years) for patients in each trial. Blank cells indicate that the treatment was not compared in that trial. All trial arms had Methotrexate in addition to the placebo or active treatment.

Study Name	Placebo	CZP	Adalimumab	Etanercept	Infliximab	Rituximab	Tocilizumab	Mean disease duration (years)
Kim 2007	9/63		28/65					6.85
DE019	19/200		81/207					10.95
ARMADA	5/62		37/67					11.65
RAPID 1	15/199	146/393						6.15
RAPID 2	4/127	80/246						5.85
START	33/363				110/360			8.1
ATTEST	22/110				61/165			7.85
Abe 2006*	0/47				15/49			8.3
Weinblatt 1999	1/30			23/59				13
Strand 2006	5/40					5/40		11.25
CHARISMA*	14/49						26/50	0.915
OPTION	22/204						90/205	7.65

* ACR50 at 3 months

4.2. DESCRIPTION OF SYNTHESIS METHODOLOGY

In the interests of transparency, reviewers and general readers must be provided with sufficient information that would allow them to exactly reproduce the analyses, if they had access to the data. If possible journal citations for the precise model of the data being assumed, and/or citation of the source of software code, must be provided. Otherwise the statistical model for the synthesis should be set out fully in algebraic form. The software code used for the synthesis should be annotated and made available, along with the data used, although in some situations confidentiality requirements may prevent the latter from being released.

A clear discussion of the underlying statistical and clinical assumptions implied by the model, and their impact on the final decision should also be provided. In particular, reasons for

choosing to model the outcomes on a particular scale (e.g. odds ratio, hazard ratio, risk difference etc) and the assumptions implied in any transformation from the relative to the absolute effects should be clearly presented (see TSD2⁷, Section 2).

4.3. PRESENTATION OF SYNTHESIS RESULTS

Although the parameters required by the CEA tend to be the absolute treatment effects of each treatment, it is essential that both the relative treatment effects, which are the outputs from the evidence synthesis, and the absolute effects on which the CEA is based are presented. It is important for those reviewing and evaluating submissions that there is absolute clarity and transparency about exactly what relative efficacies between treatments are being assumed, and exactly what absolute effects are going forward into the CEA. This can be achieved in a number of ways. Perhaps the simplest method is a table of the mean treatment effect with 95% Credible Intervals (CrI), of every treatment relative to placebo, or to a standard comparator. Table 2 gives an example, of the relative and absolute mean off-time reduction in patients given four dopamine agonists and Placebo as adjunct therapy for Parkinson’s disease. The treatment network is presented in Figure 3 (for further details see TSD2⁷, Example 5).

Table 2 Parkinson example: posterior mean, standard deviation (sd) and 95% Credible interval (CrI) of the mean off-time reduction for the fixed effects models for the treatment effects of Treatments 2 to 5 relative to Placebo, and absolute mean off-time reduction for Placebo and treatments 2 to 5.

		mean	sd	95% CrI
<i>X</i>	<i>Y</i>	Relative effects of treatment <i>Y</i> compared to <i>X</i>		
<i>Placebo</i>	<i>Treatment 2</i>	-1.81	0.33	(-2.46,-1.16)
<i>Placebo</i>	<i>Treatment 3</i>	-0.47	0.49	(-1.43,0.49)
<i>Placebo</i>	<i>Treatment 4</i>	-0.52	0.48	(-1.46,0.43)
<i>Placebo</i>	<i>Treatment 5</i>	-0.82	0.52	(-1.84,0.22)
		Absolute treatment effects		
	<i>Placebo</i>	-0.73	0.22	(-1.16,-0.30)
	<i>Treatment 2</i>	-2.54	0.40	(-3.32,-1.76)
	<i>Treatment 3</i>	-1.21	0.53	(-2.25,-0.15)
	<i>Treatment 4</i>	-1.25	0.53	(-2.28,-0.21)
	<i>Treatment 5</i>	-1.55	0.57	(-2.66,-0.43)

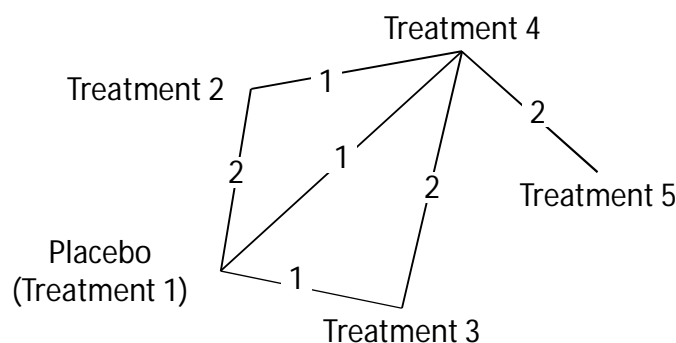


Figure 3 Parkinson network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison.

Another format (Table 3) allows investigators to contrast results of pair-wise meta-analyses with the results of a network synthesis. Although this table does not constitute a formal analysis of inconsistency in the network (see TSD4²¹), if the direct estimates are very close to their network counter-parts, there may be no need to proceed with further consideration of inconsistency. Similarly, graphical displays such as forest plots which summarise the results in the tables along with the raw effect estimates from each trial informing that treatment comparison, can also be presented. TSD3¹⁹ shows an example of these plots (Example 2 in the Appendix).

Table 3 Parkinson example: posterior mean, standard deviation (sd) and 95% Credible interval (CrI) of the relative effect of treatment Y compared to X for all possible treatment comparisons, for the network meta-analysis and separate pairwise meta-analyses with fixed effects.

X	Y	Network Meta-analysis			Pairwise Meta-analyses		
		mean	sd	95% CrI	mean	sd	95% CrI
Placebo	Treatment 2	-1.81	0.33	(-2.46,-1.16)	-1.83	0.34	(-2.49,-1.17)
Placebo	Treatment 3	-0.47	0.49	(-1.43,0.49)	-0.31	0.67	(-1.62,1.00)
Placebo	Treatment 4	-0.52	0.48	(-1.46,0.43)	-0.90	0.69	(-2.26,0.46)
Placebo	Treatment 5	-0.82	0.52	(-1.84,0.22)	-	-	-
Treatment 2	Treatment 3	1.34	0.54	(0.28,2.41)	-	-	-
Treatment 2	Treatment 4	1.29	0.52	(0.27,2.32)	1.40	0.70	(0.03,2.77)
Treatment 2	Treatment 5	0.99	0.56	(-0.10,2.10)	-	-	-
Treatment 3	Treatment 4	-0.04	0.32	(-0.68,0.59)	0.00	0.35	(-0.68,0.68)
Treatment 3	Treatment 5	-0.34	0.38	(-1.10,0.41)	-	-	-
Treatment 4	Treatment 5	-0.30	0.21	(-0.71,0.11)	-0.30	0.21	(-0.71,0.11)

A number of authors tabulate the probability that each treatment is best, which is an output available from Bayesian or other simulation-based approaches. This should be treated with

great caution, however, particularly when there are many treatment alternatives. A treatment whose mean effect ranks quite low may still have a high probability of being best if there is relatively more uncertainty in its mean effect. This is misleading because for a given set of expected (mean) treatment effects, greater uncertainty may flatter a treatment. Differences between treatments in probability of being best of less than 90% cannot be given much credence. A more reliable indicator is a plot of the rankings of each treatment. When there are multiple outcomes, for example remission, relapse, discontinuation due to side effects, “rankograms” for each treatment, which plot the ranks on each outcome, can be very informative. Readers are referred to Cipriani et al.²⁹, where these plots first appeared, and to Ades et al.³⁰ for a further application. Salanti et al.³¹ suggest a number of graphical presentations of results that may be useful in different circumstances.

4.4. MODEL DIAGNOSIS AND SELECTION

It is essential that the model that is chosen is compatible with the evidence. While questions of statistical inference are less relevant in decision making, where a “balance of evidence” approach, takes precedence over “significance testing”, the question of whether the model fits the data is still relevant. Decisions must be based on an internally coherent model, which is compatible with all the available evidence. Methods for assessing model fit are discussed in TSD2,⁷ and there is a discussion of heterogeneity and outlier detection in TSD3.¹⁹ Where several synthesis models were examined, an explanation must be provided for the choice of model that is put forward into the CEA. Choice of fixed or random effects models for synthesis of relative treatment effects is a common example (see TSD2⁷). Where the evidence is compatible with more than one model, a sensitivity analysis can be used to establish whether or not the CEA results are sensitive to model choice.

5. USING THE TECHNICAL SUPPORT DOCUMENTS

TSD2⁷ presents a general framework for pairwise and network meta-analysis for different data types, with details of fixed and random effects models for different outcome types, assumptions and applications. It highlights the importance of using the appropriate model for the type of data available and shows how to combine data given in different formats.

TSD3¹⁹ considers outlier detection and covariate and bias adjustment in detail. Models for different types of regression in both pairwise and network meta-analysis are given with a

discussion of the assumptions and risks of meta-regression, as well as its implications for decision-making.

TSD4²¹ defines inconsistency in a network meta-analysis. It describes how to detect and locate inconsistency, its likely causes and steps that should be taken to avoid it.

TSD5⁶ provides guidance on how a baseline model can be informed and built, discusses sources of evidence which can be used to inform a baseline model, and provides examples of baseline models.

TSDs 2, 3, 4 and 5 include an extensive set of worked examples of analyses using Bayesian Markov chain Monte Carlo methods in WinBUGS and provide generic code applied to illustrative examples. The WinBUGS code is also available for download as WinBUGS system (.odc) files (users are advised to download these files instead of copying and pasting code from the TSDs).

Although the models, worked examples and WinBUGS code in TSDs 3, 4 and 5 are for binary data, TSD2 shows how they can be changed to suit other types of data.

TSD6¹⁸ describes some of the software which can be used to implement the models described in TSDs 2 to 5 and to incorporate the results of the evidence synthesis into a decision model.

Finally, TSD7³² provides a reviewer's checklist. It focuses on a series of questions aimed at revealing the assumptions underlying the synthesis, the adequacy of the arguments in support of these assumptions, and aims to inform a decision on the need for further analyses.

6. REFERENCES

1. National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal (updated June 2008). 2008.
2. Brennan, A., Bansback, N., Nixon, R., Madan, J., Harrison, M., Watson, K. et al. Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. *Rheumatology* 2007; 46:1345-1354.
3. Woolacott, N.F., Bravo Vergel, Y., Hawkins, N., Kainth, A., Khadjesari, Z., Misso, K. et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2006; 10(31).
4. Woolacott, N.F., Jones, L., Ford, C.A., Mather, L.C., Sowden, A.J., Song, F.J. et al. The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation. *Health Technology Assessment* 2002; 6(16):1-245.
5. Wang, D., Connock, M., Barton, P., Fry-Smith, A., Aveyard, P., Moore, D. 'Cut down to quit' with nicotine replacement therapies in smoking cessation: a systematic review of effectiveness and economic analysis. *Health Technology Assessment* 2008; 12(2).
6. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
7. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
8. Eccles, M., Freemantle, N., Mason, J. Using systematic reviews in clinical guideline development. In: Egger M., Davey Smith G., Altman D.G., eds. *Systematic Reviews in Health Care: meta-analysis in context*. Second ed. BMJ; London: 2001; 400-409.
9. Rubin, D.B. A new perspective. In: Wachter K.W., Straf M.L., eds. *The Future of Meta-analysis*. Russell Sage Foundation; New York: 1990.
10. Greenland, S. Invited commentary: A critical look at some popular meta-analytic methods. *American Journal Of Epidemiology* 1994; 140(3):290-296.
11. Higgins, J.P.T., Green, S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 (updated February 2008). The Cochrane Collaboration, Wiley, Chichester; 2008.
12. Russell, K., Kiddoo, D. The Cochrane library and nocturnal enuresis: an umbrella review. *Evidence-Based Child Health: a Cochrane Review Journal* 2006; 1(1):5-8.

13. Wells, G.A., Sultan, S.A., Chen, L., Khan, M., Coyle, D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. *Cochrane Database of Systematic Reviews* 2009. Ottawa, Canadian Agency for Drugs and Technologies in Health.
14. Iglehart, J.K. Prioritizing comparative-effectiveness research - IOM recommendations. *New England Journal of Medicine* 2009; 361:325-328.
15. Glasziou, P.P., Irwig, L.M. An evidence-based approach to individualising treatment. *British Medical Journal* 1995; 311:1356-1359.
16. Doubilet, P., Begg, C.B., Weinstein, M.C., Braun, P., McNeill, B.J. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. *Medical Decision Making* 1985; 5:157-177.
17. Critchfield, G.C., Willard, K.E. Probabilistic analysis of decision trees using Monte Carlo simulation. *Med Decis Making, MED* 1986; 6:85-92.
18. 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; last updated April 2012; available from <http://www.nicedsu.org.uk>
19. Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
20. Cooper, N.J., Peters, J., Lai, M.C.W., Juni, P., Wandel, S., Palmer, S. et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value in Health* 2011; 14:371-380.
21. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Lu, G., Ades, A.E. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
22. Welton, N.J., Willis, S.R., Ades, A.E. Synthesis of Survival and Disease Progression Outcomes for Health Technology Assessment of Cancer Therapies. *Research Synthesis Methods* 2010; 1:239-257.
23. Welton, N.J., Cooper, N.J., Ades, A.E., Lu, G., Sutton, A.J. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Statistics In Medicine* 2008; 27:5620-5639.
24. Burch, J., Paulden, M., Conti, S., Stock, C., Corbette, M., Welton, N.J. et al. Antiviral drugs for the treatment of influenza: a systematic review and economic evaluation. *Health Technology Assesment* 2010; 13(58):1-290.
25. Batagelj, V., Mrvar, A. Pajek - Program for Large Network Analysis. *Connections* 1998; 21(2):47-57.

26. R: A Language and Environment for Statistical Computing [Windows. Vienna, Austria: R Foundation for Statistical Computing; 2010.
27. Salanti, G. Multiple-treatments meta-analysis of a network of interventions. <http://www.mtm.uoi.gr/> 2011; Available from <http://www.mtm.uoi.gr/>
28. National Institute for health and Clinical Excellence. Certolizumab pegol for the treatment of rheumatoid arthritis. 2010; TA186. National Institute for Health and Clinical Excellence. NICE technology appraisal guidance.
29. Cipriani, A., Furukawa, T.A., Salanti, G., Geddes, J.R., Higgins, J.P.T., Churchill, R. et al. Comparative efficacy and acceptability of 12 new generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373:746-758.
30. Ades, A.E., Mavranzouli, I., Dias, S., Welton, N.J., Whittington, C., Kendall, T. Network meta-analysis with competing risk outcomes. *Value in Health* 2010; 13(8):976-983.
31. Salanti, G., Ades, A.E., Ioannidis, J.P.A. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 2011; 64:163-171.
32. Ades, A.E., Caldwell, D.M., Reken, S., Welton, N.J., Sutton, A.J., Dias, S. NICE DSU Technical Support Document 7: Evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. 2012; available from <http://www.nicedsu.org.uk>