

**NICE DSU TECHNICAL SUPPORT DOCUMENT 6:
EMBEDDING EVIDENCE SYNTHESIS IN PROBABILISTIC
COST-EFFECTIVENESS ANALYSIS: SOFTWARE CHOICES**

REPORT BY THE DECISION SUPPORT UNIT

May 2011

(last updated April 2012)

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

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

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

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

Tel (+44) (0)114 222 0734

E-mail dsuadmin@sheffield.ac.uk

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University.

The DSU is commissioned by The National Institute for Health and Clinical Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

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

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

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

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

Dr Allan Wailoo

Director of DSU and TSD series editor.

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

Competing interests

Alex Sutton has received financial reimbursement when working as an advisor on the development of the Comprehensive Meta-Analysis software package.

Acknowledgements

The DSU thanks Rachael Fleurence, Jeroen Jansen, Alec Miners, Jaime Peters, Mike Spencer and the team at NICE, led by Gabriel Rogers, for reviewing this document. The editor for the TSD series is Allan Wailoo.

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

This report should be referenced as follows:

Dias, S., Sutton, A.J., Welton, N.J. & Ades, A.E. NICE DSU Technical Support Document 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>

EXECUTIVE SUMMARY

This document sets out software options for evidence synthesis that are compatible with probabilistic cost-effectiveness analysis, which is the preferred methodology for the NICE reference case. Four possibilities are discussed:

1. Evidence synthesis by Bayesian posterior estimation, and posterior sampling. Other parameters of the cost-effectiveness models can be incorporated into the same software platform. Bayesian Markov chain Monte Carlo simulation methods with WinBUGS software are the most popular choice.
2. Evidence synthesis by Bayesian posterior estimation. Posterior samples are exported to another package where the other parameters are generated and the cost-effectiveness model evaluated.
3. Frequentist methods of parameter estimation followed by forward Monte Carlo simulation from the maximum likelihood estimates and their variance-covariance matrix
4. Bootstrap re-sampling – a frequentist simulation approach to parameter estimation.

When multiple parameters are estimated from the same synthesis model, we emphasise the need to choose a method that propagates the parameter correlation structure through the cost-effectiveness model. A table is provided that shows the possible approaches and the restrictions on their use.

Software packages for evidence synthesis are listed. Technical issues relating to software are covered in an Appendix. Finally we mention software suitable for transferring data between different software packages, and software that provides user-friendly interface for integrated software platforms. These offer investigators a flexible way of examining alternative scenarios.

CONTENTS

1. INTRODUCTION	7
2. EMBEDDING EVIDENCE SYNTHESIS IN PROBABILISTIC CEA	8
2.1. BAYESIAN POSTERIOR SIMULATION: ONE-STAGE APPROACH.....	8
2.2. BAYESIAN POSTERIOR SIMULATION: TWO-STAGE APPROACH	9
2.3. FREQUENTIST ESTIMATION WITH MONTE CARLO SAMPLING.....	9
2.3.1. <i>Stand-alone meta-analysis packages</i>	10
2.3.2. <i>General statistical packages with support for meta-analysis</i>	10
2.3.3. <i>Software for network meta-analysis</i>	10
2.3.4. <i>Parameterisation of treatment effects</i>	11
2.4. FREQUENTIST ESTIMATION WITH BOOTSTRAPPING	11
2.5. SUMMARY OF METHODS AND RESTRICTIONS ON THEIR USE	12
3. USE OF MULTIPLE SOFTWARE PLATFORMS	12
3.1. COMMUNICATION BETWEEN SOFTWARE PACKAGES.....	13
3.2. INTEGRATED USE OF SOFTWARE PLATFORMS.....	13
4. REFERENCES	15
APPENDIX: PARAMETERISATION OF THE BASELINE TERM IN NETWORK META-ANALYSIS MODELS.....	19

TABLES

Table 1 Summary of methods and their properties and restrictions.....	12
---	----

Abbreviations and Definitions

CEA	cost-effectiveness analysis
FE	Fixed effects
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
RE	Random effects
VCV	variance-covariance

1. INTRODUCTION

Probabilistic methods in decision analysis were introduced in the 1980s.^{1,2} Their defining feature is that they allow for a full expression of the uncertainty in model parameters. There are two main reasons for advocating probabilistic methods in decision making. The first is that they can provide a form of sensitivity analysis which allows investigators to easily see the joint impact of the uncertainty in multiple parameters on the expected costs, benefits and on decision uncertainty. For this reason use of these methods is often called probabilistic sensitivity analysis. A second reason is that, faced with uncertainty in the vector/matrix of model parameters θ , decision makers generally choose the decision option, D , that delivers the highest expected net benefit. In other words, the decision maker selects decision D^* , such that:

$$D^* = \underset{D}{\text{Max}} E_{\theta}[NB(D, \theta)] \quad (1)$$

This “expectation” requires an integration of the net benefit function, $NB(D, \theta)$ over the joint distribution of parameters θ . There are a wide range of methods for achieving this integration, and the appropriate choice of method depends on the algebraic structure of the net benefit function. It must be emphasised that the expected Net Benefit is *not* the same as the Net Benefit at the expected value of the parameters, except in the cases where Net Benefit is linear in all its parameters, *and* there are no correlations between parameters. This is relatively rare, as most evidence synthesis is performed on log or logit scales, and many cost-effectiveness analysis (CEA) models include Markov models, which are notoriously non-linear. Further, modern methods of evidence synthesis tend to generate estimates of several parameters from a common dataset, in most cases leading to correlations between parameters. It is therefore essential that software solutions are adopted that ensure that the complex uncertainty structure in parameter estimates is faithfully propagated through the decision model.³ This document provides guidance on the appropriate choice of software to deliver probabilistic cost-effectiveness analysis in any situation.

Monte Carlo (MC) simulation from the joint parameter distribution is not only the simplest way to evaluate the expected Net Benefit, but for any form of model it also delivers other crucial tools of probabilistic CEA such as: plots of the cost-effectiveness plane, cost-effectiveness acceptability curves and estimates of the probability that a decision is cost-effective.^{4,5} MC simulation is also the easiest approach to Expected Value of Information analysis.⁶⁻⁸ Probabilistic methods have been recommended in a range of leading textbooks

and tutorial papers, and are the preferred option for submissions to re-imbursement agencies such as the National Institute of Health and Clinical Excellence (NICE) in the UK. We therefore recommend MC simulation-based approaches for *all* analyses, even for strictly linear models.

The document confines itself to the question of which computational approaches correctly preserve the properties of the evidence synthesis within a probabilistic cost-effectiveness analysis. No advice is given on the relative merits of cohort models compared to individual patient simulation. Indeed, no advice is given in this document on how to choose the best model, it is restricted to providing guidance on how to implement the model of choice.

This document is organised as follows. The main body of the document (Section 2) sets out the different analysis options. These include Bayesian Markov chain Monte Carlo (MCMC) methods (2.1, 2.2), and frequentist methods, either sampling from maximum likelihood estimates and their variance-covariance (VCV) matrices (2.3), or bootstrapping (2.4). We then provide a brief summary of software tools that can be used to help interface between different software, and review some recent developments in user-friendly “front ends”, to assist the integrated use of multiple software platforms. These offer ways in which investigators can conduct scenario analyses, not just with individual parameters but also with different datasets or different synthesis models and quickly see the impact on cost-effectiveness results.

2. EMBEDDING EVIDENCE SYNTHESIS IN PROBABILISTIC CEA

2.1. BAYESIAN POSTERIOR SIMULATION: ONE-STAGE APPROACH

When estimation of the synthesis parameters is via sampling from a Bayesian posterior distribution of the relevant parameters, this can be integrated with the CEA as a single process within a single programming package, in what has been referred to as “Comprehensive Decision Analysis”.⁹⁻¹¹

Bayesian MCMC simulation,¹² using WinBUGS,¹³ OpenBUGS¹⁴ or other MCMC packages, provides the obvious example. The advantage of this approach is that it not only estimates a Bayesian posterior distribution, but that it is simulation-based, so that its outputs are perfectly compatible with the MC sampling approach that has become the standard modelling method in so many areas of science. Samples from the joint posterior distribution can be put directly through the decision analysis, so that Net Benefit and other outputs can be evaluated for each

set of parameter samples, without requirements for assumptions about its distributional form. Distributions of additional parameters and costs can be readily incorporated.

Development of MCMC algorithms and sampling schemes is a specialised area of research. Although users need not have a detailed knowledge of the precise working of MCMC software, a good understanding of the fundamentals and of Bayesian data analysis is essential. For completeness it is worth mentioning that a broad range of *non*-MCMC simulation-based Bayesian updating schemes have also been proposed, including the Sample Importance Re-sampling algorithm,¹⁵ Bayesian Melding,^{16,17} and Bayesian Monte Carlo.¹⁸ All these have the same properties as Bayesian MCMC in that they all feature both Bayesian estimation and sampling from joint posterior distributions. The latter two were specifically designed for evidence synthesis.

2.2. BAYESIAN POSTERIOR SIMULATION: TWO-STAGE APPROACH

If investigators have a preferred software for CEA, either general software packages such as R, STATA, SAS, or spread-sheet or decision tree packages such as EXCEL or TreeAGE, a further option is to take the posterior samples from the Bayesian MCMC, or other posterior sampling scheme, and use them as input to the CEA package. This has the same technical properties as the Bayesian one-stage approach since the full posterior distribution is preserved. From WinBUGS, the CODA output, which lists all values generated from the full posterior distribution, can be exported into a spreadsheet-based program such as EXCEL, using BUS – BUGS Utility for Spreadsheets.¹⁹ When using the CODA output it is important that the correlations in the parameter estimates are preserved. This is done by ensuring that *all parameter values are sampled from the same MCMC iteration*. If the CODA output is in spreadsheet format, for example in EXCEL, this would correspond to sampling all the parameter values in one row, each time. The CODA output can also be converted to the freely available statistical software R²⁰ for convergence diagnostics, further analysis and plotting using add-on packages such as BOA – Bayesian Output Analysis Program²¹ or CODA – Convergence Diagnostics and Output Analysis.²²

2.3. FREQUENTIST ESTIMATION WITH MONTE CARLO SAMPLING

If evidence synthesis can be performed using frequentist software (which may use a variety of methods of estimation including, methods of moments, iterative weighted least-squares or

(restricted) maximum likelihood ((RE)ML)), a two-stage approach is also possible. The first step is estimation, which produces parameter estimates and their variance-covariance (VCV) matrix. In the second step, these are used to populate a multivariate normal distribution which can be used for forward MC sampling (in the same or in a different package) along with the other CEA parameters. When single parameters are of interest, such as in simple pair-wise meta-analysis, this may be the simplest option as only one variance needs to be estimated (there is no VCV matrix). In such situations it is particularly easy to find the treatment effect parameter and its variance using either specific meta-analysis software or implementing meta-analysis routines in standard statistical software packages. A systematic and comprehensive review of all software options capable of evidence synthesis is beyond the scope of this document, but noteworthy options are described below and more detailed reviews and comparisons are available elsewhere.²³⁻²⁶

2.3.1. *Stand-alone meta-analysis packages*

Numerous packages have been developed over the years, but probably those which include the most extensive and up-to-date feature sets, are

- Comprehensive meta-analysis (commercial);²⁷
- Meta-Analyst (free);²⁶
- MIX,²⁸ which is an add-on to EXCEL (commercial and free versions available);
- RevMan,²⁹ which is the official software of the Cochrane Collaboration (free).
- EXCEL: Simple meta-analysis can be carried out with a small amount of programming.

2.3.2. *General statistical packages with support for meta-analysis*

While it would be possible to program most standard meta-analysis models in any reasonably powerful statistics package, probably the most extensive freely available software routines which allow meta-analysis to be conducted and numerous graphical outputs produced are available for STATA³⁰ and for R, such as the *meta*³¹ and *rmeta*³² packages.

Although all the above options will conduct pair-wise meta-analysis, not all have the capability to do meta-regression,³³ which may be a decisive factor in choosing between them.

2.3.3. *Software for network meta-analysis*

Specific comments on frequentist software options for network meta-analysis can be found in TSD2.³⁴ For indirect comparison networks, separate syntheses can be carried out for AB

trials, and AC trials. Then the estimates obtained and their variances can be entered into the simulation package used for CEA, and the relative cost-effectiveness of A, B and C readily determined (covariances are not involved). Although this is an acceptable approach in principle, the Bayesian MCMC approach to indirect comparisons (see TSD2³⁴) may be preferable in cases where one or more of the pair-wise comparisons is represented by a very small number of trials. This is because, MCMC has the flexibility to allow “shared variance” Random Effects (RE) models, while with conventional methods it may be necessary to have some estimates from RE models and others from Fixed Effects (FE) models, which seems a less natural solution.

In many cases the use of frequentist estimates and their VCV matrix with RE models is likely to produce parameter distributions with a little less uncertainty, because Bayesian methods take uncertainty in variance parameters into account. The extent of the difference is unlikely to be critical, although investigators should always check that posterior distributions of variance parameters are sensible.

2.3.4. Parameterisation of treatment effects

Whatever software is used, model parameterisation requires care, as a number of apparently innocuous variations may give very different or wrong results. To date, a few coding problems have come to light and these are set out in the Appendix.

2.4. FREQUENTIST ESTIMATION WITH BOOTSTRAPPING

The final option of estimation and then Bootstrapping,³⁵ has been used from time to time in cost-effectiveness analysis.³⁶ In its original form, bootstrapping is a technique in which one generates a series of “new” datasets by repeatedly re-sampling *with replacement* from the original data, each time producing a new set of parameters estimates. This stream of estimates can then be treated in the same way as samples from Bayesian posterior distributions. However, this procedure is not always straightforward, particularly with small sample sizes and zero cells.

There are, nonetheless, a very wide range of variant bootstrap procedures which can mitigate these and other problems. In the parametric bootstrap, for example a model is fitted to the data by maximum likelihood and is then used to generate a series of datasets with the same size and structure as the original. The analysis procedure is applied to each of these datasets

to generate a stream of parameter values. Data analysis based on resampling is a rich area with an extensive literature. Readers are referred to texts for further information.^{35,37,38}

2.5. SUMMARY OF METHODS AND RESTRICTIONS ON THEIR USE

Table 1 summarises the methods and the restrictions on their use. Readers are referred to the text for further explanation.

Table 1 Summary of methods and their properties and restrictions

Estimation	Output to CEA Software	Restrictions
Bayesian MCMC	None: CEA within MCMC Software	None
Bayesian MCMC	MCMC chains exported	None
Bayesian MCMC	Posterior means, variances, correlations	None, but assumes multi-variate normality in posterior distribution
Bayesian MCMC	Posterior means and variances	Only suitable if no correlation between parameters*
Estimation by non-Bayesian (frequentist) methods	Parameter estimates and Variance-Covariance matrix	None, but assumes multi-variate normality of treatment effect estimates
Estimation by non-Bayesian (frequentist) methods	Parameter estimates and their variances	Only suitable if no correlation between parameters*
Estimation by non-Bayesian (frequentist) methods	Bootstrap resampling	None, but special methods are necessary for sparse data

* Users should ensure that the data structure and analysis methods do not imply correlations between parameters, before using these methods.

3. USE OF MULTIPLE SOFTWARE PLATFORMS

In recent years interfaces have become available that let different software applications communicate with each other. These facilities allow for the integration of the components of a CEA which may have been conducted in different packages. The motivations and advantages of an integrated approach across software applications are potentially multifaceted. Firstly, it allows multidisciplinary teams who have different software skills and preferences to produce an integrated analysis. For example, statisticians may wish to use general statistical software, whereas decision modellers may wish to use EXCEL or specific decision modelling software. It also allows the best software for each component of the analysis to be used therefore producing an “optimal” mix. For example, if a network meta-analysis is required, WinBUGS may be the best software to implement this in, but it has

limited graphical capabilities. Therefore, it may be desirable to present the results of the synthesis in a package with advanced graphical capability such as R. Furthermore, the original dataset may have been prepared in spreadsheet software such as EXCEL. Although use of multiple pieces of software to conduct different components of the analysis is common historically, few have been integrated.

Section 3.1 gives details of how to allow communication between different platforms, with the primary aim of transferring data between them. Section 3.2 explains how different software can be used to fully integrate data input, analysis and the display of results using multiple packages, into a single step.

3.1. COMMUNICATION BETWEEN SOFTWARE PACKAGES

To facilitate communication, transparency and future data updates, it is good practice to keep all data collected for the analysis, including all annotations and details of any corrections, in a single file, for example an EXCEL workbook with multiple worksheets. If the analysis is to be carried out in WinBUGS, data columns can be copied directly from spreadsheet software into WinBUGS and pasted by selecting *Paste Special* from the WinBUGS *Edit* menu and choosing the *Plain text* option. Alternatively, *XL2BUGS*³⁹ is an EXCEL add-in which converts EXCEL data into WinBUGS vector format, and *BAUW*⁴⁰ converts data in text format into WinBUGS vector or matrix format.

If data is stored in R, *R2WinBUGS*⁴¹ can be used to convert R objects into WinBUGS list data using the *bugs.data* function.

See <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/remote14.shtml> for details.

3.2. INTEGRATED USE OF SOFTWARE PLATFORMS

Integrated platforms reduce the need to copy data and intermediate results from one screen/system to another, and thereby reduce the risk of transcription errors. Further advantages of integrating the analysis (which also exist if the Bayesian one-stage approach is conducted since in that approach the analysis is integrated by definition) include facilitating the modification and updating of any aspect of the analysis, conducting sensitivity analyses and, more generally, promoting transparency. For example, if a new trial is reported that is to be added to the evidence synthesis, then in an integrated approach the CEA would automatically be updated. This goes some way to ensuring the appropriate uncertainty is

propagated through to the decision model. If part of this integrated approach is the inclusion of a user-friendly interface, then this can also make the exploration of the synthesis and CEA accessible to non-technical experts including clinical experts and even NICE decision makers themselves, allowing them to interrogate the analysis.

To this end, a Transparent Interactive Decision Interrogator (TIDI)⁴² which integrated syntheses conducted in WinBUGS with graphical displays and the decision model conducted in R and a “point and click” interface in EXCEL was developed for a recent Single Technology Appraisal (STA) at NICE. This pilot “proof of concept” initiative allowed members of the appraisal committee to request re-runs of the CEA using alternative parameter values in real time in the committee meetings.

Several (freely available) code routines have been developed for commonly used packages in Health Technology Assessment which allow them to communicate with other packages and these can be utilised in the creation of integrated analyses. For example, *RExcel*,⁴³ an add-on to EXCEL, provides communication between EXCEL and R and *R2WinBUGS* is one of several packages which allow the controlling of WinBUGS through R. Thus, if both of these linking packages are used in combination, then WinBUGS can be controlled through EXCEL (via R), and a Visual Basic interface can be written in EXCEL to facilitate this (which is the software setup used in the TIDI project described above). Similar control of WinBUGS through STATA⁴⁴ and several other packages is also possible, as is the embedding of OpenBUGS in R through *rbugs*⁴⁵ and the linking of many packages to each other.

4. REFERENCES

1. Critchfield, G.C., Willard, K.E. Probabilistic analysis of decision trees using Monte Carlo simulation. *Medical Decision Making* 1986; 6(2):85-92.
2. Doubilet, P., Begg, C.B., Weinstein, M.C., Braun, P., McNeil, B.J. Probabilistic sensitivity analysis using Monte Carlo simulation: A practical approach. *Medical Decision Making: an International Journal of the Society for Medical Decision Making* 1985; 5(2):157-177.
3. Ades, A.E., Claxton, K., Sculpher, M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Economics* 2005; 14:1-9.
4. 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 Economics* 2005; 14(4):339-347.
5. National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal (updated 2008). 2008.
6. Thompson, K.M., Evans, J.S. The value of improved national exposure information for perchloroethylene (perc): a case study for dry cleaners. *Risk Analysis* 1997; 17(2):253-271.
7. Claxton, K., Posnett, J. An economic approach to clinical trial design and research priority setting. *Health Economics* 1996; 5(6):513-524.
8. Felli, J.C., Hazen, G.B. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* 1998; 18(1):95-109.
9. Cooper, N.J., Sutton, A.J., Abrams, K.R., Turner, D., Wailoo, A. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Economics* 2003; 13(3):203-226.
10. Parmigiani, G., Samsa, G.P., Ancukiewicz, M., Lipscomb, J., Hasselblad, V., Matchar, D.B. Assessing uncertainty in cost-effectiveness analyses: Application to a complex decision model. *Medical Decision Making* 1997; 17(4):390-401.
11. Spiegelhalter, D.J., Myles, J.P., Jones, D.R., Abrams, K.R. Bayesian methods in health technology assessment: a review. *Health Technology Assessment* 2000; 4(38).
12. Gilks, W.R., Richardson, S., Spiegelhalter, D.J. Markov chain Monte Carlo in practice. Chapman & Hall/CRC, 1996.
13. Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; 10(4):325-337.
14. Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine* 2009; 28(25):3049-3067.

15. Rubin, D.B. Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics* 1988; 3:395-402.
16. Raftery, A.E., Givens, G.H., Zeh, J.E. Inference from a Deterministic Population Dynamics Model for Bowhead Whales. *Journal of the American Statistical Association* 1995; 90(430):402-416.
17. Poole, D., Raftery, A.E. Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal of the American Statistical Association* 2000; 95(452):1244-1255.
18. Brand, K.P., Small, M.J. Updating uncertainty in an integrated risk assessment: conceptual framework and methods. *Risk Analysis* 1995; 15(6):719-731.
19. Hahn, G. BUGS Utility for Spreadsheets. Version 1.0.1. 2001; available from <http://faculty.salisbury.edu/~edhahn/bus.htm>
20. R Development Core Team. R: A Language and Environment for Statistical Computing. 2010. Vienna, Austria, R Foundation for Statistical Computing.
21. Smith, B.J. The boa Package. Version 1.1.5. 2005; available from <http://www.public-health.uiowa.edu/boa/>
22. Plummer, M., Best, N., Cowles, K., Vines, K. CODA: Convergence diagnosis and output analysis for MCMC. *R News* 2006; 6(1):7-11.
23. Bax, L., Yu, L.M., Ikeda, N., Moons, K.G.M. A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology* 2007; 7(1):40.
24. Sterne, J.A.C., Egger, M., Sutton, A.J. Meta-Analysis Software. In: Egger M., Davey-Smith G., Altman D.G., eds. *Systematic Reviews in Health Care: Meta-analysis in context*. second edn ed. BMJ Books; London: 2001; 336-346.
25. Sutton, A.J., Lambert, P.C., Hellmich, M.A.G., Abrams, K.R., Jones, D.R. Meta-analysis in practice: a critical review of available software. In: Stangl D.K., Berry D.A., eds. *Meta-analysis in medicine and health policy*. Marcel Dekker; New York: 1998; 315-339.
26. Wallace, B.C., Schmid, C.H., Lau, J., Trikalinos, T.A. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Medical Research Methodology* 2009; 9(1):80.
27. Borenstein, M., Hedges, L., Higgins, J., Rothstein, H. Comprehensive meta-analysis version 2. 2005; available from <http://www.meta-analysis.com/>
28. Bax, L., Yu, L.M., Ikeda, N., Tsuruta, H., Moons, K.G.M. Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC Medical Research Methodology* 2006; 6(1):50.
29. The Nordic Cochrane Centre. Review Manager (RevMan). Version 5.0. 2008; available from <http://ims.cochrane.org/revman>

30. Sterne, J.A.C. Meta-analysis in Stata: an updated collection from the Stata Journal. CRC PRESS, 2009.
31. Schwarzer, G. meta: Meta-Analysis with R. Version 1.6-1. 2010; available from <http://cran.r-project.org/web/packages/meta/>
32. Lumley, T. rmeta:Meta-analysis. 2009; available from <http://cran.r-project.org/web/packages/rmeta/>
33. Thompson, S.G., Sharp, S.J. Explaining heterogeneity in meta analysis: a comparison of methods. *Statistics in Medicine* 1999; 18(20):2693-2708.
34. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
35. Efron, B., Tibshirani, R.J. An introduction to the bootstrap. Chapman & Hall, New York; 1993.
36. Lord, J., Asante, M.A. Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis1. *Health Economics* 1999; 8(4):323-333.
37. Davison, A.C., Hinkley, D.V. Bootstrap methods and their application. Cambridge Univ Pr, 1997.
38. Lunneborg, C.E. Data analysis by resampling: Concepts and applications. Duxbury, 2000.
39. Misra, S. XL2BUGS: Excel Add-In to convert data for WinBUGS. 2011; available from <http://www.simon.rochester.edu/fac/misra/software.htm>
40. Zhang, Z., Wang, L. Use BAUW to convert data. 2006; available from <http://www.psychstat.org/us/article.php/52.htm>
41. Sturtz, S., Ligges, U., Gelman, A. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* 2005; 12.
42. Bujkiewicz, S., Jones, H.E., Lai, M.C.W., Cooper, N.J., Hawkins, N., Squires, H. et al. Development of a Transparent Interactive Decision Interrogator to facilitate the decision making process in health care. *Value in Health* 2011; (in press).
43. Heiberger, R.M., Neuwirth, E. R through Excel: a spreadsheet interface for statistics, data analysis, and graphics. Springer-Verlag New York Inc, 2009.
44. Thompson, J., Palmer, T., Moreno, S. Bayesian analysis in Stata using WinBUGS. *The Stata Journal* 2006; 6(4):530-549.
45. Yan, J., Prates, M. Package rbugs: Fusing R and OpenBugs. Version 0.4-9. 2011; available from <http://cran.r-project.org/web/packages/rbugs/index.html>

46. Higgins, J.P.T., Whitehead, A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; 15:2733-2749.
47. Caldwell, D.M., Ades, A.E., Higgins, J.P.T. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331:897-900.
48. Lu, G., Ades, A. Assessing evidence consistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; 101:447-459.
49. Ades, A.E., Sculpher, M., Sutton, A., Abrams, K., Copper, N., Welton, N.J. et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics, PharmacoEcon* 2006; 24(1):1-19.
50. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
51. Woolacott, N., Bravo Vergel, Y., Hawkins, N., Kainth, A., Khadjesari, Z., Misso, K. et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2006; 10(31).
52. Woods, B.S., Hawkins, N., Scott, D.A. Network meta-analysis on the log-hazard scale, combining counts and hazard ratio statistics accounting for multi-arm trials: A tutorial. *BMC Medical Research Methodology* 2010; 10(54).
53. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>

APPENDIX: PARAMETERISATION OF THE BASELINE TERM IN NETWORK META-ANALYSIS MODELS

Two ways of specifying network meta-analysis models have appeared in the literature. The main area of difficulty lies in the precise way in which the relative effects have been parameterised, particularly in RE models. One way to write the model for binomial data is:

$$\begin{aligned}
 r_{ik} &\sim \text{Binomial}(p_{ik}, n_{ik}) \\
 \text{logit}(p_{ik}) &= \mu_i && \text{if } k = X \\
 \text{logit}(p_{ik}) &= \mu_i + \delta_{i,XY} && \text{if } k = Y \\
 \delta_{i,XY} &\sim N(d_{1Y} - d_{1X}, \sigma^2) \\
 \mu_i &\sim N(0, 100^2)
 \end{aligned} \tag{2}$$

Here, for trial i and treatment k the numerators r_{ik} are have binomial distributions with parameter p_{ik} and denominator n_{ik} . A logistic regression model says that, for a trial comparing treatments X and Y , the linear predictor consists of the “baseline” term μ_i if the treatment is X , and the baseline plus a relative treatment effect $\mu_i + \delta_{i,XY}$ if the treatment is Y . The key point is that the actual treatment that μ_i refers to can be any treatment (except the last one). The trial-specific relative effects are from distributions, whose means can be expressed in terms of the mean treatment effects relative to treatment 1 (through the consistency relations). The baseline terms are usually given vague priors.⁴⁶⁻⁵⁰

In (2) it is clear that the term μ_i refers to treatment X , but it is also clear that this is not necessarily treatment 1. In this formulation, one arm of a two arm trial informs a trial-specific “baseline”, while the other informs the baseline plus a treatment effect.

Another way of writing a network model that has appeared in certain submissions to NICE⁵¹ contains this modification:

$$\begin{aligned}
 \text{logit}(p_{ik}) &= \mu_i + \delta_{i,1X} && \text{if } k = X \\
 \text{logit}(p_{ik}) &= \mu_i + \delta_{i,1Y} && \text{if } k = Y
 \end{aligned} \tag{3}$$

In a general sense, this is the same model, except that here the μ_i terms represent Treatment 1 in every trial. In a trial that includes Treatment 1, this parameterisation is adequate, as $\delta_{i,1X}$ (the effect of X relative to itself) is zero. However, if any trial does *not* include treatment 1, then the model parameters cannot all be identified because there are three parameters to be estimated from only two arms. Another way of writing an MTC model that has appeared recently⁵² has this same property. Models specified in this way should not be accepted unless

the reference treatment is the same treatment in every trial, and therefore appears in every trial.

The instability caused by this parameterisation can be avoided if a model is placed on the baseline terms, i.e. $\mu_i \sim N(m, \sigma_m^2)$ rather than $\mu_i \sim N(0, 100^2)$. However, this solution has a number of further difficulties and is not recommended (see TSD5⁵³).