NICE DSU TECHNICAL SUPPORT DOCUMENT 5: EVIDENCE SYNTHESIS IN THE BASELINE NATURAL HISTORY MODEL

REPORT BY THE DECISION SUPPORT UNIT

August 2011

(last updated April 2012)

Sofia Dias[1], Nicky J Welton[1], Alex J Sutton[2], AE Ades[1]

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

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

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

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

Dr Allan Wailoo

Director of DSU and TSD series editor.

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

Acknowledgements

The DSU thanks Julian Higgins, Alec Miners, Matt Stevenson and the team at NICE led by Jennifer Priaulx for reviewing this document. The editor for the TSD series is Allan Wailoo.

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

This report should be referenced as follows:

Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

EXECUTIVE SUMMARY

This Technical Support Document reviews synthesis issues that arise on the construction of a baseline natural history model. The intention is to cover both the absolute response to treatment on the outcome measures on which comparative effectiveness is defined, and also other elements of the natural history model, which are usually "downstream" of the shorter-term effects reported in trials.

It is recommended that the same Generalised Linear Modelling framework is used to model the absolute effects of a "standard treatment" or placebo comparator, as that proposed for synthesis of relative treatment effects in TSD2.[1] Investigators should take care to justify their choice of data sources to inform the baseline, which could include a subset of the trials identified in the systematic review of relative effect data, cohort studies, patient registers, expert opinion, or combinations of these. It is suggested that the predictive distribution, rather than the fixed effect or random effects mean, is used to represent the baseline in order to reflect the observed variation in baseline rates.

It is preferable to construct the baseline model independently from the model for relative treatment effects, in order to ensure that the latter are not affected by assumptions made about the baseline. However, simultaneous modelling of baseline and treatment effects could have some advantages, for example, when evidence is very sparse or when other research or study design give strong reasons for believing in a particular baseline model. Options for modelling effects of covariates, based either on aggregate or, preferably, individual patient data, or the use of risk equations for baselines, are also available.

The natural history model beyond the (usually) shorter-term effects may also have parameters whose values are derived from synthesis. In the simplest and most easily interpretable models, the only effect of treatment is on the short-term outcomes and there is no independent effect of treatment on subsequent outcomes in the model. However, the evidence may not support this model structure. In this case, wherever possible, all parameters which take on different values for different treatments should be based on randomised data. Just as with the shorter term outcomes, great care should be taken in introducing non-randomised evidence that impacts directly on the between-treatment comparisons on later outcomes. Additionally the increased uncertainty and potential for bias should be addressed.

Joint modelling of multiple baseline outcomes based on data from trials, or combinations of trial and observational data is recommended where possible, as this is likely to make better

use of available evidence, to produce more robust results, and to ensure that the model is internally coherent.

Special synthesis methods for state transition models are available, that can be used to synthesise information from studies reporting state transitions at different follow-up times, results presented as either risks or rates, and information from incompletely observed state transitions.

Finally, it is important to validate longer term predictions of the model against independent data wherever possible. There are several advantages to a multi-parameter synthesis approach to model validation and calibration, based on synthesis of the validating or calibrating data within the same model as the rest of the evidence.

CONTENTS

TABLES, BOXES AND FIGURES

Abbreviations and definitions

CEA	Cost-effectiveness analysis
EOGBS	Early onset Group B Streptococcus
HTA	Health Technology Assessment
RCT	Randomised controlled trial
RE	Random effects
TSD	Technical Support Document

1. INTRODUCTION

Most cost-effectiveness analyses (CEAs) consist of two separate components: a baseline model that represents the absolute natural history under a standard treatment in the comparator set, and a model for relative treatment effects. The former may be based on trial or cohort evidence, while the latter is generally based on randomised controlled trial (RCT) data.[2] The natural history under the new treatment is then obtained by putting together the baseline natural history model with the relative effect estimates based on the trial data. For example, if the probability of an undesirable event under standard care is 0.25 and the odds ratio for a given treatment compared to standard care is 0.8 (favouring the treatment), then, ignoring the uncertainty in these quantities, the absolute probability of an event on the treatment is p=0.21, obtained as

$$\text{logit}(p) = \text{logit}(0.25) + \ln(0.8)$$

where $\text{logit}(x)=\ln(x/(1-x))$ (see TSD2[1] for details). A similar approach can be used with models that are linear in log-relative risks or log-hazard rates (see TSD2[1]).

Usually, the role of trial data within an economic evaluation – whether to inform absolute or relative effects – is limited to the short- or intermediate-term outcomes. Health economists expend considerable effort in building the longer-term elements of the model, which often take the form of a Markov transition model where the relative treatment effects will be assumed to act on specific transitions.[3] However, a wide range of modelling techniques may be utilised, apart from Markov models. "Mapping" from the shorter term outcomes, or from the Markov states into utilities is a further component of the model.[3]

This Technical Support Document (TSD) focuses on the evidence synthesis issues that arise in construction of the natural history model, based on the general principles set out in the NICE Guide to Methods of Technical Appraisal.[2] There is no attempt to give recommendations or guidance on principles of model construction, or on the type of model, except in so far as this might affect synthesis issues. Patient-level simulation models, where patients are tracked individually throughout the economic model, are outside the scope of this document, which is focused on evidence synthesis. Readers are referred to the literature for more details.[3,4]

We begin, with the baseline model for the outcomes that have been compared in some form of comparative study (Section 2). For this purpose we borrow heavily from the generalised linear modelling framework developed in TSD2,[1] since the exact same link functions and likelihoods used to analyse information on relative treatment effects can, and should, be

applied to synthesise the evidence on the baseline treatment. In this section we also devote some space to the question of what sources of evidence should be used for the baseline model, and also to potential linkage between baseline and relative effect models.

Section 3 then examines synthesis issues that relate to the natural history model "downstream" of the trial data, that is how to inform parameters, other than relative treatment effects, required to inform the economic model. Typical parameters that require values could be as diverse as complication rates from the underlying condition, duration of hospital stay, duration of medication, natural history following cessation of treatment, incidence of side effects, relapse rates, mortality on and off treatment, "mappings" from surrogate to clinical end-points or from disease-specific measures to Quality of Life measures, and so on. As noted above, there is a wide range of model types and model structures that can be deployed. If state transition models are used, it is possible that the trial outcomes represent only the transition from one specific state to another, and that information on the remaining transitions will need to be sourced from elsewhere. Usually, identification of appropriate data to inform these parameters is likely to be more critical to the decision than technical issues of how to synthesise the evidence once it is selected.

However, two specific issues deserve careful consideration. In the ideal case *all* predicted differences between treatments would originate from information from RCTs. This applies as much to the downstream outcomes as it does to the more immediate short-term outcomes which are usually based on RCT data. Any use of non-randomised data that has a direct bearing on between-treatment comparisons always needs careful consideration of potential bias[2] (see also TSD3[5]). Secondly, whether information on "downstream" outcomes is based on randomised or non-randomised data, there is a potential for conflict between the observed long-term relative effects and those predicted by the short-term and natural history models. We describe the issues and suggest possible approaches, although this area has not yet been adequately researched.

Finally, Section 4 briefly reviews issues in model calibration and validation. In most accounts of validation, predictions of final outcomes from a model are compared with independent sources of data. "Calibration" is a process whereby estimates are modified in light of external data. A third possibility is an evidence synthesis approach, in which the model parameters are all estimated simultaneously using *both* the original sources of information *and* the independent data. Section 4 outlines properties of a synthesis approach to validation and calibration, and provides some references to these methods.

2. BASELINE MODELS FOR TRIAL OUTCOMES

2.1. SOURCES OF EVIDENCE FOR BASELINE OUTCOMES

Once a baseline (or reference) intervention has been defined (see TSD2[1]), a reasoned protocol for systematic study search and inclusion should be developed,[6-8] and potential sensitivity to alternative options explored, if appropriate. Since the baseline response should be as specific as possible to the population of interest,[2,3] it may be more reasonable to use only evidence from recent trials, relevant cohort studies, register studies[9] or, in certain cases, expert opinion.[8] A common approach to identifying sources of evidence for baseline outcomes has been to use the same trials that have supplied information on relative effects, but restricting attention to the trials arms that use the baseline treatment. This is a possible approach, but needs to be justified in each case. Investigators should consider whether *all* the trials used to inform the relative effects can be considered as equally representative of the absolute response that would be obtained in the target population and under current circumstances, particularly if some of the trials were carried out many years ago or had very restrictive inclusion criteria. It is also possible to combine evidence from different types of relevant randomised and non-randomised studies.

Whatever the source of evidence used to populate the decision model, this should be transparent and reported in sufficient detail to allow outside scrutiny.[2,6,8,10]

2.2. SYNTHESIS OF AGGREGATE DATA ON BASELINE RESPONSE

2.2.1. Separate models for baseline and treatment effect

TSD2[1] introduced a Generalised Linear Modelling framework for synthesis of relative effect estimates. This can be expressed as:

$$g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$

where

$$I_{\{u\}} = \begin{cases} 1 & \text{if } u \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

$g()$ is the link function (for example the logit link), and θ_{ik} is the linear predictor, consisting of a trial-specific baseline effect in a trial i, μ_i (for example a log odds), and $\delta_{i,1k}$ a trial-specific treatment effect of the treatment in arm k relative to the treatment in arm 1 (the log odds ratio). In the Bayesian framework adopted throughout these documents in models for

the relative treatment effect, the μ_i are given unrelated vague priors. To model baseline effects the following formulation can be adopted:

$$\begin{aligned} g(\gamma) &= \theta_{ik} = \mu_i \\ \mu_i &\sim N(m, \sigma_m^2) \end{aligned} \tag{1}$$

in which the study-specific baselines are drawn from a distribution of effects with a common mean and variance. To complete the model, in a Bayesian framework, vague priors can be put on the mean and on the variance, for example $m \sim N(0,100^2)$, and $\sigma_m \sim Uniform(0,5)$ or $1/\sigma_m^2 \sim Gamma(10^{-3},10^{-3})$.

The proposal is, therefore, that a separate model is run to summarise the relevant baseline data. One option is to run this code at the same time as the model for the relative treatment effect, ensuring that the information in the baseline model does not propagate to the relative treatment effects model. This can be done in WinBUGS using the "cut" function.[11] The advantage of this approach is that both models are contained in a single file and can be run simultaneously thereby ensuring that any new data added to the baseline model automatically updates the absolute effects generated from the relative effects model. It also ensures that the samples from the posterior distribution of the baseline effect are used directly. A simpler alternative is to run separate models, and then, assuming normality of the posterior distribution of the baseline effect, take the appropriate posterior summaries (the mean and uncertainty), and insert them into the relative effect code. This will of course rely on the approximate normality of the posterior distribution of the baseline effect – this should always be checked, but usually holds, in our experience. Alternatively, the samples from the posterior distribution of the baseline effect can be fed into the separate relative effects model.

In each of the examples presented in the Appendix for TSD2,[1] where WinBUGS code for various models and outcome types is given, results from a separate external analysis are "plugged in" to generate predictions for absolute response rates or probabilities. For illustrative purposes, the external analysis generated a predictive distribution of the baseline in a "new" study, based on a random effects model. Note that this is different from simply calculating the unweighted mean of the baseline arms in the relative effects model, an approach which is not recommended under any circumstances.

The reasons for keeping the baseline and relative treatment effects models separate are to avoid the assumptions made on the baseline model affecting the relative treatment effects and because they are often based on different data sources. However, if there are strong reasons to believe in a particular baseline model, joint modelling of baseline and relative effects should

be considered (see Section 2.2.2). Issues related to correlation between outcomes and baseline by treatment interactions are discussed in TSDs 2[1] and 3,[5] respectively.

Program 1 in the Appendix to this document includes code which takes the 19 'No intervention' arms from a smoking cessation dataset[12] (see TSD4,[13] Section 4.2.1), and implements the model in equation (1).

There are two ways the results of this analysis can be used. The simplest approach is to use the posterior mean of *m* and its posterior standard deviation to represent the baseline response. But, it could be argued that this under-represents the variation observed in the data: if we were to gather more and more data on the baseline arm, our estimate of the mean would become more and more precise, but the variation would remain unchanged. An alternative, therefore, is to use the predictive distribution of a new baseline,

$$\mu_{new} \sim N(m, \sigma_m^2) \tag{2}$$

where *m* and σ_m^2 are sampled from the posterior distribution. This predictive distribution for a new baseline incorporates the uncertainty about the value a new observation might take, as well as the observed variation in the data. It is however important to ensure that the uncertainty conveyed by the predictive distribution reflects genuine uncertainty in the baseline. Therefore, we reiterate the need for careful evaluation of what studies should be used to inform the baseline model and whether the exchangeability assumption between the baseline effect in the included studies and the "new" baseline (equations (1) and (2)) holds.

Both approaches are illustrated in the Appendix (Program 1). The first column of Table 1 shows the results obtained in the smoking cessation example, using separate RE models for baseline and treatment effects. All results are based on 50,000 iterations from 3 independent chains, after discarding 20,000 burn-in iterations to ensure convergence. Using the posterior distribution of the mean produces a mean baseline smoking cessation probability of 0.07 with credible interval (0.05, 0.09). By contrast, if the predictive distribution is used, the mean is approximately the same but the wider credible interval (0.02, 0.20) better reflects the range of variation in the observed data, under the assumption of normally distributed random effects (Table 1).

Table 1 Posterior mean, standard deviation (sd) and 95% Credible Interval (CrI) of the mean and predictive log-odds of smoking cessation on 'No contact' (m and μ_{new}), absolute probabilities of smoking cessation based on the posterior and predictive distributions of the baseline log-odds, and of the log-odds ratio of response relative to 'No Contact' (log-odds ratios > 0 favour the active treatment). Posterior median, sd and 95% CrI for the between-trial heterogeneity in baseline (σ_m) and in treatment effects (σ) for RE meta-analysis with separate or simultaneous baseline and treatment effects modelling; and measures of model fit: posterior mean of the residual deviance (resdev), effective number of parameters (pD) and DIC.

	Separate Models			Simultaneous modelling		
	mean/median	sd	95% CrI	mean/median	sd	95% CrI
	Baseline model parameters					
m	-2.59	0.16	(-2.94,-2.30)	-2.49	0.13	(-2.75,-2.25)
σ_m	0.54	0.16	(0.32,0.93)	0.45	0.11	(0.29,0.71)
μ_{new}	-2.59	0.60	(-3.82,-1.41)	-2.49	0.49	(-3.48,-1.52)
	Absolute probabilities of response based on the posterior distribution of the baseline probability					
No contact	0.07	0.01	(0.05,0.09)	0.08	0.01	(0.06,0.10)
Self-help	0.12	0.05	(0.05,0.23)	0.13	0.04	(0.07,0.21)
Individual Counselling	0.15	0.04	(0.09,0.24)	0.15	0.03	(0.11,0.21)
Group Counselling	0.19	0.07	(0.08,0.37)	0.20	0.05	(0.11,0.31)
	Absolute probabilities of response based on the predictive distribution of the baseline probability					
No contact	0.08	0.05	(0.02,0.20)	0.08	0.04	(0.03,0.18)
Self-help	0.13	0.08	(0.03,0.34)	0.14	0.07	(0.04,0.30)
Individual Counselling	0.17	0.09	(0.05,0.39)	0.16	0.07	(0.06,0.33)
Group Counselling	0.21	0.12	(0.05,0.50)	0.21	0.09	(0.07,0.43)
	Relative treatment effects compared to 'No contact'					
Self-help	0.49	0.40	(-0.29,1.31)	0.53	0.33	(-0.11,1.18)
Individual Counselling	0.84	0.24	(0.39,1.34)	0.78	0.19	(0.41,1.17)
Group Counselling	1.10	0.44	(0.26,2.01)	1.05	0.34	(0.39,1.72)
σ	0.82	0.19	(0.55,1.27)	0.71	0.13	(0.51,1.02)
resdev*	54.1			47.4		
pD	45.0			40.1		
DIC	99.1			87.5		

* compare to 50 data points

Note that the choice of posterior or predictive distribution will have very little effect on the differences between treatments, but the latter will contribute greater uncertainty in the natural history model. The probabilities of smoking cessation for the four treatments calculated using both the posterior and predictive uncertainties are shown in Table 1. Figure 1 illustrates the impact of the two approaches on the uncertainty in treatment effects, which results in the

wider credible intervals when the predictive distribution is used, representing the true level of uncertainty.

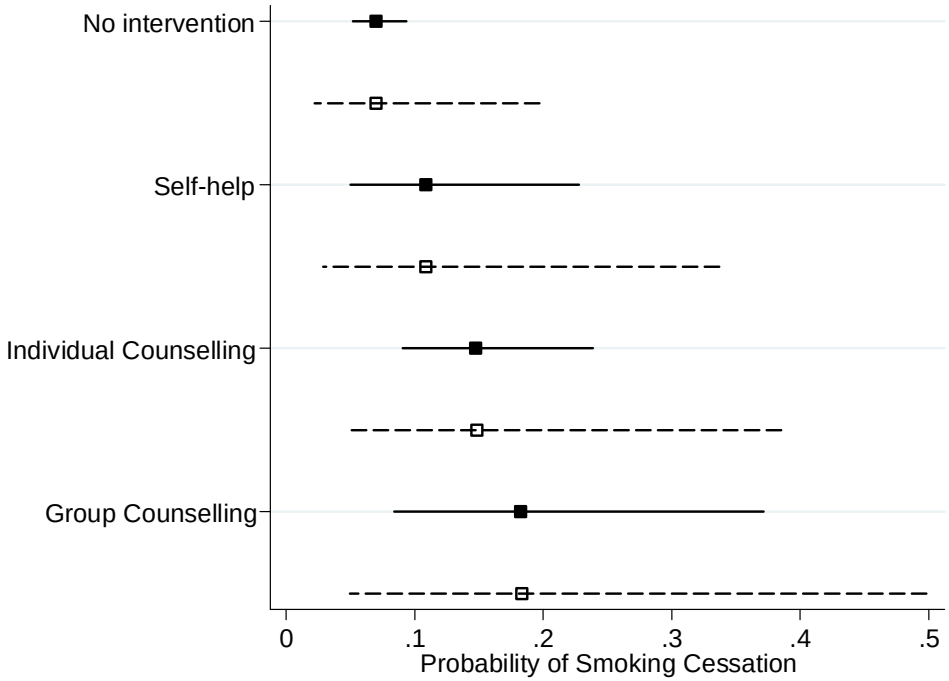

Figure 1 Absolute probabilities of smoking cessation for all treatments with 95% CrI based on the posterior (filled squares, solid lines) and predictive distributions (open squares, dashed lines) of the mean log-odds for "No intervention" for separate RE models for the baseline and treatment effects.

2.2.2. Simultaneous modelling of mean and treatment effects

The separation of absolute and relative treatment effects may seem artificial. Nevertheless, it is the recommended method because it means that the treatment effects are unaffected by any assumptions made about the baseline. However, there may be reasons for modelling the baseline and treatment effects together. One reason would be that this can increase the stability of the model when there are zero cells (see TSD2[1], Section 6.3). Another may be that, based on other research, there are strong reasons for believing in a particular model for the baseline, for example when modelling results from cluster randomised[14,15] or multicentre trials (whereas usually no model for baseline is assumed – see TSD2[1]).

To carry out such an analysis, it is only necessary to replace the "unrelated" priors for μ_i in the standard meta-analysis code presented in TSD2,[1] with a "random effects" prior with a mean and variance, and to supply priors for the mean and between-study variance of the baseline effects. In a network meta-analysis where not all trials include the baseline

(reference) treatment, it is necessary to ensure that the μ_i being modelled always refer to the baseline treatment, i.e. treatment 1. Note that simultaneous modelling of the baseline and treatment effects will have considerable impact on the relative effect estimates, and always needs to be justified. A sensitivity analysis to show the effect on the relative treatment effects should be carried out, if possible. WinBUGS code for simultaneous modelling of baseline and treatment effects is supplied in Program 2 in the Appendix. Once again, we would recommend that the predictive distribution of a "new" baseline (equation (2)) is taken forward for decision modelling.

The second column in Table 1 shows the posterior and predictive probabilities of smoking cessation for the four treatments from a simultaneous model of baselines and treatment effects. Having a model that simultaneously estimates baseline and treatment effects greatly impacts on the estimated between-trial heterogeneity (posterior median of σ=0.82 for separate models and 0.71 in the joint model) and consequently on the uncertainty around the mean treatment effects. This in turn produces less uncertainty in the absolute treatment effects based on the predictive distribution.

The heterogeneity in the observed baselines σ_m is also smaller when modelling baseline and treatment effects simultaneously than when separate analyses are performed, which has an impact on the variability of the predictive distribution for the baseline, given by the standard deviation of μ_{new} in Table 1.

2.3. BASELINE MODELS WITH COVARIATES

2.3.1. Using aggregate data

Covariates may be included in the baseline model by including terms in the linear predictor. For a covariate *C*, which could either be a continuous covariate or a dummy covariate, we would have, for arm *k* of trial *i*

$$\theta_{ik} = \mu_i + \beta C_i + \delta_{ik} I_{\{k \neq 1\}}$$

An estimate of the covariate effect β, like the estimate of μ, could be obtained from the trial data or externally. Govan et al[16] give an example where the covariate on the baseline is estimated from aggregate trial data with the purpose of reducing aggregation bias.[17] This is a phenomenon in which the presence of a strong covariate, even if balanced across arms, and even if it is not a relative effect modifier, causes a bias in the estimation of the relative

treatment effects, towards the null. Govan et al[16] also show a method for dealing with missing data on covariates. See TSD3[5] for further discussion.

2.3.2. Risk Equations for the baseline model based on Individual Patient Data

A far more reliable approach to informing a baseline model which expresses difference in baseline progression due to covariates such as age, sex, disease severity at onset of treatment, is to use individual patient data. This is considered superior to aggregate data as the coefficients can be estimated more precisely and with less risk of ecological bias. The results are often presented as "risk equations" based on multiple regression from large trial databases, registers, or cohort studies. Natural histories for each treatment are then generated by simply adding the treatment effects based on trial data to the risk equations as if they were another risk factor. The main difficulty facing the cost-effectiveness analyst here is in justifying the choice of data source and its relevance to the target population. Analyses should be presented that explore the different characteristics of the populations in these alternative studies, and their relation to the target population for the decision. If necessary, sensitivity analyses should be presented to show sensitivity of results to the choice of data source used to inform these parameters.

3. SYNTHESIS ISSUES IN THE REST OF THE NATURAL HISTORY MODEL

Choice of evidence sources and statistical model for the natural history model beyond the immediate short-term trial outcomes is beyond the scope of this document. However, we provide some comments on the origin of treatment differences, or implied treatment differences, in longer term outcomes, as this touches on synthesis issues, on the internal coherence of models and their consistency with the evidence.

3.1. SOURCE OF INFORMATION FOR NATURAL HISTORY PARAMETERS AND IMPLICATIONS FOR RELATIVE TREATMENT EFFECTS

Generally, the source of evidence used for each natural history parameter should be determined by a protocol driven review.[2,6,8] Previous CEAs are an important source of information on the data sources that can inform natural history.

A common modelling strategy is to assume that there are no differences between treatments in the "downstream" model, conditional on the shorter term trial outcomes. We can call this the "single mapping hypothesis" as the implication is that, given information on the short-term differences, longer-term differences can be obtained by a single mapping applicable to all treatments. For example, in a model to assess cost-effectiveness of various antiviral drugs for the treatment of influenza, the base-case analysis assumed that use of antivirals only affected short-term outcomes and had no additional impact on longer term complication and hospitalisation rates.[18] Models with this property are attractive, although they make strong assumptions. The assumptions are natural if the alternative active comparators can be considered to be a single class, but may be less plausible if they are not. Such assumptions have to be justified clinically and physiologically, and for each outcome "mapped" available data, for example on length of hospital stay, time on treatment, complications rates, mortality, and all other downstream outcomes, should be reviewed, examined and interpreted. This review should also include the empirical and statistical literature on adequacy of surrogate outcomes, and in particular whether the evidence supports the view that treatment effects on the shorter-term "surrogate" translate into the same longer-term benefits for all treatments. This review might usefully extend beyond the class of products being considered, because the wider the range of treatment for which a "single mapping" hypothesis can be sustained, the more robust it is likely to be. Eventually, however, it may be decided that the relation between surrogate and clinical outcomes is only relevant for the subset of treatments within the decision. The use of "surrogate endpoint" arguments in health technology assessment (HTA) extends far beyond the outcomes classically understood as "surrogates" in the clinical and statistical literature.[19] HTA literature makes frequent use of "mapping" from short-term to longer-term outcomes, as this allows modellers to base the modelled treatment differences on short term evidence.

If the assumption that all downstream differences between treatments outcomes are due exclusively to differences in shorter term trial outcomes is not supported by the evidence, then the first option is to use available randomised evidence to drive longer-term outcomes. This necessarily implies different "mappings" for each treatment.

The second and least preferred option is the use of non-randomised evidence. However, as with short-term outcomes, it is essential that any use of non-randomised data that directly impacts on differential treatment effects within the model is carefully justified, and that the increased uncertainty and the possibility of bias is recognised and addressed.[2]

3.2. JOINT SYNTHESIS OF MULTIPLE OUTCOMES TO INFORM NATURAL HISTORY

The natural history model usually consists of a succession of "states" or sub-processes and involves a series of parameters which may impact on life-times costs, quality and length of life. It is preferable for these parameters to be estimated simultaneously from all the available data, as this is likely to allow more information to be incorporated and more validation to be carried out on the agreement between the model predictions and the evidence. The simplest examples of a coherent modelling of multiple outcomes are provided by the competing risk and the ordered probit analyses described in TSD2[1] (Sections 3.3 and 3.6, respectively). For example, use of the ordered probit model for the baseline and treatment effects guarantees coherent prediction of the probability that patients will achieve the different levels of response on categorical scales such as the PASI or ACR. By contrast, if ACR 20, ACR 50, and ACR 70 response are analysed separately, it is possible to end up with a model that makes impossible predictions, for example that more patients experience a 50% improvement on ACR than experience a 20% improvement.

However, use of advanced modelling techniques may not have a substantial impact on cost-effectiveness, and the usual approach in which each natural history parameter is sourced independently from data is more commonly adopted.

Joint modelling of multiple trial outcomes to obtain the relative treatment effects has particular advantages, as is being seen increasingly in HTA. As well as reflecting a "coherent" view of the different outcomes, and correctly capturing the correlations between them, these methods address the frequently encountered problem of different outcomes being reported by different trials. The option of choosing a single outcome as the basis for the between-treatment comparison, may result in a high proportion of the information being discarded. It may be preferable, and lead to more robust results, if a model can be devised that expresses the relationships between the different outcomes, and thus allows *all* the evidence on treatment efficacy to be incorporated. Examples of models of treatment effects on multiple outcomes include treatment effects at multiple follow-up times,[20,21] and multivariate models for continuous outcomes.[22-24] It is also possible to synthesise two separate trial outcomes and parameters that link the outcomes, but which are based on observational data.[25,26]

Somewhat more complex examples have arisen in the analysis of influenza treatments[18,27] which included a model of the relation between "time to end of fever" and "time to end of

symptoms", or synthesis of outcomes on tumour response, time to progression and overall survival in advanced breast cancer.[28,29] However, model structures vary across different diseases and, even within types of conditions, the structure of the evidence available to inform models can vary considerably. For these reasons it is difficult to provide general recommendations, other than to note that a single model encompassing several outcomes, as long as its assumptions are clear, and reflect a consensus view among clinical experts, is likely to provide a more robust basis for cost-effectiveness modelling.

3.2.1. Synthesis of state transition models

As with other natural history models, state transition model parameters may each be estimated independently from different sources, or they may be modelled jointly, although, as before, there are advantages in using methods that are capable of incorporating available information from all relevant sources. However, synthesis of state transition model parameters raises some special considerations because of the great variety of forms in which information is made available, for example:

1. Data in study j may be reported as the probability of state transitions during a time interval T_j while the modeller may wish to use these data in a model with a cycle time T_0. It is important to note that the standard adjustment[30] is only valid for 2-state models.
2. Information may be available on risks, or on rates.
3. Information may be available on Hazard Ratios, but these cannot be easily converted into Relative Risks (or vice versa) in multi-state models, as the Relative Risk depends on the cycle time.
4. Information may be available on state transitions from state A to state B, where individuals may have visited other states in between. This is sometimes referred to as an incompletely observed Markov process.

Methods are available for synthesising a wide range of information on transitions, reported in different ways, over different time periods, and between different states in a model.[31] Further, these methods can be used to simultaneously model natural history and treatment effect parameters,[32] as before. Such methods also provide examples of a synthesis approach to calibration, described below. To date, these methods have all been limited to the case where all transition times are exponentially distributed. It remains to be seen how and under what conditions, the methods can be extended to other distributions.

4. MODEL VALIDATION AND CALIBRATION THROUGH MULTI-PARAMETER SYNTHESIS

Natural history models should be validated against independent data wherever possible. For example, in CEAs comparing a new cancer treatment to a standard comparator, the survival rates predicted in the standard arm could be compared to published survival rates, perhaps after suitable adjustment for age or other covariates. With other conditions, given an initial estimate of incidence or prevalence, together with statistics on the size of the population, the natural history model may deliver predictions on absolute numbers admitted to hospital with certain sequelæ, complications, or mortality. Once again these predictions could be checked against independent data to provide a form of validation.

A more sophisticated approach is to use this external data to "calibrate" the natural history model. This entails changing the "progression rate" parameters within the model so that the model accurately predicts the independent calibrating data. Calibration, in a Bayesian framework particularly, can also be seen as a form of evidence synthesis.[33] In this case the calibrating data is characterised as providing an estimate of a complex function of model parameters. This approach offers a remarkably simple form of calibration because, in principle, all that is required is that the investigator specifies the function of model parameters that the calibrating data estimates, and that a term for the likelihood for the additional data is added to the model. The information then propagates "backwards" through the model to inform the basic parameters. There are many advantages of this method over standard methods of calibration, which have recently been reviewed by Vanni et al:[34]

1. it gives an appropriate weight to the calibrating data, taking account of sampling error;
2. it avoids the "tweaking" of model parameters until they "fit" the calibrating data, a procedure that fails to capture the uncertainty in the data;
3. It avoids forcing the investigator to decide *which* of several natural history parameters should be changed (see below);
4. Assessment of whether the validating data conflicts with the rest of the model and the data supporting it can proceed using standard model diagnostics, such as residual deviance, DIC, or cross-validation.[1,5,25]

Examples of this approach have appeared in descriptive epidemiology[35-38], and also in screening applications. In a model of early-onset neonatal group B streptococcus disease (EOGBS), the natural history model involved a series of parameters: probability of maternal carriage of group B streptococcus disease, probability of transmission to the newborn given

maternal carriage and probability of EOGBS given transmission. While information was available on each of these probabilities, the model was "calibrated" to data on the numbers of cases of EOGBS that had been reported in the British Isles through a paediatric clinical surveillance scheme.[39] The effect of this form of calibration in this case is to put extremely weak constraints on the individual progression parameters, but to place quite strong constraints on their product.

This kind of approach could potentially be applied in a number of clinical areas where independent data on long-term follow-up, registration of disease, or cause-specific mortality are available, though more research is needed before clear recommendations can be made.

5. REFERENCES

1. 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. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

2. National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal (updated June 2008). 2008.

3. Briggs, A., Claxton, K., Sculpher, M. Decision modelling for health economic evaluation. Oxford University Press, Oxford; 2008.

4. Barton, P., Bryan, S., Robinson, S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services & Research Policy* 2004; 9:110-118.

5. Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

6. Sculpher, M., Fenwick, E., Claxton, K. Assessing Quality in Decision Analytic Cost-Effectiveness Models. A Suggested Framework and Example of Application. *Pharmacoeconomics* 2000; 17(5):461-477.

7. Weinstein, M.C., O'Brien, B., Hornberger, J., et al. Principles of Good Practice for Decision Analytic Modeling in Health Care Evaluation: Report of the ISPOR task force on Good Research Practices-Modeling studies. *Value Health* 2003; 6:9-17.

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

9. Golder, S., Glanville, J., Ginnelly, L. Populating decision-analytic models: The feasibility and efficiency of database searching for individual parameters. *International Journal of Technology Assessment in Health Care* 2005; 21:305-311.

10. Nuijten, M.J.C. The selection of data sources for use in modelling studies. *Pharmacoeconomics* 1998; 13:305-316.

11. Spiegelhalter, D., Thomas, A., Best, N., Lunn, D. WinBUGS User Manual Version 1.4 January 2003. Upgraded to Version 1.4.3. 2007; available from http://www.mrc-bsu.cam.ac.uk/bugs

12. Hasselblad, V. Meta-analysis of multi-treatment studies. *Med Decis Making, MED* 1998; 18:37-43.

13. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Lu, G., Ades, A.E. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

14. Brown, J., Welton, N.J., Bankhead, C., Richards, S.H., Roberts, L., Tydeman, C. et al. A Bayesian approach to analysing the cost-effectiveness of two primary care

interventions aimed at improving attendance for breast screening. *Health Economics* 2006; 15:435-445.

15. Welton, N.J., Ades, A.E., Caldwell, D.M., Peters, T.J. Research Prioritisation Based on Expected Value of Partial Perfect Information: a Case Study on Interventions to Increase Uptake of Breast Cancer Screening. *Journal of the Royal Statistical Society (A)* 2008; 171:807-841.

16. Govan, L., Ades, A.E., Weir, C.J., Welton, N.J., Langhorne, P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. *Statistics In Medicine* 2010; 29:1340-1356.

17. Rothman, K.J., Greenland, S., Lash, T.L. Modern Epidemiology. 3 ed. Lippincott, Williams & Wilkins, Philadelphia; 2008.

18. Burch, J., Paulden, M., Conti, S., Stock, C., Corbette, M., Welton, N.J. et al. Antiviral drugs for the treatment of influenza: a systematic review and economic evaluation. *Health Technology Assesment* 2010; 13(58):1-290.

19. Taylor, R.S., Elston, J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK health technology assessment reports. *Health Technology Assesment* 2009; 13(8).

20. Lu, G., Ades, A.E., Sutton, A.J., Cooper, N.J., Briggs, A.H., Caldwell, D.M. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics In Medicine* 2007; 26(20):3681-3699.

21. Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M.C., Schomig, A. et al. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *Lancet* 2007; 370:937-948.

22. Nam, I.-S., Mengerson, K., Garthwaite, P. Multivariate meta-analysis. *Statistics In Medicine* 2003; 22:2309-2333.

23. Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J., Thompson, J.R. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics In Medicine* 2007; 26:78-97.

24. Riley, R.D., Thompson, J.R., Abrams, K.R. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008; 9:172-186.

25. Ades, A.E. A chain of evidence with mixed comparisons: models for multi-parameter evidence synthesis and consistency of evidence. *Statistics In Medicine* 2003; 22:2995-3016.

26. Epstein, D., Sutton, A. Modeling correlated outcomes in health technology appraisals. *Value in Health* 2011; in press.

27. Welton, N.J., Cooper, N.J., Ades, A.E., Lu, G., Sutton, A.J. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of

antivirals for treatment of influenza A and B. *Statistics In Medicine* 2008; 27:5620-5639.

28. Welton, N.J., Willis, S.R., Ades, A.E. Synthesis of Survival and Disease Progression Outcomes for Health Technology Assessment of Cancer Therapies. *Res Synth Method* 2010; 1:239-257.

29. National Institute for health and Clinical Excellence. Advanced Breast Cancer: diagnosis and treatment. NICE Clinical Guidelines. 2009.

30. Miller, D.K., Homan, S.M. Determining transition probabilities: confusion and suggestions. *Med Decis Making* 1994; 14:52-58.

31. Welton, N.J., Ades, A.E. Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis and Model Calibration. *Med Decis Making, MED* 2005; 25:633-645.

32. Price, M., Welton, N., Ades AE. Parameterisation of treatment effects for meta-analysis in multi-state Markov models. *Statistics In Medicine* 2011; 30:140-151.

33. Ades, A.E., Sutton, A.J. Multiparameter evidence synthesis in epidemiology and medical decision making: current appoaches. *Statistics in Society, JRSS(A)* 2006; 169(1):5-35.

34. Vanni, T., Karnon, J., Madan, J., White, R.G., Edmunds, W.J., Foss, A.M. et al. Calibrating Models in Economic Evaluation: A Seven-Step Approach. *Pharmacoeconomics* 2011; 29:35-49.

35. Goubar, A., Ades, A.E., De Angelis, D., McGarrigle, C.A., Mercer, C.H., Tookey, P.A. et al. Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data. *Journal of the Royal Statistical Society A* 2008; 171:541-580.

36. Presanis, A., De Angelis, D., Spiegelhalter, D., Seaman, S., Goubar, A., Ades, A. Conflicting evidence in a Bayesian synthesis of surveillance data to estimate HIV prevalence. *Journal of the Royal Statistical Society (A)* 2008; 171:915-937.

37. Sweeting, M.J., De Angelis, D., Hickman, D., Ades, A.E. Estimating HCV prevalence in England and Wales by synthesising evidence from multiple data sources: assessing data conflict and model fit. *Biostatistics* 2008; 9:715-734.

38. Presanis, A.M., De Angelis, D., Goubar, A., Gill, O.N., Ades, A.E. Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men. *Biostatistics* 2011; 13:666-681.

39. Colbourn, T., Asseburg, C., Bojke, L., Phillips, Z., Claxton, K., Ades, A.E. et al. Prenatal screening and treatment strategies to prevent Group B Streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analysis. *Health Technology Assessment* 2007; 11(29):1-226.

APPENDIX: WINBUGS CODE FOR ILLUSTRATIVE EXAMPLES

Below we set out code for a baseline model (Program 1) and a model which estimates baseline and treatment effects simultaneously (Program 2), with random effects, a binomial likelihood and logit link function. In TSD2[1] a generalised linear model framework was introduced, with explanations and examples of how the code for the binomial/logit model could be adapted for other likelihoods and link functions, including Poisson/log, Normal/identity and others. The baseline models below can be adapted in exactly the same way.

All programming code is fully annotated. The code below is fully general and Program 2 will work for pairwise or network meta-analysis with any number of trials with any number of arms.

The program codes are printed here, but are also available as WinBUGS system files from http://www.nicedsu.org.uk. Users are advised to download the WinBUGS files from the website instead of copying and pasting from this document. We have provided the codes as complete programs. However, the majority of the code for Program 2 is identical to program 1(c) in TSD2,[1] with new lines of code identical to code in Program 1, the separate baseline model. We have therefore highlighted the common lines of code between Programs 1 and 2, in blue and bold, to emphasise the modular nature of the code.

PROGRAM 1. SMOKING CESSATION: BINOMIAL LIKELIHOOD, BASELINE RE MODEL WITH PREDICTIVE DISTRIBUTION

```
# Binomial likelihood, logit link
# Baseline random effects model
model{                                  # *** PROGRAM STARTS
for (i in 1:ns){                        # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])               # Likelihood
   logit(p[i]) <- mu[i]                 # Log-odds of response
   mu[i] ~ dnorm(m,tau.m)               # Random effects model
 }
mu.new ~ dnorm(m,tau.m)                 # predictive dist. (log-odds)
m ~ dnorm(0,.0001)                      # vague prior for mean
var.m <- 1/tau.m                        # between-trial variance
tau.m <- pow(sd.m,-2)                   # between-trial precision = (1/between-trial variance)
sd.m ~ dunif(0,5)                       # vague prior for between-trial SD
}
```

Absolute probabilities of response can be calculated for any treatment by inputting the estimates for baseline predictive mean and uncertainty from the analysis above (i.e. the posterior mean and variance obtained from monitoring mu.new) into the treatment effects model, as detailed in the Appendix to TSD2.[1]

Alternative prior distributions can be used for the baseline random effects variance (see TSD2,[1] Section 6.2, for a discussion of prior distributions). For example, the last two lines of code in Program 1 can be replaced by a vague Gamma prior on the precision parameter, which is sometimes also referred to as a vague inverse Gamma prior on the variance:

```
tau.m ~ dgamma(0.001,0.001)
sd.m <- sqrt(var.m)
```

Additional code can be added before the closing brace to estimate the probabilities of response on the baseline treatment, based on the posterior (R) or predictive (R.new) distributions of the mean baseline log-odds of response.

```
logit(R) <- m                    # posterior probability of response
logit(R.new) <- mu.new           # predictive probability of response
```

The data structure has two components: a list specifying the number of studies ns and the main body of data in vector format, in the order r[] then n[], the numerators and denominators for all of the trial arms containing the baseline treatment. Both data components need to be loaded into WinBUGS for the program to run.

```
# Data (Smoking Cessation: baseline arms only)
list(ns=19)  # ns=number of studies

r[]     n[]     #       Study ID
9       140     #       1
75      731     #       3
2       106     #       4
58      549     #       5
0       33      #       6
3       100     #       7
1       31      #       8
6       39      #       9
79      702     #       10
18      671     #       11
64      642     #       12
5       62      #       13
20      234     #       14
0       20      #       15
8       116     #       16
95      1107    #       17
15      187     #       18
78      584     #       19
69      1177    #       20
END

# Initial values
#chain 1
list(mu=c(0,0,0,0,0,   0,0,0,0,0,   0,0,0,0,0,   0,0,0,0), sd.m=1, m=0)
#chain 2
```

```
list(mu = c(-1,-1,-1,-1,-1,   -1,-1,-1,-1,-1,   -1,-1,-1,-1,-1,   -1,-1,-1,-1,-1), sd.m=2, m= -1)
#chain 3
list(mu = c(1,1,1,1,1,    1,1,1,1,1,    1,1,1,1,1,    1,1,1,1), sd.m = 0.5, m = 1)
```

PROGRAM 2. SMOKING CESSATION: BINOMIAL LIKELIHOOD, SIMULTANEOUS BASELINE AND TREATMENT EFFECTS RE MODEL WITH PREDICTIVE DISTRIBUTION

This code implements the simultaneous modelling of baseline and treatment effects described in Section 2.2.2. Inclusion of a model for the baseline effect has a strong impact on the posterior distributions of the relative treatment effect. Therefore, we ***do not*** recommend this model unless under very special circumstances, such as those discussed in Section 2.2.2. Use of this model should be justified in detail.

```
# Binomial likelihood, logit link
# Simultaneous baseline and treat effects model for multi-arm trials
model{                                          # *** PROGRAM STARTS
for(i in 1:ns){                                 # LOOP THROUGH STUDIES
   w[i,1] <- 0                                  # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0                              # treatment effect is zero for control arm
   mu[i] ~ dnorm(m,tau.m)                       # model for trial baselines re treatment 1
   for (k in 1:na[i]) {                         # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])              # binomial likelihood
      logit(p[i,k]) <- mu[i] + delta[i,k]       # model for linear predictor
      rhat[i,k] <- p[i,k] * n[i,k]              # expected value of the numerators
      dev.NA[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution including NAs
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
      dev[i,k] <- dev.NA[i,k]*(1-equals(n[i,1],1))   #Deviance contribution with correction for NAs
     }
   resdev[i] <- sum(dev[i,1:na[i]])             #  summed residual deviance contribution for this trial
   for (k in 2:na[i]) {                         # LOOP THROUGH ARMS
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])     # trial-specific LOR distributions
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]   # mean of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k               # precision of LOR distributions (with multi-arm trial correction)
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])   # adjustment for multi-arm RCTs
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)          # cumulative adjustment for multi-arm trials
     }
 }
totresdev <- sum(resdev[])                      # Total Residual Deviance
d[1]<-0                                         # treatment effect is zero for reference treatment
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }       # vague priors for treatment effects
sd ~ dunif(0,5)                                 # vague prior for between-trial SD
tau <- pow(sd,-2)                               # between-trial precision = (1/between-trial variance)
mu.new ~ dnorm(m,tau.m)                         # predictive dist. for baseline (log-odds)
m ~ dnorm(0,.0001)                              # vague prior for mean (baseline model)
var.m <- 1/tau.m                                # between-trial variance (baseline model)
tau.m <- pow(sd.m,-2)                           # between-trial precision = (1/between-trial variance)
sd.m ~ dunif(0,5)                               # vague prior for between-trial SD (baseline model)
}                                               # *** PROGRAM ENDS
```

Alternative prior distributions can be used for the baseline random effects variance as before.

Additional code can be added before the closing brace to produce estimates of absolute effects of each treatment based on the posterior or predictive distributions of the mean baseline log-odds of response for treatment 1 (the baseline/reference treatment).

```
# Provide estimates of treatment effects T[k] on the natural (probability) scale based on posterior distr of baseline model
# and T.new[k] based on predictive distr of baseline model
for (k in 1:nt) {
   logit(T[k]) <- m + d[k]
   logit(T.new[k]) <- mu.new + d[k]
 }
```

The data structure is similar to that presented in TSD2.[1] Briefly, `ns` is the number of studies in which the model is to be based, `nt` is the number of treatments, and in the main body of data `r[,1]` and `n[,1]` are the numerators and denominators for the first treatment; `r[,2]` and `n[,2]`, the numerators and denominators for the second listed treatment; `r[,3]` and `n[,3]`, the numerators and denominators for the third listed treatment; `t[,1]`, `t[,2]` and `t[,3]` are the treatments being compared in the trial arms, and `na[]` gives the number of arms in the trial. Text is included after the hash symbol (`#`) for ease of reference to the original data source.

No Contact was chosen as the baseline/reference treatment because it was the current practice. However, in this example some trials do not include the baseline treatment 1 (trials 2 and 21 to 24 in the data list below). To ensure that the model is put on the correct baseline parameter `mu`, an extra arm containing treatment 1 was added to these trials, with `r[,1]=NA` and `n[,1]=1` and the number of arms in the trial amended accordingly.

```
# Data (Smoking Cessation)
# nt=no. treatments, ns=no. studies
list(nt=4,ns=24 )
```

r[,1]	n[,1]	r[,2]	n[,2]	r[,3]	n[,3]	r[,4]	n[,4]	t[,1]	t[,2]	t[,3]	t[,4]	na[]	#	ID
9	140	23	140	10	138	NA	NA	1	3	4	NA	3	#	1
NA	1	11	78	12	85	29	170	1	2	3	4	4	#	2
75	731	363	714	NA	NA	NA	NA	1	3	NA	NA	2	#	3
2	106	9	205	NA	NA	NA	NA	1	3	NA	NA	2	#	4
58	549	237	1561	NA	NA	NA	NA	1	3	NA	NA	2	#	5
0	33	9	48	NA	NA	NA	NA	1	3	NA	NA	2	#	6
3	100	31	98	NA	NA	NA	NA	1	3	NA	NA	2	#	7
1	31	26	95	NA	NA	NA	NA	1	3	NA	NA	2	#	8
6	39	17	77	NA	NA	NA	NA	1	3	NA	NA	2	#	9
79	702	77	694	NA	NA	NA	NA	1	2	NA	NA	2	#	10
18	671	21	535	NA	NA	NA	NA	1	2	NA	NA	2	#	11
64	642	107	761	NA	NA	NA	NA	1	3	NA	NA	2	#	12
5	62	8	90	NA	NA	NA	NA	1	3	NA	NA	2	#	13
20	234	34	237	NA	NA	NA	NA	1	3	NA	NA	2	#	14
0	20	9	20	NA	NA	NA	NA	1	4	NA	NA	2	#	15
8	116	19	149	NA	NA	NA	NA	1	2	NA	NA	2	#	16
95	1107	143	1031	NA	NA	NA	NA	1	3	NA	NA	2	#	17
15	187	36	504	NA	NA	NA	NA	1	3	NA	NA	2	#	18
78	584	73	675	NA	NA	NA	NA	1	3	NA	NA	2	#	19

```
69      1177    54      888     NA      NA      NA      NA      1       3       NA      NA      2       #       20
NA      1       20      49      16      43      NA      NA      1       2       3       NA      3       #       21
NA      1       7       66      32      127     NA      NA      1       2       4       NA      3       #       22
NA      1       12      76      20      74      NA      NA      1       3       4       NA      3       #       23
NA      1       9       55      3       26      NA      NA      1       3       4       NA      3       #       24
END


# Initial values
#chain 1
list(sd=1, m=0, sd.m=1, d=c(NA,0,0,0), mu.new=0, mu=c(1,1,1,1,1,   1,1,1,1,1,   1,1,1,1,1,   1,1,1,1,1,   1,1,1,1) )
#chain 2
list(sd=1.5, m=2, sd.m=2, d=c(NA,2,1,2), mu.new=1, mu=c(-1,1,-1,1,-1,   2,1,-2,1,2,   1,1,2,1,-2,   1,2,1,-2,1,   1,2,1,2) )
#chain 3
list(sd=3, m=.5, sd.m=.5, d=c(NA,-2,5,-5), mu.new=-1, mu=c(-1,5,-3,1,-1,   5,1,2,3,2,   1,5,2,1,-5,   1,2,-5,-3,1,   5,2,1,-5) )
```