RFP Topic: Use of Network Meta-analysis to Inform Clinical Parameters in Economic Evaluations

Nicola J Cooper[1]
Alex J Sutton[1]
Felix Achana[2]
Nicky J Welton[3]

June 2015

[1] Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK
[2] Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry CV8 7AL, UK
[3] School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK

TABLE OF CONTENTS

Table

Figures

ABBREVIATIONS

CrI	credible interval
DIC	deviance information criteria
FE	fixed effects
IPD	individual participant data
MA	(pairwise) meta-analysis
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MTC	mixed treatment comparison
NMA	network meta-analysis
RCT	randomised controlled trial
RE	random effects
SR	systematic review

PREFACE

It is fascinating to note that perhaps as many as six different groups independently[1-6] it seems, started publishing extensions to pairwise meta-analysis methods to include more than two intervention options. However, it was perhaps in the context of economic evaluation, particularly to inform appraisals by the National Institute for Health and Care Excellence (NICE) in the UK, which really "forced the issue" and accelerated their uptake and development.[7] New methods were simply required if intervention decisions regarding the cost-effectiveness of more than two alternatives were to be coherent and correctly incorporate parameter uncertainty.[7,8] That is not to say there was not some concern regarding the assumptions network meta-analysis (NMA) models make and potential danger ahead if such methods were used incorrectly and uncritically. Within a few years, extensive documentation became available,[6,9-13] including much needed specific code for the required non-standard statistical software WinBUGS,[14] greatly facilitating the implementation of the methods and support for its use in a decision making context gained momentum.[15] More recently still, user-friendly add-ons to more general use statistical[16,17] and spreadsheet software[18] as well as interfaces which run WinBUGS "behind the scenes"[19-21] have been developed. As we write this early in 2015, there seems to be an increasing number of methodology papers being published in the area, suggesting this wave of synthesis methodology research has not yet peaked. With this in mind, we predict these guidelines will need to be updated quite quickly as we anticipate refinements and new methods will continue to appear.

However, NMA was not the only development in evidence synthesis to take place over the last two decades. A review of recent developments in meta-analysis published in 2008[22] covering the previous decade noted many other extensions to standard meta-analysis models to address complexities in specific contexts, such as for diagnostic data, survival data, multiple outcome data, individual participant data (IPD), and covariates had been developed. It was noted that many of these extensions had been implemented using Markov Chain Monte Carlo (MCMC) simulation methods, presumably because of the power and flexibility such an approach has when fitting "non-standard" statistical models.

What is crucial to appreciate is that these "other" synthesis developments and NMA are not mutually exclusive. Specific complexities, such as multiple outcomes and competing risks, are just as likely to be present in a NMA context where more than two intervention options are of interest. Hence, after the initial wave of papers describing the "basic" NMA methodology were published, many subsequent papers have outlined extensions to NMA which also incorporate other synthesis extensions, for example, NMA with multiple time points,[23] NMA including baseline risk as a covariate.[24] We are not in the position where all complications in evidence synthesis have been considered in a NMA context, let alone where non-MCMC software is available to implement the methods. Therefore, for what could be considered "advanced" methods, at this date in time, some knowledge of MCMC software will be required and statistical expertise essential if a bespoke synthesis model is required in a given context.

We give this background to help clarify what these guidelines do and do not aim to do. This guide focuses on the use of NMA to inform clinical parameters in economic decision modelling; although it is difficult to define a dividing line between methods and issues for NMA *per se* and those specific to economic evaluation (not least due to the fact that much use of NMA was driven by its need to inform decision models!). To this end we have navigated a course which we hope is accessible to provide necessary background to support some of the more technical topics.

1. INTRODUCTION

A key element of evidence-based healthcare evaluations is to assess the effectiveness and cost-effectiveness of all relevant competing interventions based on the available evidence. Ideally, effectiveness data are obtained from well-conducted randomised controlled trials (RCTs)[25] identified through the application of transparent systematic review methods.[26] Where multiple RCTs exist, it is well-established practice to summarise the effectiveness data by applying evidence synthesis to quantitatively combine the data in order to obtain overall pooled estimates of effectiveness. For comparisons between two healthcare interventions, it is common practice to apply pairwise meta-analysis (MA) methods[27] to obtain pooled effectiveness estimates which may be used to inform associated economic evaluation(s). However, for healthcare decision making it is necessary to consider comparisons of all relevant competing interventions to answer policy-relevant questions which may require the comparison of more than two interventions and/or interventions not previously trialled against one another.[28] In fact, where more than two interventions are being compared it is unlikely that RCTs exist that compare all the interventions of interest directly. Network meta-analysis (NMA)[2] (also known as, and equivalent to, mixed treatment comparisons (MTC)[29,30] and multiple treatment meta-analysis (MTM)[31]) extends the standard pairwise meta-analysis framework, to allow the simultaneous estimation of comparative effectiveness of multiple interventions using an evidence base of trials that individually may not compare all intervention options, but form a connected network of comparisons.

The remainder of the report is structured as follows: The next section outlines the methods used to identify the relevant literature used to compile these guidelines, followed by Section 3 which discusses the role and types of economic evaluation within health technology assessment (HTA) and Section 4 which provides a non-technical overview of NMA. Section 5 introduces a general four-step framework for interfacing NMA with economic decision models and Section 6 focuses specifically on how to use estimates from published NMAs as inputs of clinical effectiveness in economic models. To consolidate the two previous sections, Section 7 presents a decision tool to inform the optimal approach to NMA for a given decision problem. Section 8 discusses the issues associated with applying NMA to synthesise adverse/sparse event study data and Section 9 provides an overview of advanced and emerging NMA methods relevant to economic decision modelling. In Section 10 two case studies are presented which illustrate the application of standard and advanced NMA methodology to inform clinical parameters in economic evaluations. Finally, Section 11 offers some concluding remarks.

2. METHODS

This report, on the use of NMA to inform clinical parameters in economic evaluations, was compiled through a targeted review of the literature. While not exhaustive in the way a search for studies going into a systematic review/meta-analysis of a substantive topic should be, we adopt a "berry picking"[32] approach. This approach relies heavily on expanding citation searches of known published methods articles (including those referenced in the book by Welton et al.[33] and the National Institute for Health and Care Excellence [NICE] evidence synthesis technical support documents available from http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series%282391675%29.htm and also published as a series in edited form in Medical Decision Making[9-13,34,35]) as well as articles iteratively identified by "generations" of citation searches on those articles identified by the previous generation of citation searches. In addition, we have scrutinised two comprehensive lists of NMA methods publications compiled i) for the "indirect and mixed treatment comparisons" course, developed and run by some of the authors of this

report (http://www.bristol.ac.uk/social-community-medicine/projects/mpes/courses/treatment-comparisons/), and ii) by Georgia Salanti available on-line (http://www.mtm.uoi.gr/index.php/tutorial). The two case studies presented as part of this guidance were selected from the literature identified.

3. ECONOMIC EVALUATION WITHIN A HEALTH TECHNOLOGY ASSESSMENT

Healthcare decision makers worldwide are faced with the problem of how best to allocate resources within limited budgetary constraints. To enable resources to be allocated efficiently and equitably, when making decisions about which interventions to fund, decision makers need to apply an explicit framework taking into account both clinical and economic considerations. Economic evaluation, which combines both clinical outcomes and resource use in order to estimate the costs and benefits associated with competing interventions, offers such a framework.[36-38]

Economic evaluations may be conducted alongside RCTs or through decision modelling. The former addresses questions relevant to a specific RCT population whereas the latter is able to bring together all relevant evidence to make decisions for a wider population (e.g. national) level over a long-term time horizon. In this guidance we focus on the latter which provides an explicit quantitative approach to synthesise evidence from multiple sources to enable the evaluation of the cost-effectiveness of competing interventions that may not have been directly considered within a single RCT and/or where there may be limited or non-existent data on, for example, long-term costs and effects.[33,39] Decision models form an important component of health technology assessments, where decision making bodies such as Canadian Agency for Drugs and Technologies in Health (CADTH) and NICE in the UK, need to decide on which interventions to fund based on evidence-based analyses of both the clinical effectiveness and cost-effectiveness.

Decision models may be evaluated deterministically or stochastically. The former method allows the model to be evaluated analytically without the need to randomly sample from parameter distributions but ignores parameter uncertainty. However, when a model is non-linear (which many are (see Section 5.4)) the deterministic approach may calculate the point estimates of expected costs and benefits incorrectly.[40] To assess the sensitivity of the decision, obtained using this method, to the uncertainty in the model parameter values, sensitivity analysis methods such as scenario analysis may be implemented in which the decision is re-calculated based on “extreme” parameter values obtained from the literature expert opinion.[41] In contrast, the stochastic approach allows model parameters to vary randomly according to statistical or empirical distributions specified to reflect the uncertainty in their population mean value and this uncertainty is propagated through the model to determine the resulting uncertainty in the expected model outcomes.[42] The advantage of this method (often referred to as probabilistic sensitivity analysis[43,44]) is that it allows the joint impact of uncertainty in multiple model parameters on the expected costs and benefits to be investigated thus allowing a global analysis of uncertainty in the decision.[13,40] If the parameters are correlated (i.e. not independent of one another), and this has been properly specified within the model, then the probabilistic approach propagates correlations automatically and correctly computed expected costs and benefits.[40,41] Given the above, in the remainder of this report we focus exclusively on stochastic evaluation of decision models where parameter uncertainty and correlations are appropriately accounted for in the modelling.

It is beyond the remit of this report to discuss the relative merits of different types of economic decision models; readers are referred to the ISPOR Modeling Good Research Practices Task Force series for further information.[45] This report focuses on how to use results from NMA to inform parameters for stochastic cost-effectiveness analysis and thus evaluate the model of choice. Whichever model structure is being used, where possible, clinical effectiveness

parameters (i.e. relative effect(s) compared with a reference intervention) should be estimated from comparative studies, and where multiple studies exist, synthesis should be considered to obtain estimates together with appropriate uncertainty. Therefore, NMA (as described in the next section) may well be the most appropriate synthesis model to produce relative measures of clinical effect. However, most decision models require absolute effect parameters (e.g. probability of an event for each intervention strategy in the decision space); methods for deriving such parameters are considered in Section 5.3.

4. NETWORK META-ANALYSIS (NMA)

NMA extends the standard pairwise meta-analysis framework to allow the simultaneous comparison of three or more interventions using an evidence base of trials that individually may not compare all intervention options. As in pairwise meta-analysis, fixed effect (FE) and random effects (RE) models[27] may be fitted. FE models assume that all studies included in the NMA estimate the common true intervention effects and only differ due to sampling error. In contrast, RE models assume that the included studies estimate different true intervention effects but that these study-specific intervention effects are *similar* (i.e. come from a common random effects distribution). The RE models allow for, and quantify, variability beyond that expected by chance alone. Such variability is often referred to as heterogeneity and is defined as the variation in the intervention effects between trials *within* a pairwise comparison. Heterogeneity, *within* pairwise contrasts, is an issue for both pairwise and NMA and therefore it is often recommended that random effects terms be considered in NMA models.[9] To decide on the most appropriate NMA model, model fit statistics, such as the Deviance Information Criteria (DIC), (see Section 5.2) can be used. (Note that it is common in NMA to assume that heterogeneity between each pairwise contrast is equal and thus estimated by a single common heterogeneity parameter,[46] although advanced methods can relax this assumption.[29])

The additional assumptions of an NMA compared to a pairwise analysis are:

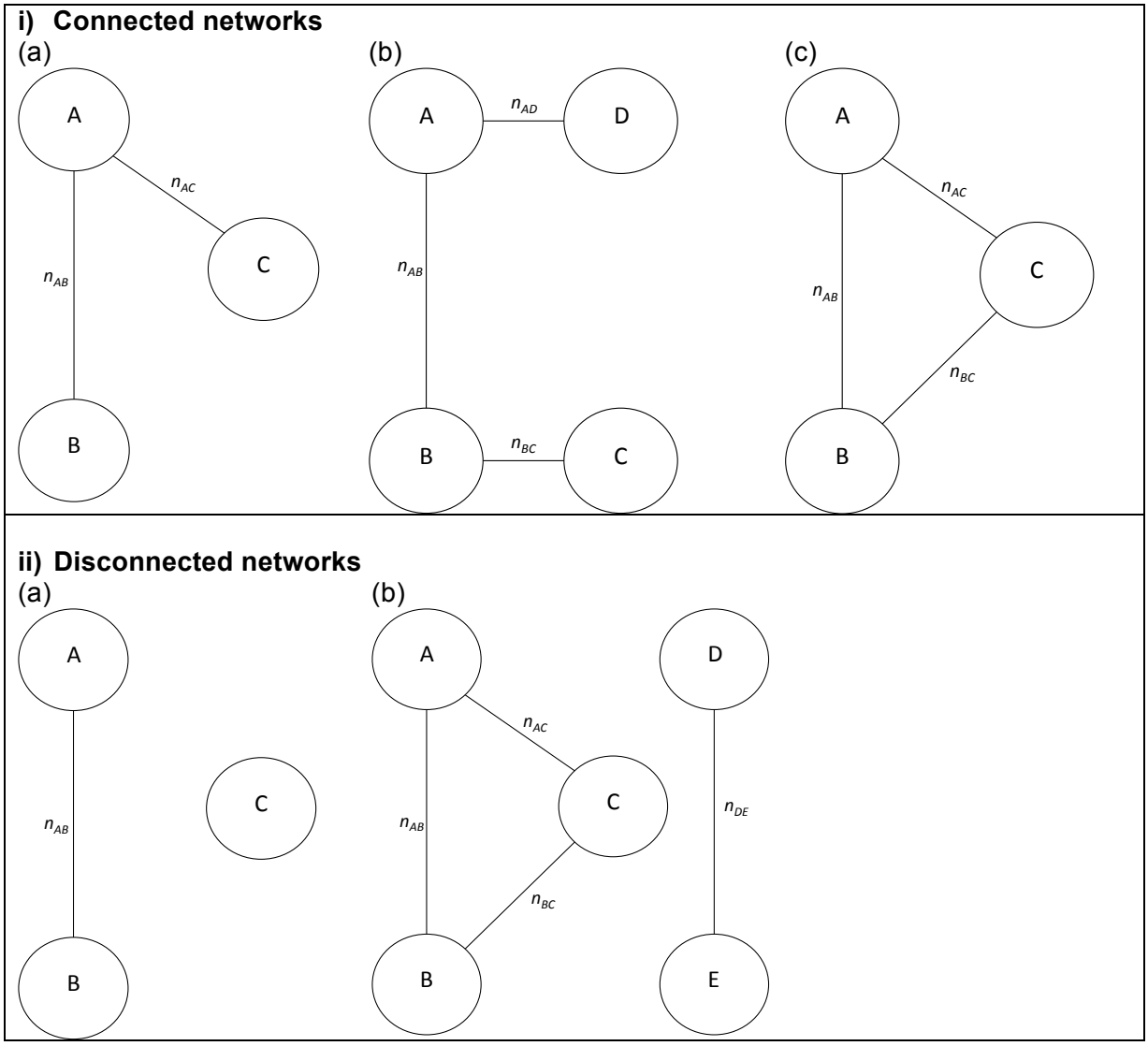
i) The trials to be included in the synthesis must form a connected network. Figure 1 (i) shows connected networks. In Figure 1 (i)(a) the lines connecting the intervention nodes indicate there are n_{AB} trials comparing interventions A versus B and n_{AC} trials comparing interventions A versus C (and no trials comparing interventions B versus C). This is considered a connected network because pathways exist between all intervention nodes along the connecting lines (i.e. in this example, the comparison of B versus C can be estimated via comparisons A versus B and A versus C). Such a network has been referred to in the literature as providing an indirect comparison[1] because an estimate of the relative effect of B versus C can be obtained despite no direct comparative trials of B versus C. Similarly in Figure 1 (i)(b) the nodes connecting the intervention nodes indicate there are n_{AB} trials comparing interventions A versus B, n_{BC} trials comparing interventions B versus C and n_{AD} trials comparing interventions A versus D (and no trials comparing interventions A versus C, B versus D and C versus D). In Figure 1 (i)(c) lines exist connecting all three interventions to each other directly. Such a network could be made up of a collection of two-arm trials (i.e. n_{AB} A versus B trials, n_{AC} A versus C trials, and n_{BC} B versus C trials). Alternatively, a single or multiple three-arm trial(s) could define this network (note, a single three-arm trial would contribute to all three pairwise comparisons in the network) or the network could be defined by a combination of two- and three-arm trials. Also, note that the network may include interventions not of direct interest to the decision problem being addressed (Section 5.1).

   Finally Figure 1 (ii) shows disconnected networks. Figure 1 (ii)(a) shows a network with an isolated node, where intervention C is not connected to interventions A or B; that is, no trials exist that compare intervention A to C or B to C. In such situations standard NMA models are not valid (although some software options, including WinBUGS, produce parameter estimates and therefore care needs to be taken!). Figure 1 (ii)(b) shows another example of a disconnected network. In this example, two connected sub-networks exist that are not connected to each other. In this case, comparisons may be made within each sub-network but not between them.

ii) There is consistency across the evidence base. Consider the three-intervention network displayed in Figure 1 (i)(c) and, for simplicity, assume the evidence base consists of only two-arm trials. In this situation NMA assumes that if the two-arm trials comparing B versus C had a third arm A, then they would produce an estimate of A versus C and A versus B that was consistent (i.e. assumes the underlying effects to be identical or sampled from the same distribution depending on whether fixed or random effects are assumed in the synthesis model) with any A versus C and A versus B trials that may actually exist. That is, the intervention effect measured using an indirect comparison is valid and equivalent to the intervention effect measured using a direct comparison. Specifically, in the case of the three-intervention network, the effect of B versus C is equal to the effect of A versus B plus the effect of A versus C. If this consistency (also known as transitivity[47]) assumption is not satisfied then this indicates variation in the intervention effects *between* pairwise contrasts in the network; that is, there is *inconsistency*. Note, inconsistency is a property of evidence loops (in this simple network there is one evidence loop containing intervention nodes A, B and C) and therefore scrutinising any of the three comparisons in the network would produce the same result. Further, it is possible to allow for inconsistency (which can be viewed as a type of heterogeneity *between* pairwise contrasts) by the incorporation of further parameters in the NMA model[48] which relaxes this assumption. This is not considered further in this report because it is difficult to interpret the results from such models in a meaningful way that can then be used as inputs into decision models.

It is important to note that any time the effectiveness of two interventions, not directly compared in a pairwise meta-analysis, is of interest or the relative effectiveness of more than two interventions, which have not all been compared in all trials, is required, then assumption ii) above will usually be assumed *implicitly* when interpreting the evidence base. NMA makes this assumption *explicit* and offers a framework to assess its validity.[11]

FIGURE 1: CONNECTED AND DISCONNECTED NETWORK DIAGRAMS



This section has provided a non-algebraic overview of NMA outlining the underlying modelling assumptions, as well as discussing the advantages of NMA over standard pairwise meta-analysis. For more details about NMA methodology and its implementation we recommend:

- Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 1 to 7: 2011; last updated April 2014; available from http://www.nicedsu.org.uk (published as a series in edited form in MDM July 2013[9-13,34,35]).

The latter includes extensive software code for the WinBUGS package to implement NMA methods for a wide variety of outcome measures.

As for the decision modelling considered in the previous section, emphasis on correctly accounting for parameter uncertainty and correlations is given throughout the remainder of this report (in particular, see Sections 5.3 and 5.4).

5. INTERFACING NETWORK META-ANALYSIS WITH ECONOMIC DECISION MODELS

This section focuses on the situation where a single absolute effect (e.g. probability) is derived from each NMA to inform the clinical parameters for each intervention in the economic decision model. The situation will be more complex in a number of instances including the following.

The decision model may consider more than one outcome for each intervention and when these outcomes cannot be assumed to be independent, these need to be taken account of in the NMA model. For example, consider induction of labour studies which report having a Caesarean section and failure to achieve vaginal delivery in 24 hours. These two outcomes are not independent because some of those who fail to have vaginal delivery in 24 hours will have vaginal delivery after 24 hours, whereas the rest will have Caesarean section. Another instance is where there are competing risks, or where the data are reported in a way that induces competing risks (i.e. number of first events); for example, some atrial fibrillation studies report the number of first events (i.e. stroke and haemorrhage), whereas others report the total number of events, or number of events but only the first of each type. The emerging methods to address multiple outcomes in NMA by jointly modelling the outcomes, including the specific challenges of competing risks, multiple time points and informing Markov transition matrices, are considered in Sections 9.2 and 9.5.

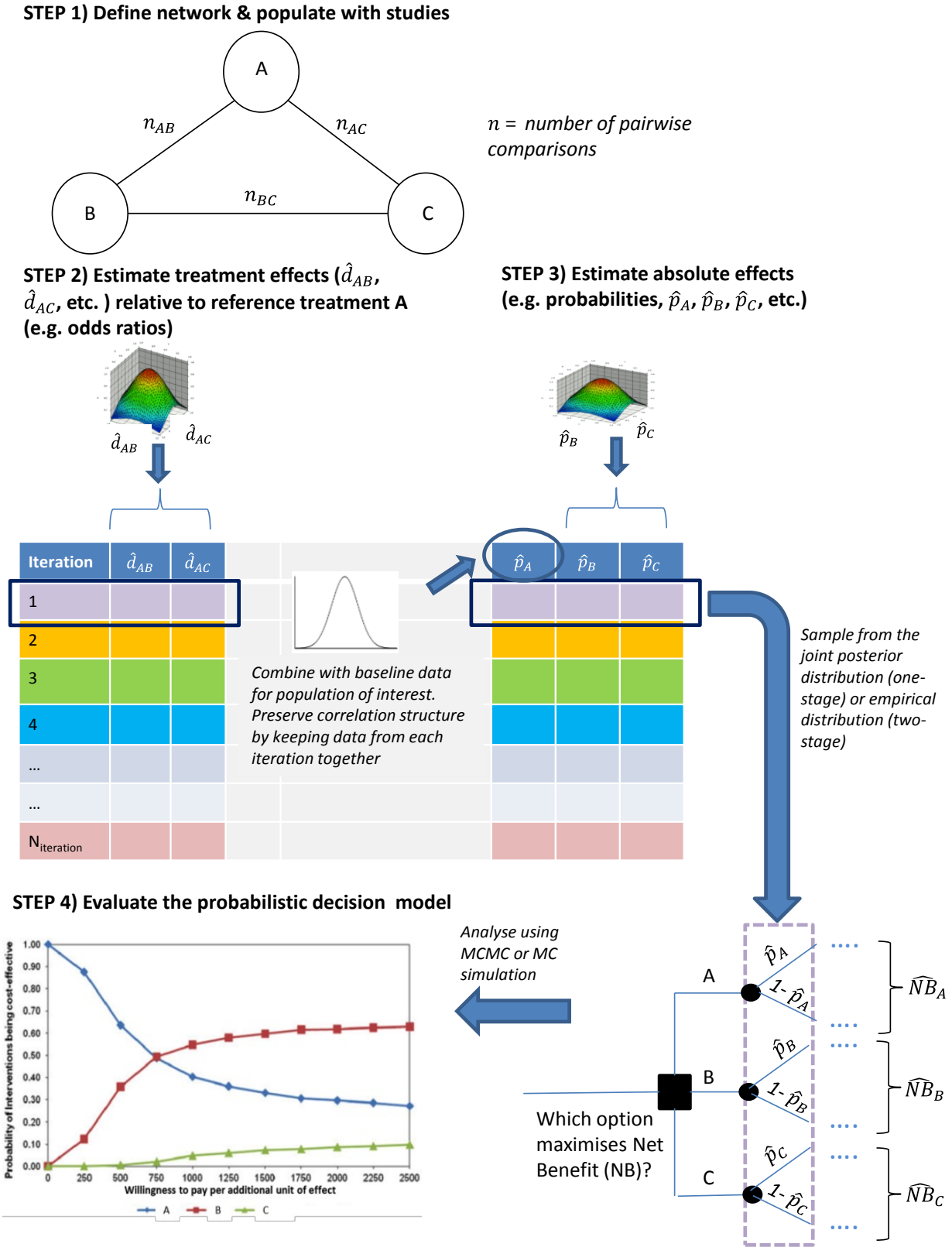
Another complexity is when relevant outcome data are available from multiple studies but the format of the data varies between studies. For example, consider an evaluation of interventions to promote weight loss. If the outcome of interest is weight loss this may be presented in study reports for individual arms, in terms of the mean and variance in each intervention group being compared, or as the mean difference (with corresponding variance) *between* intervention groups.[49] Sections 9.1, 9.3 and 9.4 consider this and other situations where data may be available in different formats, including time to event outcomes and simultaneous use of individual participant data (IPD) and aggregate data.

A third complexity can arise where studies report different but related outcomes; for example, time points reported may differ between studies. If only one of the array of outcomes reported is used in the NMA a large proportion, even the majority, of relevant evidence may be excluded from the analysis. NMA methods which consider multiple outcomes and model the relationship between these outcomes facilitate the incorporation of such disparate data and in doing so include more of the relevant evidence base. These are considered further in Sections 8 and 9, and Case study 2 (Section 10). Figure 2 presents a schematic diagram outlining the four steps required when interfacing NMA with economic decision models. The four steps are:

1) Define scope of decision problem and associated evidence network for NMA
2) Estimate intervention effects relative to reference intervention using NMA
3) Estimate absolute effects derived from NMA and baseline data
4) Evaluate the probabilistic decision model utilising absolute effects

These steps are broadly applicable to instances of single and multiple outcomes but, for clarity, we focus on the situation where a single outcome per intervention is required for the decision model. Where multiple outcomes per intervention or other complexities outlined above are included in the NMA, advanced modelling will be necessary, as considered in Section 9.

FIGURE 2: A SCHEMATIC DIAGRAM OUTLINING HOW TO INTERFACE NETWORK META-ANALYSIS WITH ECONOMIC DECISION MODELS



5.1 Step 1) Define scope of decision problem and associated evidence network for NMA

The first step when undertaking an economic evaluation within a health technology assessment is to define the scope of the decision problem including clear definitions of the target population and interventions in the decision space. In addition, pre-specified study inclusion and exclusion criteria (including study design and comparators of interest – which may include interventions not in the decision space) are essential for the NMA of effectiveness as well as clear definitions of the outcome(s) of interest ensuring these are compatible with the economic decision model. Doing this will ensure i) consistency of evidence included in the network, and ii) focused policy-relevant results are provided by the synthesis rather than a summary of the (possibly disparate) evidence base.

These areas are considered in more detail below under the following headings a) Target population(s), b) Interventions in the comparator set, c) Study design and d) Outcomes.

5.1.1 Target population

It is important to explicitly define the target population for the decision of interest. Where possible, this should include consideration of potential intervention effect modifiers, such as the severity of the disease, comorbidities, intervention history, race, age, gender, socio-economic status and other demographic characteristics.[50] There may be more than one target population, in which case it may be appropriate to consider separate economic decision models.[51] For example, in an evaluation of interventions to prevent stroke in individuals with atrial fibrillation, the clinical benefit gained by individuals may differ depending on whether they have already experienced a stroke previously and therefore separate economic models may be appropriate. This in turn may result in different inclusion criteria being specified, and hence different studies included, for the NMAs informing each population subgroup economic model. Even if the same studies are used in the NMA for both subgroups, it may be desirable to include covariates in the NMA modelling (e.g. percentage of the population having suffered a previous stroke) and careful thought is needed on which estimates of effectiveness from the NMA model are used to inform the subgroup specific decision models (see Section 5.2). It is acknowledged that the more specifically the target population is defined, the less studies may be considered relevant for the NMA thus resulting in sparse networks. Therefore, it may be somewhat of a trade-off between how specifically the population is defined and how much of the evidence base is included. The validity of an analysis using less restrictive inclusion criteria will depend on whether the differences in populations are intervention effect modifiers as to whether one can generalise from the evidence population to the target population. As suggested above, it may be possible to incorporate intervention modifying covariates into the NMA to adjust for population differences but, especially when using summary data, this should not be seen as a panacea solution (see Section 5.2).

5.1.2 Interventions in the comparator set

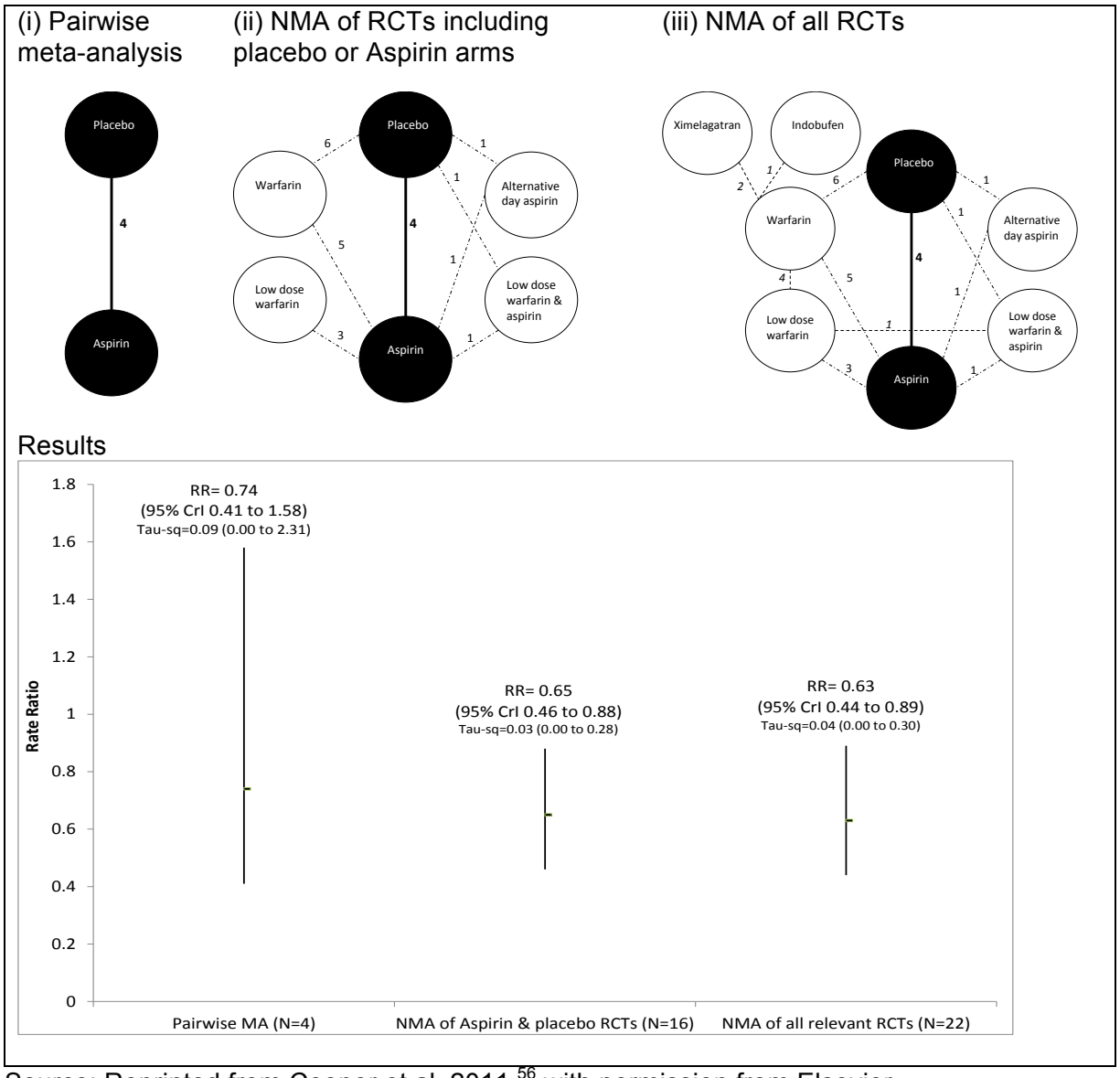
All candidate interventions relevant to the decision problem need to be identified and, where possible, explicitly defined in terms of dose, co-interventions, healthcare setting, timing of delivery and so on. Using the terminology of Ades et al.,[35] the *decision* comparator set (Figure 3(i)) includes all candidate interventions relevant to the target population within the decision space. The mapping of interventions used in the trials to those in the decision space may not be completely clear cut (e.g. doses or timing of interventions may vary but may be considered similar enough to be dealt with as equivalent). The debates on lumping (i.e. combining similar but distinct interventions) versus splitting (i.e. maintaining a higher degree of differentiation in

intervention definitions) in MA have a long history and we cannot provide a definite answer here other than recommend clinical expert judgement and potentially exploring alternative NMA models such as hierarchical or structured models (e.g. allowing for dose response effects).[52]

The broader *synthesis* comparator set may include additional interventions beyond those included in the *decision* comparator set (e.g. placebo, interventions not licensed or older interventions no longer recommended). Including additional interventions in the synthesis comparator set may make a disconnected network connected. Other advantages of extending the network include[35,50,53,54] i) an increase in the potential to check consistency, ii) potential reduction in uncertainty by the inclusion of more evidence, and iii) the final results will be more robust and less sensitive to the inclusion of any individual study. However, the expansion of the network may lead to increased risk of inconsistency and heterogeneity, and thus increased uncertainty in the pooled estimates from the RE model. For example, the inclusion of "older" trials comparing subsequently superseded interventions may have been carried out in an era where trial methodology/practice/populations were different to more recent trials, increasing the propensity for such trials to be inconsistent with the recent ones (but the authors are not aware of any empirical evidence to support this at present and seeking clinical input maybe helpful in such situations). Although Figure 3 makes little sense clinically, it does clearly illustrate many of the concepts discussed above; that is, panel (i) shows the decision comparators and the direct evidence available for deciding between placebo and Aspirin for the prevention of stroke in individuals with non-rheumatic atrial fibrillation, panel (ii) extends the network to include a broader set of synthesis comparator interventions now including further trials in which either decision comparator is compared to another intervention, and panel (iii) extends the network further to include all connected trials in non-rheumatic atrial fibrillation conducted at the time the analysis was performed. The bottom panel in Figure 3 shows that extending the network as displayed in panel (ii) reduces the uncertainty in the estimated relative intervention effect between placebo and Aspirin; however, you get diminishing returns as the network expands (panel (iii)) away from the decision-relevant interventions.[53-55] It is therefore important to carefully consider the structure and scope of the network.

FIGURE 3: AN ILLUSTRATIVE EXAMPLE EVALUATING THE USE OF ASPIRIN COMPARED TO PLACEBO FOR THE PREVENTION OF STROKE IN INDIVIDUALS WITH NON-RHEUMATIC ATRIAL FIBRILLATION

As the network is extended to include all RCTs in the clinical area containing either placebo or Aspirin, and thus additional interventions, it can be observed that the uncertainty is reduced considerably (panel ii). However, when the NMA is expanded to include all RCTs (panel iii) the uncertainty in the pooled risk ratio is increased slightly compared to the NMA of placebo or Aspirin RCTs despite the inclusion of more information. This is because of the increase in the between-study variance (i.e. the measure of within-intervention comparisons between-study heterogeneity) that in turn reduces the absolute weight given to each study in the synthesis.



Source: Reprinted from Cooper et al. 2011,[56] with permission from Elsevier.

5.1.3 Study design

Study designs to be included in the NMA need to be pre-specified in the inclusion criteria. For an informed decision making process it is important to make optimal use of the available evidence. Typically, evidence on relative effectiveness is restricted to RCTs (thus implicitly assigning a weight of zero to all other available evidence). RCTs are considered to be the gold standard of evidence for relative effects because their controlled setting minimizes potential bias due to lack of internal validity, although it has been argued that their experimental setting limits their external validity.[57,58] On the other hand, observational studies may reflect the target population and setting more closely, but are often prone to internal biases such as selection bias, non-response bias and/or confounding.[59] Additionally, in some contexts, randomised evidence may be extremely limited or unavailable for ethical reasons, which may lead to the use/inclusion of observational data (despite its limitations). However, when this is done, it is important to consider and adjust for possible biases that may exist in the evidence (see below). Open-label extension studies are a particular type of observation study that are usually undertaken to assess the long-term safety and efficacy of an intervention following-up patients previously enrolled in an efficacy RCT. Such studies pose their own challenges as both the patient and the physician are aware of which groups are receiving what type of intervention (lack of blinding), and usually no standard intervention or placebo is utilized as a comparator. Also, only a proportion of the patients initially recruited into the RCT will likely agree to take part in the open-label part of the study[60] and this population will potentially be unrepresentative of the original trial. For example, the open-label study has a greater propensity to include those participants who responded well to the intervention or switched interventions after not responding well on the comparator.[60]

Given the above, analysts are faced with a choice about whether to limit the NMA to "best available evidence", thus restricting the evidence base to well-conducted "low risk of bias" RCTs, or include "all available evidence", in which all available evidence is included. The latter may increase precision but at the expense of an increased risk of bias if studies at high risk of bias are not "down-weighted" in the analysis.[61] Methods to enable the incorporation of different study designs within the NMA framework (Step 2) by adjusting for study-specific biases have been proposed within the meta-analysis[61-64] and NMA[57] literature. Such methods are still at the developmental stage and more rigorous evaluation is required before they can be applied routinely as the primary analysis. However, as for any systematic review-based endeavour, it is good practise to assess the risk of bias within individual studies[65] in terms of allocation concealment, blinding of outcome assessment, completeness of follow-up for RCTs[66] and blinding of outcome assessment, and completeness of follow-up and balance of confounders between study arms for non-randomised studies.[67,68] These measures of risk of bias can then be used to assess their potential influence on the results as part of a sensitivity analysis.

5.1.4 Outcomes

It is important to clearly define the outcomes required to inform the clinical parameters of the cost-effectiveness model, as they may differ from the outcomes required for the clinical review of any health technology assessment.[69] That is to say, the synthesis approach used to "just" combine and summarise trial outcome results, as required in a typical systematic review/meta-analysis for any clinical review, may not produce quantities which are a "good fit" for the effectiveness parameters, as defined, in the economic model. For the clinical review, very often, the most commonly reported outcomes from the primary studies will be used as the primary outcomes of the meta-analysis as to maximise the data used. Such an approach may well be sensible from an inference-driven perspective, but in a specific decision making context such an approach may not be optimal (see Case study 2, Section 10). Burch et al.[70] evaluate the effectiveness and cost-effectiveness of the use of anti-virals for the treatment of influenza. This

is a good example where the synthesis estimates from the clinical review, performed on the median time to the alleviation of symptoms as reported in the trials (since a proportion of patients were still ill at the end of the trials and hence their outcome times unknown and censored), were not appropriate to inform the economic evaluation, where mean time to alleviation of symptoms was required.[71] Another situation is when time is an important factor. For example, if the studies report effectiveness of the intervention of interest at different time points but not all studies report all time points. In this situation if only one time point was of interest for the economic evaluation the options available include i) only include trials which report data at the time point of interest – but this may exclude a lot of trials; ii) assume effectiveness is similar at different times and combine across different time points – but this makes strong assumptions; or iii) fit a synthesis model which allows for different time points in the modelling[72,73] (see Case study 2, Section 10).

Having defined the scope and specified the inclusion/exclusion criteria as discussed above, the next stage is study identification. Typically, the studies (usually RCTs) to populate the network are identified by means of a systematic literature search with relevant studies identified as those that adhere to the pre-specified study inclusion and exclusion criteria including target population, comparator interventions, study design inclusions and outcomes discussed above. A published systematic review(s) (SR) in the area of interest may already exist. If this is the case, then it may be possible to use this review, and any evidence synthesis within it, to obtain estimates of the clinical relative effect parameters required for the economic decision model as long as it meets the scope and pre-specified inclusion/exclusion criteria; however, it may require updating or expanding to include all comparators in the synthesis set (see Section 6).

Before proceeding to fitting the NMA to the data (Step 2) it is imperative to check that the network of identified study data is connected (see Section 4) and issues, such as sparse data (see Section 8), are identified.

5.2 Step 2) Estimate intervention effects relative to reference intervention using NMA

The second step is to fit an NMA model to the data from the studies in the network of intervention comparisons, to obtain estimates of all intervention effects relative to a reference intervention (for example, standard care or placebo). If a published NMA relevant to the decision problem already exists, it may be possible to use the results to inform the clinical parameters for the economic decision model without further analysis or modifying existing analyses. Situations where this is and is not possible are considered in Section 6.

Firstly, it is important to decide on the appropriate outcome statistic for the data being combined. For binary data (i.e. number of events out of the total number of study participants per intervention arm) the most common outcomes are relative risk or odds ratio, however, for continuous data (e.g. time to event data) appropriate outcomes include the mean difference or hazard ratio. Rate data may also be available where the number of events in a specified follow-up time is presented; such a data format allows for multiple (assumed independent) events per person. For details on how to choose between outcome measures and their respective model specifications and WinBUGS code to implement the NMA models see Dias et al.[9] See Section 9 for more complex NMA methods, as models of the type described within this section will be relevant for some applications.

Both heterogeneity and inconsistency can be considered to be due to factors that interact with the intervention effect and vary between trials. Such factors may be numerous, unknown and/or

unmeasured, but when they are known and measured then they can (in theory) be incorporated into the synthesis model to reduce heterogeneity and/or inconsistency.[10,74]

Patients' event rate risk on the reference intervention (more commonly termed baseline risk) may be considered as a proxy for unmeasured patient-level characteristics which may be modifiers of intervention effect and therefore a potential source of heterogeneity and/or inconsistency.[10,24] It may therefore be appropriate to adjust for patients' event rate risk on the reference intervention in the NMA. In the context of informing an economic model, knowing that the effectiveness of an intervention varies by unknown patient characteristics complicates the estimation of relative effectiveness because identifying the appropriate baseline level for the target population will generally not be possible without knowing which patient characteristics influence it without having trials in the specific population of interest. A strong argument can be made for obtaining the IPD from the included studies in an attempt to identify which patient-level characteristics are important predictors of effectiveness. In reality, fitting meta-regression models (to summary data) will be restricted by the number of trials in the network.[75] If known individual participant-level intervention effect-modifying covariates exist (e.g. disease severity), and the distribution of these varies both within and between comparisons, a strong case can be made for attempting to obtain and use IPD (see Section 9.4) in which regression methods can be used to adjust for such effects. Alternatively, it may be possible to restrict the study inclusion criteria for NMA to focused populations removing the variability in intervention modifying covariate(s) effects (see Section 5.1); for example, analysing early and late-stage disease populations separately. Additionally, unlike pairwise meta-regression analysis, alternative assumptions regarding the regression effects on the different interventions can be made; that is, regression effects can be assumed to be the same, different or exchangeable across all or subsets of the interventions.[74]

It is important that the relative effects of the interventions and their uncertainty be appropriately incorporated into economic models.[76] For FE NMA the pooled intervention effects are the appropriate summary from which to calculate the absolute effects in Step 3) to input in the decision model. However, for RE models there are a number of measures that may be appropriate depending on the interpretation of the heterogeneity in the studies included in the NMA and how this relates to the target population (see Case study 1, Section 10). Welton et al.[76] identify five possible summaries that may be used as inputs in the economic model:

(i) *Random effects (posterior) mean* (i.e. the mean of the random effects distribution) – this assumes that the decision setting is exactly equal to the average setting from the studies included in the NMA;

(ii) *Predictive distribution* – this assumes that the target setting for the decision is "similar" to those in the studies included in the NMA; that is the relative intervention effects (compared to the reference intervention) in the decision setting comes from the same distribution of intervention effects but we do not know where in the distribution it lies. Therefore, predictive distribution incorporates both the uncertainty about the value of a new observation as well as the observed variation in the data;

(iii) *Shrunken study-specific estimate* – considers the decision setting to be *similar* to those in the studies included in the NMA (as for the predictive distribution above) but it is most closely represented by a single study population. When estimates are combined using RE models, the model updates estimates of the individual studies, taking into account the results from all the other studies in the analysis; that is, the study-specific intervention effects are "shrunk" towards the overall random effects mean assuming the borrowing of

information across studies (i.e. exchangeability).[27,76] These study-specific intervention effects estimates (known as "shrunken" estimates) will be more precisely estimated than the study estimates alone.

(iv) *Random effects distribution* – considers the decision setting to be made up of those included in the studies in the NMA and therefore the heterogeneity estimated in the NMA setting is expected to be seen in the decision setting. In this context it is necessary to integrate over the entire RE distribution for the economic model; and

(v) *Independent study-specific estimate* – considers the decision setting to be represented by a single study population where information from all other study populations is irrelevant (e.g. poor quality leading to a high risk of bias and cause of heterogeneity). In this case the effect estimate from the study alone (not the NMA!) is the appropriate input for the decision model.

For more technical details, including obtaining inputs for the economic model when covariates are incorporated into the NMA, see Welton et al.[76]

Having fit the NMA it is important to identify and investigate any inconsistencies in the data;[65] that is, where estimates for the same pairwise effect differ between different evidence sources (see Section 4). Methods for identifying inconsistency include node splitting as described in detail by Dias et al.[11] This method may detect inconsistencies in the network and identify the problematic loops but will not detect which data within the identified loops of evidence in the network are inconsistent; this requires the analyst, ideally with clinical input, to re-examine the studies in the data together with the inclusion criteria. Note that the detection of inconsistency suffers from low power and therefore failure to find inconsistency in the network does not mean that there is no inconsistency. It is also important to check the fit of the model to the data both relative to other possible NMA model formulations (where appropriate) and in absolute terms.[11] Goodness of model fit statistics include the DIC[77] which can be used to choose between model formulations that utilise the same data (e.g. fixed and random effects), and the posterior mean residual deviance statistic which assesses how well the model fits the data.[77] In addition, analysts should examine the between-study variance terms of different model specifications (note, often adding covariates to the NMA model only changes the DIC or posterior mean residual deviance statistic marginally but the value of the between-study variance parameter may reduce considerably indicating that intervention-covariate interactions may explain a worthwhile proportion of the between-study variability).

5.3 Step 3) Estimate absolute effects derived from NMA and baseline data

Most economic decision models require clinical effectiveness inputs expressed in absolute terms rather than relative; that is, separate probabilities for each intervention option of interest (e.g. the probability of intervention A preventing stroke, probability of intervention B preventing stroke, probability of intervention C preventing stroke, etc.). This can be achieved by combining the relative effects estimated in Step 2) with baseline data (i.e. the natural history in absolute terms under the reference intervention in a comparator set) for the population of interest to obtain the absolute intervention effect under the different intervention regimens of interest. As the relative intervention effects have been estimated jointly from a single NMA synthesis model, in most cases this will induce correlations between the parameters and therefore it is essential to maintain this correlation structure when translating relative effects to absolute effects for the decision model (see Step 4).

There are a number of ways of estimating the natural history under the reference intervention and which is the most appropriate will depend on the context. This baseline data can be estimated from some or all of the trials in the NMA that include a reference intervention arm, however, care needs to be taken to ensure the data are representative of the target population under current circumstances (e.g. may want to limit the analysis to the most recent trials or country-specific trials, say, Canadian-based). If this approach is adopted, and the synthesis and baseline model fitted simultaneously, care should be taken to ensure neither influences the other and both analyses remain independent.[12] An alternative approach, that may provide a better representation of the population of interest, is to use baseline natural history data that is independent of the relative effects data obtained, for example, from registries and/or cohort studies[12] which may have better external validity. A separate baseline synthesis model can then be fitted to estimate the natural history under the reference intervention. Technical details on both of the above methods, as well as further topics regarding the estimation of the baseline model (including incorporation of covariates, use of IPD, use of multiple outcome measures, synthesis of state transition models, and model validation and calibration), are presented elsewhere.[12]

The clinical outcomes under each intervention option in the decision set can be obtained by putting the results from the baseline model together with the relative effect estimates from the NMA to obtain absolute effects. As discussed in Step 2), there a number of ways to summarise RE models. The random effects mean together with its variance can be used to represent the baseline response; however, Dias et al.[12] argue that this under-represents the variation observed in the data (i.e. as we gather more baseline data the mean estimate will become more precise but the variation will remain constant). Instead, they recommend the use of the predictive distribution for a new baseline which incorporates both the uncertainty about the value of a new observation and the observed variation in the data, but currently this is rarely used in practice although may be the most relevant parameter in more instances than is appreciated. Alternatively, in some situations, it may be more appropriate to use the *shrunken* estimates of the study(s) relevant to the decision population (see Step 2).[27,76]

5.4 Step 4) Evaluate the probabilistic decision model utilising absolute effects

Finally, the absolute effects, along with all other model parameters, can be input into the specifically developed economic decision model and the decision model evaluated stochastically by simulating from the joint distribution of all model inputs to ensure parameter uncertainty is appropriately propagated through to the expected model outcomes (see Section 3). Allowing for the uncertainty in the model parameters, the best decision option is that which maximises expected net benefit.[78] It is important to point out that care needs to be taken when calculating the expected net benefit, as the expected net benefit is not the same as the net benefit at the expected values of the model parameters, unless net benefit is linear in all its parameters and there are no correlations between parameters[13] (this implies that net benefit needs to be calculated at each iteration when using MCMC or MC [Monte Carlo] methods as described in i) to iv) below). As most evidence synthesis models, that are often used to inform model parameters, are fitted on the logarithmic or logit scale and the required parameters for the model are usually probabilities on the natural scale, the transformation from one to the other will be non-linear. With NMA there is also the added complication that when multiple effectiveness parameters are estimated within a single synthesis model, in most cases, this will induce correlations between the parameters. When intervention options are compared within a stochastic framework, these correlations may affect the uncertainty in the incremental net benefit, and for this reason it is crucial that the joint parameter uncertainty, including correlation structure, be propagated through the economic decision model.[13,41] Note that if only a single

pairwise comparison from the network is of interest (i.e. estimated using the whole network of evidence e.g. atrial fibrillation in Figure 3) then correlation may not be an issue.

Four approaches to incorporating evidence synthesis (including NMA) results into probabilistic economic decision models, which will correctly propagate the uncertainty and correlation structure in the evidence synthesis in any situation, have been outlined by Dias et al.[13] and are described below.

i) *Bayesian posterior simulation – One-stage approach* (combining Steps 1 to 4 in Figure 2): This integrated approach simultaneously estimates the relative intervention effects from the NMA (Step 2) and the baseline rates using the relevant data for the population of interest as discussed above (or the baseline effects may be specified directly as statistical distributions if the baseline data are analysed elsewhere). Expressions for the absolute effects are specified and thus derived directly (Step 3) by sampling from the Bayesian posterior distributions of the specified parameters. The samples from these posterior distributions, in turn, inform the relevant parameters of the economic decision model (Step 4). By specifying all this modelling within a single coherent framework (also known as comprehensive decision modelling[79-81]), utilising only a single piece of software code, it ensures that the joint parameter uncertainty, including correlation structure, is maintained. As samples from the joint posterior distribution of the absolute effects are fed directly into the decision model, and net benefit evaluated for each set of parameter samples, no distributional assumptions for these absolute effects are required. This may be particularly important for adverse effects or other rare event data where posterior distributions may be non-symmetrical (see Section 8). This is a simulation-based method which uses Markov Chain Monte Carlo (MCMC) sampling (i.e. propagates evidence uncertainty "back" from the data onto the parameters and then forward through the decision model[81]) as implemented in several freely available statistical packages such as WinBUGS,[14] OpenBUGS,[82] JAGS[83] and STAN.[84]

ii) *Bayesian posterior simulation – Two-stage approach:* As above, stage one of this two-stage approach simultaneously estimates the relative intervention effects from the NMA (Step 2) and the baseline rates using the relevant data, and expressions for the absolute effects are specified and thus derived directly (Step 3) by sampling from the Bayesian posterior distributions of the specified parameters (alternatively the relative effects only may be estimated in the first stage and baseline effects specified along with the other decision model parameters in stage 2 of the modelling (see below)). In stage 2, these parameter estimates, expressed as an array of simulated values generated from the full posterior distribution (which together form an empirical distribution), are exported into another software package (e.g. Excel, TreeAge Pro,[85] R[21]) where the decision model and data for other model parameters are specified. This decision model is then evaluated using Monte Carlo simulation (Step 4) which draws sample values from the array of simulated values (i.e. an empirical distribution) for the effectiveness parameters (and whatever distributional forms are specified for the other model parameters). Usually, many values are sampled (i.e. 1000s) to obtain accurate empirical distributional representations of the outcomes of interest (i.e. net benefit) for each intervention option with associated uncertainty. When implementing this two-stage approach, it is imperative that, at each iteration of the Monte Carlo sampler, the sample value drawn for each effectiveness parameter comes from the same iteration in the MCMC output (i.e. in practical terms this means sampling from the same "row" of simulated data array for all effectiveness parameters). If this is not adhered to, the correct correlation structure between effectiveness parameters will not be correctly maintained. But when implemented correctly, this two-stage approach has the same technical properties as the one-stage approach described; that is, no distributional

assumptions are specified for the effectiveness parameters and correlations are maintained. Exporting the array of simulated values can be achieved manually by cutting and pasting the values from, say WinBUGS[14] using the CODA output, or automated by using such packages as R2WinBUGS[86] and/or RExcel,[19,87] or NetMetaXL.[18]

iii) *Frequentist estimation with Monte Carlo simulation*: Modules have recently been developed in STATA,[16,17] R and SAS[88] that enable the user to fit NMA models. These methods are typically not simulation-based and thus do not provide empirical distribution samples (as described and utilised in the two-stage approach above) for the effectiveness parameters of interest. Instead they produce parameter estimates, together with their variance-covariance matrix. By making assumptions about the distributional form of these relative effectiveness parameter estimates (usually multivariate normal), it is possible to specify parametric distributions for them, along with the other model parameters, that can be sampled from in order to evaluate the stochastic economic decision model (using Monte Carlo simulation for this second stage). Hence, the limitation of this approach relates to the accuracy of the assumed distributional form of effectiveness parameters, the implications of which, to our knowledge, have not been fully investigated. Note also that many frequentist software routines for random effects evidence synthesis modelling (including NMA) do not allow for the uncertainty in the variance parameters, unlike the Bayesian methods, and thus may provide overly precise estimates; however, this underestimation is likely to be relatively small[89] and frequentist methods which incorporate such parameter uncertainty[89,90] have been developed.

iv) *Frequentist estimation with bootstrapping*: Bootstrapping is a statistical technique whereby "new" datasets are generated by resampling with replacement from the original data[91] and the analysis is repeated on each sample dataset. These analyses produce a set of parameter estimates similar to those samples from a Bayesian posterior distribution which can be used as part of a one- or two-stage approach as described above. Note that this approach may be problematic when sample sizes and/or number of studies are small, and/or zero cells are present, and is relatively unexplored in an NMA context.[92]

In summary, where NMA is required to provide parameter estimates for multiple intervention effects in an economic decision model, then the Bayesian MCMC simulation methods of synthesis (i.e. methods i) and ii) outlined above) are likely to be the most convenient because they allow the full joint posterior uncertainty in the model parameters, including correlation structure, to be propagated with ease through the decision model as part of either a one- or two-stage approach. It is appreciated that Bayesian MCMC software packages, such as WinBUGS, can be difficult to learn as they operate quite differently from general purpose statistical packages such as SAS, STATA, SPSS, etc. This factor has probably limited their use in the past; however, more user-friendly software front-ends to MCMC packages[18-20] are becoming available, and some even focus on fitting NMA models specifically, and may provide an appealing alternative. As outlined above, frequentist methods are also available and if used carefully and with their limitations noted, offer a reasonable alternative in certain situations, although they lack the flexibility to be able to incorporate shared parameter models (see Section 9.1) and joint modelling of multiple parameters (see Section 9) that may be necessary for decision modelling. But their performance, particularly with respect to the limiting assumptions they require, needs further evaluation before they can be fully endorsed.

As with any economic decision model, it is important to perform a thorough sensitivity analysis to assess the influence of the model structural and input parameter assumptions on the cost-effectiveness results.

6. USING ESTIMATES FROM PUBLISHED NMAs TO INFORM CLINICAL PARAMETERS IN AN ECONOMIC EVALUATION

If a published SR in the area of interest already exists then it may be possible to use study data from this review, and any evidence synthesis within it, to obtain estimates of the clinical parameters required for the economic decision model as long as the review meets the pre-specified scope and exclusion/inclusion criteria set out in Step 1 (Figure 2). Where the published SR includes an up-to-date NMA relevant to the target population of interest and contains all relevant comparators, the results may be used to obtain absolute effects for use in the economic decision model (omitting Step 2). However, unlike pairwise meta-analysis where it may be possible to extract the pooled estimate and its variance from the published analysis for use in a decision model, multiple effectiveness parameters from an NMA will be correlated and therefore the variance-covariance matrix will be required to maintain the joint parameter uncertainty (as noted in Steps 3 and 4); in our experience this is rarely, if ever, published. Also, if a RE NMA model has been fitted, there is also the issue of whether the appropriate measure from the NMA has been reported (e.g. random effects (posterior) mean, predictive distribution, etc.; see Step 2), Section 5). However, provided the individual study summary data are published it will be possible to reanalyse the NMA for purposes of informing the decision model as described in Steps 2), 3) and 4) above.

Where the published SR includes pairwise meta-analysis(es) then the evidence base may contribute to the network, but it is likely that additional studies either addressing omitted comparisons of interest or studies published since the review will need to be identified (via appropriate systematic search strategies) and included(Figure 4 and Figure 5). To fit the evidence base within the restrictive pairwise meta-analysis framework it is likely that interventions may have been "lumped" together (e.g. same intervention but different doses or differences in intensity of control regimes defined as "usual care", etc.); however, NMA provides a framework that (dependent on the data available) potentially allows such interventions to be kept distinct, permitting decisions to be made not only about the most effective intervention but also the most effective intervention dose or regime (Figure 4 and Figure 5) for the target population.

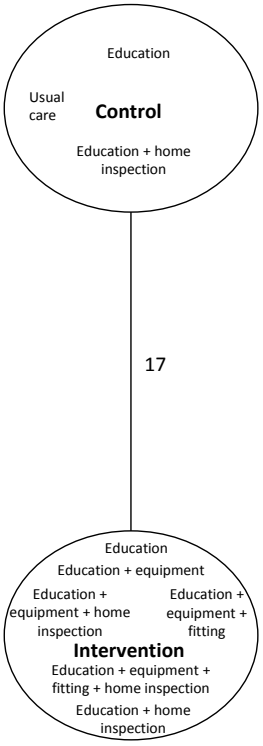
As initially highlighted in Section 5, even within NMA, the degree of "lumping/splitting" will often require expert input as a level of subjectivity is inevitable, especially where evidence bases are sparse (increasing the likelihood of a disconnected network)(see Section 8), and there may be a necessity for "lumping" in order to produce "coherent" estimates of relative intervention effects. Note that "lumping" brings increased precision but at the risk of increased heterogeneity that may be important if we are to identify the optimal regimen of an intervention. An exploratory tool to help inform the decision of whether to lump interventions together has been developed.[52,93]

Finally, to aid the analysts, reviewer's checklists have been developed which present a series of questions to help assess the credibility and applicability of the results from an NMA to the decision setting of interest.[35,50]

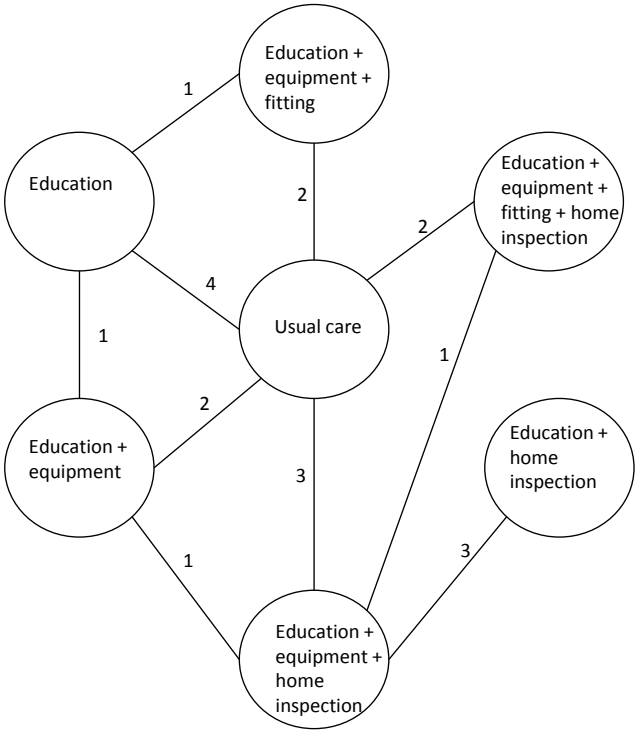
FIGURE 4: AN EVALUATION OF THE EFFECTIVENESS OF INTERVENTIONS TO INCREASE THE PREVALENCE OF SMOKE ALARMS IN HOUSEHOLDS WITH CHILDREN

The network diagrams below show how the "intervention" arm of the pairwise meta-analysis presented in the Cochrane review[94] can be "split out" into distinct intervention strategies using an NMA framework[95] which enables all interventions to be compared with one another and the most effective intervention strategies to be identified (note the remit of the NMA was extended to include intervention versus invention studies to the network).

i) Pairwise meta-analysis of any interventions to increase the prevalence of smoke alarms versus usual care[94]



ii) NMA including all interventions to increase the prevalence of smoke alarms versus each other and usual care[95]

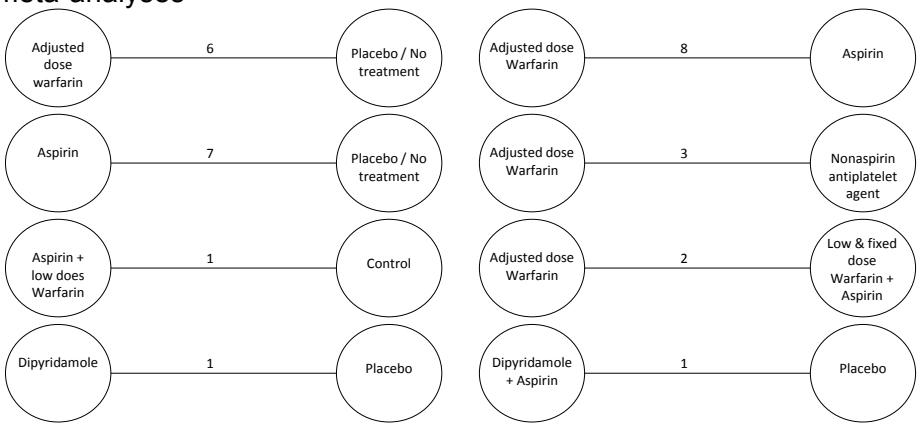


Source: Reprinted from Achana et al. 2014,[96] with permission from Elsevier.

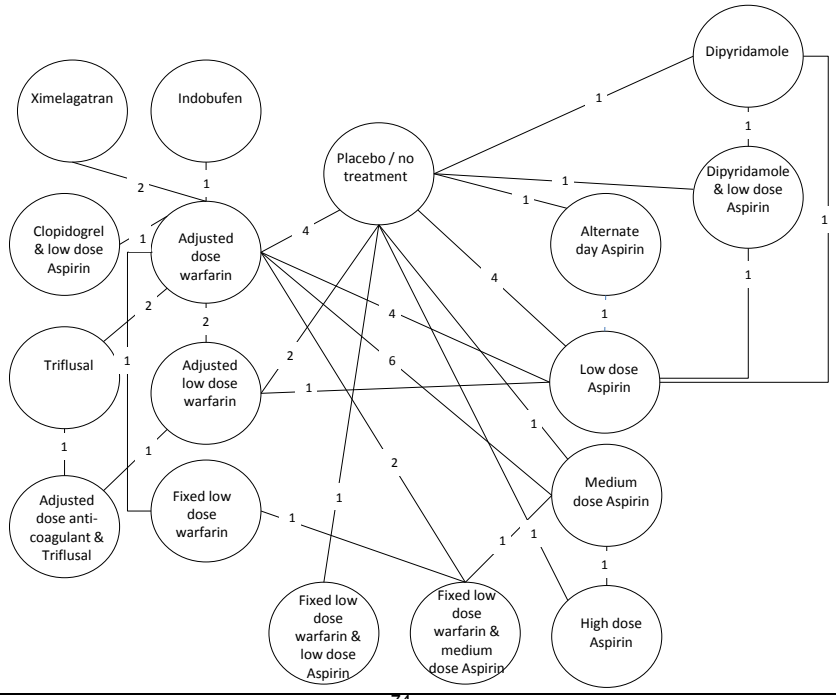
FIGURE 5: AN EVALUATION OF ANTITHROMBOTIC THERAPY TO PREVENT STROKE IN PATIENTS WHO HAVE NONVALVULAR ATRIAL FIBRILLATION

Hart et al.[97] published eight pairwise meta-analyses to evaluate the effectiveness of different antithrombotic therapies to prevent stroke in patients who have nonvalvular atrial fibrillation (panel i). These eight pairwise meta-analyses were combined by Cooper et al.[74] to form a network of evidence (panel ii) and NMA was undertaken to allow all interventions to be compared to one another (even where no direct comparative evidence existed) and to estimate the most effective intervention(s) for preventing stroke.

i) Pairwise meta-analyses[97]



ii) Network meta-analysis *(using all the trials identified in the Hart et al.[97] systematic review plus one additional study, BAFTA, which compared adjusted anti-coagulant vs low-dose Aspirin)*[74]



ii) Source: Reprinted from Cooper et al. 2009,[74] with permission from Wiley.

7. CAN PREVIOUS SYSTEMATIC REVIEWS HELP INFORM AN NMA SYNTHESIS FOR ECONOMIC EVALUATION? A DECISION TOOL

Figure 6 presents a decision tool we have developed to inform the optimal approach to NMA for a given decision problem. The tool primarily relates to Steps 1 and 2 of the schematic diagram outlining how to interface NMA with economic decision models presented in Figure 2, although all steps have been included for completeness.

As described in Section 5.1 (Step 1) above, when undertaking an economic evaluation within a health technology assessment it is essential to first define the scope of the decision problem, including pre-specified study inclusion and exclusion criteria clearly defining target population(s), interventions included in the comparator set, and outcomes; the first stage in the decision tool. The next stage is to identify whether an existing SR(s) has already been published addressing the question of interest.

If NO => The analyst proceeds down the right-hand side of the flow diagram, first performing a new SR for the decision problem of interest to identify relevant studies by applying the pre-specified inclusion criteria, and then using this evidence base to define the network. The next stage is to check whether the network is connected and if it is, then the analyst may proceed to conducting the NMA (Step 2). At this stage it is important to explore inconsistency, heterogeneity and model fit of the NMA before combining the results with baseline data to obtain absolute effects (Step 3). These absolute effects, along with other parameter estimates, can then be input into the probabilistic economic decision model and evaluated using one of the methods defined in Section 5.4 (Step 4) above. If the network is not connected, then the analyst must question whether it is clinically justifiable to "lump" interventions together or proceed omitting isolated disconnected interventions. If the analyst can answer YES to either of these then they may proceed to the NMA and progress as outlined above. If NO, then the analyst cannot proceed as defined in the scope and must choose one of the following options: (i) revise the proposed evaluation by limiting the decision space or expanding the network scope; (ii) adopt an alternative analysis method (see Section 9 for non-standard NMA models); or (iii) abandon the quantitative synthesis and thus the economic evaluation due to lack of data.

If YES => The analyst proceeds down the left-hand side of the flow diagram, first carefully checking that the scope and inclusion criteria of the published SR(s) match those pre-specified for their analysis. If the published review contains an NMA then the analyst must answer a series of questions to assess whether it is appropriate to use the NMA to inform the clinical parameters in the economic decision model. These are:

Is the NMA up to date?

Are interventions in the network distinct and if not, has any "lumping" of interventions been clinically justified?

Are the appropriate measure (e.g. random effects (posterior) mean, predictive distribution, etc., Section 5.3) and corresponding variance-covariance matrix from the NMA presented in the published review?

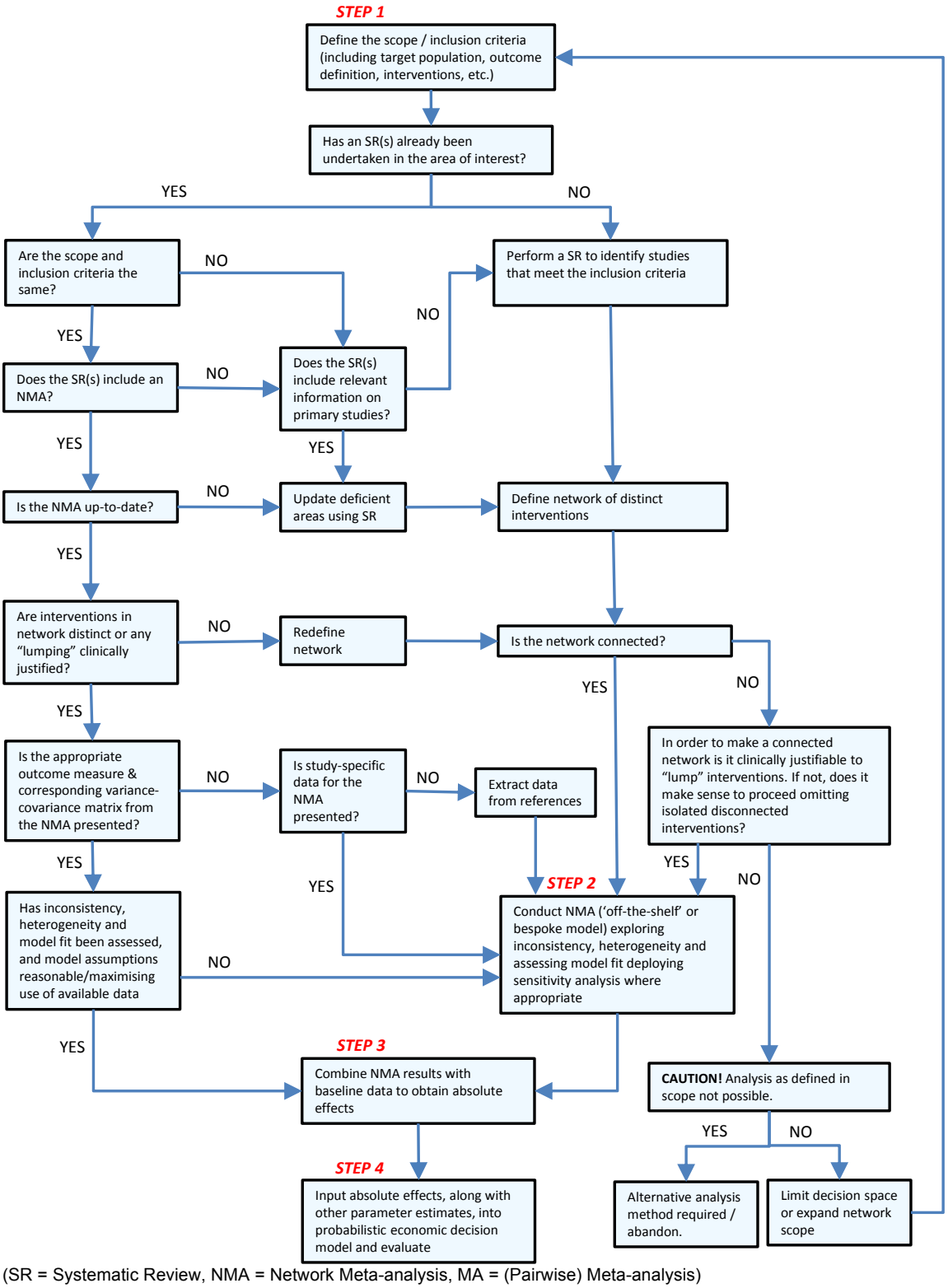
Has inconsistency, heterogeneity and model fit been assessed, and model assumptions reasonable and/or maximise use of available data?

Answering NO to any of the above questions results in the analyst switching to the example-specific SR and NMA pathway (i.e. the right-hand side of the flow diagram).

Throughout the above process, to ensure transparency and reproducibility, it is important that decisions made at each step of the tool be carefully documented and references (such as published SR(s), NMA models used, etc.) cited where applicable.

FIGURE 6: DECISION TOOL TO INFORM THE OPTIMAL APPROACH TO NMA FOR A GIVEN DECISION PROBLEM

To ensure transparency and reproducibility, decisions made at each step of the tool should be carefully documented and references cited where applicable.



(SR = Systematic Review, NMA = Network Meta-analysis, MA = (Pairwise) Meta-analysis)

8. ADVERSE EVENTS/SPARSE DATA

In addition to informing the efficacy parameters in an economic evaluation, it may be appropriate to use NMA to synthesise information on adverse events relating to interventions. Below we consider exclusively the synthesis of RCT adverse event/sparse data and predominantly the unique technical challenges a quantitative synthesis of such data presents for economic evaluation, although we acknowledge this may not always be the optimal choice of data to use.[98] For a more general coverage of the evaluation of adverse events, including the challenges of non-standardised reporting practices observed for adverse events in RCTs and the use of different sources of data (Section 5.1), the interested reader is referred elsewhere.[99]

Over the past two decades, a sizeable literature has grown around the use of evidence synthesis – and specifically pairwise meta-analysis of RCTs – for identifying and quantifying the risk of adverse events. The main issue relevant to both MA and NMA relates to sparse data. This is particularly pertinent to adverse event data but may also be relevant to sparse efficacy data. A particular challenge often faced when synthesising adverse event data (or sparse efficacy data) is that arms of certain studies will observe zero events in one or more study arms, producing problems with the use of many commonly used synthesis methods, and, in the context of NMA, may result in an apparently connected network becoming effectively disconnected. A routine approach when dealing with zero events is to apply a continuity correction to the data to enable outcomes of interest, such as the odds ratio or relative risk to be calculated and combined using standard (inverse variance weighting) methods. However, as discussed in The Cochrane Collaboration Handbook (Section 16.9)[100,101] and informed by simulation studies,[102-104] such methods are to be avoided in favour of methods that avoid the need for continuity corrections including "exact" likelihood methods that model the data directly. Such methods also remove the reliance on the assumptions that the effect estimates are normally distributed or that the effect sizes' standard errors are known when they are in fact estimated (as is done using standard inverse variance methods).[105] Due to the often sparse nature of adverse event data, the between-study heterogeneity parameter will often be very poorly estimated, and frequently estimated to be 0 (i.e. reducing the random effects model to a fixed effect one) even when heterogeneity is present.[103] Due to this, and the fact that exact likelihood methods incorporating random effects were, until recently, difficult to implement, the Cochrane Handbook advises that: "incorporation of heterogeneity into an estimate of a intervention effect should be a secondary consideration when attempting to produce estimates of effects from sparse data – the primary concern is to discern whether there is any signal of an effect in the data."[100] However, in the context of an economic evaluation where the aim is to include as much variability and uncertainty of parameter estimates as possible, as well as use the most "appropriate" estimate of effect, the argument for using random effects is perhaps more persuasive when heterogeneity is present. Fortunately the implementation of exact methods, such as logistic regression including random effects, is becoming more commonplace in standard statistical software,[105] as well as development of methods specifically targeted at the random effect sparse data meta-analysis context.[106] Bayesian MCMC models usually use exact likelihood specifications and hence provide an appealing alternative to frequentist methods,[9,81,103] although care needs to be taken when specifying prior distributions which are intended to be vague as, particularly for the heterogeneity variance, these can be more informative than intended when data are sparse.[107] A further issue relates to how to deal with "double-zero" studies; that is, studies with no events of interest in *either* arm of a two-arm trial. Until recently, the view was that such trials should be excluded from the analysis of relative measures of effect, such as the odds ratio or relative risk[100] (however, recent publications have challenged this view[106,108] and argued that they should be included and these claims need further investigation). Given the challenges of parameter estimation in sparse networks there

may be “pressure” to “lump” interventions together in order to make networks estimable (see Section 5.1).

There are further issues which need consideration when undertaking NMA of adverse event (and sparse) data. Fortunately, as for MA, many Bayesian NMA models (both fixed and random effects) do use exact likelihood specifications (e.g. those outlined in the NICE Technical Support Documents[109]). A frequentist exact (logistic regression) random effect NMA model has also recently been described by Simmonds and Higgins.[105] A further consideration is the distribution of the single- (i.e. zero observations in one arm of the trial) and double-zero studies throughout the network. While there would appear to be some controversy as to whether double-zero studies should be included in a pairwise meta-analysis (on a relative scale), this issue remains pertinent for NMA (and studies with more than two arms and multiple zeros need further investigation), but raises associated concerns regarding network node connections (see Section 4). For example, i) Can double-zero studies alone be considered to form connections between intervention nodes when no other studies exist to link a node into a network? And ii) Can an intervention be included when all studies that include the intervention observe zero events for that intervention? There is clearly a need for further work to decisively answer such questions, but in the meantime the authors advise that at least one event needs to be observed for any intervention node included in the analysis and double-zero studies should be initially excluded, with the effect of including them being explored as a sensitivity analysis. Even if these principles are adhered to, our experience is that for networks with very sparse numbers of events, problems of estimation can exist and care needs to be exercised to ensure “sensible” answers from converged estimation algorithms are obtained (note also that zeros can create problems when applying model fit statistics). A further issue relates to adverse events which are essentially impossible under some intervention options; for example, febrile neutropenia can only occur in trial arms of cancer studies in which chemotherapy is given. In such situations estimating relative effects between chemotherapy and non-chemotherapy arms will not be possible, and estimating absolute rates using single trial arms may be the best solution; however, this needs further research.

A small, but highly innovative, literature has started to appear for advanced methods for NMA of sparse data.[110-113] A recurring theme is the desire to make estimation more robust, and reduce uncertainty by including and/or borrowing strength across further information sources including multiple (safety) outcomes,[110] drugs within the same class,[111] varying doses of drugs,[111,112] observational studies[113] and formally elicited expert opinion.[113] Soares et al.[113-115] present an interesting application on leg ulcer healing over several publications which are recommended reading as they include a complex evidence synthesis[113] with sensitivity analysis incorporating elicited expert opinion[113,115] which informs an economic model[114] and value of further information.[114]

Developments thus far have focused on the assumption that the risk of adverse events is constant over time. Assessing time-dependent effects, using summary data alone, will be challenging, although possible if enough data are reported[9] but greatly facilitated by the availability of IPD,[110] as illustrated elsewhere.[116] Along with this further development of methods comes an ever-expanding number of modelling options.[110] Many authors stress the importance of the use of sensitivity analysis to explore the robustness of estimation to the specific modelling choices/assumptions made[110,111,113] and many of the above innovations were put forward in this spirit, that is, as alternatives to simpler modelling approaches. Even without the use of these advanced methods many modelling options exist (although the sparsity of data means often little power will be available to choose between alternative models and check modelling assumptions including consistency of network loops, etc.), and we recommend sensitivity

analysis exploring the impact of analysis choices including statistical model specification/method of parameter estimation, any prior distributions used and the inclusion of double-zero studies.

A guideline paper on Bayesian NMA for drug safety has recently been published[110] which is recommended reading and goes into (more) depth on many important issues (some are outlined in the next section) and provides a helpful reporting checklist. In this drug safety guideline much emphasis is placed on the need for sensitivity analysis, the importance of which is discussed above, and the need for further work consolidating what appears to be output from multiple groups, working relatively independently, is necessary before more prescriptive guidelines can be developed.

9. OTHER ISSUES AND ADVANCED TOPICS

This section covers advanced and emerging topics relating to NMA and its use in HTA and economic evaluation. Sections 9.1 through 9.4 focus on model extensions to synthesise specific data types and structures, often when the outcomes and data available from the relevant studies are not all the same. Many of these could be considered bespoke NMA models motivated and developed as solutions to specific HTA contexts; allowing for the complexities in data structure that were encountered. Regarding utilising these methods in future HTAs, it may be that one of these approaches is a perfect fit for the topic in hand, but, more likely, a degree of modification will be required; perhaps utilising ideas from multiple papers (in addition to further innovations). This section does no more than provide a brief overview of the area, but does include many references that are recommended further reading. Many of the cited papers utilise WinBUGS and supply the code to fit the described models which will greatly facilitate the process of utilising the models described, although it should be acknowledged that this level of modelling is advanced and only recommended for those with previous knowledge and experience.

The second case study (Section 10) highlights a particularly interesting advanced bespoke modelling approach which was chosen because it utilises multiple cutting-edge ideas in an effort to improve estimation when data are few and sparse (a relatively common situation in our experience).

9.1 Shared parameter models

Relevant outcome data may be available from multiple studies but the format of the data varies between studies. As outlined in Section 5, continuous outcomes (e.g. weight loss, change in blood pressure, etc.) may be presented in study reports for individual arms, in terms of the mean and variance in each intervention group being compared, or as the mean difference (with corresponding variance) *between* intervention groups.[49] Using MCMC methods it is relatively straightforward to directly synthesise different data formats by coding the associated likelihoods (required for the different data formats) but ensuring these likelihoods include appropriate parameters in common. These models are sometimes referred to as shared parameter models (see Section 4 of Dias et al. 2013[9] and relevant examples cited therein[71,73,117-120]). Specific models of this type are also considered briefly below in Sections 9.3 and 9.4 for time to event outcomes and simultaneous use of IPD and aggregate data respectively.

9.2 Multiple outcomes

Commonly in meta-analysis different outcomes have been analysed separately (or outcomes have been standardised before combining, but use of such outcomes in a decision modelling context is problematic at best), but there is a considerable literature on the joint synthesis of multiple outcomes.[121,122] In a pairwise meta-analysis context, this can provide efficiency gains, in terms of parameter estimation, if the correlations between outcomes are known[123] or if some studies do not measure/report all outcomes.[124] In the latter case, even if only one of the outcomes is required for a decision model, strength can be borrowed from including studies which only report the other outcome through the estimated relationship between outcomes. This could be relevant, for example, in instances where a long-term outcome is of interest but a surrogate outcome is more commonly measured in the trials.[125] NMA models to fit multiple outcomes with binary, continuous, time to event (survival) or mixed outcomes have recently been described.[126,127] If an exchangeability assumption for intervention effects across outcomes is made, then estimates for intervention effects for outcomes for which no data are available can be obtained.[128] Very recently an alternative approach to synthesising multiple outcomes has

been considered[129] which maps the outcome measures onto one another and allows intervention effects to be expressed on any of the outcome scales that have been used; this alternative approach needs more evaluation but shows promise.

a) Competing risks

A special case of multiple outcomes occurs when there are several different failure time outcomes that are considered mutually exclusive; these are competing risk outcomes. Once a patient has reached any one of these end points (or is censored), they are considered to be out of the risk set.[130] While it may appear that separate univariate NMAs could be conducted on each outcome, a joint synthesis is required to take into account the statistical dependencies induced by the competing nature of the outcomes. In order to achieve this, a multinomial likelihood and appropriate link function are incorporated into the standard NMA model[130] and the likely effect of this compared to multiple univariate analyses is discussed by Trikalinos and Olkin 2008.[131]

b) Multiple time points

Studies may report outcomes at multiple time points and potentially different time points from each other. While separate analyses could be carried out at each time point, this would be inefficient as information could be borrowed across time points and thus used to inform the required parameters for the decision model. This approach has been considered for binary outcomes[73] and has even been shown to connect multiple networks which were disconnected at specific time points.

9.3 Time to event (survival) data

In the simplest instance relative intervention effects based on time to event data can be summarised using reported (ln) hazard ratios (derived from parametric or non-parametric survival analyses)[132] and combined using standard NMA methods.[9] But numerous complications can exist which require further consideration. Firstly, trials may not report hazard ratios but cumulative number of events (i.e. the total number of subjects who have experienced an event by a specific time point) and a method to combine this sort of data with hazard ratios has been developed.[133] A further outcome that is sometimes reported is median survival times, and approaches to combine this with reported cumulative number of events and (ln) hazard ratios have been considered.[134] This work presents this extension in the context of competing risk outcomes (see Section 9.1.1) and shows how the inputs required for the associated economic decision model – mean progression time and mean survival time – can be derived. The mean time to an event is also sometimes reported and a further paper combines this type of data with median time and count data using a parametric survival model (see below).[71]

A further issue is that working with hazard ratios assumes the intervention effect is constant for the duration of the trial data. Often a certain parametric survival function is assumed for the baseline intervention (e.g. Weibull) and the intervention-specific hazard ratios from the NMA are applied to this (including for extrapolation to time points beyond the included trials when required by decision models). This implies that intervention effects only impact on one of the parameters of the survival distribution (e.g. the scale parameter for a Weibull distribution) when the survival distribution is described by multiple parameters.[135] A further concern is that since the tail of the survival function can have considerable impact on the expected survival, violations of the constant hazard ratio assumption can lead to large changes in parameter estimates used in the economic decision model.[136] As an alternative to using hazard ratios Ouwens et al.[135] developed an NMA model based on fitting parametric survival curves illustrated by fitting a model in which the intervention effect can impact both parameters of a Weibull distribution

(although a Gompertz, log-logistic or log-normal distribution could be used) and thus if the shape parameters for two different interventions differ, the hazard ratio will not be constant over time. In order to utilise this more flexible model, cumulative number of event data at multiple time points per trial is required (this was derived from published Kaplan-Meier curves in the example presented). In a related paper (using data in the same format as above) Jansen[137] developed an NMA model for survival data using fractional polynomials to model the hazard function to relax the constant hazard ratio assumption (instead of multi-parameter survival distributions described above) and model intervention effect using several parameters. The aim of both approaches is that the intervention effects expressed more closely fit the available data. A further option, allowing flexible modelling, would be to use individual patient data (see below).

9.4 Individual participant data (IPD)

This report has focused on the synthesis of aggregate/summary effectiveness data essentially assuming that data for the NMA is being extracted from the published literature. However, if IPD was available then this could be used in the analysis instead. The benefits of conducting an IPD synthesis over one using summary data generally are well reported[138] so focus here is on specific advantages for economic decision modelling. If there is interest in patient-specific covariates, either to explain between-study heterogeneity/inconsistency[75,139] or to explore cost-effectiveness for subgroups of patients an IPD analysis can have much more power than one relying on summary statistics and aggregated study level covariates (e.g. using average patient age as an intervention by covariate interaction in a regression [Section 5.2] compared to the individual ages of every patient etc.). A further instance when IPD modelling would be beneficial is when the decision model utilises patient-level simulation (Section 3). While there is a history of meta-analysing IPD in a two-stage process[140] by first reducing the IPD into summary statistics, a single one-stage process is recommended here so correlations between interactions and main effects can be propagated easily through to the decision modelling.[10] NMA IPD models which allow the inclusion of patient-level covariates have been developed for binary outcome data[139,141] and time to event (survival data).[142] Since IPD may not be available from all relevant studies, several modelling extensions allowing the simultaneous synthesis of IPD and summary/aggregate data have also been described.[141-143]

9.5 Informing Markov transition matrices

One model type often used to model chronic conditions which reoccur, such as asthma or epilepsy, is multi-state Markov models. At the heart of such models are transition matrices which contain the probability estimates individuals move between the defined health states (e.g. well, mild symptoms, major symptoms, etc.) for the different interventions under evaluation. Perhaps the most commonly used approach to estimate these transition probabilities is through the use of single individual patient datasets. However, philosophically, there is no reason why multiple relevant data sources should not be used and evidence synthesis methods adopted to synthesise them (as has been advocated through this document). Following initial work to estimate the transition matrix for the baseline intervention, using multiple studies reporting data in multiple formats,[144] further work also estimating intervention effects using NMA has been described.[145] An interesting aspect of this work is the exploration of the fit of alternative intervention models regarding which transitions the interventions are assumed to work on. Taking this a step further, it has been demonstrated how uncertainty in the most appropriate intervention model can be accommodated in the analysis by averaging results over alternative candidate models.[146]

9.6 Bias models

Concerns about variable study quality and various types of publication biases are common threats to the validity of meta-analyses in general. While many instruments exist for assessment of trial quality no routine methods for incorporating the uncertainty induced by sub-optimal trials in a pairwise meta-analysis exist, although experimental approaches have been considered using both expert opinion[61,64] and empirical estimates derived from the published literature.[61,147] This is a topic requiring further research generally, and specifically for NMA.

Several methods for testing and adjusting pairwise meta-analyses for publication bias/small study effects have been developed[89] and these are starting to be translated across into an NMA context.[121,147]

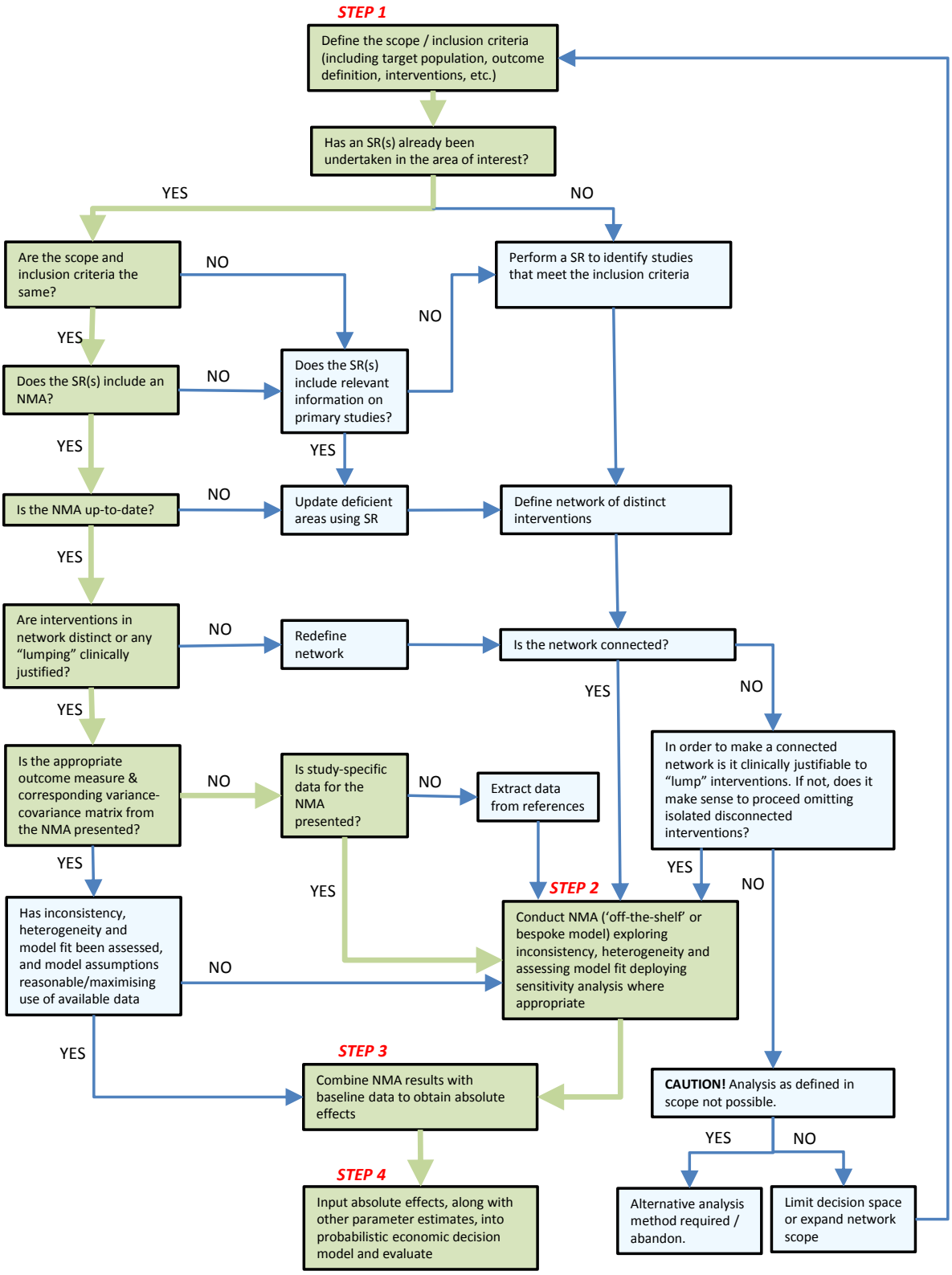
10. CASE STUDIES

10.1 Case study 1: An evaluation of cost-effectiveness of medicinal poisoning prevention practise in households with pre-school children

This case study evaluates the cost-effectiveness of medicinal poisoning prevention practise in households with pre-school children. As discussed in Step 2) of the general framework for interfacing NMA with economic decision models (Section 5), there are a number of measures from a RE NMA that may be input in an economic model and the appropriate measure depends on the interpretation of heterogeneity in the studies included and how this relates to the target setting of the decision.[76] This case study illustrates how using different summary measures from the synthesis models (NMA for the relative effects, and MA for absolute effects under the reference intervention (in this example, Usual care)), may affect the cost-effectiveness results and subsequently the overall decision (See Section 5.2). In this example, two evaluations are presented – one using the random effects (posterior) mean and the second using the predictive distribution. (For other possible measures that could have been adopted see Section 5.2).

A summary of the decision pathway (illustrated through completing the tool presented in Figure 6) for conducting NMA to inform the clinical parameters in this evaluation of the cost-effectiveness of medicinal poisonings prevention in households with pre-school children is presented in Figure 7. Here the green boxes and arrows indicate the path taken through the tool. A more detailed description, split up into the four steps outlined in Figure 2**,** is given below.

FIGURE 7: DECISION PATHWAY FOR CONDUCTING THE NMA TO INFORM CLINICAL PARAMETERS IN THE ECONOMIC DECISION MODEL TO EVALUATE THE COST-EFFECTIVENESS OF MEDICINAL POISONING PREVENTION PRACTISES IN HOUSEHOLDS WITH PRE-SCHOOL CHILDREN



Step 1: Define scope of decision problem and associated evidence network for NMA
For this case study the scope of the decision problem was defined as follows:

a) **Target population:** Pre-school children aged 4 years or below.

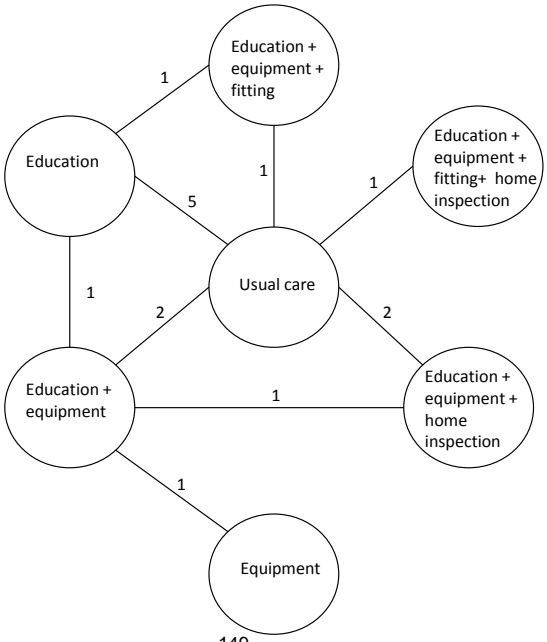
b) **Interventions:** Medicinal poisoning prevention interventions for households – i) Usual care, ii) Education (e.g. parent information leaflet), iii) Equipment (e.g. free or low-cost cupboard locks), iv) Education + Equipment, v) Education + Equipment + Fitting, vi) Education + Equipment + Home safety inspection, and vii) Education + Equipment + Fitting + Home safety inspection.

c) **Study design**: RCTs and observational studies (note, for purposes of simplicity for this case study, potential differential biases by study design have not been considered in the NMA model).

d) **Outcomes:** Proportion of households practising safe storage of medicines.

A systematic review was identified in the literature that matched the scope outlined above[94,148] and based on this data an NMA had been published.[149] Figure 8 presents the connected network diagram depicting the evidence base to which the published NMA was fitted. The number of studies for each pairwise comparison in the network is presented on the linking lines between intervention nodes. Note that all the studies in the network, except one, were two-arm studies. There was one three-arm study which compared Usual care, Education + Equipment and Education + Equipment + Home safety inspection.

FIGURE 8: NETWORK DIAGRAMS FOR THE SAFE STORAGE OF MEDICINES INTERVENTIONS FOR HOUSEHOLDS WITH PRE-SCHOOL CHILDREN



Source: Reprinted from Achana et al. 2015.[149]

Step 2: Estimate intervention effects relative to reference intervention using NMA

The NMA published by Achana et al.[149] fitted a standard NMA random effects model with a binary outcome[9] to the study data taking into account the correlation structure induced by the study with three arms.[9] Unfortunately, the published NMA[149] did not include the summary measures and corresponding variance-covariance matrix from the NMA but did provide the study data by intervention arm, which allowed the NMA model to be fitted utilising the same study data. For this case study, as mentioned above, both the random effects (posterior) mean distribution of the intervention effects relative to Usual care (and each other) and the predictive distribution, were obtained from NMA. For comparison, both distributions are presented in Table 1(a). Note the mean estimates are very similar for both measures (only differing by MC error) but the uncertainty is greater (i.e. 95% credible intervals [CrI] wider) for the predictive distribution as it incorporates both the uncertainty about the value of a new observation as well as the observed variation in the data.

The between-study variability, a measure of heterogeneity (i.e. the variability in intervention effects within pairwise comparisons above that expected by chance[150]) was estimated to be 0.33 (95% CrI, 0.01 to 1.24) suggesting a reasonable degree of heterogeneity.[81]

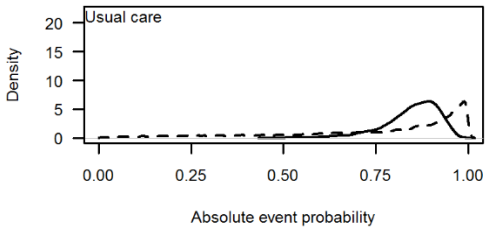
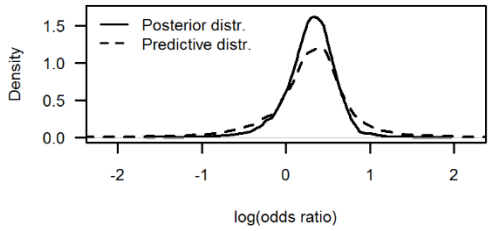
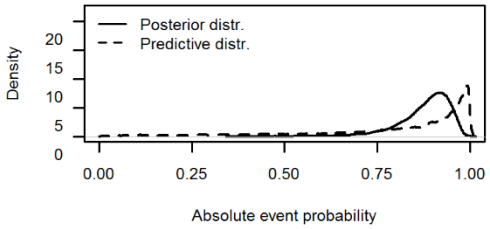
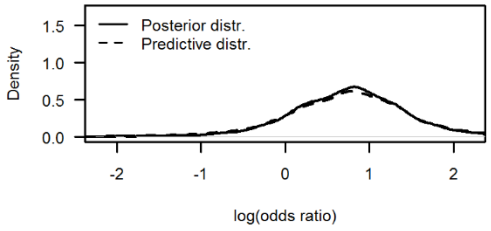
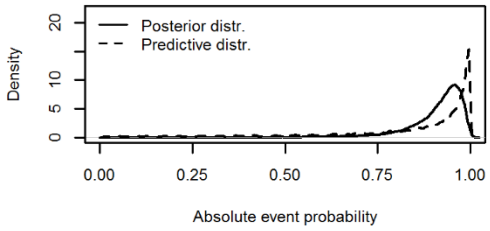
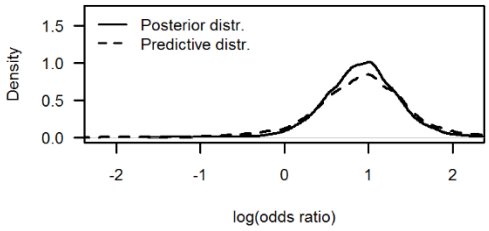
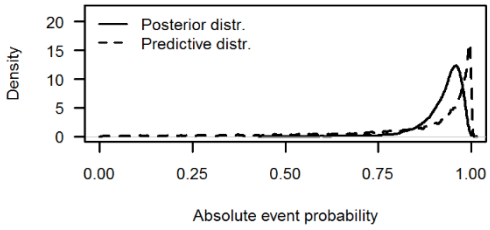
There was no evidence of inconsistency between the direct and indirect evidence in all networks. This was assessed, where both direct and indirect evidence was available, for closed loops (excluding loops formed by multi-arm studies) in the network, using the node-split method (but the potentially low power of the assessment should be appreciated).[151]

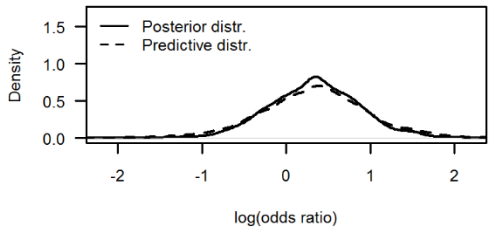
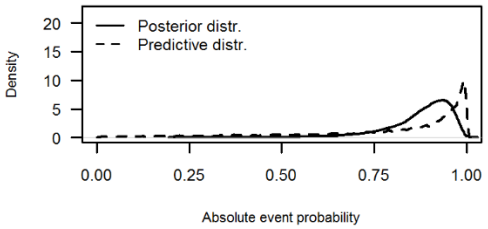
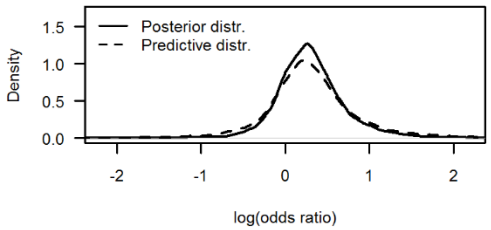
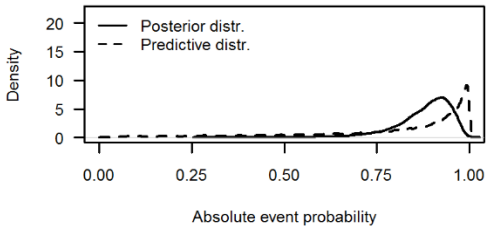
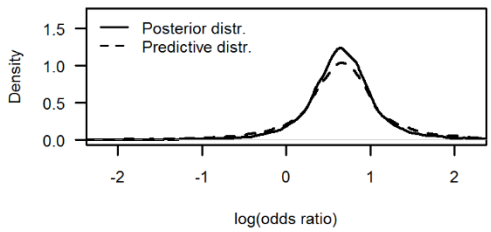
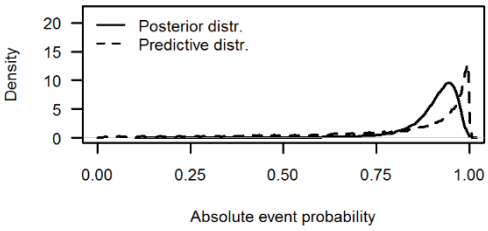
The goodness of fit of the model to the data was assessed by calculating the posterior mean residual deviance.[77] Under the null hypothesis, the model provides an adequate fit to the data if the posterior mean residual deviance equals the number of unconstrained data points;[152,153] in this case the mean residual deviance is approximately equal to the number of unconstrained data points (i.e. residual deviance = 23.5 compared to 24 unconstrained data points).[77,153]

Step 3: Estimate absolute effects derived from NMA and baseline data

The probability of a household practising safe storage of medicines on the reference intervention (in this case, Usual care) was calculated by fitting a separate synthesis to the Usual care arm data from studies included in the NMA which included a Usual care arm.[12] As for the relative effects, absolute effects under the reference intervention (i.e. the probability of a household practising safe storage of medicines under Usual care) was represented by both the random effects (posterior) mean distribution or the predictive distribution (Table 1 (b) Usual care). Estimates of the probability of a household practising safe storage of medicines under each of the different intervention strategies was then derived from the NMA relative effect estimates for each intervention (compared to Usual care) and the probability of a household practising safe storage of medicines under Usual care (i.e. absolute effects under the reference intervention), maintaining the correlation structure as discussed in Section 5.4. Table 1(b) (which presents both distributions). Again it can be observed that the means are similar but the uncertainty associated with the predictive distribution is much wider due to the incorporation of uncertainty about the value of a new observation as well as the observed variation in the data in both the relative effects and absolute effects under Usual care.

TABLE 1: POSTERIOR MEAN AND PREDICTIVE DISTRIBUTIONS FOR RELATIVE AND ABSOLUTE EFFECTS ESTIMATED IN STEPS 2) AND 3)

Interventions	(a) Relative effects: Odds Ratios compared to Usual care	(b) Absolute effects: Probability of practising safe storage of medicines
Usual care	N/A	 Posterior mean = 0.86 (0.66, 0.95) Predictive mean = 0.87 (0.13, 1.00)
Education	 Posterior mean = 1.37 (0.71, 2.22) Predictive mean = 1.39 (0.44, 3.59)	 Posterior mean = 0.89 (0.70, 0.97) Predictive mean = 0.90 (0.15, 1.00)
Equipment	 Posterior mean = 2.53 (1.06, 5.96) Predictive mean = 2.56 (0.7, 8.52)	 Posterior mean = 0.94 (0.80, 0.98) Predictive mean = 0.94 (0.24, 1.00)
Education + Equipment	 Posterior mean = 1.41 (0.48, 4.29) Predictive mean = 1.42 (0.35, 5.38)	 Posterior mean = 0.90 (0.65, 0.98) Predictive mean = 0.90 (0.15, 1.00)

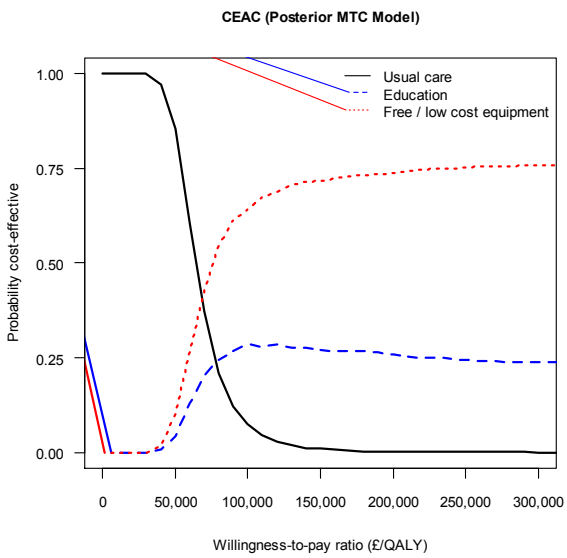
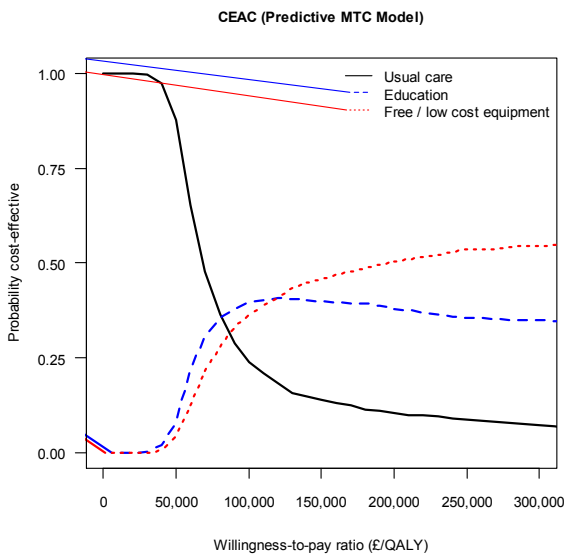
Education + Equipment + Home Inspection	 Posterior mean = 1.32 (0.64, 3.58) Predictive mean = 1.31 (0.46, 5.40)	 Posterior mean = 0.89 (0.68, 0.97) Predictive mean = 0.90 (0.15, 1.00)
Education + Equipment + Fitting	 Posterior mean = 1.94 (0.77, 5.28) Predictive mean = 1.94 (0.53, 7.54)	 Posterior mean = 0.92 (0.74, 0.98) Predictive mean = 0.93 (0.2, 1)
Education + Equipment + Fitting + Home Inspection	 Posterior mean = 2.17 (0.5, 8.56) Predictive mean = 2.15 (0.39, 11.07)	 Posterior mean = 0.93 (0.69, 0.99) Predictive mean = 0.93 (0.19, 1)

Step 4: Evaluate the probabilistic decision model utilising absolute effects

Finally, the absolute effects estimated in Step 3), along with other model parameters, were input into the purposely developed economic model and evaluated using approach i) outlined in Section 5.4 (i.e. Bayesian posterior simulation one-stage approach). The evaluation (including the NMA, calculation of absolute effects (derived from NMA and baseline data) and economic model) were performed within a single WinBUGS[14] programme "controlled" from R[21] using R2WinBUGS.[86]

Figure 9 displays the cost-effectiveness acceptability curves calculated from both analyses. In both of these, Usual care has overwhelmingly the highest probability of being the most cost-effective at low willingness-to-pay values, but between £50,000 and £100,000 (UK sterling) the probabilities of being most cost-effective for education and free/low-cost equipment strategies start to overtake Usual care. Notice, due to the additional uncertainty associated with the predictive distribution, the cost-effectiveness acceptability curves are "bunched" closer together in panel b).

FIGURE 9: COST-EFFECTIVENESS ACCEPTABILITY CURVES UTILISING THE A) RANDOM EFFECTS (POSTERIOR) MEAN DISTRIBUTION AND A) PREDICTIVE DISTRIBUTION FOR RELATIVE INTERVENTION EFFECTS (OR) AND ABSOLUTE EFFECTS UNDER USUAL CARE IN THE ECONOMIC MODEL

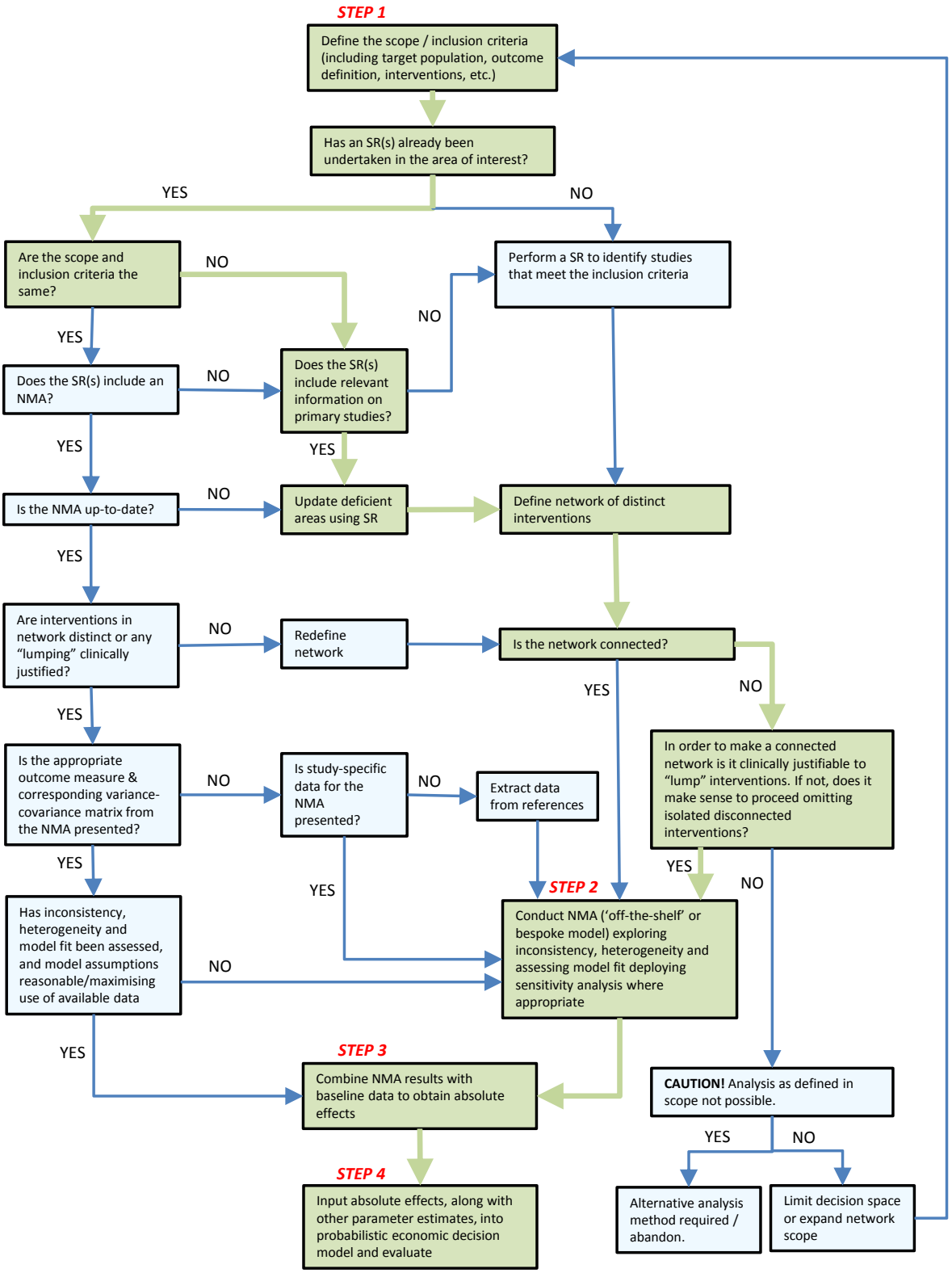
a) Deriving probability of practising safe storage of medicines from random effects (posterior) mean distribution of ORs and baseline probability	b) Deriving probability of practising safe storage of medicines from predictive distribution of ORs and baseline probability
	

10.2 Case study 2: An evaluation of the effectiveness of smoking cessation interventions: A bespoke NMA developed specifically to inform a cost-effectiveness decision model

This case study presents an evaluation of the effectiveness of smoking cessation interventions. This recently published evaluation[73] provides an excellent example of the state-of-the-art for developing and using an NMA to inform economic decision models. While we highlight features of this work below, and cross reference to where related issues are discussed in this report, we encourage the reader to obtain the original detailed description of the evidence synthesis[73] and read it in full (online appendices to the paper are also available which provide the trial data and the specific WinBUGS code used). The overarching aims of the analysis were to include as much of the trial evidence as possible – while modelling it as realistically as possible – and using statistical approaches to assess the fit of competing alternative model specifications. The full HTA report,[154] which utilises this synthesis, is also available and includes a more "traditional" review of effectiveness which summarises pairwise comparisons between interventions using standard pairwise meta-analysis. Having results of both a standard meta-analysis and bespoke NMA of the same literature, for which an economic model is required, clearly illustrates some of the advantages of using an NMA with a customised structure (see below) for informing the economic evaluation. For purposes of this case study we have concentrated on the synthesis model presented by Madan et al.[73]

A summary of the decision pathway for conducting NMA to inform the clinical parameters in this evaluation of the cost-effectiveness of smoking cessation interventions is presented in Figure 10. A more detailed description, split up into the four steps outlined in Figure 2, is given below.

FIGURE 10: DECISION PATHWAY FOR CONDUCTING THE NMA TO INFORM CLINICAL PARAMETERS IN THE ECONOMIC DECISION MODEL TO EVALUATE THE COST-EFFECTIVENESS OF SMOKING CESSATION INTERVENTIONS



Step 1) Define scope of decision problem and associated evidence network for NMA

The case study considers the use of electronic aids (such as websites or text messaging services) as an adjunct to standard care for smoking cessation with the aim of identifying whether such aids are cost-effective, and if so, which ones are the most cost-effective. The cost-effectiveness model took the form of a decision tree, in which a successful outcome was defined as 12-month continuous abstinence from smoking.

a) **Target population:** Two distinct adult populations i) those making a committed quit attempt using the electronic aid as an adjunct to pharmacological interventions, and ii) those at an earlier stage in the quitting process who are less motivated and not using pharmacological interventions.

b) **Interventions in the comparator set:** The first thing to note is that for non-pharmacological interventions such as these, the definitions of the interventions may be diverse, multifaceted and difficult to comprehensively categorise (see Welton et al.[76] for a further example which specifically considers the modelling of complex packages of care, and Achana et al.[96] for consideration of the specific challenges of synthesis for decision making in a public health context). In the case study the evidence base included trials evaluating computer-generated printed materials, stand-alone computer programs, text services, e-mails, static and interactive websites, bulletin boards, chat rooms and on-line forums. In such situations there is a tension between the desire to draw fine distinctions between specific technologies with limited data, while producing generalizable conclusions on the different types of electronic interventions. The authors, having taken advice from smoking cessation experts on the key dimensions that would influence the effectiveness of smoking cessation programmes, developed a system with five categories based on consideration of the two dimensions: i) whether the intervention provided generic advice or tailored feedback; and ii) whether the intervention used a single or multiple channels. Similarly, the non-electronic interventions that could be given concurrently, or in control arms, were categorised into five groups also. Along with a sixth category in both groupings for no-electronic or no non-electronic component, this categorisation system still led to 36 possible intervention combinations that could be defined for any trial arm in the evidence base.

c) **Study design inclusions:** RCTs

d) **Outcomes:** Although the cost-effectiveness model required effects on 12-month continuous abstinence, the times at which the trials recorded smoking status varied from one to 24 months and studies reported these at between one and three follow-up times. *Of the 58 studies in the evidence base, only four reported the outcome that was required for the economic model (!).* As the authors note, using traditional approaches such as excluding the other 54 trials, or conducting the synthesis at the most commonly reported time point (six months, reported in 12 studies) and basing the 12 month abstinence on extrapolation assumptions, would be extremely wasteful in terms of excluding data that is potentially relevant, but not estimating the precise quantity of interest. Faced with this, the authors developed a model based on survival curves that incorporated all reported time points from all identified trials.

   However, a further complication was that only 28 of the studies reported continuous abstinence while the remaining 30 only reported point abstinence (whether or not a participant is smoking at a certain point of time rather than completely refrained for the whole follow-up period). But 26 of the 28 studies that reported continuous abstinence also

reported point abstinence. From this data, the nature and strength of the relationship between the two outcomes could be established and this, in turn, allowed the incorporation into the synthesis model of the 30 studies which did not report continuous abstinence. Hence, although not reporting the outcome of interest, these studies now contributed information to its estimate of effect through the estimated relationship between outcomes (see Section 9.2 for an overview of multiple outcome approaches to NMA). This example acutely highlights the difficulty in defining *a priori* exactly what data are "*relevant"* for a given application. It is for this reason that in these guidelines we have avoided simply suggesting that "all relevant data/studies should be included" (which has been suggested elsewhere[155]) while acknowledging we cannot provide a prescriptive description of what the "relevant evidence" is that is broadly generalizable.

A systematic review identified 58 relevant RCTs with a total of 151 arms, and this evidence base was used to derive the estimates of efficacy used in the decision model.

Step 2) Estimate intervention effects relative to reference intervention using NMA
(Full model details and WinBUGS code are available in the original paper.)

The model was a bespoke NMA incorporating multiple (and surrogate) outcomes and follow-up times and fitted using MCMC methods in the WinBUGS software with vague prior distributions specified throughout (and sensitivity analysis to their impact conducted). Alternative statistical models for time to event (relapse) were explored for continuous abstinence as were models relating point to continuous abstinence. When estimating the efficacy of the different interventions, data were not available to estimate all 36 categories of intervention (defined above) so models with fewer parameters, but making further assumptions, were explored assuming a) the effect of the electronic and non-electronic components were additive (implying that the effect of an intervention with both active electronic and conventional interventions is equal to the sum of the effect of each intervention given in isolation); and b) the effect of each electronic intervention was the same irrespective of its category. Consistency of the NMA model was explored by fitting an inconsistency model. The DIC[77] was used to choose between the alternative model specifications.

Model estimation suggested that electronic aids to smoking cessation are likely to be effective in reducing relapse rates among those making an attempt to quit. Including information on a surrogate marker (point abstinence) only slightly reduced uncertainty in the main outcome (continuous abstinence), and for simplicity was ultimately excluded for simplicity in the final cost-effectiveness modelling.

A bespoke synthesis model allowed a coherent synthesis of studies which varied in terms of outcome measure, follow-up period and characteristics of the intervention. These factors prevented a more conventional synthesis of the evidence base, which would have had to exclude the majority of the studies. Bayesian MCMC methods are highly adaptable to complex evidence structures and offer specific benefits where "off-the-shelf" models are inadequate.

This case study is advanced, and a "tall order" to try and emulate for a less experienced analyst. But it is included here to show what is possible if the preconceived ideas of what a meta-analysis is are broken down; our hope is it inspires others to start thinking about what can be achieved by more realistic modelling of the data and taking initial steps to do this by adapting models as required by specific situations.

Step 3) and 4) Estimate absolute effects derived from NMA and baseline data, and evaluate the probabilistic decision model utilising absolute effects

As mentioned above, this case study focuses on the bespoke synthesis model developed specifically by Madan et al.[73] to inform the economic model presented elsewhere.[154] For completeness, Steps 3) and 4) in the framework are very briefly presented below, and the reader is referred to the HTA report[154] for more in-depth technical details. The estimates of baseline continuous abstinence to 12 months were taken from previous economic evaluations of conventional therapies. It was assumed that effectiveness of the electronic aid interventions was the same regardless of the control arm intervention (i.e. the effect of the electronic aid intervention is additive when used as an adjunct to pharmacological and/or counselling control interventions). In this way, probabilities of relapse at 12 months – the outcome required for the decision model – were derived from hazard ratios estimated from the NMA. More details can be found in Chen et al.[154] The synthesis was fitted in WinBUGS but the HTA report provides no details about how the estimates from the synthesis model were input, along with other input parameters, into the economic model and evaluated.

11. CONCLUDING REMARKS

Evidence synthesis to inform input parameters in economic decision models for health technology assessments has made great methodological advances in recent years, as has been demonstrated throughout these guidelines. However, it is still an evolving area with many unresolved issues and complications beyond simply including multiple intervention options. These include inconsistency of i) time of outcome reported, ii) population studied, and iii) outcome definitions. Therefore, it is hard to be completely prescriptive as to the best synthesis model for a given question as “off-the-shelf” solutions may not yet have been developed and further bespoke synthesis models to comprehensively model the complexities of the problem and accommodate in-cohesive study data are still required.

ACKNOWLEDGEMENTS

Dr Nicky Welton is funded by the ConDuCT-II Medical Research Council (MRC) Hub for Methodology Research, UK.

REFERENCES

1. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997 Jun;50(6):683-91.

2. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002 Aug 30;21(16):2313-24.

3. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med. 1996 Dec 30;15(24):2733-49.

4. Eddy DM, Hasselblad V, Shachter RD. Meta-analysis by the confidence profile method: the statistical synthesis of evidence. Boston: Academic Press; 1992.

5. Hasselblad V. Meta-analysis of multitreatment studies. Med Decis Making. 1998 Jan;18(1):37-43.

6. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis, 2nd ed. Boston: Academic Press; 1994. p. 357-76. Chapter 19.

7. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol. 2014 Feb;67(2):138-43.

8. Cooper NJ, Spiegelhalter D, Bujkiewicz S, Dequen P, Sutton AJ. Use of implicit and explicit Bayesian methods in health technology assessment. Int J Technol Assess Health Care. 2013 Jul;29(3):336-42.

9. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making [Internet]. 2013 Jul [cited 2015 Jul 28];33(5):607-17. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704203

10. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making [Internet]. 2013 Jul [cited 2015 Jul 28];33(5):618-40. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704206

11. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making [Internet]. 2013 Jul [cited 2015 Jul 28];33(5):641-56. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704208

12. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making [Internet]. 2013 Jul [cited 2015 Jan 20];33(5):657-70. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704201

13. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. Med Decis

Making [Internet]. 2013 Jul [cited 2015 Aug 12];33(5):671-8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704202

14. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS user manual [Internet]. Cambridge, United Kingdom: MRC Biostatistics Unit, Institute of Public Health; 2003 Jan. [cited 2015 Aug 26]. Available from: http://www.uclouvain.be/cps/ucl/doc/stat/documents/manual14.pdf

15. Malone DC. Using indirect comparisons in pharmacoeconomic studies--time for implementation. Clin Ther. 2007 Nov;29(11):2454-5.

16. White IR. Network meta-analysis. Stata J. 2015;15(4):951-85.

17. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PLoS One [Internet]. 2013 [cited 2015 Jul 28];8(10):e76654. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3789683

18. Brown S, Hutton B, Clifford T, Coyle D, Grima D, Wells G, et al. A Microsoft-Excel-based tool for running and critically appraising network meta-analyses--an overview and application of NetMetaXL. Syst Rev [Internet]. 2014 [cited 2015 Jul 28];3:110. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4195340

19. Bujkiewicz S, Jones HE, Lai MC, Cooper NJ, Hawkins N, Squires H, et al. Development of a transparent interactive decision interrogator to facilitate the decision-making process in health care. Value Health [Internet]. 2011 Jul [cited 2015 Jul 28];14(5):768-76. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3161376

20. van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. Res Synth Methods. 2012 Dec;3(4):285-99.

21. R Development Core Team. R: a language and environment for statistical computing [Internet]. Vienna, Austria: The R Foundation for Statistical Computing; 2011. [cited 2015 Aug 26]. Available from: http://www.gbif.org/resource/81287

22. Sutton AJ, Higgins JP. Recent developments in meta-analysis. Stat Med. 2008 Feb 28;27(5):625-50.

23. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. Stat Med. 2007 Sep 10;26(20):3681-99.

24. Achana FA, Cooper NJ, Dias S, Lu G, Rice SJ, Kendrick D, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Stat Med. 2013 Feb 28;32(5):752-71.

25. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ [Internet]. 2004 Jun 19 [cited 2015 Jul 30];328(7454):1490. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC428525

26. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess. 2003;7(1):1-76.

27. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for meta-analysis in medical research. Chichester, England: John Wiley & Sons, Ltd.; 2000. (Wiley series in probability and statistics).

28. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics. 2008;26(9):753-67.

29. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23(20):3105-24.

30. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ [Internet]. 2005 Oct 15 [cited 2015 Aug 4];331(7521):897-900. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1255806

31. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011 Feb;64(2):163-71.

32. Bates MJ. The design of browsing and berrypicking techniques for the online search interface. Online Review. 1989;13(5):407-24.

33. Welton NJ, Sutton AJ, Cooper N, Abrams KR, Ades AE. Evidence synthesis for decision making in healthcare. Chichester, England: Wiley; 2012 May.

34. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1. Med Decis Making [Internet]. 2013 Jul [cited 2015 Aug 26];33(5):597-606. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704205/pdf/10.1177_0272989X13487604.pdf

35. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making [Internet]. 2013 Jul [cited 2015 Aug 5];33(5):679-91. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704204

36. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford, United Kingdom: Oxford University Press; 2006.

37. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. BMJ. 2011;342:d1766.

38. Gray A. Applied methods of cost-effectiveness analysis in health care. Oxford: Oxford University Press; 2011. (Handbooks in health economic evaluation series).

39. Baio G. Bayesian methods in health economics. Boca Raton (FL): Chapman & Hall/CRC Press; 2012. (Chapman & Hall/CRC biostatistics series).

40. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005 Apr;14(4):339-47.

41. Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Econ. 2006 Apr;15(4):373-81.

42. Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU technical support document 15: cost-effectiveness modelling using patient-level simulation [Internet]. Sheffield, UK: School of Health and Related Research, University of Sheffield; 2014 Apr. [cited 2015 Aug 26]. Available from: http://www.nicedsu.org.uk/TSD15_Patient-level_simulation.pdf

43. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. Comput Biomed Res. 1986 Jun;19(3):254-65.

44. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Med Decis Making. 1985;5(2):157-77.

45. Drummond M, Helfand M, Mullins CD, editors. Special issue: recommendations of the ISPOR-SMDM Joint Modeling Good Research Practices Task Force. Med Decis Making [Internet]. 2012 [cited 2015 Aug 26];32(5):653-743. Available from: http://mdm.sagepub.com/content/32/5.toc

46. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. Med Decis Making. 2005 Nov;25(6):646-54.

47. Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. Syst Rev [Internet]. 2014 [cited 2015 Aug 12];3:109. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4183945

48. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc. 2006;101(474):447-59.

49. Gray LJ, Cooper N, Dunkley A, Warren FC, Ara R, Abrams K, et al. A systematic review and mixed treatment comparison of pharmacological interventions for the treatment of obesity. Obes Rev. 2012 Jun;13(6):483-98.

50. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014 Mar;17(2):157-73.

51. Stevenson M, Gomersall T, Lloyd JM, Rawdin A, Hernandez M, Dias S, et al. Percutaneous vertebroplasty and percutaneous balloon kyphoplasty for the treatment of osteoporotic vertebral fractures: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2014 Mar;18(17):1-290.

52. Del Giovane C, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. Stat Med. 2013 Jan 15;32(1):25-39.

53. Hawkins N, Scott DA, Woods B. How far do you go? Efficient searching for indirect evidence. Med Decis Making. 2009 May;29(3):273-81.

54. Dequen P, Sutton AJ, Scott DA, Abrams KR. Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery. Value Health. 2014 Jun;17(4):416-23.

55. Caldwell DM, Dias S, Welton NJ. Extending treatment networks in health technology assessment: how far should we go? Value Health. 2015 Jul;18(5):673-81.

56. Cooper NJ, Peters J, Lai MC, Juni P, Wandel S, Palmer S, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value Health. 2011 Mar;14(2):371-80.

57. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. Stat Med. 2013 Jul 30;32(17):2935-49.

58. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005 Jan 1;365(9453):82-93.

59. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ [Internet]. 1996 May 11 [cited 2015 Aug 6];312(7040):1215-8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2350940

60. Taylor WJ, Weatherall M. What are open-label extension studies for? J Rheumatol. 2006 Apr;33(4):642-3.

61. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A Stat Soc. 2009;172(1):119-36.

62. Tritchler D. Modelling study quality in meta-analysis. Stat Med. 1999 Aug 30;18(16):2135-45.

63. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med. 2003 Dec 15;22(23):3687-709.

64. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc [Internet]. 2009 Jan [cited 2015 Aug 5];172(1):21-47. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2667303

65. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015 Jun 2;162(11):777-84.

66. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996 Feb;17(1):1-12.

67. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health [Internet]. 1998 Jun [cited 2015 Aug 5];52(6):377-84. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756728

68. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. Ottawa: Ottawa Hospital Research Insititute; 2014. [cited 2015 Aug 26]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

69. Drummond MF, Iglesias CP, Cooper NJ. Systematic reviews and economic evaluations conducted for the National Institute for Health and Clinical Excellence in the United Kingdom: a game of two halves. Int J Technol Assess Health Care. 2008;24(2):146-50.

70. CRD/CHE Technology Assessment Group, Burch J, Paulden M, Conti S, Stock C, Corbett M, et al. Antiviral drugs for the treatment of influenza: a systematic review and economic evaluation [Internet]. London: National Institute for Health and Care Excellence; 2009 Feb. [cited 2015 Aug 26]. (NICE technology appraisal 168). Available from: http://www.nice.org.uk/guidance/ta168/documents/influenza-zanamivir-amantadine-and-oseltamivir-review-assessment-report2

71. Welton NJ, Cooper NJ, Ades AE, Lu G, Sutton AJ. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. Stat Med. 2008 Nov 29;27(27):5620-39.

72. Stettler C, Allemann S, Wandel S, Kastrati A, Morice MC, Schomig A, et al. Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis. BMJ [Internet]. 2008 [cited 2015 Aug 6];337:a1331. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2527175

73. Madan J, Chen YF, Aveyard P, Wang D, Yahaya I, Munafo M, et al. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. J R Stat Soc Ser A Stat Soc. 2014;177(1):295-314.

74. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009 Jun 30;28(14):1861-81.

75. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol. 2002 Jan;55(1):86-94.

76. Welton NJ, Soares MO, Palmer S, Ades AE, Harrison D, Shankar-Hari M, et al. Accounting for heterogeneity in relative treatment effects for use in cost-effectiveness models and value-of-information analyses. Med Decis Making [Internet]. 2015 Jul [cited 2015 Aug 6];35(5):608-21. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4471065

77. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J Roy Stat Soc Ser B Stat Met. 2002;64(4):583-639.

78. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998 Apr;18(2 Suppl):S68-S80.

79. Cooper NJ, Sutton AJ, Abrams KR, Turner D, Wailoo A. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. Health Econ. 2004 Mar;13(3):203-26.

80. Parmigiani G. Modeling in medical decision making: a Bayesian approach. New York: Wiley; 2002.

81. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester, England: Wiley; 2004.

82. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. R News [Internet]. 2006 [cited 2015 Aug 27];6(1):12-7. Available from: https://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf

83. Plummer M. JAGS: a program for analysis of Bayesian graphical modeling using Gibbs sampling. In: The 3rd International Workshop on Distributed Statistical Computing (2003); 2003 Mar 20 -22; Vienna. Vienna, Austria: The R Foundation; 2003.

84. Stan Development Team. Stan modeling language. User's guide and reference manual [Internet]. Version 2.5.0. [place unknown]: The Stan Development Team; 2014 Oct 20. [cited 2015 Aug 27]. Available from: https://github.com/stan-dev/cmdstan/releases/tag/v2.5.0

85. TreeAge Pro [Internet]. Williamstown (MA): TreeAge Software, Inc.; 2015. [cited 2015 Aug 27]. Available from: https://www.treeage.com/

86. Gelman A, Ligges U, Sturtz S. R2WinBUGS: a package for running WinBUGS from R. J Stat Softw. 2005;12(3):1-16.

87. Heiberger RM, Neuwirth E. R through Excel: a spreadsheet interface for statistics, data analysis, and graphics. Dordrecht, Netherlands: Springer; 2009.

88. Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC, et al. Statistical approaches for conducting network meta-analysis in drug development. Pharm Stat. 2011 Nov;10(6):523-31.

89. Rothstein H, Sutton AJ, Borenstein M, editors. Publication bias in meta-analysis : prevention, assessment and adjustments. Chichester, England: Wiley; 2005.

90. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Stat Med. 1997 Apr 15;16(7):753-68.

91. Lord J, Asante MA. Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. Health Econ. 1999 Jun;8(4):323-33.

92. Van Den Noortgate W, Onghena P. Parametric and nonparametric bootstrap methods for meta-analysis. Behav Res Methods. 2005 Feb;37(1):11-22.

93. Ahern MJ, McFarlane AC, Leslie A, Eden J, Roberts-Thomson PJ. Illness behaviour in patients with arthritis. Ann Rheum Dis [Internet]. 1995 Apr [cited 2015 Aug 6];54(4):245-50. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1005568

94. Kendrick D, Young B, Mason-Jones AJ, Ilyas N, Achana FA, Cooper NJ, et al. Home safety education and provision of safety equipment for injury prevention. Cochrane Database Syst Rev. 2012;9:CD005014.

95. Cooper NJ, Kendrick D, Achana F, Dhiman P, He Z, Wynn P, et al. Network meta-analysis to evaluate the effectiveness of interventions to increase the uptake of smoke alarms. Epidemiol Rev [Internet]. 2012 [cited 2015 Aug 27];34:32-45. Available from: http://epirev.oxfordjournals.org/content/34/1/32.full.pdf+html

96. Achana F, Hubbard S, Sutton A, Kendrick D, Cooper N. An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial. J Clin Epidemiol. 2014 Apr;67(4):376-90.

97. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med. 2007 Jun 19;146(12):857-67.

98. Loke YK, Golder SP, Vandenbroucke JP. Comprehensive evaluations of the adverse effects of drugs: importance of appropriate study selection and data sources. Ther Adv Drug Saf [Internet]. 2011 Apr [cited 2015 Aug 6];2(2):59-68. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4110807

99. Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. BMJ [Internet]. 2014 [cited 2015 Aug 6];348:f7668. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898583

100. Higgins JPT, Deeks JJ, Altman DG, editors, Cochrane Statistical Methods Group. Special topics in statistics: rare events (including zero frequencies) [Internet]. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. [London]: The Cochrane Collaboration; 2011 Mar. Chapter 16.9 [cited 2015 Aug 27]. Available from: http://handbook.cochrane.org/chapter_16/16_9_rare_events_including_zero_frequencies.htm.

101. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions [Internet]. Version 5.1.0. [London]: The Cochrane Collaboration; 2011 Mar. [cited 2015 Aug 27]. Available from: http://handbook.cochrane.org/

102. Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. Stat Med. 2007 Jan 15;26(1):53-77.

103. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med. 2004 May 15;23(9):1351-75.

104. Sweeting M, Sutton A, Lambert P. Correction. Stat Med [Internet]. 2006 [cited 2015 Aug 27];25(15):2700. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sim.2536/epdf

105. Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. Stat Methods Med Res. 2014 May 12. Epub ahead of print.

106. Kuss O. Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. Stat Med. 2015 Mar 30;34(7):1097-116.

107. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med. 2005 Aug 15;24(15):2401-28.

108. Friedrich JO, Adhikari NK, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. BMC Med Res Methodol [Internet]. 2007 [cited 2015 Aug 10];7:5. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1783664

109. Dias S, Welton NJ, Sutton AJ, Ades AE, Decision Support Unit. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials [Internet]. Sheffield, England: NICE Decision Support Unit; 2014 Apr. [cited 2015 Aug 27]. Available from: http://www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%2015April2014.pdf

110. Ohlssen D, Price KL, Xia HA, Hong H, Kerman J, Fu H, et al. Guidance on the implementation and reporting of a drug safety Bayesian network meta-analysis. Pharm Stat. 2014 Jan;13(1):55-70.

111. Warren FC, Abrams KR, Sutton AJ. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. Stat Med. 2014 Jun 30;33(14):2449-66.

112. Fu H, Price KL, Nilsson ME, Ruberg SJ. Identifying potential adverse events dose-response relationships via Bayesian indirect and mixed treatment comparison models. J Biopharm Stat. 2013;23(1):26-42.

113. Soares MO, Dumville JC, Ades AE, Welton NJ. Treatment comparisons for decision making: facing the problems of sparse and few data. J Roy Stat Soc Ser A Stat Soc. 2014;177(1):259-79.

114. Soares MO, Dumville JC, Ashby RL, Iglesias CP, Bojke L, Adderley U, et al. Methods to assess cost-effectiveness and value of further research when data are sparse: negative-pressure wound therapy for severe pressure ulcers. Med Decis Making. 2013 Apr;33(3):415-36.

115. Soares MO, Bojke L, Dumville J, Iglesias C, Cullum N, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Stat Med. 2011 Aug 30;30(19):2363-80.

116. Askling J, Fahrbach K, Nordstrom B, Ross S, Schmid CH, Symmons D. Cancer risk with tumor necrosis factor alpha (TNF) inhibitors: meta-analysis of randomized controlled trials of adalimumab, etanercept, and infliximab using patient level data. Pharmacoepidemiol Drug Saf. 2011 Feb;20(2):119-30.

117. Salanti G, Higgins JP, White IR. Bayesian synthesis of epidemiological evidence with different combinations of exposure groups: application to a gene-gene-environment interaction. Stat Med. 2006 Dec 30;25(24):4147-63.

118. Minelli C, Thompson JR, Abrams KR, Lambert PC. Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. Stat Med. 2005 Dec 30;24(24):3845-61.

119. Welton NJ, Johnstone EC, David SP, Munafo MR. A cost-effectiveness analysis of genetic testing of the DRD2 Taq1A polymorphism to aid treatment choice for smoking cessation. Nicotine Tob Res [Internet]. 2008 Jan [cited 2015 Aug 12];10(1):231-40. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2257987

120. Govan L, Ades AE, Weir CJ, Welton NJ, Langhorne P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. Stat Med. 2010 May 30;29(12):1340-56.

121. Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. Stat Med. 2013 Jan 15;32(1):51-66.

122. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. Stat Med [Internet]. 2011 Sep 10 [cited 2015 Aug 12];30(20):2481-98. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3470931

123. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. Stat Med. 1998 Nov 30;17(22):2537-50.

124. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC Med Res Methodol [Internet]. 2007 [cited 2015 Aug 12];7:3. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1800862

125. Hughes D, McGuire A. A review of the economic analysis of obesity. Br Med Bull. 1997;53(2):253-63.

126. Efthimiou O, Mavridis D, Riley RD, Cipriani A, Salanti G. Joint synthesis of multiple correlated outcomes in networks of interventions. Biostatistics [Internet]. 2015 Jan [cited 2015 Aug 12];16(1):84-97. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4481542

127. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bagos P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. Stat Med. 2014;33(13):2275-87.

128. Achana FA, Cooper NJ, Bujkiewicz S, Hubbard SJ, Kendrick D, Jones DR, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. BMC Med Res Methodol [Internet]. 2014 [cited 2015 Aug 12];14:92. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142066

129. Lu G, Kounali D, Ades AE. Simultaneous multioutcome synthesis and mapping of treatment effects to a common scale. Value Health [Internet]. 2014 Mar [cited 2015 Jul 28];17(2):280-7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3991420

130. Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. Value Health. 2010 Dec;13(8):976-83.

131. Trikalinos TA, Olkin I. A method for the meta-analysis of mutually exclusive binary outcomes. Stat Med. 2008 Sep 20;27(21):4279-300.

132. Collett D. Modelling survival data in medical research. 2nd ed. Boca Raton (FL): Chapman & Hall/CRC Press; 2003.

133. Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. BMC Med Res Methodol [Internet]. 2010 [cited 2015 Jul 28];10:54. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2906500

134. Welton NJ, Willis SR, Ades AE. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. Res Synth Methods. 2010 Jul;1(3-4):239-57.

135. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. Res Synth Methods. 2010 Jul;1(3-4):258-71.

136. Guyot P, Welton NJ, Ouwens MJ, Ades AE. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. Value Health. 2011 Jul;14(5):640-6.

137. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC Med Res Methodol [Internet]. 2011 [cited 2015 Jul 28];11:61. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3112194

138. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. Stat Med. 1995 Oct 15;14(19):2057-79.

139. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: individual patient data may be beneficial if only for a subset of trials. Stat Med. 2013 Mar 15;32(6):914-30.

140. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2(3):209-17.

141. Jansen JP. Network meta-analysis of individual and aggregate level data. Res Synth Methods. 2012 Jun;3(2):177-90.

142. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Stat Med. 2012 Dec 10;31(28):3516-36.

143. Saramago P, Chuang LH, Soares MO. Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data. BMC Med Res Methodol [Internet]. 2014 [cited 2015 Jul 28];14:105. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4236567

144. Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Med Decis Making. 2005 Nov;25(6):633-45.

145. Price MJ, Welton NJ, Ades AE. Parameterization of treatment effects for meta-analysis in multi-state Markov models. Stat Med. 2011 Jan 30;30(2):140-51.

146. Price MJ, Welton NJ, Briggs AH, Ades AE. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. Value Health. 2011 Mar;14(2):205-18.

147. Trinquart L, Chatellier G, Ravaud P. Adjustment for reporting bias in network meta-analysis of antidepressant trials. BMC Med Res Methodol [Internet]. 2012 [cited 2015 Jul 28];12:150. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3537713

148. Wynn P, Zou K, Young B, Majsak-Newman G, Hawkins A, Kay B, et al. Prevention of childhood poisoning in the home: overview of systematic reviews and a systematic review of primary studies. Int J Inj Contr Saf Promot. 2015 Sep 24;1-26. Epub ahead of print.

149. Achana FA, Sutton AJ, Kendrick D, Wynn P, Young B, Jones DR, et al. The effectiveness of different interventions to promote poison prevention behaviours in households with children: a network meta-analysis. PLoS One [Internet]. 2015 [cited 2015 Jul 28];10(3):e0121122. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4404249

150. Borenstein M. Introduction to meta-analysis. Chichester, England: John Wiley & Sons; 2009.

151. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010 Mar 30;29(7-8):932-44.

152. Congdon P. Applied Bayesian modelling. Hoboken (NJ): Wiley; 2003.

153. McCullagh P, Nelder JA. Generalized linear models. London: Chapman and Hall; 1989.

154. Chen YF, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess [Internet]. 2012 [cited 2015 Jul 28];16(38):1-205. Available from: http://www.journalslibrary.nihr.ac.uk/hta/volume-16/issue-38

155. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 [Internet]. London: NICE; 2013 Apr 4. [cited 2015 Jun 17]. (Process and methods guides). Available from: http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf