STATISTICAL AND GRAPHICAL EVIDENCE SYNTHESIS METHODS IN HEALTH TECHNOLOGY ASSESSMENT

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Sze Huey Tan BEng MSc
Department of Health Sciences
University of Leicester

September 2015

STATISTICAL AND GRAPHICAL EVIDENCE SYNTHESIS METHODS IN HEALTH TECHNOLOGY ASSESSMENT

Sze Huey Tan BEng MSc

Abstract

This thesis focusses on the challenges relating to clinical- and cost-effectiveness analysis in Health Technology Assessment (HTA). It includes methodological developments, both statistical and presentational, in evidence synthesis aiming to address those challenges.

In HTA, analysts often face problems with limited availability of data required to inform economic model. This thesis proposes innovative evidence synthesis approaches to address this challenge, illustrated in two examples. Bivariate random-effects meta-analysis (BRMA) and network meta-analysis (NMA) were used to synthesise all available evidence to predict progression-free survival (PFS), in metastatic prostate cancer. This enabled the specification of a three-state Markov model previously limited to two states when PFS was not recorded. In the second example, a scenario in multiple sclerosis is considered where utility data for the trials included in a HTA were not available and external utility data from a single study was used instead. This thesis illustrates how BRMA can be applied to include all available evidence to inform utility estimates for use in a cost-effectiveness analysis.

NMA, allowing for a simultaneous and coherent comparison of multiple interventions, is increasingly used in HTA. However, due to the inherent complexity of presenting NMA results, it is important to ease their interpretability. A review of existing methods of presenting NMA results in HTA reports revealed that there is no standardised presentational tool for their reporting. Novel presentational approaches were developed which are presented in this thesis.

The original contributions of this thesis are the innovative approaches to incorporate historical data to predict and increase the precision of parameter estimates for cost-effectiveness analysis to better inform health policy decision-making; and three novel graphical tools to aid clear presentation and facilitate interpretation of NMA results. Ultimately, the hope is that the graphical tools developed will be recommended in updated guidance setting the standards for future HTAs.

ACKNOWLEDGEMENTS

Foremost, I will like to express my sincere gratitude to my primary supervisor, Dr Sylwia Bujkiewicz, for the continuous guidance, support and encouragement over the entire duration of my PhD study. My special thanks also go to my other supervisors, Professors Keith Abrams, Alex Sutton and Nicola Cooper for devoting a substantial amount of their time to guiding me. It is my great fortunate to be guided and supported throughout the duration of this research by very talented and committed supervisors, who also provided me the opportunities to broaden my network and horizons by attending meetings and conferences with them.

I would like to thank everyone in the Department of Health Sciences and in particular, Professors Nuala Sheehan and David Jones for their insightful comments and encouragement at various stages of my PhD, and my fellow mates in the department for making my three years in the UK an enjoyable and enriching experience. I am also grateful to Professor David Machin for his encouragement and advice during my study and for patiently proof-reading chapters of my thesis.

Last but not the least, I would like to thank my parents, sister and friends for being there always to support and encourage me during the course of my PhD study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BRMA	Bivariate random-effects meta-analysis
CEAC	Cost-effectiveness acceptability curve
CI	Confidence interval
CrI	Credible interval
DIC	Deviance information criterion
EDSS	Expanded Disability Status Scale
EQ-5D	EuroQol 5-Dimensions
GRADE	Grading of Recommendations Assessment Development and Evaluation
HR	Hazard ratio
HRQoL	Health related quality of life
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio
ICMA	Indirect comparison meta-analysis
IPD	Individual patient data
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MCMC	Markov Chain Monte Carlo
mHRPC	metastatic hormone-refractory prostate cancer
MS	Multiple sclerosis
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NPV	Net present value
OR	Odds ratio
OS	Overall survival
PFS	Progression-free survival
PI	Prediction interval
PWMA	Pair-wise meta-analysis
QALY	Quality-adjusted life-year
RCT	Randomised controlled trial
SD	Standard deviation
SE	Standard error
SFP	Summary forest plot
STA	Single technology appraisal
TTP	Time to progression
UK	United Kingdom
URMA	Univariate random-effects meta-analysis

1 Introduction

1.1 Aims of the thesis

In conducting Health Technology Assessment (HTA) for health policy decision making it is necessary to collate information on the clinical- and cost-effectiveness of the technologies of interest. Information should ideally be comprehensive and obtained from all relevant, well-managed and documented data sources. This can be from clinical trials, observational studies, expert opinion, bench research and/or secondary analyses (such as meta-analysis) by extracting data from any of these sources. Hence, synthesis methods that enable direct and/or indirect comparisons of competing treatment or intervention technologies and jointly amalgamate evidence from multiple outcomes are required to quantify how the different technologies compare in terms of their effectiveness.

The aims of this thesis were to:

1. Review and identify methodological and presentational issues related to evidence synthesis and economic modelling in HTA;
2. Formulate recommendations on how to approach the identified issues; and
3. Develop methods to address the issues using motivating case studies from published HTA reports

The background to these three areas of methodological and presentational challenges related to evidence synthesis and economic modelling in HTA are discussed in Sections 1.1, 1.3 and 1.4 and the structure of this thesis is outlined in Section 1.5.

1.2 Presentational challenges for evidence synthesis in HTA

Evidence synthesis of clinical effectiveness of treatment interventions has largely been evaluated using standard pairwise meta-analysis of all studies that compared the same (two) options of interest (Borenstein et al., 2009, DerSimonian and Kacker, 2007). Although this method may be useful for judgement in situations where the new intervention is to be compared with either placebo or standard care, it is limited in

scope. Thus, decision makers are often interested in assessing the comparative effectiveness of more than two competitor interventions, in settings where randomised controlled trials (RCTs) comparing all technologies of interest may not be available; and even when head-to-head RCTs do exist, there is increasing interest to incorporate data from other related studies to further inform the effectiveness estimate of the comparisons.

A methodology to address this issue, which has increasingly been applied, is network meta-analysis (NMA) (Ades, 2003, Ades et al., 2006, Caldwell et al., 2005, Higgins and Whitehead, 1996, Lu and Ades, 2004, Lumley, 2002). However, NMA is a complex statistical method, the results from which need to be made accessible to non-statistical experts, including policy-makers, in order to maximize their usefulness in HTA.

Although the UK National Institute for Health and Care Excellence (NICE) guidance states that all data used for estimates of effectiveness should be presented in tabular form with the source of the data clearly stated (NICE, 2004, NICE, 2008, NICE, 2013) and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force provided detailed guidance for the conduct of NMA analyses through a series of good practice documents (Hoaglin et al., 2011, Jansen et al., 2011), specific details and recommendations on presentational formats, particularly of the data and results, are limited. It is therefore the intent of this thesis to develop novel graphical tools to facilitate clear, consistent and transparent reporting of NMA in HTA for health economic decision making. A review of the existing guidelines and current practice on the presentation of NMA results together with the new graphical tools developed for improved presentation are included in Chapter 3.

1.3 Methodological issues concerning economic modelling in HTA

Health policy decision makers are constantly faced with the challenge of how best to allocate resources for a wide range of health interventions within their limited budgetary constraints. Hence, there is a need to take into account the clinical and economic considerations of the health interventions for different disease conditions as a whole when making decisions about which health interventions to fund. This is to

ensure that scarce resources are efficiently allocated to achieve the maximum healthcare gains for the public.

Decision-analytical modelling to estimate the cost-effectiveness of health interventions, which compares costs and effects of two or more interventions, is widely used to guide the prioritisation of scarce healthcare resources. The Markov model is one that is commonly used in economic modelling due to its flexibility in allowing the disease pathway of patients to be represented using distinct states. It is therefore imperative that the disease progression of patients is appropriately represented using multiple states in the model as inappropriate modelling of the natural pathway of the disease may produce inappropriate cost-effectiveness results. However, data for the full set of parameters required to specify a multi-state model for cost-effectiveness evaluation are not always available.

The motivating case study for this section was a 2007 NICE HTA report on the treatments for metastatic hormone-refractory prostate cancer (Collins et al., 2007). This report relied on a two-state (stable disease and death) Markov model due to the availability of sufficient data only on overall survival. The thesis seeks to illustrate how evidence synthesis methods, incorporating historical data describing the relationship of clinical endpoints, enabled the specification of a three-state (stable disease, progression and death) Markov model to better inform the cost-effectiveness analysis for health policy decision-making in this context. The background to the 2007 NICE HTA report and the methodologies applied to incorporate historical data using evidence synthesis under a Bayesian framework are described in Chapter 4.

1.4 Methodological issues concerning utility used in cost-effectiveness analysis in HTA

Cost-effectiveness analysis for health economic decision making typically requires the estimating of resource costs and health effects resulting from the use of one or more new interventions compared to the management of the disease under the usual standard of care. Resources consumed often include the costs of, for example, drug acquisition and administration, inpatient clinic or hospitalisation care, and outpatient follow up.

The health effects are obtained from either disease-specific or generic measures by means of questionnaire instruments.

Disease-specific measures, as the name implies, are specific to a particular disease or health condition and examples of these are presented in Chapter 5. Generic measures (also referred to as utilities) are non-disease specific and can be used across most diseases. An example is the EuroQol 5-Dimensions (EQ-5D) Questionnaire (Rabin and de Charro, 2001), which is the recommended health related quality of life (HRQoL) instrument for cost-effectiveness analyses by NICE. However, this utility measure is not always reported in RCTs and often mapping techniques are used to estimate EQ-5D from disease-specific measures.

A 2011 NICE single technology appraisal (STA) of Fingolimod for the treatment of relapsing remitting multiple sclerosis (MS) in adults (Asaria et al., 2011, Novartis Pharmaceuticals UK Ltd, 2011) used a mapping model developed by Orme and colleagues (Orme et al., 2007) to estimate the EQ-5D utility values from the data collected on the Expanded Disability Status Scale (EDSS) scores from two RCTs. Using this STA as a motivating case study, this thesis aims to illustrate how advanced meta-analytic methods could be used to include all available evidence on the relationship between disease-specific and generic measure of health effects to inform mapping models used in cost-effectiveness analysis. Background on the NICE STA and Bayesian evidence synthesis techniques used for the estimation of health utility values (EQ-5D) from EDSS scores in patients with MS are described in Chapter 5.

1.5 Structure of the thesis

This thesis is structured into six chapters, where the first chapter presents the objectives and background to the methodological and presentational issues in HTA that motivate the need for this PhD project.

Chapter 2 introduces the fundamental concept and methodologies of Bayesian statistics, evidence synthesis and economic evaluation that are applied to the analyses performed. All modelling in Chapter 3 to Chapter 5 is conducted under the Bayesian framework, which is particularly useful for performing clinical- and cost-effectiveness analyses in a

coherent fashion as it allows the incorporation of historical data into the evidence synthesis and hence inform health policy decision making.

Chapter 3 highlights the presentational issues and challenges in reporting NMA results in HTA reports. It begins with a review of existing guidelines for the presentation of NMA and by reviewing published HTA reports, what has been done in practice in the UK. Based on the results of the reviews, recommendations to improve reporting and novel graphical tools developed to aid clear presentation and facilitate interpretation of NMA results in HTA are presented.

Chapter 4 discusses the many challenges in the economic modelling of cancer treatments. It presents, with the use of a motivating HTA report, a novel method of using Bayesian evidence synthesis techniques to predict clinical effectiveness of cancer treatments for the specification of multi-state Markov model when the full set of parameters required for the multi-state economic model were not available. The network of RCTs and historical data for the construction of prior distribution for the estimation of the clinical endpoints using Bayesian evidence synthesis techniques are presented.

Chapter 5 demonstrates how advanced Bayesian meta-analytic methods are able to include all available evidence in estimating EQ-5D from disease specific measures beyond the data from a single study were used to inform estimates of utility required in cost-effectiveness analysis. A motivating example using a NICE STA of Fingolimod for the treatment of relapsing remitting MS in adults was used as the case study for presenting this novel approach. Search strategy and selection criteria of studies for constructing the prior distributions for the coefficients in linear regression models estimating EQ-5D using EDSS was developed.

Finally, Chapter 6 summaries the conclusions of the preceding chapters and discusses how this thesis has applied existing evidence synthesis techniques to address modern-day challenges in the analysis and reporting of clinical- and cost-effectiveness in HTA. Strength and limitations in the application of statistical methodologies encountered throughout this thesis and opportunities for further work are also included.

2 Statistical Methodology

2.1 Chapter Overview

This chapter gives an introduction to the statistical methods and software used in this thesis. Basic statistical sampling distributions and concept of Bayesian statistics which are applied throughout this thesis are presented. Software which facilitate Bayesian analysis performed in this thesis are also introduced. Methodologies for evidence synthesis, comprising of standard pairwise meta-analysis (PWMA), network meta-analysis (NMA) and bivariate random-effects meta-analysis (BRMA) are covered. These methodologies are for the synthesis of aggregate summary data across trials as synthesis of individual patient data (IPD) from different trials is not considered in this thesis. Specification of economic models for cost-effectiveness analysis and the terminologies used for the interpretation of results and evaluation of health technologies are also described.

2.2 Bayesian Methods

2.2.1 Bayesian statistics

Its origin dates back to 1763 when work by Thomas Bayes, an English minister, was published posthumously. This work is today known as the Bayes' Theorem. In Bayesian analysis, if θ is the unknown parameter of interest and Y represents a piece of data that describes θ, Bayes' theorem takes the form:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)\,d\theta}$$

and is commonly expressed as:

$$p(\theta|Y) \propto p(Y|\theta)p(\theta)$$

where $p(Y|\theta)$ is the likelihood of θ, $p(\theta)$ is the probability density for the degree of belief (known in the Bayesian context as the prior probability) for θ and $p(\theta|Y)$ is the resulting probability density (known as the posterior probability) for θ after

supplementing the likelihood with the prior. Thus, Bayes' theorem can also be understood as:

$$Posterior \propto Prior\ X\ Likelihood$$

Here, probability statements are made about the unknown parameter, θ. This is different from the classical or frequentist approach to statistics where θ is said to be a fixed but unknown number. In Bayesian statistics, there is no difference between a fixed but unknown number and a random number, so θ is treated as a random parameter in an analysis and hence probability distributions (based on subjective belief or external information) about its value can be defined.

The Prior, Likelihood and Posterior

A defining characteristic of Bayesian statistics is its capability to incorporate external information or subjective beliefs in the form of a prior distribution for the unknown parameter of interest, θ. Prior distributions can take two forms: informative or non-informative. Informative prior distribution is often constructed based on external information such as historical data or subjective beliefs about θ elicited from 'experts' with knowledge about the plausible range of values that θ can take. In contrast to this, non-informative prior distribution doesn't express any prior belief or information and should ideally have large variance such that all possible parameter values for θ are almost equally likely.

As the posterior distribution is a function of the likelihood and prior distribution, the prior distribution will have some degree of influence on the posterior, depending on the type of prior and the precision of the data. A non-informative prior should ideally exert no or minimal influence on the posterior as shown in Figure 2.1. Hence, the posterior essentially represents the likelihood but with minimal increase in precision.

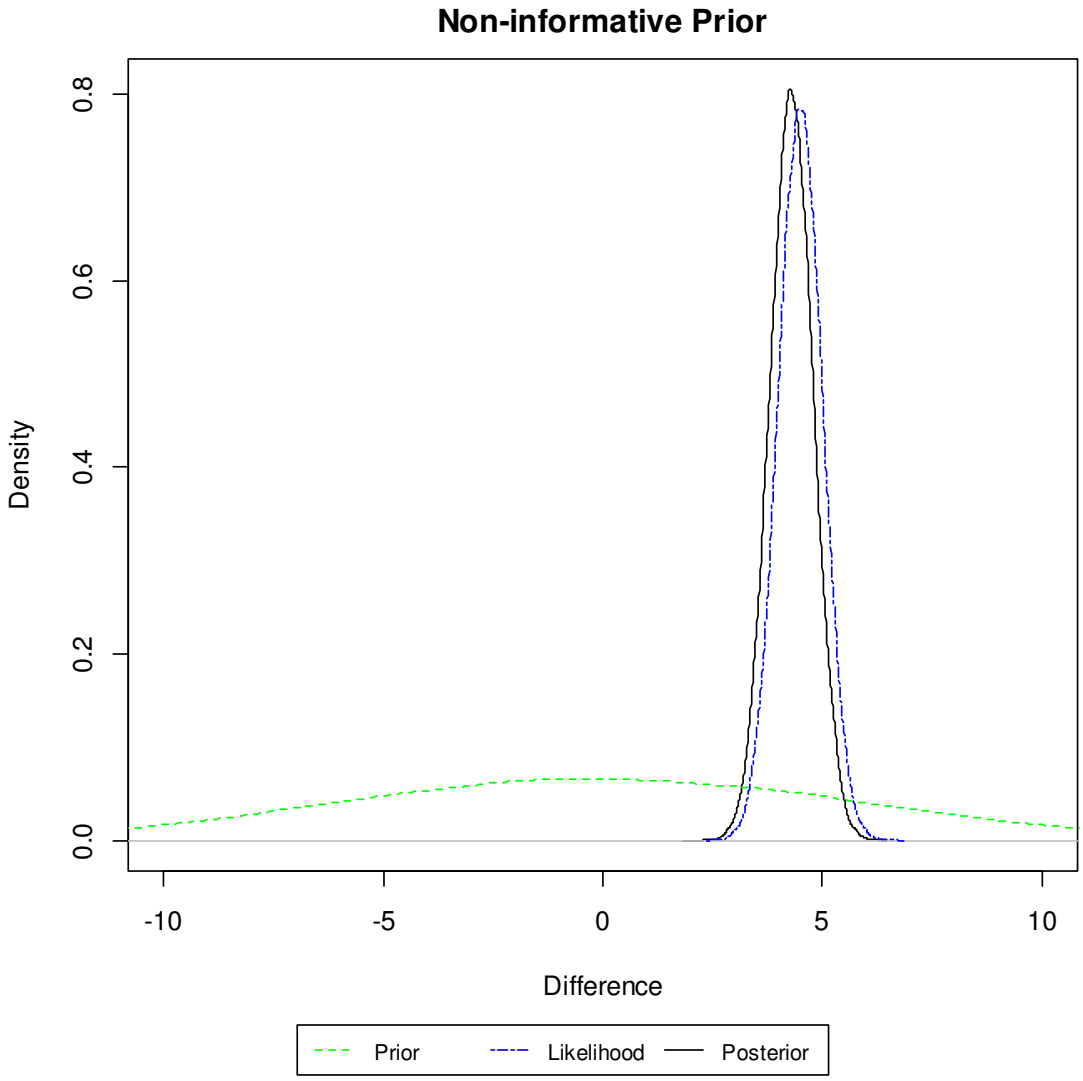

Figure 2.1: Prior, likelihood and posterior distributions where the prior is non-informative

The degree of influence that an informative prior has on the posterior depends on the precision and location of the prior relative to that of the data. Different informative prior distributions and their corresponding influences on the posterior distributions are presented in Figure 2.2.

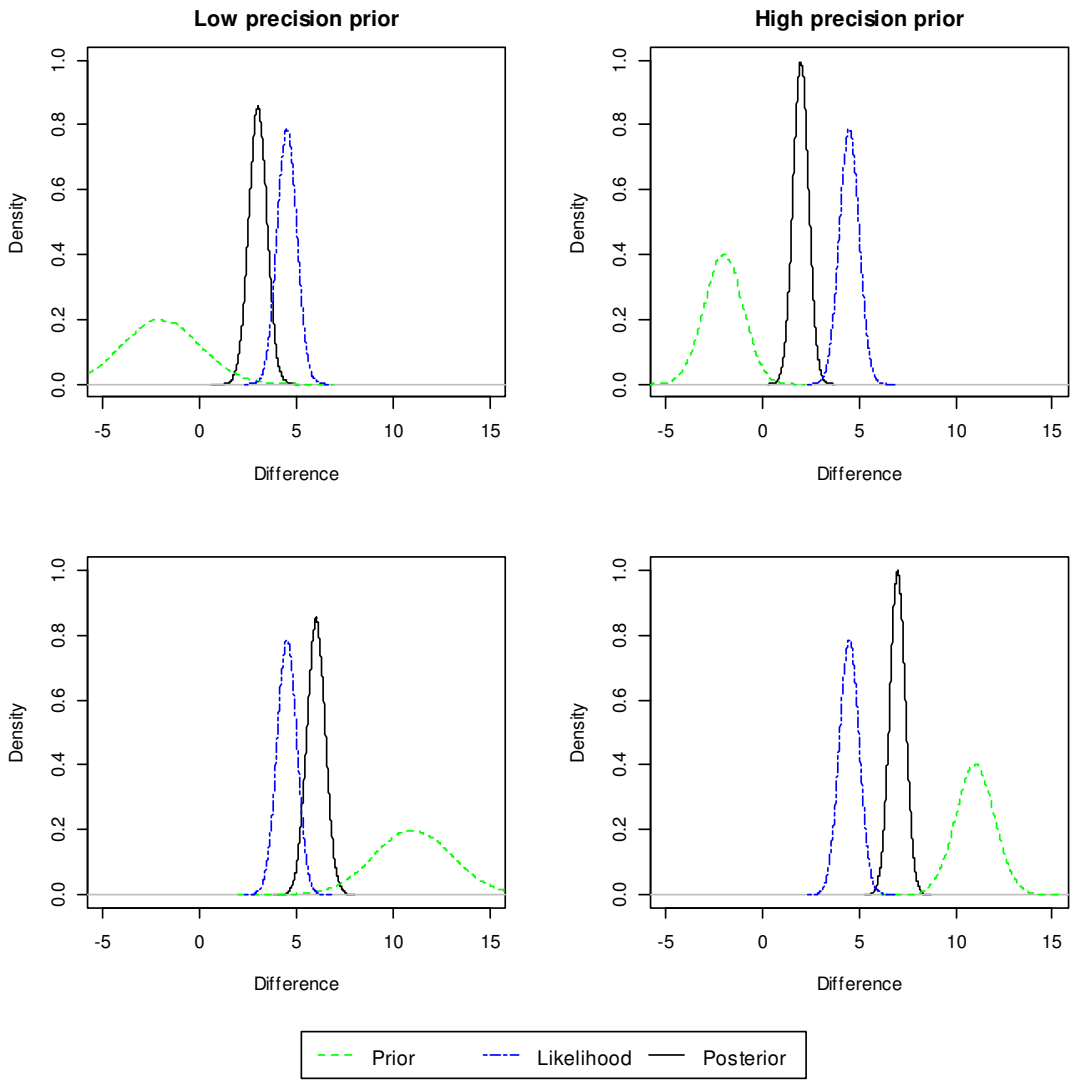

Figure 2.2: Prior, likelihood and posterior distributions where the priors are informative

When conducting any Bayesian analysis, sensitivity analysis should be performed to investigate the influence that the prior distribution has on the posterior distribution. Prior distributions that are either non-informative or informative can be applied to the analysis to assess how the posterior distributions differ between the priors. Informative priors can be either optimistic or pessimistic priors. Optimistic and pessimistic informative prior distributions are defined based on the context of the analysis involved and display positive and negative evidence/beliefs about the effectiveness of the outcome measure respectively. Usage and definition of optimistic and pessimistic prior distributions in Bayesian analysis are illustrated in Chapter 4. Non-informative prior

distributions specified in this thesis are defined using the statistical distributions described in Section 2.2.2.

Advantages and disadvantages of Bayesian statistics

A major deterrent for the use of Bayesian statistics by many is the need to specify appropriate priors for the analysis. This can be challenging as there may be no useful historical data that can be utilised to define the informative prior distribution, and where it is available, there are no standardised guidelines for the appropriate construction of prior distributions. Although non-informative prior distributions can be utilised where no informative prior distribution can be constructed, defining an appropriate "non-informative prior distribution can be difficult in itself as it has been shown that the choice of "non-informative" prior distributions can also result in differential results, in particular for variance parameters (Lambert et al., 2005). Statistical distributions used for defining prior distributions are discussed in Section 2.2.2.

Where useful historical data from a number of sources exist and there may be a need to collate the information for the construction of an informative prior distribution, Bayesian meta-analysis can be used to synthesise the data. The posterior distribution from the meta-analysis can be directly utilised to form the informative prior distribution for the primary analysis. In the light of new data, the new data can now be treated as the likelihood and the posterior distribution from the meta-analysis can then be used as the prior distribution for the primary analysis. This sequential use of the Bayes theorem is one of the major advantages of Bayesian statistics.

The Bayesian approach to analysis is utilised in this thesis in a number of ways. Firstly, Bayesian network meta-analyses using non-informative prior distributions were performed as illustrated in Chapter 3 on the development of novel graphical tools for the reporting of NMA results. Secondly, bivariate meta-analyses under the Bayesian framework were utilised for prediction of outcomes (when not reported) as illustrated in Chapter 4 and for the construction of prior distributions for subsequent Bayesian analyses as illustrated in Chapter 5. Lastly, Bayesian multivariable linear regression analyses were conducted using informative prior distributions constructed using the

Bayesian bivariate meta-analysis and this sequential use of Bayes is illustrated in Chapter 5.

2.2.2 Sampling distributions

A number of sampling distributions are utilised in this thesis for the construction of the prior distributions and likelihood functions. Prior distributions specified include the uniform, normal, half-normal, log-normal, beta and gamma distributions. Distributions utilised in the simulation of data to form the likelihood include the binomial, multinomial, log-normal and beta-distributions.

Uniform distribution

A uniform distribution is characterised by its constant probability over a range of values (a, b). It is commonly adopted for an unknown parameter $Y \sim Uniform(a, b)$ to indicate that Y has an equal probability of taking any value from a to b. The distribution for Y has the following statistical properties:

$$p(y|a, b) = \frac{1}{b - a}; y \in (a, b)$$

$$E(Y|a, b) = \frac{a + b}{2}$$

$$V(Y|a, b) = \frac{(b - a)^2}{12}$$

The uniform distribution is generally used for the specification of prior distributions in Bayesian analyses to convey indifference about the prior probability over a range of plausible values. Applications of its use include the specification of non-informative prior distributions for standard deviation and correlation parameters in this thesis.

Binomial distribution

The binomial distribution is a discrete probability distribution of the number of successes in n independent experiments, each of which has a probability of success of θ and a probability of failure of $1-\theta$. Each of the experiments which can only be a success or a failure is defined as a Bernoulli trial. The likelihood $\theta^y(1-\theta)^{n-y}$ gives the probability of having y successes and $n-y$ failures. However, as the y successes can occur anywhere among the n trials, there are $\binom{n}{y}$ different ways of distributing y successes in a sequence of n trials. Hence, a discrete binomial variable Y, denoted as $Y \sim Binomial(n, \theta)$ represents a binomial distribution with statistical properties:

$$p(y|n,\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}; \quad y = 0,1,\dots,n$$

$$E(Y|n,\theta) = n\theta$$

$$V(Y|n,\theta) = n\theta(1-\theta)$$

When $n = 1$, the binomial distribution is effectively a Bernoulli distribution, which is denoted as $Y \sim Bern(\theta)$. In this thesis, the binomial distribution is used as a sampling distribution for simulating data that occur as proportions, such as gender in patient demographics.

Multinomial distribution

The multinomial distribution is a generalization of the binomial distribution, where it is used to compute the probabilities in situations where there are *k* possible outcomes. For n independent trials, each of which leads to a success for exactly one of the k possible outcomes where each outcome has a fixed probability of success, θ_i $(i = 1, 2, \dots, k)$, and all the probabilities $(\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k))$ sum to 1 $\left(\sum_{i=1}^{k}\theta_i = 1\right)$. When *k*=2, it reduces to the binomial distribution. Now, let Y_i denote the number of times outcome i is observed in the n independent trials and the vector of $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_k)$ therefore follows a multinomial distribution, denoted as $\boldsymbol{Y} \sim Multinomial(n, \boldsymbol{\theta})$ and has the following statistical properties:

$$p(y|n,\theta) = \frac{n!}{y_1! \dots y_k!}\theta_1^{y_1} \dots \theta_k^{y_k}$$

$$E(Y_i|n,\theta_i) = n\theta_i; \quad i = 1, \dots, k$$

$$V(Y_i|n,\theta_i) = n\theta_i(1-\theta_i)$$

$$Cov(Y_i; Y_j) = -n\theta_i\theta_j; \;\; i \neq j$$

While the n trials are independent, the Y_i are dependent on one another as they must sum to n. Hence, there is a covariance term for each pair of Y_i.

Normal distribution

The normal distribution is the most commonly used continuous probability distribution in statistics for real-valued parameters whose distribution is unknown. This distribution has a bell-shape with its peak at the mean value, μ, and describes the probability of any real observation that lies between any two real limits (numbers). The spread of the distribution around the mean value is determined by the variance parameter, σ^2. The larger the value of the variance, the further apart the two real limits get. Hence, $Y \sim Normal(\mu, \sigma^2)$ represents a normal distribution with statistical properties:

$$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-\mu)^2}{2\sigma^2}}; \quad y \in \mathbb{R}$$

$$E(Y|\mu,\sigma^2) = \mu$$

$$V(Y|\mu,\sigma^2) = \sigma^2$$

It is useful as a prior distribution for any real-valued parameters and non-informative prior distributions are easily specified using this distribution by having the mean to be zero and the variance to be large.

Half-normal distribution

The half-normal distribution, as the name implies, represents a distribution created by truncating the normal distribution at zero. Effectively, if $X \sim Normal(0, \sigma^2)$, then

$|X| \sim HNormal(\sigma^2)$ and takes only positive values. Therefore the half-normal distribution is useful for specifying prior distribution with support for positive values, with σ controlling the upper range of support. As with the normal distribution, the higher the values of σ, the more non-informative the prior distribution gets. This distribution is often used to specify the prior distribution for standard deviations. The half-normal distribution for a random variable Y, denoted as $Y \sim HNormal(\sigma^2)$ has these statistical properties:

$$p(y|\sigma^2) = \sqrt{\frac{2}{\pi\sigma^2}}\, e^{\frac{-y^2}{2\sigma^2}}; \quad y \in (0, \infty)$$

$$E(Y|\sigma^2) = \sqrt{\frac{2}{\pi}}\, \sigma$$

$$V(Y|\sigma^2) = \sigma^2 \left(1 - \frac{2}{\pi}\right)$$

Log-normal distribution

The log-normal distribution is a distribution that takes only positive real values similarly as the half-normal distribution and is also a variant of the normal distribution. It is a continuous probability distribution of a random variable Y, whose logarithm is normally distributed. Hence, if $Y \sim LN(\mu, \sigma^2)$, then $\log(Y) \sim Normal(\mu, \sigma^2)$, and $Y \sim LN(\mu, \sigma^2)$ represents a distribution with these properties:

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma y} e^{\frac{-[\log(y)-\mu]^2}{2\sigma^2}}; \quad y \in (0, \infty)$$

$$E(Y|\mu, \sigma^2) = e^{\mu + \frac{\sigma^2}{2}}$$

$$V(Y|\mu, \sigma^2) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

This distribution is useful as a sampling distribution for parameters that take positive values, such as cost and ratios (such as odds ratio and hazard ratio).

Beta distribution

The beta distribution describes a family of continuous probability distributions constrained to lie between 0 and 1 and parameterised by two positive shape parameters (a and b). The shape parameters allow flexible specifications of different probability distributions between 0 and 1; hence making the beta distribution useful as a prior distribution for unknown proportions. $Y \sim Beta(a, b)$ represents a beta distribution with properties:

$$p(y|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{a-1}; \quad y \in (0,1)$$

$$E(Y|a,b) = \frac{a}{a+b}$$

$$V(Y|a,b) = \frac{ab}{(a+b)^2(a+b+1)}$$

where $\Gamma(a)$ represents the gamma function, and $\Gamma(a) = (a-1)!$ if a is an integer.

Gamma distribution

Gamma distributions are useful for quantities constrained to be positive, such as cost data in economic modelling. A random variable Y that follows a gamma distribution, denoted as $Y \sim Gamma(a, b)$, has a distribution with these statistical properties:

$$p(y|a,b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}; \quad y \in (0,\infty)$$

$$E(Y|a,b) = \frac{a}{b}$$

$$V(Y|a,b) = \frac{a}{b^2}$$

Alternative use of the gamma distribution is as a prior distribution for the precision parameter (1/variance) of a normal distribution.

2.3 Software and Computational Issues

2.3.1 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods are a class of algorithms for drawing random samples from a probability distribution by constructing a Markov chain whose 'equilibrium distribution' represents the desired distribution. A Markov Chain is a sequence of random variables $\theta^{(1)}$, $\theta^{(2)}$, $\theta^{(3)}$, … such that the future depends on the past only through the present ($\theta^{(n+1)}|\,\theta^{(n)}$). Under regularity conditions, the distribution of $\theta^{(n)}$ tends to an 'equilibrium distribution' as n tends to ∞, regardless of the starting value $\theta^{(1)}$. A sample ($\theta^{(n)}$) of the desired distribution is therefore obtained when the Markov chain reaches its 'equilibrium' state after a number of steps in the algorithm (often termed as the convergence of the Markov chain).

Due to its simplicity to draw random samples from a desired distribution, it has become an imperative tool for Bayesian analysis, especially when there is no simple analytical solution for the posterior distribution. In complex hierarchical problems, the joint posterior and marginal posterior distributions are often of complex form but the conditional posterior distributions are of simple form. MCMC allows the desired complex posterior distributions to be derived by sampling from the conditional posteriors. An example using the Gibbs sampling is presented in the next paragraph to illustrate this property of MCMC. The Gibbs sampler is utilised here as it is the MCMC method used in the WinBUGS (Bayesian inference Using Gibbs Sampling) software that is used for all the Bayesian analyses in this thesis.

A conditional posterior distribution is the posterior of one parameter given the value of the other parameters and is obtained from the joint posterior distribution by treating the other parameters as fixed. Considering the situation of three parameters θ, γ, φ and data $\boldsymbol{y}$, where the parameter of interest is θ. If the joint posterior distribution is $p(\theta,\gamma,\varphi|\boldsymbol{y})$, the conditional posterior for parameters θ, γ, φ are $p(\theta|\gamma,\varphi,\boldsymbol{y})$, $p(\gamma|\theta,\varphi,\boldsymbol{y})$ and $p(\varphi|\theta,\gamma,\boldsymbol{y})$ respectively. The Gibbs sampler starts with initial values for all the parameters, θ^0, γ^0, φ^0, and a new value for each parameter is generated from the joint posterior conditional on all the other current parameter values, that is, θ^1 is generated from $p(\theta|\gamma^0,\varphi^0,\boldsymbol{y})$. The other parameters are generated in turn, γ^1 from $p(\gamma|\theta^1,\varphi^0,\boldsymbol{y})$ and φ^1 from $p(\varphi|\theta^1,\gamma^1,\boldsymbol{y})$. Subsequent random observations are

generated by repeating the above for a number of iterations. As mentioned previously, the desired distribution will arise when the convergence of the Markov chain is reached. Once this happens, for example at the k-th iteration, a further n iterations are performed to generate random observations for the parameters. This will produce a series of observations, $\theta = (\theta^k, \theta^{k+1}, \theta^{k+2}, \ldots, \theta^{k+n})$, that form the desired posterior distribution for θ.

Markov Chain Monte Carlo analyses can be performed using various software, many of which are "user-designed" macros for commercial software (for example Excel and STATA) to serve the analyses they are intended for. In economic modelling, MCMC analyses are sometimes conducted using macros developed in Excel. However, the WinBUGS software is more widely used for the purpose of MCMC analyses in a variety of applications, such as Bayesian analysis, evidence synthesis and economic modelling. The WinBUGS software is used for the analyses in this thesis and a description of the software is provided in Section 2.3.2.

2.3.2 WinBUGS

WinBUGS (Lunn et al., 2000, Spiegelhalter et al., 2003) is a statistical software developed for Bayesian analysis using MCMC method. It implements the Gibbs sampling MCMC method and requires a specification of the probability model in a straightforward and natural way. The flexibility of this software allows complex modelling to be performed easily. In this thesis, WinBUGS is used extensively for all Bayesian evidence syntheses and economic modelling.

One major drawback of the software is that it assumes that users are Bayesian statisticians and does not provide any 'caution' warnings when inappropriate prior distributions and models are fitted. It is therefore important to use it with considerable care and understand the need to assess the priors and likelihood, fit of the model, assignment of appropriate initial values for the parameters, convergence of the Markov chain and problems of autocorrelation between simulations.

2.3.3 R Software

The WinBUGS software discussed in the Section 2.3.2 allows analysts to perform Bayesian analyses with ease. However, it has limited graphical output facilities which make presentation of results through the use of custom plots very difficult. In this thesis, the open source software R (R Core Team, 2012) is used extensively together with the WinBUGS software for the analysis and presentation of evidence synthesis and economic modelling results.

The R software appears to be a good choice as it has an inbuilt functionality that enable R to connect to WinBUGS. Tapping into the R2WinBUGS (Sturtz et al., 2005) functionality in R, it is possible to (i) exploit R software data management functionality to organise data for analysis in WinBUGS; (ii) use posterior distributions estimated from WinBUGS to construct prior distributions for subsequent analysis easily in one R file; (iii) construct an empirical prior distribution (for a subsequent analysis) using the posterior distribution (from the current analysis) computed in WinBUGS by saving the MCMC simulation data points of the posterior distribution in R; (iv) generate custom plots in R using the results computed from WinBUGS; (v) develop novel presentational graphs using R software plot functions and results from WinBUGS.

2.4 Meta-Analysis

2.4.1 Pairwise meta-analysis

Meta-analysis is a statistical method that combines the results from several independent studies, with the aim to improve the precision of the estimate of a treatment effect over that obtained from individual studies and also to explain heterogeneity between the results of individual studies. Although the first use of meta-analysis dates back to 1904 when the statistician Karl Pearson (Pearson, 1904) combined the results of several studies of typhoid inoculation and published the pooled results in the British Medical Journal, the term "meta-analysis" was coined by Gene V. Glass (Glass, 1976), a statistician, in 1976 to represent the analytic approach used to combine results from multiple independent studies.

Traditionally, meta-analysis has primarily involved the assessment of treatment effect between two interventions for one outcome of interest. With the advent of new meta-analytic methodologies to involve more than two interventions or multiple outcomes of interest (Higgins and Whitehead, 1996), meta-analysis performed in the former context is often termed pairwise meta-analysis (PWMA) in recent years. When it involves the synthesis of evidence from more than two interventions simultaneously, the meta-analytic method designed for this is called network meta-analysis (NMA), which is discussed in Section 2.4.2. An alternative method that estimates the treatment effects between two interventions but for two outcomes of interest simultaneously is bivariate meta-analysis which is described in Section 2.4.3.

2.4.1.1 Fixed-effect and random-effects meta-analysis

There are two types of statistical models commonly used in meta-analysis, namely: the fixed-effect model and the random-effects model (Borenstein et al., 2009). The choice of which model to use depends on how the variability between the results of the studies included in the synthesis is defined. In this section, these two statistical models are explained in the context of PWMA; however, the concept is similar when extended to NMA and bivariate meta-analysis.

In a fixed-effect model, it is assumed that all studies in the PWMA are estimating a common true underlying effect size and all variation observed in the studies only reflects sampling error (also referred to as within-study variability). Algebraically, the fixed-effect model is defined as follows:

$$y_i \sim Normal(d, s_i^2) \quad i = 1, 2, \ldots, N$$

where y_i is the outcome effect size for study i, d is the underlying mean effect size common to all studies and s_i^2 is the within-study variance for study i.

In the field of medicine, the assumption of a common true underlying effect size may not hold true due to differences in patient populations or study locations (Kriston, 2013). In this situation, a random-effects model may be more appropriate than the fixed-effect model. This is because in the random-effects model, the effect sizes of the studies in the meta-analysis are assumed to be sampled from a distribution of true

effects which allow for between-study variability (also known as heterogeneity). The random-effects model (DerSimonian and Laird, 1986) is defined as follows:

$$y_i \sim Normal(\delta_i, s_i^2) \quad i = 1, 2, \ldots, N$$

$$\delta_i \sim Normal(d, \tau^2) \qquad (2.1)$$

where y_i and s_i^2 are the outcome effect size and within-study variance for study i respectively. y_i is an estimate of the true effect size δ_i specific to study i, assumed to be drawn from a normal distribution with overall population mean d and between-study variance τ^2.

2.4.1.2 Heterogeneity

Under the random-effects model, the true effect sizes for studies in the meta-analysis are allowed to vary from study to study. This between-study variability in effect sizes is commonly termed statistical heterogeneity (Borenstein et al., 2009, Sutton et al., 2000). Both the heterogeneity and variability from sampling error (within-study variability) make up the overall variability in observed effect sizes from study to study.

The extent of heterogeneity may be measured in several ways, the estimated τ^2 in equation (2.1) gives the magnitude of the heterogeneity across studies. However, its interpretation is specific to the particular effect size metric measured. Other measures of heterogeneity include the Cochran $Q\ statistic$ and $I^2\ statistic$ (Higgins and Thompson, 2002, Higgins et al., 2003) which are standardised measures not affected by the metric of the effect size index. Cochran $Q\ statistic$ also provides a test of homogeneity of the effect sizes across studies and is sensitive to the number of studies. The $I^2\ statistic$ gives the proportion of total variability in the effect size that is due to heterogeneity, expressed as a ratio from 0% to 100%, and is not sensitive to the number of studies. When I^2 is close to 0%, almost all the variability observed is spurious; however when I^2 is large, most of the variability is due to heterogeneity.

While random-effects meta-analysis can be performed to allow for heterogeneity and the Q and I^2 statistics can be used to assess and quantify heterogeneity, they cannot explain the source of the heterogeneity. If heterogeneity cannot be explained by

differences in the characteristics of the studies such as differences in study design, patient population, treatment duration or drug dose administrated, technique such as subgroup analysis or meta-regression (Thompson and Higgins, 2002) may be used to investigate the source of the heterogeneity.

2.4.1.3 Publication Bias

Another important consideration when performing meta-analysis is publication bias, which exists when the studies identified for the meta-analysis are systematically different from all studies that should have been included. This happens because studies reporting relatively large effect sizes or significant results are more likely to be published than studies reporting smaller effect sizes and non-significant results (Dickersin et al., 1987), possibly leading to an overestimation of the true effect size.

Methods have been developed for assessing, quantifying and adjusting for publication bias. The presence of publication bias may be assessed using the funnel plot (Egger et al., 1997), where the effect size of each study is plotted (on the x-axis) against the inverse of its standard error (on the y-axis). If studies are distributed symmetrically about the mean effect size, then there is no publication bias. When asymmetry of the plot is observed, it may indicate the presence of publication bias or even small-study effect (Sterne and Egger, 2001). To quantify the extent of publication bias observed, various tests have been proposed (Peters et al., 2008, Peters et al., 2010, Rucker et al., 2008). Other methods for adjusting publication bias include the 'trim and fill' method (Duval and Tweedie, 2000) and regression-based methods proposed by Moreno and colleagues (Moreno et al., 2009).

2.4.2 Network meta-analysis

Network meta-analysis is a recent development in evidence synthesis that extends the functionality of standard PWMA to allow for a simultaneous and coherent comparison of multiple interventions using an evidence base of trials that individually may not compare all the treatment options of interest. NMA can take the form of a fixed-effect model or random-effects model, similar in concept to that in PWMA described in

Section 2.4.1.1. However, due to the inclusion of more randomised controlled trials (RCTs) in a NMA compared to a PWMA for the same research outcome of interest, random-effects model is more commonly applied to account for heterogeneity between the RCTs. NMA is the term used to refer to two evidence synthesis techniques: Indirect Treatment Comparisons (IC) and Mixed Treatment Comparisons (MTC), which are described in the next two sections.

2.4.2.1 Indirect Treatment Comparisons

Naïve comparison of individual arms from different RCTs as if they are from the same RCT is inappropriate and should not be used. This is because the advantage of randomisation in the RCTs is completely disregarded, causing the evidence from this naïve comparison approach to be equivalent to that from observational studies and is prone to bias.

Evidence synthesis method that allows for the comparisons of interventions when there is strictly no head-to-head RCT that compared them directly by evaluating the difference between the interventions through at least one common comparator was proposed (Bucher et al., 1997, Lumley, 2002). This model evaluates the relative effectiveness between the two interventions using two sets of RCTs that compare each of the interventions with the common comparator separately; the inclusion of the common comparator preserves the randomization of the originally assigned patient groups.

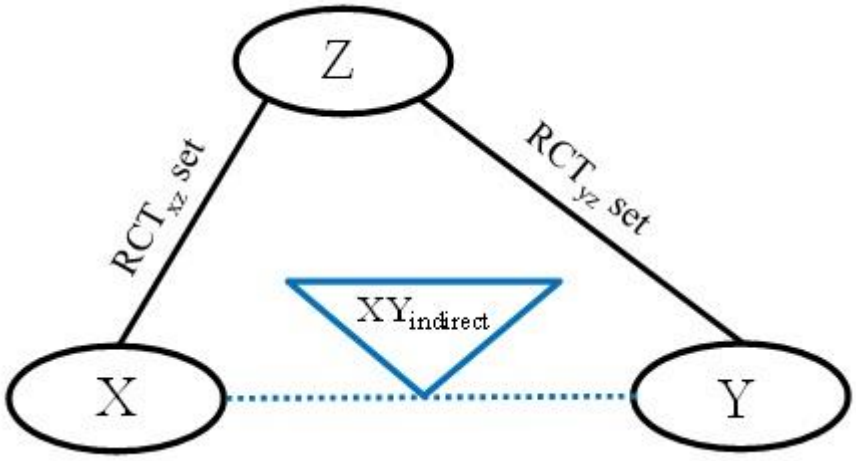

Figure 2.3: Indirect treatment comparisons

For example, in Figure 2.3, there is no RCT that directly compares Interventions X and Y but there are RCTs that compared Interventions X and Z (RCT_{xz} set) and Interventions Y and Z (RCT_{yz} set), indicated by the solid lines that connect the interventions. Hence, relative effectiveness of X versus Y can be estimated indirectly using IC by contrasting trials of X versus Z with trials of Y versus Z. Algebraically, the *indirect* relative effectiveness of X versus Y is estimated as follows:

$$XY_{indirect} = XZ_{direct} - YZ_{direct}$$

where XZ_{direct} and YZ_{direct} represent the direct relative effectiveness of X versus Z estimated from RCT_{xz} set and the direct relative effectiveness of Y versus Z estimated from RCT_{yz} set respectively.

2.4.2.2 Mixed Treatment Comparisons

Mixed treatment comparisons is an evidence synthesis method that allows for the simultaneous comparison of three or more different interventions in one meta-analysis (Ades, 2003, Caldwell et al., 2005, Lu and Ades, 2004). Taking the simplest case of three interventions and using the example in Figure 2.3, consider now that there are direct head-to-head RCTs that compare interventions X and Y. The relative effectiveness between interventions X and Y estimated using MTC is the "summation" of the indirect relative effect, $XY_{indirect}$, estimated from the sets of RCTs of X and Y

with common comparator Z (RCT_{xz} set and RCT_{yz} set) and the direct relative effect, XY_{direct} estimated from the set of head-to-head RCTs (RCT_{xy} set) as shown in Figure 2.4; hence, the name mixed treatment comparisons. When pooling direct and indirect pairwise contrast evidence in MTC, there should be consistency between the direct and indirect evidence. This is an important assumption for MTC which is discussed in Section 2.4.2.4.

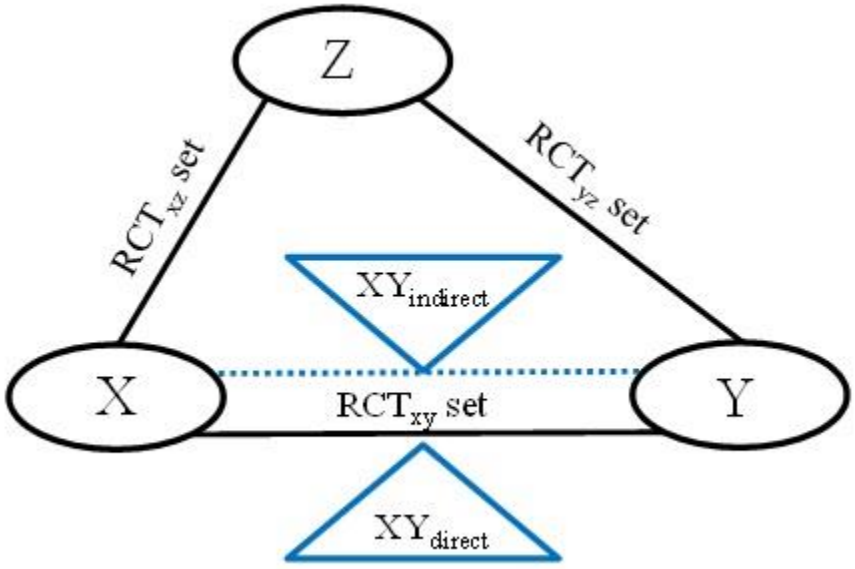

Figure 2.4: Simplest case of Mixed Treatment Comparisons – 3 interventions

Figure 2.4 presents the simplest structure of a MTC with the minimum required interventions (of three) to conduct a MTC analysis. Often, to address the research question of interest, numerous interventions are identified and included in a MTC analysis. A more complex structure of interventions and RCTs that compares the interventions is shown in Figure 2.5. This structural diagram of evidence is termed a network diagram.

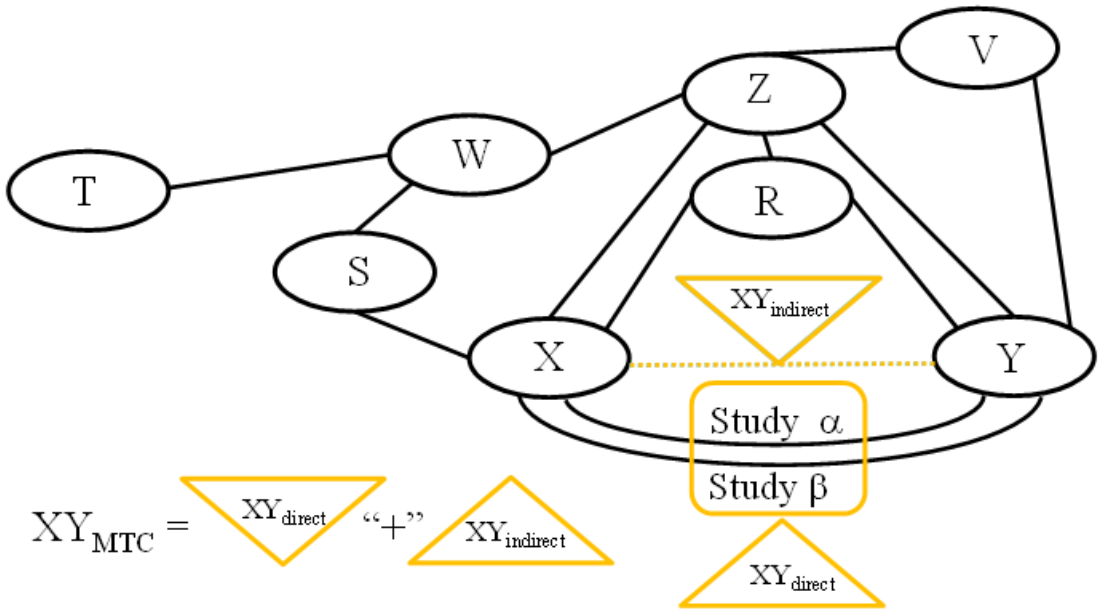

Figure 2.5: Mixed Treatment Comparisons – Network Diagram

When there are more than three interventions, there will be several direct and indirect comparisons, hence, it is important to note that the indirect relative effect, $XY_{indirect}$, in Figure 2.5 is not the same as the indirect relative effect, $XY_{indirect}$, in Figure 2.4. This is because the estimate of $XY_{indirect}$ in Figure 2.5 is no longer solely from the two set of RCTs (RCT_{xz} set and RCT_{yz} set) but from many other RCTs in the network.

Considering binary outcome measure for the MTC analysis, the Bayesian random-effects MTC model is specified as in Formula (2.2). For treatment k in RCT i, it is assumed that the occurrence of $r_{i,k}$ events from $n_{i,k}$ individuals follows a binomial distribution with event probability $p_{i,k}$.

Likelihood: $r_{i,k} \sim Binomial(p_{i,k}, n_{i,k})$

Model: $logit(p_{i,k}) = \begin{cases} \mu_{ib} & if\ k = b \\ \mu_{ib} + \delta_{ibk} & if\ k\ 'after'\ b \end{cases};\ b = A, B, etc.$

$\delta_{ibk} \sim Normal(d_{bk} = d_{Ak} - d_{Ab}, \sigma^2)$

Priors: $d_{Ak} \sim Normal(0, 10^2), \sigma \sim Uniform(0,2)$ (2.2)

For RCT i that compares interventions b and k, the parameter μ_{ib} is the effect of the baseline intervention b and the parameter δ_{ibk} is the study-specific effects of intervention k relative to b, which is normally distributed with mean d_{bk} and between-study variance σ^2. The mean, d_{bk}, is the difference between d_{Ak} (effect of interventions k relative to A) and d_{Ab} (effect of interventions b relative to A). The random-effects model can be reduced to a fixed-effect model by setting $\sigma^2 = 0$, and hence $\delta_{ibk} = d_{bk}$.

2.4.2.3 Additional summary statistics of NMA

As NMA allows for the simultaneous and coherent comparison of multiple interventions instead of two interventions in standard PWMA, it is no longer confined to answering if Intervention X is *better* than Intervention Y in terms of efficacy. It has the added advantage of answering which intervention is the best, second best and so on by ranking the interventions in terms of their efficacy.

This can be conducted with ease within a Bayesian framework. All interventions in the analysis can be ranked using probabilities rather than crude methods. Ranks may be presented as summary statistics (e.g., mean/median rank, surface under the cumulative ranking curve (SUCRA) (Salanti et al., 2011)), or graphical representations of the distribution of ranks (e.g. rankograms / barplots) indicating the probability that a given intervention is 1st, 2nd or 3rd best when compared to all other interventions in the network. Probability that each intervention is the best (ranked first) can also be estimated by calculating the probability that the intervention is ranked first.

2.4.2.4 Assumptions, Advantages and Limitations

An important assumption when using NMA is that the RCTs for the different intervention comparisons are similar in all ways other than the interventions being compared. Simply, it means that respective pairwise contrast effect sizes estimated are assumed to be the same (in fixed-effect model) or exchangeable (in random-effects model) in all RCTs in the analysis regardless of whether or not the pairwise contrast exists in the RCTs. Using the RCTs in Figure 2.4 for illustration, intervention effect of

X versus Y in RCT_{xy} set is assumed to be the same as/exchangeable with the intervention effect of X versus Y in RCT_{xz} set and RCT_{yz} set even though intervention contrast of X versus Y does not exist in either of the RCT sets.

Another assumption is consistency in the direct and indirect effect estimates within each pairwise contrast in the network of trials when pooling direct and indirect effects in a NMA. Validity of NMA is of major concern when there is the possibility of inconsistency. Empirical studies have been conducted to address this concern through comparison of direct with indirect sources of evidence (Gartlehner and Moore, 2008, Song et al., 2011). Statistical approaches for assessing consistency in NMA have also been proposed (Dias et al., 2010, Lu et al., 2011). If there are inconsistencies in the indirect and direct estimates within pairwise contrasts in the NMA, the assumption of exchangeability is unlikely to be met as well, and it is questionable if the dissimilar sources of evidence should be pooled together.

There are several advantages of using NMA for comparing multiple interventions. One advantage is the preservation of within-trial randomisation when combining RCTs evidence (i.e. NMA is performed using the relative effectiveness results of randomised arms of interventions from each trial included in the network – hence there is no breaking of randomisation when synthesising the results). There is greater transparency of the framework, that is, there is no need for 'back of the envelope' indirect comparisons based on a series of PWMAs. It also provides potential reduction of uncertainty due to the inclusion of more data.

2.4.3 Bivariate random-effects meta-analysis

Network meta-analysis extends the traditional meta-analysis to the synthesis of evidence involving multiple interventions to investigate effect size between the multiple pairs of interventions on one outcome of interest. Multivariate meta-analysis, on the other hand, extends the traditional meta-analysis by allowing evidence synthesis that investigates the comparative effect size between two interventions to be conducted for multiple correlated outcomes jointly in one analysis.

When it is of interest to evaluate multiple outcomes in evidence synthesis, separate traditional meta-analyses are often utilised to synthesise the evidence for each outcome independently, even when the outcomes are known to be correlated. Considering that the outcomes are correlated, a more intuitive approach would be to employ a multivariate random-effects meta-analysis that allows the synthesis of multiple outcomes jointly while accounting for their correlations. In this section, a version of the multivariate random-effects meta-analysis that only involves two correlated outcomes - called the bivariate random-effects meta-analysis (BRMA) will be discussed. The BRMA model specified is for the meta-analysis of continuous outcome data.

2.4.3.1 Specification of the general normal BRMA model

Suppose that a systematic review was conducted to review studies that reported either one or both of the outcomes of interest (Y_1 and Y_2), and n studies were identified. Now, let $Y_{h,i}$ denote the summary measure, such as treatment effect, and $\sigma_{h,i}$ the corresponding standard error for outcome h $(h = 1\ or\ 2)$ of the ith $(i = 1, 2, \dots, n)$ study.

Under a random-effects framework, each summary measure $Y_{h,i}$ is assumed to be an estimate of a true value $\mu_{h,i}$ in each study, where the $\mu_{h,i}$ is further assumed to be drawn from a distribution with mean value of β_h and between-study variance of τ_h^2. Assuming that both $Y_{h,i}$and $\mu_{h,i}$ are normally distributed, the general normal BRMA model (van Houwelingen et al., 2002) is defined as follows:

$$\begin{pmatrix} Y_{1,i} \\ Y_{2,i} \end{pmatrix} \sim Normal\left(\begin{pmatrix} \mu_{1,i} \\ \mu_{2,i} \end{pmatrix}, \Sigma_i\right) \qquad (2.3)$$

$$\begin{pmatrix} \mu_{1,i} \\ \mu_{2,i} \end{pmatrix} \sim Normal\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathrm{T}\right) \qquad (2.4)$$

with

$$\Sigma_i = \begin{pmatrix} \sigma_{1,i}^2 & \sigma_{1,i}\sigma_{2,i}\rho_{w,i} \\ \sigma_{2,i}\sigma_{1,i}\rho_{w,i} & \sigma_{2,i}^2 \end{pmatrix}$$

$$\mathrm{T} = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_b \\ \tau_2\tau_1\rho_b & \tau_2^2 \end{pmatrix}$$

The BRMA model can be viewed as comprising of two hierarchical parts, one at the within-study level and the other at the between-study level. The model defined in (2.3) represents the within-study model of the BRMA model, where $Y_{1,i}$ and $Y_{2,i}$ are treatment effect estimates that follow a bivariate normal distribution with study-level mean true effects, $\mu_{1,i}$ and $\mu_{2,i}$, which are correlated with corresponding within-study covariance matrices Σ_i. The covariance matrices (Σ_i) is formulated by the within-study standard errors of the estimates, $\sigma_{1,i}$ and $\sigma_{2,i}$, and the within-study correlation of the outcomes, represented by $\rho_{w,i}$.

The model defined in (2.4) represents the between-study model component of the BRMA model, where the correlated study-level mean effects for the 2 outcomes, $\mu_{1,i}$ and $\mu_{2,i}$, follow a bivariate normal distribution with summary means, β_1 and β_2, and corresponding between-study covariance matrix T. The covariance matrices (T) is formulated by the between-study standard deviations, τ_1 and τ_2, and the between-study correlation, represented by ρ_b.

Here, the BRMA model differs from the combination of two independent univariate random-effects meta-analysis (URMA) models by the inclusion of the within-study and between-study correlations, $\rho_{w,i}$ and ρ_b respectively. If the within-study and between-study correlations are both zero ($\rho_{w,i}$=ρ_b= 0), the BRMA model reduces to two independent URMA models as shown in equation (2.1).

2.4.3.2 Parameterisation of the between-study model

In order to place prior distribution on the between-study correlation, the between-study model in Equation (2.4) is parameterised using a product normal formulation as described by Bujkiewicz and colleagues (Bujkiewicz et al., 2013), specified as:

$$\mu_{1,i} \sim Normal(\eta_1, \psi_1^2)$$

$$\mu_{2,i}|\mu_{1,i} \sim Normal(\eta_{2,i}, \psi_2^2)$$

$$\eta_{2,i} = \lambda_0 + \lambda_1(\mu_{1,i} - \overline{\mu_{1,i}}) \qquad (2.5)$$

whereby the parameters of the model can then be estimated as follows:

$$\left.\begin{aligned} \beta_1 &= \eta_1 \\ \tau_1^2 &= \psi_1^2 \\ \beta_2 &= \lambda_0 \\ \tau_2^2 &= \psi_2^2 + \lambda_1^2\psi_1^2 \\ \rho_b &= \frac{\lambda_1\psi_1^2}{\sqrt{\psi_1^2(\psi_2^2 + \lambda_1^2\psi_1^2)}} \end{aligned}\right\} \qquad (2.6)$$

It can be seen in Equation (2.6) that the parameters (ρ_b, τ_1, τ_2 and λ_1) in the product normal formulation of the BRMA model are inter-dependent. Due to the inter-dependencies between the parameters, care must be taken when placing prior distributions on the parameters to ensure that they take values within plausible range.

In this thesis, prior distributions are placed on the correlation (ρ_b) and either the between-study standard deviations (τ_1 and τ_2) or the hyper parameters ψ_1 and ψ_2. This allows the regression coefficient (λ_1) to be evaluated using the relationship defined in Equation (2.6) as follows:

$$\lambda_1 = \frac{\rho_b\tau_2}{\tau_1} \qquad (2.7)$$

or

$$\lambda_1 = \frac{\psi_2\rho_b}{\psi_1\sqrt{1 - p_b^2}} \qquad (2.8)$$

As the other regression coefficient λ_0 (the constant term in the regression equation) is not related to the correlation (ρ_b), non-informative prior distribution is assigned to it. Construction of informative prior distributions for the between-study model is discussed in Section 2.4.3.7.

2.4.3.3 Concept of "borrowing of strength"

In the situation of two independent URMAs for outcome 1 and 2 individually, the estimate of the summary mean, β_1, for outcome 1 depends solely on the data for outcome 1 (i.e. the $Y_{1,i}$, $\sigma_{1,i}$, and τ_1). The same applies for outcome 2. There is no sharing of information between the two outcomes for the estimation of the summary statistics of the outcomes, even when they are correlated.

When using a BRMA for the two correlated outcomes (where $\rho_{w,i} \neq 0$; $\rho_b \neq 0$), the estimate of the summary mean, β_1, for outcome 1 now also incorporates the data from outcome 2 (i.e. the $Y_{2,i}$, $\sigma_{2,i}$, and τ_2) through the within-study and between-study covariances (i.e. $\sigma_{1,i}\sigma_{2,i}\rho_{w,i}$ and $\tau_1\tau_2\rho_b$). The same applies for outcome 2 as there is bilateral sharing of data between the outcomes. This concept of information sharing in hierarchical data analysis is commonly termed "borrowing of strength", and in this case across outcomes as well as studies.

This "borrowing of strength" in the context of meta-analysis implies that (i) each study's weight for the pooled estimates in the BRMA may be different from that in the individual URMA for each outcome; (ii) with the inclusion of information from the correlated outcome, the standard error for the pooled estimates in the BRMA are likely to be smaller than that in the individual URMAs. It has been shown that for the situation when the variances, τ_h^2 $(h = 1,2)$, are the same for both the BRMA and URMAs, the variance of the pooled estimate, β_1, in the BRMA is always less than or equal to that in the URMA and it applies regardless of whether the within-study and between-study correlations are negative or positive (Riley et al., 2007).

Clearly, there is no "borrowing of strength" when the correlations are zero ($\rho_{w,i} = 0$; $\rho_b = 0$) and the BRMA therefore is equivalent to two independent URMA. However, even if there are large within-study and between-study correlations, there is no "borrowing of strength" from outcome 2 for outcome 1 if the within-study covariance matrices Σ_i are the same for all the studies. Riley and colleagues demonstrated this through the use of a modified dataset where the within-study covariance matrices Σ_i were set to be the same for all the studies (Riley et al., 2007).

In reality, it is very unlikely that the within-study covariance matrices Σ_i for all studies in a BRMA to be the same; hence "borrowing of strength" is bound to exist. It has been proven that the degree of "borrowing of strength" is directly proportional to the magnitude of (i) the between-study correlation; (ii) the differences between the within-study standard errors; and (iii) the difference between the within-study covariances (Riley et al., 2007). In the next section, issue of missing data for BRMA and how it exploits the "borrowing of strength" functionality of the BRMA model will be discussed.

2.4.3.4 Issues of missing data

When performing BRMA for two correlated outcomes, only one set of studies that report either or both the outcomes is used for the analysis. Usually, this set of studies is larger than (or sometimes equal to) each of the individual sets of studies when performing separate URMAs each of the two outcomes. However, an important point to note is that at least some studies in the evidence set must report both outcomes in order to estimate the between-study correlation in a BRMA.

As in all mixed effect models, in BRMA assumption is made that data are "missing at random". However, it is not always easy to justify this in the context of meta-analysis as it is subjected to publication bias, dissemination bias, and within-study selective reporting (Sterne et al., 2001).

A further issue with missing data in BRMA is the missing within-study correlations for the studies in the BRMA. It is usually not reported in published articles and thus cannot be extracted together with the summary statistics $Y_{h,i}$ and $\sigma_{h,i}$. Obtaining it requires requesting for individual patients' data from the corresponding author of the studies and is challenging and time-consuming. However, there are situations where it is possible to assume that the within-study correlations for all the studies are zero, such as for diagnostic screening analysis where the two outcomes, sensitivity and specificity, are measured using two separate sets of patients. Where is it not possible to assume that $\rho_{w,i} = 0$, the issue of missing within-study correlations can be handled in a Bayesian framework with the use informative prior distributions for $\rho_{w,i}$ constructed using data

external to the BRMA set of studies. Although between-study correlation, ρ_b, can be estimated so long as some studies included in the BRMA report both outcomes, informative prior distributions for ρ_b can also be incorporated in the Bayesian analysis using external data. Bayesian approach for BRMA and methods of constructing informative prior distributions for $\rho_{w,i}$ and ρ_b are discussed in sections 2.4.3.5 to 2.4.3.7.

2.4.3.5 Bayesian approach for the BRMA model

Although BRMA has been proposed (Riley et al., 2007) as a more appropriate approach when synthesizing evidences from two correlated outcomes, it is not commonly applied in practice as obtaining the within-study correlation to fit the BRMA model is often difficult. However, a Bayesian framework provides the flexibility to conduct a BRMA by utilising external data about the within-study correlation through the use of prior distributions (discussed in Section 2.2.1). A Bayesian approach for multivariate meta-analysis of mixed outcomes has been proposed by Bujkiewicz and colleagues (Bujkiewicz et al., 2013) that allows both the within-study and between-study correlations to be constructed using external evidences. This model for multivariate meta-analysis can be reduced to a BRMA by reducing the dimension of the model so that the number of outcomes is two.

In a Bayesian framework, a distribution is placed on each of the two outcomes regardless of whether they are missing for some studies, so that the missing summary measures can be estimated from the model though MCMC simulation (described in Section 2.3.1). Although any missing values for $Y_{1,i}$ or $Y_{2,i}$ can be predicted from the BRMA model using MCMC simulation, the corresponding missing standard errors ($\sigma_{1,i}$ or $\sigma_{2,i}$ respectively) need to be estimated. To estimate the standard errors, it is assumed that the corresponding population variances come from the same distribution under the assumption of exchangeability of the variances. The variances are defined to follow half-normal distributions as follow:

$$var_{1,i} \sim HNormal(g_1)$$

$$var_{2,i} \sim HNormal(g_2)$$

$$g_1 \sim \Gamma(1.0, 0.01)$$

$$g_2 \sim \Gamma(1.0, 0.01)$$

and the standard errors squared are calculated as:

$$\sigma_{1,i}^2 = \frac{var_{1,i}}{N_i}$$

$$\sigma_{2,i}^2 = \frac{var_{2,i}}{N_i}$$

By specifying distributions for the population variances, it allows the population variances for the studies with "missing" standard error to be predicted from the MCMC simulation, conditional upon both the data and posterior estimates of the model parameters. This in turn allows the computation of the corresponding standard errors required for defining the within-study covariance matrices Σ_i. Non-informative and informative prior distributions for $\rho_{w,i}$ can be used together with the standard errors to define the within-study covariance matrices Σ_i.

The Bayesian BRMA model is applied to an example on prostate cancer in Chapter 4 for the prediction of progression-free survival (PFS) outcome estimates using the BRMA model on two correlated outcomes: overall survival (OS) and PFS. The next two sections give detailed descriptions on the construction of informative priors for the within-study and between-study correlations. The Bayesian BRMA model is also used in Chapter 5; however, non-informative prior distributions are defined for the parameters in the model as the purpose for using the model was to construct prior distributions for a subsequent analysis.

2.4.3.6 Construction of prior distributions for the within-study correlations

As correlations are constrained between -1 and +1, non-informative prior distribution for the within-study correlations between the two outcomes for the ith $(i = 1, 2, \dots, n)$ study can be defined using a uniform distribution as follows:

$$\rho_{w,i} \sim Uniform(-1, 1)$$

External evidence for the within-study correlations can be obtained from published articles reporting it or through the use of external IPD on the two outcomes. Methods of constructing informative prior distributions using external IPD are discussed in details by Bujkiewicz and colleagues who used a double bootstrap method to estimate the correlation with uncertainty (Bujkiewicz et al., 2013). In this thesis, informative prior distributions for the within-study correlations are constructed using correlation with corresponding 95% confidence intervals (CIs) or study sample sizes reported in published articles. In order to sample from a prior distribution for correlation which is constrained between -1 and +1, Fisher transformation method is used to convert the correlation to a corresponding Fisher correlation parameter, *z*, which follows a normal distribution for sampling. The Fisher transformation is defined as:

$$z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

where z is the Fisher correlation parameter and ρ is the correlation. Back-transformation can then be performed to convert z back to ρ by using the equation:

$$\rho = \frac{\exp(2z)-1}{\exp(2z)+1}$$

Denoting the estimates obtained from the published study as ρ_s(95% CI: $\rho_{s(LB)}$ to $\rho_{s(UB)}$), the corresponding Fisher correlation parameter, z_s, $z_{s(LB)}$ and $z_{s(UB)}$ are estimated and utilised to define the prior distribution (of the transformed correlation) for sampling as follows:

$$z_{w,i} \sim Normal(z_s, \sigma_s^2)$$

$$\sigma_s = \frac{z_{s(UB)} - z_{s(LB)}}{2 \times 1.96}$$

Where the 95% CI is not reported, the study sample size, n, is used to estimate the standard error for z_s, which is approximated using the equation (Altman, 1991) (page 293):

$$\sigma_s = \frac{1}{\sqrt{n-3}}$$

The within-study correlation for the ith $(i = 1, 2, \dots, m)$ study is in turn estimated using the back-transformation equation as follows:

$$\rho_{w,i} = \frac{\exp(2z_{w,i}) - 1}{\exp(2z_{w,i}) + 1}$$

Method of constructing the prior distribution for the between-study correlation is discussed in the next section.

2.4.3.7 Construction of prior distributions for the between-study correlation

To construct prior distributions for the between-study correlation, external data in the form of summary data ($Y_{h,i}$ and $\sigma_{h,i}$) from published studies identified (from a systematic review) need to be extracted. It is important to note that the data format is different from that for the construction of prior distribution for the within-study correlations which uses IPD.

Informative prior distribution for the between-study correlation of the BRMA model is to be constructed using the same BRMA model described in Section 2.4.3.2 but using "external data" as the likelihood data and non-informative prior distributions as the priors in the Bayesian analysis.

Hence, to fit the model to the external data, non-informative prior distributions have to be placed on the correlation and standard deviations. Non-informative prior distribution for the between-study correlations can be defined using a uniform distribution, similar to that for the within-study correlation shown in Section 2.4.3.6, as follows:

$$\rho_b \sim Uniform(-1, 1)$$

Non-informative prior distribution for the between-study standard deviations can be defined using half-normal distributions as follows:

$$\tau_{1,i} \sim HNormal(\sigma^2)$$

$$\tau_{2,i} \sim HNormal(\sigma^2)$$

where σ^2 is large.

The posterior distribution for the between-study correlation from this analysis forms the informative prior distribution for the between-study correlation of the BRMA model with the primary analysis data as the likelihood data. The prior distribution (which is the posterior distribution from the external BRMA) can be utilised in the BRMA model in the form of empirical distribution using WinBUGS. Alternatively, the parameters of the posterior distribution can be used to construct prior distribution on the Fisher transformation scale (as discussed in Section 2.4.3.6).

2.5 Economic Modelling

2.5.1 Markov Models

Markov models are widely used in economic modelling of medical conditions because of their ability to model both effects and costs of treatments simultaneously for economic decision making (Sonnenberg and Beck, 1993, Briggs and Sculpher, 1998). As Markov model is good for monitoring random processes that occur over time, it is particularly suited for the modelling of progressive diseases. It allows the various stages of disease progression to be represented using distinct states in the model and the potential disease progression pathway to be modelled using transition probabilities between the states. The disease pathway of patients are then modelled over time using discrete time period, termed as a 'cycle' to characterise the time unit for modelling the disease processes (for example, from diagnosis to progression to death).

An example of a Markov model used in the evaluation of treatments in some cancers, such as metastatic breast cancer or metastatic prostate cancer, is presented in Figure 2.6. Here, the states of the disease are denoted by ovals and the transitions between states presented using arrows indicating the potential paths of disease progression over time. The model is a three-state model as illustrated by the three ovals (states) in the model.

The first state of the disease pathway of cancer in the Markov model is stable disease, which is the 'asymptomatic' disease state where the patient has the disease but has not experienced the ill effects of the disease or is not at any increased risk of mortality compared to another individual with the disease. This is the state where a patient is

undergoing active cancer treatment. The second state is the 'progressive disease' state where the patient experienced tumour progression and stopped the existing cancer treatment regimen. The third and last state of the model is the 'death' state which is commonly termed the absorbing state of a Markov model as it is impossible to leave this state upon entering. As shown in Figure 2.6, there are arrows that bend back into the state that they left; this shows that patients remained in the states that they were in during the previous cycle.

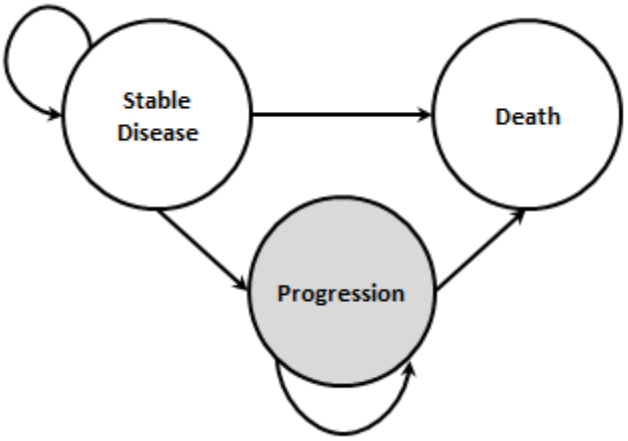

Figure 2.6: Markov model for cancer

The representation of various stages of the disease progression using distinct states allows costs and utility outcomes to be elegantly specified for each state and at each cycle and the clinical effectiveness of health technologies to be incorporated in the model through the transition probabilities between states, at each cycle. This therefore enables long-term cost and utility values to be estimated for all health technologies under evaluation. In this thesis, the three-state Markov model presented in Figure 2.6 is used in the decision model developed for the cost-effectiveness evaluation of health technologies for metastatic prostate cancer.

Markov models have many structural advantages that make them ideal for decision modelling. It is easy to incorporate 'time' in the model; the elapse time from one state to another state can be determined using the transition probabilities between states and the time horizon to model the disease processes can be specified. This allows long-term costs and outcomes to be estimated using appropriate discounting rates (Section 2.5.7).

Not all patients diagnosed with the disease will die of the disease; Markov models allow competing risk of death to be included in the model with ease by having a direct transition pathway (arrow) from asymptomatic state to death state. The Markov model in Figure 2.6 has three states, however, for other chronic diseases with recurring relapse of events or progression to other disease, additional states for the disease may need to be modeled. They can be dynamically incorporated in the model by adding new states and transition pathways. The most attractive advantage of the Markov model is the simplicity to model clinical outcomes, costs and utilities simultaneously in one model and the ease of interpretation of the estimated results due to the simple structure of the model.

Despite the flexibility of the Markov models to represent the natural pathway of disease processes, they do have their limitations. One major limitation of the Markov models is the 'memoryless' feature of the Markov models, known as the Markov assumption, where the probability of a patient moving out of a state is not dependent on the experience that the patient had prior to entering this state. The implication of this 'memoryless' feature is that it is not possible to model a disease process where future events depend on past events. The simplicity of a Markov model to model disease processes using distinct states, as much as being an advantage, can be seen to be too simple as health processes are more fluid in real life.

2.5.2 Quality-adjusted life-years

Quality-adjusted life-years (QALYs) are used as a measure of effect in the cost-effectiveness analysis. QALY is a single measure that explains an individual's health-related quality of life and length of life gained due to the use of a health technology (Briggs et al., 2006). To estimate the QALY, utility data in the form of health-related quality of life (HRQoL) are required to quantify the potential health status of patients with the disease condition, as well as the impact that the health technology (in terms of disease progression and serious adverse effects) has on their HRQoL. QALYs are then calculated by multiplying the utility data by the life years of each individual patient.

In Health Technology Assessment (HTA), the HRQoL instrument used as recommended by NICE is the preference-based EuroQoL five-dimensional (EQ-5D)

questionnaire. Standardising the HRQoL instrument for all health evaluations, which span a wide range of health conditions, allows QALYs for all health technologies to be measured on the same scale and aid comparisons of health policies with different health conditions. The EQ-5D utility takes a value between -0.594 to 1 (Dolan and Roberts, 2002), where 1 represents good health, 0 represents death and negative values represents a health state worse than death.

The advantages of using EQ-5D instrument for computing utilities are that the questionnaire is easy to complete (by the patients) and the utilities generated represent community or societal preferences. However, as EQ-5D utilities are constrained between -0.594 and 1, modelling of EQ-5D utilities can be challenging. Difficulties in the prediction of EQ-5D utilities, with reference to a study on multiple sclerosis, are discussed in Chapter 5

2.5.3 Cost data

Cost data of the technologies under investigation need to be collected for appropriate assessment of the health technologies of interest. For the assessment of the health technologies in the UK, cost data extracted from a RCT investigating the technologies of interest and performed on the UK population is most ideal. However, very often this is not possible and resource data that most closely resemble that expected in the UK were used. Resource data often include cost relating to, for example, drug acquisition, drug administration, inpatient clinic, hospitalisation and outpatient follow up. The total cost for each technology is subsequently computed by summing the applicable resource data. Challenges in modelling resource costs for decision modelling in metastatic prostate cancer are discussed in Chapter 4.

2.5.4 Incremental cost effectiveness ratio

When comparing two or more health technologies in a cost-effectiveness analysis, the established statistic for assessment is the Incremental Cost Effectiveness Ratio (ICER). The ICER aids to quantify both the difference in costs between the health technologies and the difference in clinical effects (in terms of QALYs) of the technologies in one

single statistics for decision making. Mathematically, it is computed using the following equation:

$$ICER = \frac{\Delta\, Costs}{\Delta\, Effects} \qquad (2.9)$$

where $\Delta\, Costs$ represents the difference in costs between technologies (new technology minus existing technology); $\Delta\, Effects$ represents the corresponding difference in clinical effects. Simply, the ICER is interpreted as the additional cost per unit of clinical effect gained from the new technology over the existing technology. Currently, a threshold ratio of £30,000/life-years gained for ICER is specified by NICE (NICE, 2013) to represent the willingness-to-pay for a unit of health gain by the NHS. Generally, new technologies with ICER of at most £30,000/life-year gained are reimbursed. However, there are exceptions in the use of the threshold ratio for decision making of certain health conditions and this is illustrated in Chapter 4 for the cost-effectiveness analysis of metastatic prostate cancer.

In practice, there are uncertainties in the costs and clinical effects of the technologies under investigation and these uncertainties are incorporated in the economic Markov model probabilistically, which in turn give ICER estimate with uncertainty. As economic models are Bayesian by nature, quantifying the uncertainty using Bayesian credible interval (CrI) instead of CI is more intuitive. This can be obtained easily by taking the $\alpha/2$ and $(1-\alpha/2)$ percentiles of the simulation results of the economic model as the $(1-\alpha)100\%$ CrI for ICER. However, graphical plots such as the cost-effectiveness plane (Black, 1990) and cost-effectiveness acceptability curve (CEAC) (Briggs and Fenn, 1998, O'Brien and Briggs, 2002, Fenwick et al., 2001), which present the uncertainty in its entirety, are the preferred ways of presenting uncertainty in the ICER estimate in HTA. These two graphical tools are described in greater details in Section 2.5.5 and 2.5.6.

2.5.5 Cost-effectiveness plane

The cost-effectiveness plane (Black, 1990) presents the difference in clinical effects between two technologies (new technology minus existing technology) per patient against the difference in costs per patient. The cost-effectiveness plane, with four

quadrants created by the axes and labelled using the positioning in the compass (NW, NE, SE and SW) is shown in Figure 2.7. With the x-axis representing the difference in clinical effects and the y-axis representing the difference in costs, the gradient for any line passing through the origin of the cost-effective plane plot is equal to the ICER (see Equation (2.9) for computation of ICER).

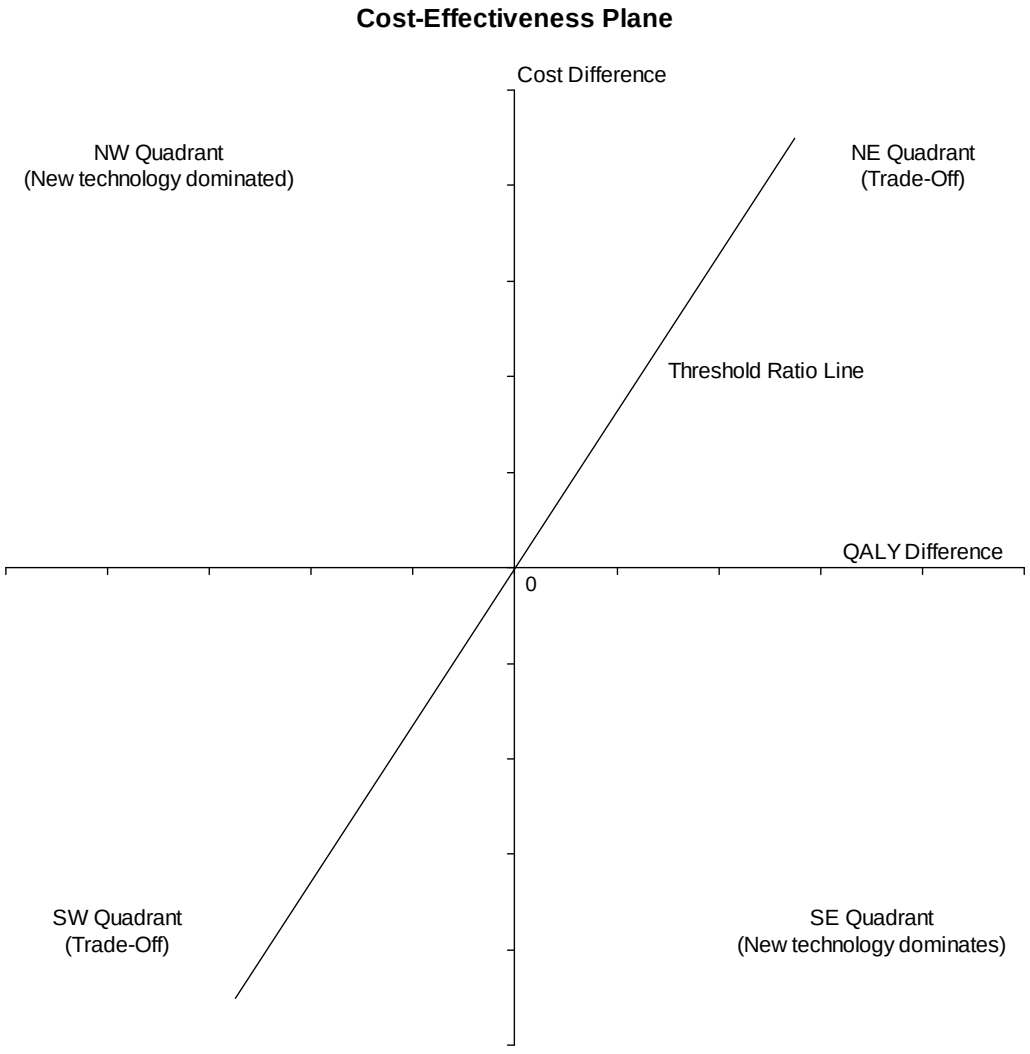

Figure 2.7: Cost-effectiveness plane for comparing new technology with existing technology (Briggs et al., 2006)(pp 122, Figure 5)

In reality, there are uncertainties in the costs and effects of the technologies as mentioned in Section 2.5.4. So, for the purpose of understanding and interpreting results using the cost-effectiveness plane, there is a need to consider the ideal situation

of certainty and knowing actually where the ICER estimate for the new technology lies on the cost-effectiveness plane.

If the ICER estimate lies in the NW quadrant, the new technology is less effective but more costly than the existing technology and is said to be 'dominated' by the existing technology. However, if the new technology is less costly but more effective, the technology will lie in the SE quadrant and is said to 'dominate' the existing technology. In both situations, decision on which technology to adopt is straight-forward, the technology that is more effective and less costly would be adopted.

It is when the estimate lies in the other two quadrants (NE and SW) that a decision must be made as to which technology should be adopted. This decision is generally made based on the ICER for the new technology and the threshold ratio (for the ICER) pre-specified by the decision maker. The cost-effectiveness plane is particularly useful in this instance as a line whose gradient representing the threshold ratio can be plotted on the cost-effectiveness plane. Any estimates that lies to the right of (or below) the threshold ratio line suggest that the new technology dominates the existing technology and should be adopted, while any estimates to the left of (or above) the threshold line indicate that the new technology is dominated. In most cases, the new technology under evaluation is more effective but also more costly and lies in the NE quadrant.

The above describes the ideal situation of knowing with certainty where the ICER estimate for the new technology lies. This is however not the case and uncertainty surrounding the costs and effects of the technologies exist in practice. Now, incorporating uncertainty in the estimates of the costs and effects using a probabilistic economic model will give outputs in the form of distribution over the cost difference, effect difference as well as the joint cost-effect distribution (Briggs et al., 2006). These outputs obtained using simulations performed on the probabilistic economic model are plotted on the cost-effectiveness plane to give a joint cost-effectiveness density plot. An example of this plot with 20000 Monte Carlo simulations is presented in Figure 2.8. A line of best-fit through the points on the plot gives the ICER for the new technology of interest. The joint cost-effectiveness density plot is used in Chapter 4 to illustrate the joint cost-effectiveness density for the different economic models developed. Another plot that presents the uncertainty of the ICER (the cost-effectiveness acceptability curve) is discussed in Section 2.5.6.

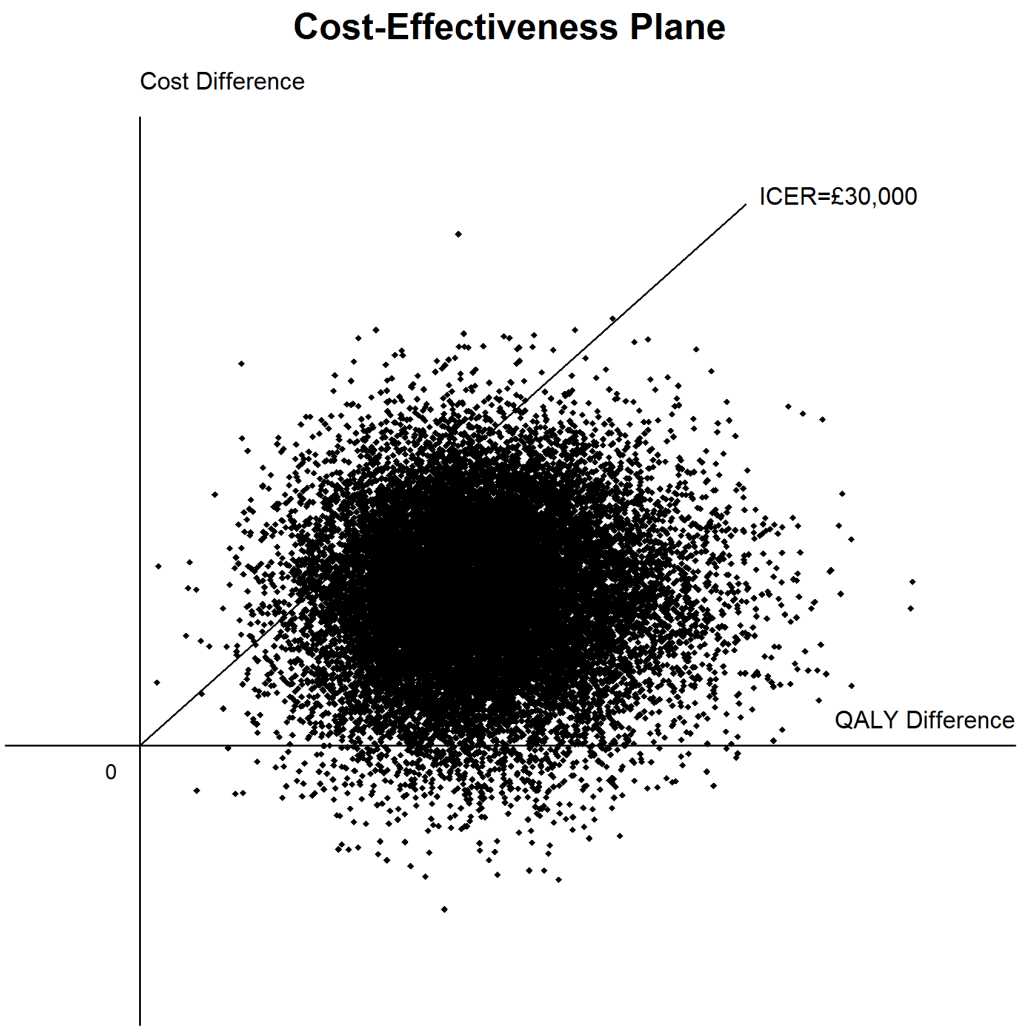

Figure 2.8: Estimated joint cost-effectiveness density plot on the cost-effectiveness plane

2.5.6 Cost-effectiveness acceptability curve

The CEAC presents the probability that each technology is cost-effective against the existing technology over a range of (maximum) values that the healthcare provider is willing to pay for an additional unit of clinical effects (eg. QALYs) gained (Fenwick et al., 2001). A CEAC can be plotted using a joint cost-effectiveness density plot such as the one presented in Figure 2.8. Lines of threshold ratios representing differential values of willingness-to-pay by decision makers can be drawn on the density plot. At each willingness-to-pay threshold, the probability that the technology is cost-effective can be computed by considering how many of the simulation points fall to the right of (and below) the corresponding 'threshold ratio' line. The CEAC can therefore be

plotted based on the probabilities calculated over the range of willingness-to-pay threshold values of the healthcare provider. Figure 2.9 shows an example of CEAC and it clearly illustrates that if cost is not an issue for the healthcare provider, the curve tends toward the probability that the technology is effective. As the CEAC is a graphical presentation of points on the joint density plot of incremental costs and effects, in reality the CEAC can take many different shapes (Fenwick et al., 2004).

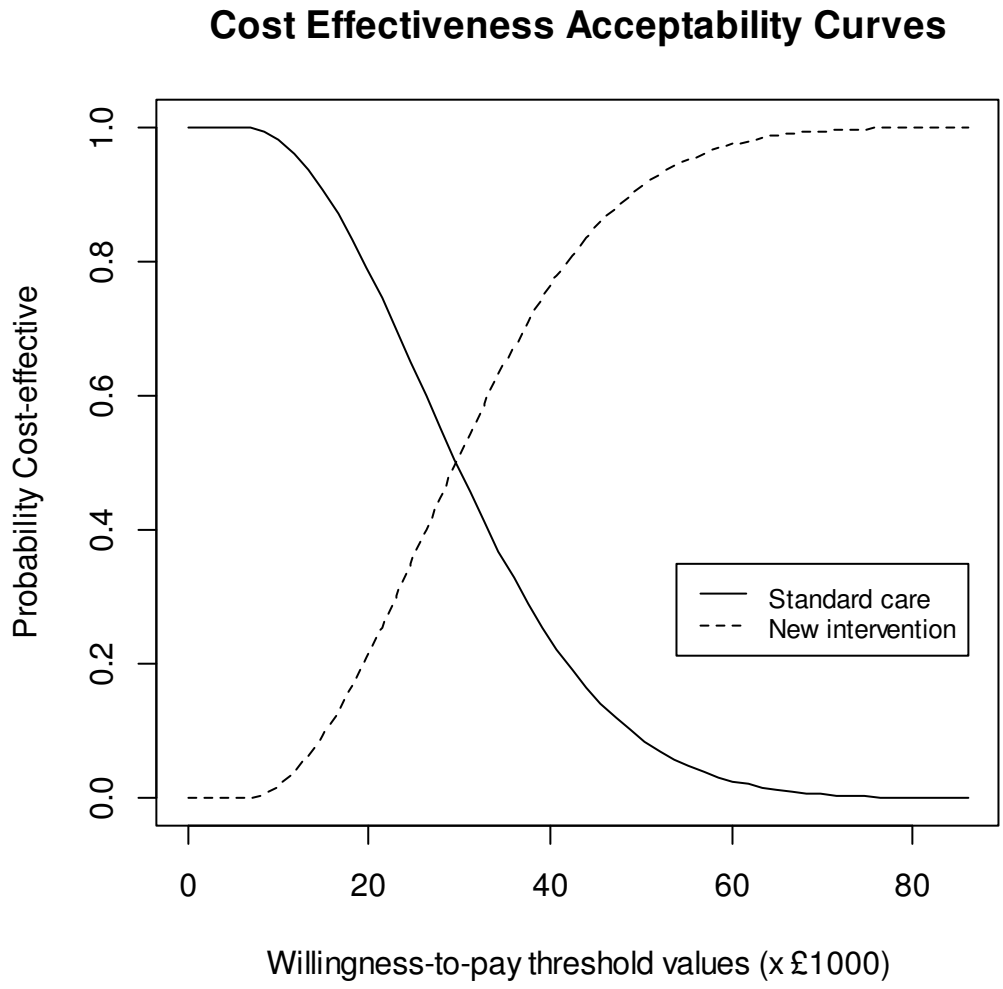

Figure 2.9: Cost-effectiveness acceptability curve

2.5.7 Discounting

When considering long-term health endpoints, such as mortality, in an economic modelling, it is important to make appropriate adjustments to costs and utilities outcomes to capture potential changes to the costs and utilities outcomes over the long

period of time frame. The rationale behind it from the viewpoint of analyst and patients alike is that as time passes, the cost of a treatment, for example in nominal amount of £1000 today may be valued more than the same nominal amount of £1000 in 10 years' time. Similarly, the degree of quality of life utilities measured today may be valued more than in 10 years' time when there are other newer treatment options available.

Using a Markov model and having the number of cycle representing years or months, costs and utilities for different timing in the model can be adjusted appropriately by applying a rate, known as discounting rate, to allow for comparison of costs and utilities in terms of a net present value (NPV).

The formula for discounting is as follows:

$$V_0 = \frac{V_t}{(1+r)^t}$$

where V_0 represents the NPV, that is the cost or utility value at time zero; V_t represents the cost or utility value at cycle t and r is the discounting rate to be applied for the adjustment. For cost-effectiveness analysis in HTA, the National Institute for Care Excellence (NICE) in the UK stipulates that a discounting rate of 3.5% should be applied for all base-case analysis (NICE, 2013). Alternative discounting rates (for example 6%) applied to cost and utility outcomes in the Markov model as sensitivity analyses are advised by NICE to allow analysts and decision-makers to explore the difference in ICER from the base-case analysis for differential discounting rates.

In the cost-effectiveness analysis presented in Chapter 4, discounting rates were applied to adjust costs and utilities in the model as the model was developed with a time horizon of 15 years.

2.6 Chapter Summary

This chapter reviewed the statistical theory and methodology that are applied in this thesis. Bayesian statistics introduced in Section 2.2 was applied throughout this thesis with all analyses conducted under the Bayesian framework using the WinBUGS and R software described in Section 2.3. Non-informative prior distributions were used in the analysis in Chapter 3, while informative prior distributions were constructed for the

analyses performed in Chapter 4 and Chapter 5 alongside the use of non-informative prior distributions.

Standard PWMA and NMA described in sections 2.4.1 and 2.4.2 are used in Chapter 3 for the development of novel graphical presentational tools to facilitate better reporting of NMA results. Bayesian BRMA introduced in Section 2.4.3 is utilized in Chapter 4 to synthesise two correlated outcomes (OS and PFS endpoints in cancer studies) jointly to predict the relative intervention effect of PFS which was not reported in one of the RCTs in the analysis. The predicted relative intervention effect is subsequently used to inform a cost-effectiveness evaluation model specified using the methodology described in Section 2.5. The Bayesian BRMA model is also used in Chapter 5 to construct informative prior distribution for the regression coefficients in linear regression models to provide a more informative estimate of the outcome of interest (EuroQol 5-Dimensions Questionnaire) for multiple sclerosis.

Other methodologies utilized in this thesis that are chapter specific are introduced in the methods section of the chapters. For example, methods related to survival analysis are introduced in Chapter 4 and statistical techniques for simulating data are introduced in Chapter 5. Details on the application of all these methodologies in one coherent analysis framework using the R and WinBUGS software are described in the chapters where they were employed.

3 Graphical presentational approaches for reporting network meta-analysis

3.1 Introduction

Until recently, systematic reviews and health technology assessments (HTA) have largely been limited to pairwise comparisons of interventions where direct evidence exists. Often there is an array of candidate interventions relevant to the clinical question of interest, thus an analysis comparing all the relevant interventions may be more appropriate and useful to decision-makers. Methodology to address this issue, which has increasingly been applied, is network meta-analysis (NMA) (Ades, 2003, Ades et al., 2006, Caldwell et al., 2005, Higgins and Whitehead, 1996, Lu and Ades, 2004, Lumley, 2002).

Due to the inherent feature of NMA to compare multiple interventions simultaneously, there has been rapid growth in the number of HTA reports that utilise NMA for the synthesis of evidence from clinical trials. NMA is a complex statistical method, the results from which need to be made accessible to non-statistical experts, including policy-makers, in order to maximize their usefulness in HTA. Therefore, it is imperative that clear, consistent and transparent reporting of this complex statistical analysis is established. The aims of this project are to (i) review the existing guidelines on the presentation of NMA analyses and to assess what has previously been done in practice in the UK by reviewing HTA reports, some of which were commissioned to inform NICE technology appraisals; (ii) formulate recommendations to improve future reporting of NMA analyses and lastly; (iii) develop graphical tools for the improved presentation of NMA analyses in the future.

Methods for the selection and review of UK National Institute for Health Research (NIHR) HTA reports that utilized NMA for clinical effectiveness and cost-effectiveness evaluations and methods for the development of graphical tools for reporting NMA results are described in Section 3.2.

Existing guidelines on the use and reporting of NMA in HTA in the UK are reviewed in Section 3.3.1. Results of the review of current practice for reporting NMA results in UK HTA reports are presented in Section 3.3.2. Based on the review results,

recommendations for improved reporting of NMA were developed and a set of important NMA results (that should be routinely reported) were identified for purpose of designing NMA graphical tools which are discussed in sections 3.3.3 and 3.3.4. Utilising the recommendations and important NMA results identified, novel graphical tools for reporting NMA were developed as part of this thesis. Expert opinions on the design and feasibility of the graphical tools for presenting important NMA results are discussed in Section 3.3.5. The final set of graphical tools developed and published is presented in Section 3.3.6. Other graphs developed and evaluated by the experts but were not developed further are presented in Section 3.3.7. Finally, Section 3.4 gives a discussion of the current reporting practice in HTA and graphical tools developed in this thesis.

3.2 Methods

Methods utilised to achieve the aims of this project are described in this section. A project flowchart depicting the steps involved from the review of presentational approaches of NMA in HTA to the development of novel graphical tools for presenting NMA results is shown in Figure 3.1.

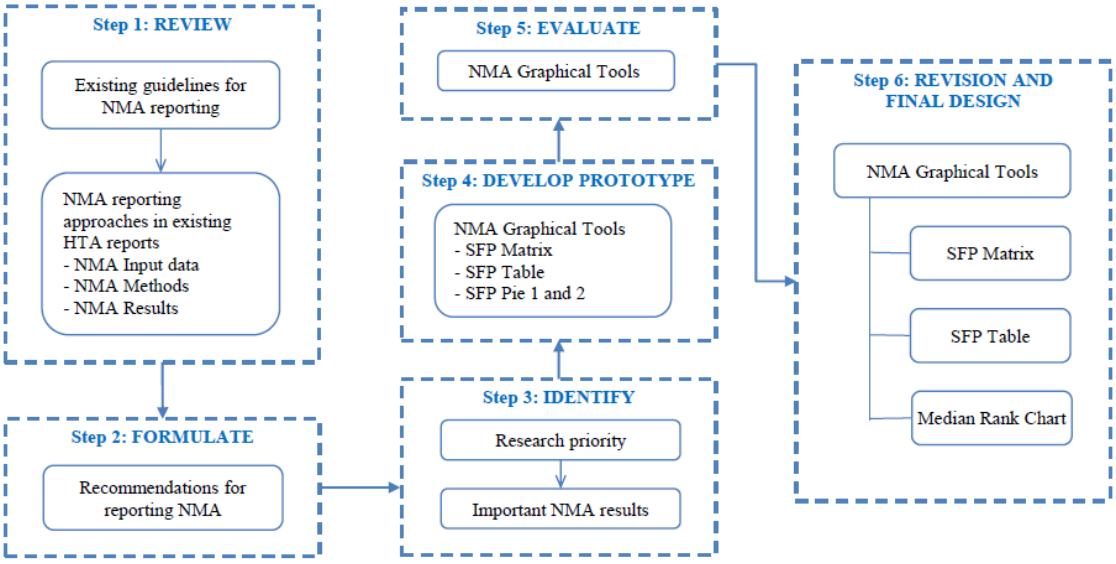

Figure 3.1: Graphical tools development flowchart

Comprehensive reviews of existing guidelines about the reporting of evidence syntheses in HTA and the reporting styles of NMA in published UK NIHR HTA reports were conducted (Step 1). Selection criteria of HTA reports and review

methodologies for establishing the current practice of reporting NMA in HTA are described in Section 3.2.1.

Taking into account the review results, recommendations for better reporting of NMA were developed (Step 2) to aid the conceptualisation of what NMA results are important and should be considered in the design of graphical tools for presenting NMA results (Step 3). Next, prototype of graphs were developed (Step 4) and evaluated by statisticians/quantitative experts with knowledge of PWMA and NMA (Step 5).

Lastly, the graphical tools were revised (Step 6) incorporating comments and suggestions provided by the statisticians/quantitative experts, as well as anonymous reviewers who reviewed the published manuscript (Tan et al., 2014). Methods for the development of graphical tools for reporting NMA results are described in Section 3.2.2.

3.2.1 Review of the reporting styles of NMA analyses in HTA reports

UK NIHR HTA programme reports, listed on their website as published between 1997 and 2011 inclusively that used NMA methodology for evidence synthesis were identified.

Using a standardized data extraction form (see Appendix A) these reports were examined to establish the approaches taken in the reporting of NMA as regards:

a. *Input data* - presentation of the number of interventions, study level data, and the relationship structure of the interventions and the studies included in the analysis;

b. *Methods* - specification of Bayesian or frequentist statistical models, software used and, where appropriate, presentation of prior distributions used (and assessment of their influence on the results via sensitivity analyses), and assessment of model convergence; and

c. *Results* - presentation of relative effects, absolute effects, probability of treatment being best, and/or ranking of interventions, with a particular emphasis on the use of tables and graphical presentational approaches. The use, if at all,

of the NMA results in informing the economic decision model was also considered.

Based on the results of this review, recommendations were formulated to improve NMA reporting. These recommendations considered input data, statistical model and results and covered (i) the qualities that make good NMA analysis reporting; (ii) the most appropriate and informative presentation methods; and (iii) the target audiences for the information. Graphical tools were subsequently designed on the basis of existing institutional guidance on NMA reporting in HTA and the recommendations made.

3.2.2 Development of graphical tools for NMA reporting

Despite the increase in the use of NMA in HTA and journal articles, there is no established graphical presentational standard for reporting the results of NMA analogous to the forest plot (Lewis and Clarke, 2001) for PWMA. This may be in part due to the complexity and magnitude of results generated when using NMA compared to PWMA. For example, a NMA including five different interventions generates 10 pairwise comparisons; and this increases to 45 pairwise comparisons when 10 different interventions are included. Presenting such large numbers of results can be challenging, especially when NMA is used to evaluate a number of different outcome measures within the area of interest.

When reporting NMA, it is therefore crucial to identify what NMA results are recommended and important. This was accomplished through understanding the existing institutional guidance on NMA reporting in HTA as discussed in Section 3.3.1 and a review of the current practice in NMA reporting in HTA reports as described in Section 3.2.1. General principles of graphical excellence for presenting data (Cleveland, 1994, Few, 2004, Tufte, 2001), in a manner that highlight and organise the data effectively, were utilised. This included reducing non-data ink, enhancing data ink, and grouping, prioritising and sequencing the data. The graphs developed for the presentation of NMA results were programmed using the R software language (R Core Team, 2012).

A preliminary version of the graphical tools developed was presented at the Methods in Meta-Analysis (MiM) meeting in London to engage statisticians with expertise in

PWMA and NMA to discuss important issues pertaining to NMA and seek their opinions about the graphs. The survey form used for the evaluation of the graphical tools developed is presented in Appendix B. Revisions to the graphs were made following an assessment of the suggestions provided at this meeting, as well as suggestions given by the anonymous reviewers of the published article.

In this thesis, the graphical tools developed are presented using a recently published study, which used NMA to investigate the use of tocolytic therapy for preterm child delivery (Haas et al., 2012), to illustrate its use for binary outcome on the odds ratio (OR) scale. This published NMA included 95 randomised controlled trials (RCTs) and considered eight classes of drugs for the treatment of preterm delivery [See Figure 2 in (Haas et al., 2012) for the network of interventions and trials included in the NMA]. The primary outcome measure was 48hr delay in delivery and the analysis was performed on the OR scale. Other secondary outcomes were also analysed in the study but for the illustration of the graphs developed, only the results of the primary outcome were displayed; graphs for the other outcomes will have a similar display format.

In the original article (Haas et al., 2012), key analysis results of the primary endpoint, such as NMA and PWMA ORs, probability best and rankogram were presented separately using tables and figures (Table 1, Figure 3 and Figure 7 in the original article). The graphical tools developed as part of this thesis are designed to consolidate the key results into a single figure that enable easier referencing of results for the authors and ease of interpretation of the results for the readers such as clinicians and academics.

The graphs developed can also be used to present other outcome measures (such as continuous data and hazard ratios). The study dataset on Parkinson's Disease available in the NICE Technical Support Document 2 ((Dias et al., 2011d) Example 5) was used to illustrate its usage for continuous outcome measure.

3.3 Results

A review of the existing guidelines for the use and reporting of NMA in UK HTA is discussed in Section 3.3.1. Results of the review on the current practice of reporting NMA in UK HTA are detailed in Section 3.3.2 and recommendations proposed for

better reporting of NMA analyses are presented in Section 3.3.3. Based on the results of the review on NMA reporting and recommendation developed, important NMA results identified are presented in Section 3.3.4 and considered in the preliminary design of graphical tools for better presentation of NMA reports. Evaluation results of the preliminary graphical tools developed are reported in Section 3.3.5 and the final set of graphical tools developed is presented in Section 3.3.6. The remaining preliminary graphical tools evaluated by the experts but were not developed further are presented in Section 3.3.7.

3.3.1 Guidance on use and reporting of NMA in HTA

NMA is now an established methodology acknowledged by HTA agencies worldwide including the National Institute for Health and Care Excellence (NICE) in England and Wales, the Canadian Agency for Drugs and Technologies in Health, the French Haute Autorité de la Santé, and the Pharmaceutical Benefits Advisory Committee in Australia, as well as emerging national agencies in Austria, Brazil, Colombia, Cuba and Ireland.

Since 2004, NICE guidance (NICE, 2004) has emphasized that, in HTA appraisals, evidence synthesis using pair-wise meta-analysis (PWMA) of head-to-head RCTs is the preferred method. If no head-to-head RCTs are available, the guidance permits the use of Indirect treatment Comparisons (ICs) so long as the potential bias in its use is appropriately explored and reported. In 2008, NICE re-emphasized the preference for synthesis results from head-to-head RCTs where available (NICE, 2008). However, it also stated that Mixed Treatment Comparison (MTC) analyses results may be included even if head-to-head trials were available, if it is justified that the MTC analysis will add information that is not available from the head-to-head trials. In the latest update of the guidance in 2013 (NICE, 2013), NICE adopts the use of the term "Network Meta-Analysis (NMA)" [(NICE, 2013) page 41] that comprises of both IC and/or MTC and further strengthens its position regarding the use of NMA in HTA appraisals. For the rest of this Chapter, the term NMA is used to refer to IC and/or MTC.

Internationally recognized guidelines for good reporting of standard PWMA of RCTs, such as the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement (formerly known as the Quality of Reporting of Meta-Analyses

(QUOROM) checklist), have been developed (Liberati et al., 2009b, Moher et al., 2009). While much of the PWMA reporting guidance is also highly relevant to the NMA context (including advice on clarity of labels and legends, and use of reasonable sizes for symbols and lines in tables and figures), the inherently greater complexity of the latter approaches presents further challenges for reporting such analyses.

Most specific NMA guidance is embedded in institutional HTA evaluation documentation. In the UK, the NICE guidance states that all data used for estimates of effectiveness should be presented in tabular form with the source of the data clearly stated (NICE, 2004, NICE, 2008, NICE, 2013). In the 2004 and 2008 guidance, it also states that, for NMA analyses the evidence may be presented in either tabular or diagrammatic form and should be reported as both relative and absolute effectiveness estimates. In the latest 2013 guidance, NICE takes on a stronger stand on the presentation of NMA results. It states that evidence from NMA must be presented using both tabular and graphical formats and the results from standard PWMAs should be presented alongside the results from NMA [(NICE, 2013) page 42, point 5.2.17]. This is re-enforced by more detailed advice that complements their guidance on evidence synthesis (though their content is non-mandatory when making submissions) (Ades et al., 2012, Dias et al., 2011b, Dias et al., 2011a, Dias et al., 2011c, Dias et al., 2011e, Dias et al., 2011d, Dias et al., 2011f). Another detailed source of guidance on conducting NMA analyses is the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force good practice documents (Hoaglin et al., 2011, Jansen et al., 2011). While these provide many of the details necessary to conduct a successful NMA analysis, specific details and recommendations on presentational formats, particularly of the data and results, are limited.

3.3.2 Review of NMA reporting in existing HTA reports

A total of 608 UK NIHR HTA reports were published between 1997 and 2011 inclusive, of which 375 contained systematic reviews of clinical effectiveness and/or cost effectiveness of the health technology under investigation. Amongst them, 205 reports contained evidence synthesis and 19 of these reports (including 10 of which informed NICE appraisals) utilized NMA methodology, and this defined our review

sample (while the remaining used only PWMA). A flowchart of this HTA report review selection is shown in Figure 3.2.

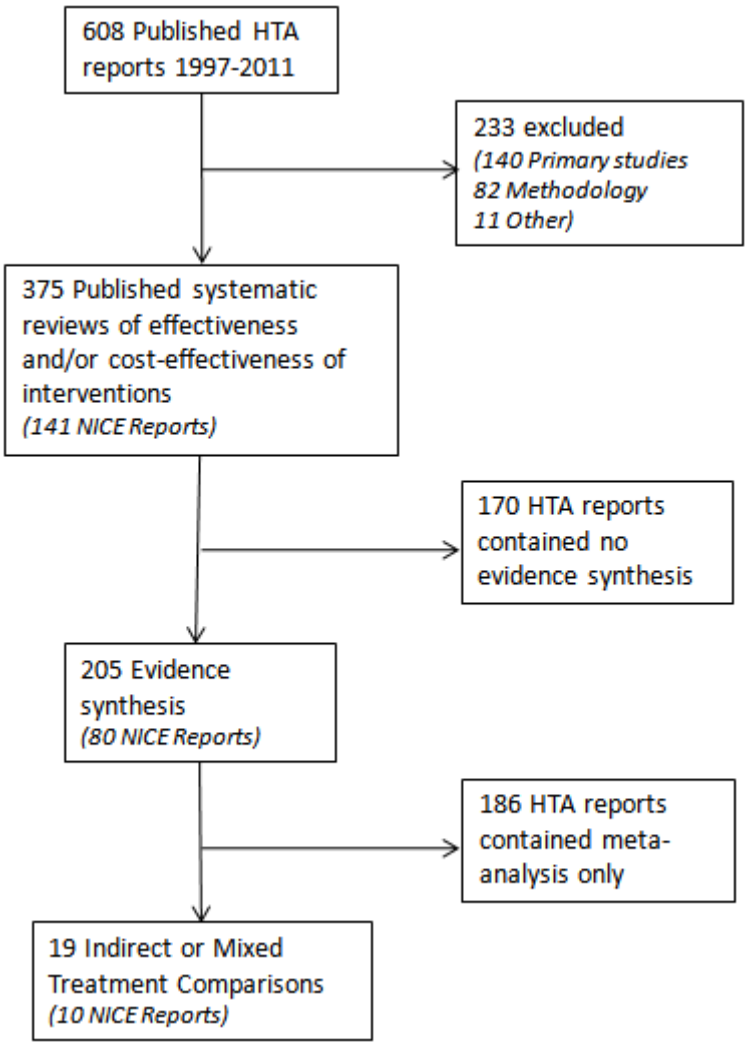

Figure 3.2: Flowchart of HTA review selection

All 19 reports included in the review were published since 2004, the year in which NICE guidance (NICE, 2004) recommended the use of IC analysis (and subsequently including MTC in the 2008 NICE guidance) when no head-to-head RCT data is available for HTA. Figure 3.3 shows the distribution of the reports published by year, sub-categorised as NICE or non-NICE commissioned reports.

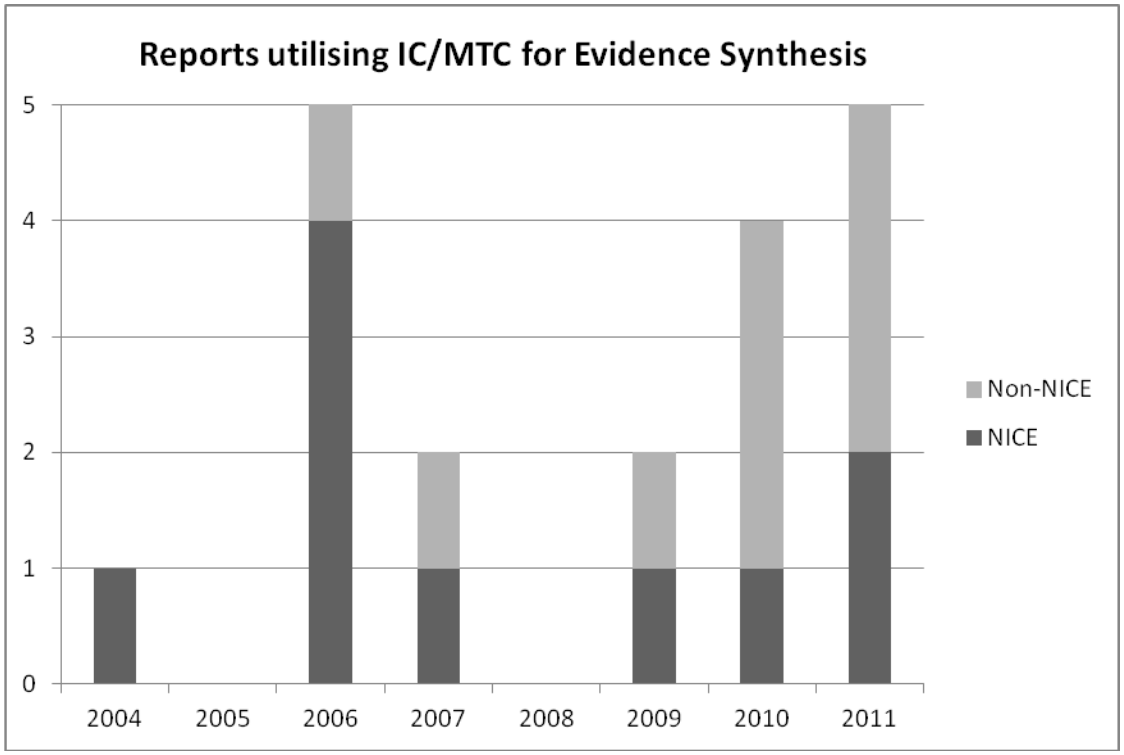

Figure 3.3: Published HTA reports utilizing IC/MTC in evidence synthesis since 2004

Of the 19 reports included in the review, eight reports used IC methods and 11 reports used MTC methods to synthesize the clinical effectiveness data. The evaluation of approaches used for presenting NMA results in this thesis is focused on the presentation of clinical effectiveness results although the use of NMA methodology for the synthesis of adverse events (AEs) was seen in six reports. Eighteen reports included only RCTs in the evidence synthesis; the remaining one report used both observational cohort studies and RCTs but performed sensitivity analysis excluding observational data. The number of interventions used in the NMA analysis ranged from 3 (minimum required for NMA analysis) to 15. There was no major `lumping' of treatment interventions by drug class although varying drug doses were grouped together in some reports. Complete study summary data (unless excluded for confidentiality reasons) used for the analysis were provided for 17 reports. The main reporting characteristics of the reports included in the review are presented in Table 3.1 and are discussed in more detail in the sections which follow.

3.3.2.1 Presentation of NMA data

Eleven of the 19 HTAs used a network table of the format presented in Figure 3.4(a) to display the treatment comparisons (columns) considered by each trial (rows). Only three included all data used in the synthesis as elements in the network table. The rest used ticks, cross marks, shading or patient numbers to indicate what treatment interventions were investigated in each RCT although, in the latter cases, data may have been presented for the relevant trials in other sections of the report.

a)

Study	Intervention			
	A	B	C	D
RCT 1	✓	✓		
RCT 2		✓		✓
RCT 3	✓	✓		
RCT 4		✓		✓
RCT 5	✓		✓	
RCT 6		✓	✓	
RCT 7		✓		✓
RCT 8			✓	✓
RCT 9	✓	✓		
RCT 10		✓	✓	
RCT 11	✓	✓		

b)

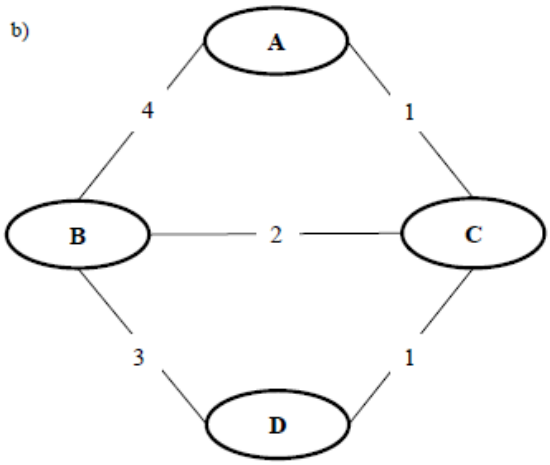

Figure 3.4: Example of: (a) Network Table and (b) Network Diagram

Four reports used network diagrams, similar to that shown in Figure 3.4(b), to display similar information on the treatment comparisons considered by the included trials. This type of diagram graphically displays all the treatment interventions included in the NMA and links these treatments with lines if the comparison of the treatments exists in at least one of the studies. Only two reports used both network table and network diagram. The remaining six reports did not show the network of trials used. Presentation of NMA data in the 19 reports are presented in Table 3.1.

Table 3.1: NMA presentation in the 19 reports

				Data Presentation			Results Presentation					
HTA Report	Appraisal	Method	No. of Interventions	Network Table	Network Diagram	Not presented	Matrix Table	Table (ratio)	Table (rate)	SFP/ CP	Text only	Unclear
Vol 15, No 40. (2011) (Squires)	NICE	MTC[b]	6	✓[@]	✓		✓			✓		
Vol 15, No 39. (2011) (Lewis)	Non-NICE	MTC[a]	15		✓		✓			✓		
Vol 15, No 31. (2011) (Greenhalgh)	Non-NICE	MTC[b*]	4		✓[#]			✓				
Vol 15, No 19. (2011) (Bhattacharya)	Non-NICE	IC[b]	4			✓				✓		
Vol 15, No 10. (2011) (Rodgers)	NICE	IC[b]	4			✓			✓			
Vol 14, No 40. (2010) (Imamura)	Non-NICE	MTC[a]	14	✓				✓		✓		
Vol 14, No 24. (2010) (McKenna)	Non-NICE	IC[c]	4	✓				✓				
Vol 14, No 17. (2010) (McDaid)	Non-NICE	MTC[b*]	4	✓[#]	✓			✓				
Vol 14, No 2. (2010) (Thompson-Coon)	NICE	IC[b]	3			✓		✓				
Vol 13, No 58. (2009) (Burch)	NICE	IC[b]	3			✓		✓				
Vol 13, No 34. (2009) (Ara)	Non-NICE	MTC[b]	5			✓		✓				
Vol 11, No 39. (2007) (Soares-Weiser)	Non-NICE	MTC[b]	8	✓					✓			
Vol 11, No 2. (2007) (Collins)	NICE	IC[b] MTC[c*]	IC: 3 MTC: 8	✓							✓	
Vol 10, No 46. (2006) (Woolacott)	NICE	MTC[b]	8	✓				✓				
Vol 10, No 38. (2006) (Brown)	Non-NICE	IC[b*]	7			✓		✓				
Vol 10, No 23. (2006) (King)	NICE	MTC[c*]	6	✓					✓			
Vol 10, No 9. (2006) (Main)	NICE	MTC[c*]	Analysis 1: 3 Analysis 2: 6	✓^								✓
Vol 10, No 31. (2006) (Riemsma)	NICE	IC[b]	3	✓[@]					✓			
Vol 8, No 19. (2004) (Bridle)	NICE	MTC[c]	6	✓[@]					✓			
Total:	10 NICE 9 Non-NICE	8 IC 11 MTC	Range: 3 to 15	11	4	6	2	9	5	4	1	1

[a]Reported in NMA Chapter; [b]Reported in Clinical Effectiveness Chapter; [c]Reported in Cost-effectiveness Chapter
*Adverse events were analysed and used the same IC/MTC unless otherwise stated
[#]: Reported in Appendix; ^: Reported on 1 table for both analyses [@]: Reported study sample sizes/outcome data/both in network tables.
Abbreviations: SFP: Summary Forest Plot; CP: Caterpillar Plot; RET: Relative Effect Table (e.g. Figure 4c); ET: Absolute Effect Table (e.g. Figure 4d)

3.3.2.2 Presentation of NMA synthesis model and its implementation

All reports discussed the method and the rationale for the use of NMA. Sixteen reports utilized the Bayesian framework for the statistical model. Seven reports presented the statistical model in the main report text while five presented it in an appendix. Four reports did not present the statistical model but referenced published models. The remaining three gave a short description of the statistical model used. Checks for inconsistencies in the NMA network(s) were assessed in two reports using informal methods (formal methods of using deviance information criteria or node-splitting technique (Dias et al., 2010) developed in 2010 were not applied). Summary of the Bayesian evidence synthesis implemented in the 16 reports are presented in Table 3.2.

Table 3.2: Bayesian Evidence Synthesis implementation in the 19 reports

		Bayesian Evidence Synthesis Presentation			
HTA Report	Appraisal	WinBUGS used	Prior defined	Prior used	Convergence checked
Vol 15, No 40. (2011) (Squires)	NICE	✓	✓	Vague & Informative	
Vol 15, No 39. (2011) (Lewis)	Non-NICE	✓	#	Vague	✓
Vol 15, No 31. (2011) (Greenhalgh)	Non-NICE	✓	✓	Vague	✓
Vol 15, No 19. (2011) (Bhattacharya)	Non-NICE				
Vol 15, No 10. (2011) (Rodgers)	NICE	✓	✓	Vague	✓
Vol 14, No 40. (2010) (Imamura)	Non-NICE	✓	✓	Vague	✓
Vol 14, No 24. (2010) (McKenna)	Non-NICE	✓	#	Vague	✓
Vol 14, No 17. (2010) (McDaid)	Non-NICE	✓	✓	Vague	✓
Vol 14, No 2. (2010) (Thompson-Coon)	NICE				
Vol 13, No 58. (2009) (Burch)	NICE	✓			✓
Vol 13, No 34. (2009) (Ara)	Non-NICE	✓	✓	Vague	
Vol 11, No 39. (2007) (Soares-Weiser)	Non-NICE	✓	✓	Vague	✓
Vol 11, No 2. (2007) (Collins)	NICE	✓			
Vol 10, No 46. (2006) (Woolacott)	NICE	✓	#	Vague	✓
Vol 10, No 38. (2006) (Brown)	Non-NICE				
Vol 10, No 23. (2006) (King)	NICE	✓	✓	Vague	
Vol 10, No 9. (2006) (Main)	NICE	✓	✓	Vague	
Vol 10, No 31. (2006) (Riemsma)	NICE	✓	✓	Vague	
Vol 8, No 19. (2004) (Bridle)	NICE	✓	✓	Vague	
	Total:	16	11	14 Vague; 1 Informative	9

#No mention of it in main report text but vague priors were used and listed in the appendix of the HTA report

In view of the computationally intensive nature of NMA analysis, the software used for the analysis was reviewed. Out of the 19 reports, 16 used the WinBUGS statistical software version 1.4 (Lunn et al., 2000, Spiegelhalter et al., 2003). The others did not specify explicitly the software used but reported that either SAS, STATA, RevMan or StatsDirect were used for the presented PWMA results that were reported together with the IC results. For the 16 reports that used WinBUGS software, 14 defined the Bayesian prior distributions. All of these used vague prior distributions, of which five conducted sensitivity analysis to assess the influence of the choice of prior distribution on the results obtained. Checking of the convergence of the MCMC sampler in the Bayesian NMA was reported by nine out of the 16 reports using WinBUGS and 14 reports included WinBUGS codes in their appendices.

3.3.2.3 Presentation of NMA results

NMA effectiveness results were presented for 18 of the 19 reports (the results for the remaining report were unclear). Two had a section specifically for the NMA effectiveness analysis while 13 presented the NMA analysis in the effectiveness and four in the cost-effectiveness chapters of the report (i.e. NMA was only used in the cost-effectiveness modelling). Approaches of the presentation of the NMA results in the 19 reports are presented in Table 3.1.

Sixteen reports presented the results of the NMA in tables, with four of these also presenting either summary forest or caterpillar plots (example given in Figure 3.5).

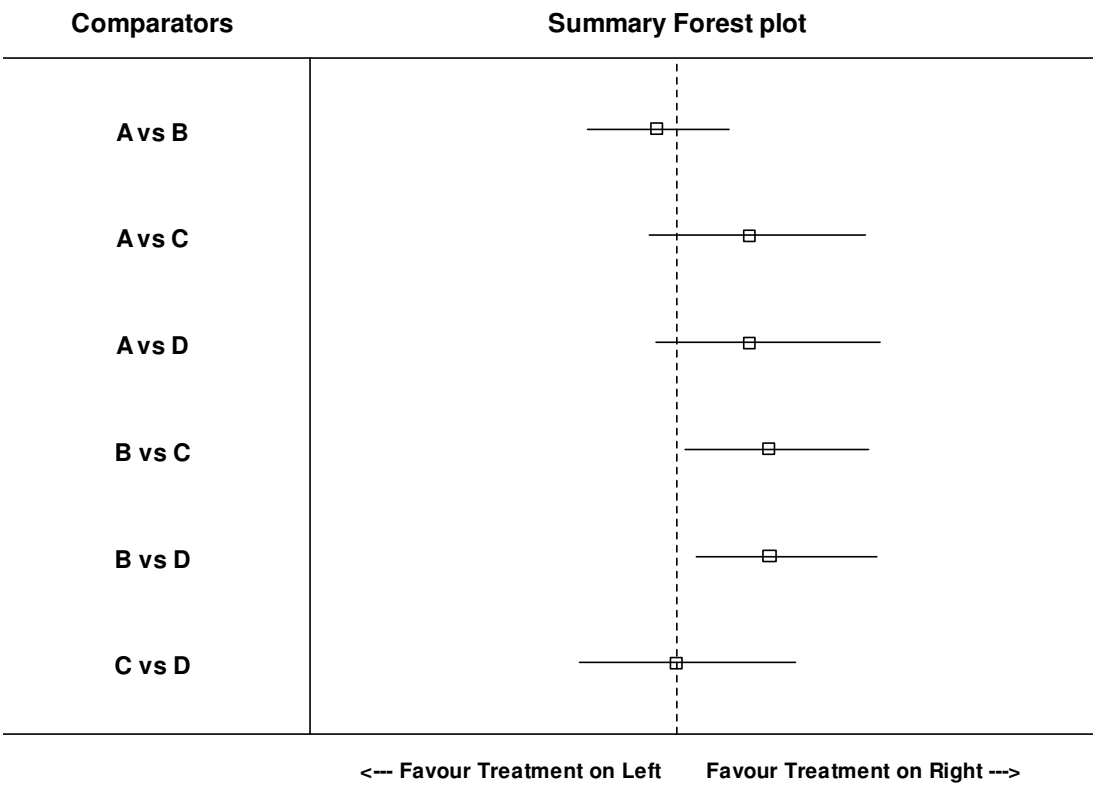

Figure 3.5: Summary Forest plot

Across the 16 reports, three different table formats were used:

- two reports used a matrix table of relative effects (Table 3.3) which contains all permutations of treatment comparisons for both NMA and PWMA, separated by the off-diagonal;
- nine used a relative effect table (Table 3.4) which summarizes pooled ratios/weighted mean differences of selected treatment comparisons relevant to the HTA; and
- five used an absolute effect table (Table 3.5) which presents the posterior probability of rates (for example: response rate) for all treatment interventions in the NMA by the use of a specified underlying baseline rate.

Table 3.3: Matrix table of relative effects

		Network Meta-Analysis		
	Intervention A	$OR_{A\text{-}B\_NMA}$ (95% CrI)	$OR_{A\text{-}C\_NMA}$ (95% CrI)	$OR_{A\text{-}D\_NMA}$ (95% CrI)
Standard Meta-Analysis	$OR_{A\text{-}B\_MA}$ (95% CrI)	Intervention B	$OR_{B\text{-}C\_NMA}$ (95% CrI)	$OR_{B\text{-}D\_NMA}$ (95% CrI)
	$OR_{A\text{-}C\_MA}$ (95% CrI)	$OR_{B\text{-}C\_MA}$ (95% CrI)	Intervention C	$OR_{C\text{-}D\_NMA}$ (95% CrI)
	Not calculable	$OR_{B\text{-}D\_MA}$ (95% CrI)	$OR_{C\text{-}D\_MA}$ (95% CrI)	Intervention D

Key: $OR_{A\text{-}B\_NMA}$= Odd ratio of A vs B using Network Meta-analysis
$OR_{A\text{-}B\_MA}$= Odd ratio of A vs B using Meta-analysis

Table 3.4: Relative effects table

Intervention Comparators		Network Meta-analysis		Standard Meta-Analysis	
Intervention		Mean	95% CrI	Mean	95% CrI
A	B	$OR_{A\text{-}B\_NMA}$	(95% CrI)	$OR_{A\text{-}B\_MA}$	(95% CrI)
A	C	$OR_{A\text{-}C\_NMA}$	(95% CrI)	$OR_{A\text{-}C\_MA}$	(95% CrI)
A	D	$OR_{A\text{-}D\_NMA}$	(95% CrI)	Not calculable	Not calculable
B	C	$OR_{B\text{-}C\_NMA}$	(95% CrI)	$OR_{B\text{-}C\_MA}$	(95% CrI)
B	D	$OR_{B\text{-}D\_NMA}$	(95% CrI)	$OR_{B\text{-}D\_MA}$	(95% CrI)
C	D	$OR_{C\text{-}D\_NMA}$	(95% CrI)	$OR_{C\text{-}D\_MA}$	(95% CrI)

Key: $OR_{A\text{-}B\_NMA}$= Odd ratio of A vs B using Network Meta-analysis
$OR_{A\text{-}B\_MA}$= Odd ratio of A vs B using Meta-analysis

Table 3.5: Absolute effects table

Interventions	Network Meta-analysis		Pair-wise Meta-Analysis	
	Mean	95% CrI	Mean	95% CrI
Intervention A	Eff_{A_NMA}	(95% CrI)	Eff_{A_MA}	(95% CrI)
Intervention B	Eff_{B_NMA}	(95% CrI)	Eff_{B_MA}	(95% CrI)
Intervention C	Eff_{C_NMA}	(95% CrI)	Eff_{C_MA}	(95% CrI)
Intervention D	Eff_{D_NMA}	(95% CrI)	Eff_{D_MA}	(95% CrI)

Key: Eff_{A_NMA}= Absolute Treatment Effects estimates of A using Network Meta-analysis
Eff_{A_MA}= Absolute Treatment Effects estimates of A using Pair-wise Meta-analysis

Thirteen HTAs reported comparative effectiveness estimates, of which eight reported all permutations of pair-wise comparison results from the NMA analysis, five concentrated either on active treatments compared with placebo (or no treatment or standard care), or on active treatments of interest compared to one another.

NMA results were reported either as pooled relative effects (ORs, hazard ratios, weighted mean differences) or rates (response rates, withdrawal rates) with either 95% confidence interval (CI) (for frequentist evidence synthesis) or 95% credible interval (CrI) (for Bayesian evidence synthesis). Five HTAs reported the probability each treatment was the most effective of which four of these presented the "best" statistic in tables and one reported it only in the main text of the report. None of the published reports provided tables or graphs that ranked the technologies in terms of effectiveness as has been presented elsewhere (Salanti et al., 2011).

All except two reports used the NMA evidence synthesis results to inform their economic decision model. Of those that did, one used a subset of RCTs used for effectiveness and another used more RCTs than those included in the effectiveness evaluation.

3.3.3 Recommendations to improve reporting of NMA

As mentioned in Section 3.2.1, recommendations for NMA reporting need to consider three domains: data, model and results. With a focus on each of the three domains, recommendations formulated based on institutional guidance (Section 3.3.1) and the review results (Section 3.3.2) are presented in Table 3.6.

In any statistical analysis, the credibility of the results depends on the quality of the data used and the appropriateness of the model adopted. It is, therefore, imperative that a clear description of the data and statistical models is presented to ensure transparency and reproducibility. Therefore, it is recommended that studies used in the NMA should be clearly presented with, at a minimum, references properly cited, and the data used in the analysis clearly stated or outcome data (e.g. for a binary outcome, number of events and number of patients in arm) from which the outcome measures used (e.g. log-odds ratios & standard errors) can be calculated. A tabular format is a good way of presenting this information. A network diagram is an excellent way of visualizing the relationships between the studies and interventions under evaluation. Since this could be derived from a tabular description of all the studies, of the form described above, it is not strictly essential, but where space allows its inclusion is encouraged. However, if any included studies have more than two treatment arms, then it is not possible to derive a network table from a network diagram, in addition to not being able to identify citations to specific studies.

For the statistical model, this should be fully described with associated algebra together with analysis code (including data with a data code sheet explaining the data structure) either in the main text or as an appendix. If report space is limited, then either citation to the model specification published elsewhere would be required or supplementary material describing the model provided online.

Table 3.6: Table of recommendations for NMA reporting

Component	Qualities	Presentation Method Recommended [Domains fulfilled]	Target audience
Data	D1. Studies included in analysis clearly presented & references cited D2. Treatments compared in each study reported D3. Treatment data / effect sizes (with a measure of uncertainty) for each included study reported	Network table, with the *structure* of Figure 3.4a, but with outcome data presented in the table cells/references to studies included. Network table provides transparency of the trial data used in the Analysis [D1-D3]	Academics Decision-makers
Model	M1. Statistical model described typically with algebraic representation M2. Citation given to the model used Analysis code presenting model (and data)	Full analysis codes including data with data code sheet. This can be provided in Appendix or as online supplementary materials. This allows reproducibility of the results and contains the model used and data included in the analysis [M1-M3]	Academics Statistical analysts
Results	R1. Results of interest (given the aim of the study) reported R2. (Relative) Comparison of all treatments R3. Probability best statistics/Ranking of treatments	Tables (see Table 3.4 for an example) and summary forest plot of results (see Figure 3.5) of interest [R1] Matrix Tables (See Table 3.3 for an example) with Summary Forest Plot/Caterpillar Plot (See Figure 3.5 for an example) [R1,R2] Tables or 'rankogram' presenting probability best statistics and ranking [R3]	Academics Decision-makers

The presentation of results from the analysis relies largely on the question of interest: which treatment is best, what is the treatment effect for a specific comparison, or perhaps all comparisons with either usual care or all treatments in the network being of interest. NMA analyses can answer these questions by providing estimates of: the treatment effects for any comparisons included in the network; the probability that each treatment is best; and the ranking distribution of each treatment. It may not be feasible and/or desirable to report all of them in a report manuscript and therefore the focus of the analysis should be well defined to guide the choice of the most appropriate statistics to report. This issue strengthens the desirability for making data/analysis code available to enable readers to obtain results for any aspect of the analysis not reported. Hence, prescriptive requirements on reporting results are not possible; however, the use of both graphical and tabular approaches to reporting is encouraged for ease of interpretation. This is consistent with the most recent NICE guidance (NICE, 2013) published in 2013.

The above recommendations should be used in conjunction with the good practice in NMA methods documents recently published by an ISPOR Task Force (Hoaglin et al., 2011, Jansen et al., 2011) and NICE's Decision Support Unit (Dias et al., 2011c, Dias et al., 2011d, Dias et al., 2011a, Dias et al., 2011f, Dias et al., 2011e, Dias et al., 2011b, Ades et al., 2012). The latter provides a checklist for reviewers' of NMAs and goes beyond the issue of methods and results reporting including sections on the definition of the decision problem and embedding the synthesis in a probabilistic cost-effectiveness model.

Finally, the choice on what type of results to report and which tools to use depends on the audience. For example, whereas academics may be interested in all three components of the NMA analysis, the statistical analysts would be expected to focus more on the model specification and decision makers on the transparency of the data and clarity of the results of interest. Using the recommendations formulated from the review of HTA guidance notes and reports, important NMA results for the design of NMA graphical tools were identified and are described in the next section.

3.3.4 Important NMA results identified for designing graphical tools

The review presented in Section 3.3.2 highlighted that the most often reported NMA results included relative effects of comparative pairs of interventions, absolute effects of interventions, and probability best; all of which are recommended in the published NMA methods guidance documents (NICE, 2004, NICE, 2008, NICE, 2013) by agencies such as the National Institute for Health and Care Excellence (NICE) or International Society For Pharmacoeconomics and Outcomes Research (ISPOR) (Hoaglin et al., 2011, Jansen et al., 2011).

Another statistic used in the reporting of NMAs, although not reported in the HTA reports reviewed, is the order of preference of an intervention among a number of interventions (i.e., the ranking of an intervention, where the probability that an intervention is rank 1 is the probability best statistic). The ranks may be presented as summary statistics (e.g., mean/median rank, surface under the cumulative ranking curve(SUCRA) (Salanti et al., 2011)), or graphical representations of the distribution of ranks (e.g. rankograms / barplots) indicating the probability that a given intervention is 1st, 2nd, 3rd best, to being the last (the worst intervention) when compared to all other interventions in the network. In addition to the above, PWMA results are reported in the HTA reports, sometimes alongside NMA results to allow informal consistency checks to be made. Prediction intervals (the interval indicating the likely location for the underlying effect in a new study), although not routinely reported, have recently been advocated (Riley et al., 2011) for the reporting of the impact of heterogeneity in evidence synthesis.

Based on this identified set of important NMA results, which should be routinely communicated, graphical tools were designed to facilitate concise and transparent reporting of NMA results. The graphical tools developed were evaluated by experts at a MiM meeting and results of the assessment are presented in Section 3.3.5.

3.3.5 Evaluation of graphical tools by experts

Graphical tools that aim to present NMA results in a clear and concise manner that combine both graphs and numerical estimates for optimal interpretation of NMA results and with built-in alternative display options to satisfy the needs of different audiences

were evaluated by experts at a MiM meeting. Opinions about the graphical tools from statisticians/quantitative experts were collected using the evaluation form presented in Appendix B. Four graphs were presented at the meeting, namely: summary forest plot matrix (SFP Matrix), summary forest plot table (SFP Table), summary forest plot pie 1 (SFP Pie 1) and summary forest plot pie 2 (SFP Pie 2). Reviewers who answered the evaluation form were requested to comment on the content and clarity of the graphs in reporting NMA results.

Sixteen experts returned the evaluation form, eight indicated that they had been involved in at least one NMA analysis while seven had not and one did not answer the question. In terms of the presentation of the content of NMA results, both the SFP Matrix and SFP Table scored the highest with a median score of 9. The SFP Pie 1 and SFP Pie 2 had median score of 6 and 7 respectively. With regards to the clarity of the results presented. The SFP Table had the highest median score of 8; the SFP Matrix, SFP Pie 1 and SFP Pie 2 had median score of 7, 6 and 5 respectively. One statistician did not score the graphs but stated that he/she preferred the SFP Table. In summary, the SFP Table scored the highest in terms of both content and clarity in presenting NMA results. Eleven experts indicated that they will consider using the SFP Table, eight for the SFP Matrix, one for the SFP Pie 1 and none for the SFP Pie 2. When asked what other important information should be considered, inconsistency was mentioned by seven experts, ranking by five, absolute effects by two and sample size and multiple outcome by one each.

Based on the comments and suggestions from the experts, the SFP Matrix and SFP Table were revised. In view of the interest in presenting rankings and the complexity of the SFP Pie graphs (which were originally designed with rankings and the probability best statistics in mind), a new graph, Median Rank Chart, to present ranking was developed to replace the SFP Pie graphs. These three graphs (SFP Matrix, SFP Table and Median Rank Chart) form the final set of graphical tools published and disseminated for use in presenting NMA results and are presented in Section 3.3.6. The SFP Pie graphs evaluated at the meeting are presented in Section 3.3.7.

3.3.6 Graphical tools for NMA reporting

Three graphical tools were developed that aim to amalgamate the important NMA results, identified in Section 3.3.4, to aid readability and maximise interpretation in NMA reports. Two of the graphs present relative effects of comparative pairs of interventions, probability best, ranking statistics (using optionally either rankogram or SUCRA percentages), and heterogeneity estimates. They also present the results of the PWMA alongside the NMA results to allow informal checks for consistency of results to be made easily. The primary aim of the third plot was to give a simple summary of the order of preference of interventions in terms of effectiveness.

The different graphical displays were developed with different target audiences in mind. With academics and statisticians/analysts in mind, the main objective was to graphically present all key NMA results on a single graph whilst ensuring interpretability through clear presentation; this also aimed to help meet restrictions on the number of tables and figures often enforced by research journals (using graphs 1 and 2 below). While completeness of NMA results presentation may be desired by the academics and analysts, clinicians and decision makers in healthcare are more likely to be interested in visualising the overall conclusions of the analysis by presenting the rankings of all interventions in terms of their effectiveness (i.e. highlighting "top-ranking" interventions) (graph 3).

Illustrative examples of graph 1, graph 2 and graph 3 presented on the OR scale, where the NMA outcome is binary, are presented in the next three sections respectively. An example of its use for the presentation of NMA results on continuous outcome data is shown in Appendix C. Functions for creating the graphs are made available for download by the public from the departmental website: https://www2.le.ac.uk/departments/health-sciences/research/biostats/sb-supplementarymaterials/nma-graphics.

3.3.6.1 Graph 1: Summary Forest Plot Matrix (SFP Matrix)

Description

The first plot, referred to as the Summary Forest Plot Matrix, is shown in Figure 3.6. The plot design is similar to a scatterplot matrix often used for the investigations of correlations. Along the diagonals, the interventions included in the network are displayed. These interventions may be ordered, for example, by their median rank, as is done here, to highlight the most relevant comparisons by placing them at the top of the graph. Below the diagonal, in the lower triangle of the plot, summary forest plots for all possible combinations of the intervention pairs analysed in the NMA - in black colour - are presented above the PWMA results - in grey colour - to aid visual assessment of consistency between the two analyses (the intervention labelled horizontally to the right of the plot is compared with the intervention labelled vertically above and clear labelling of the axes is given for each 'plot element' on the bottom of the Matrix). The summary forest plots display the point estimates of effect size (drawn as a square) with 95% CrIs (or CIs when only one RCT was used for presenting the PWMA results) and 95% prediction intervals (shown by two-tiered error bars). Any summary plot without a grey-coloured estimate indicates a comparison for which no head-to-head trials exist. The corresponding numerical estimates of comparative effectiveness are presented above the diagonal in the upper triangle and are presented as a "mirror image" to the summary forest plots taking the diagonal as the mirror-line. To assist in understanding the heterogeneity of the studies in the network, the numerical estimate of between-study variance (i.e. heterogeneity) is reported below the matrix. Alternating shadings of each plot element is used to improve readability (a technique often used in rail/bus timetables). Also included in the matrix, along the diagonal, are the median ranks together with rankograms which provide the full probability distribution of rankings for each intervention.

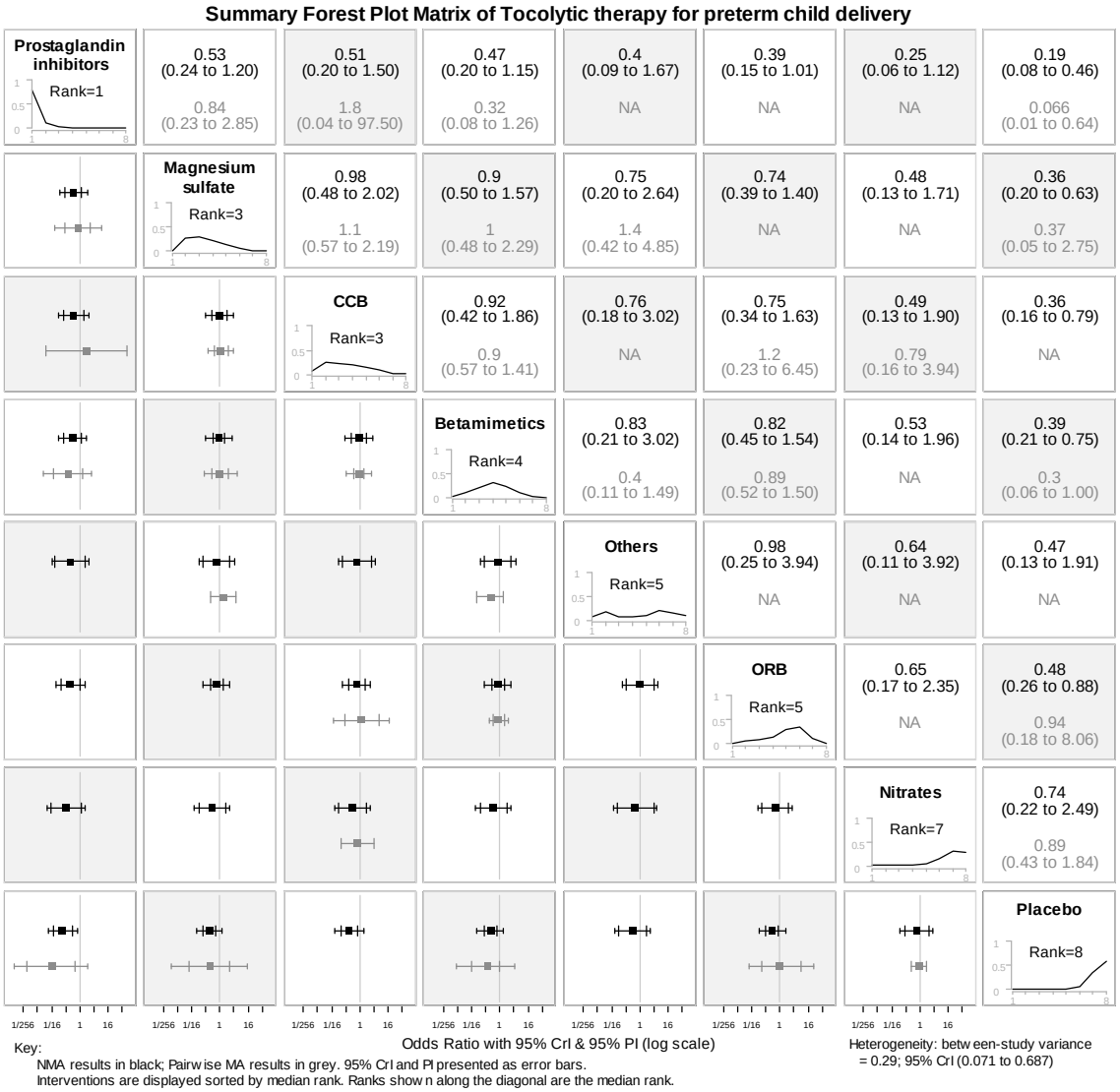

Figure 3.6: Summary forest plot matrix

For the example in Figure 3.6, the drug class of prostaglandin inhibitors (row 1) is most likely the best intervention with the highest median rank (1) and probability best statistic as shown on the rankogram (0.80 - the height of the density at rank 1 (x-axis)). Its effectiveness relative to the interventions which are ranked 2nd and 3rd based on median rank (Magnesium sulfate and Calcium channel blockers (CCB)) is given by the OR of 0.53 (95% CrI: 0.24 to 1.20) and 0.51 (95% CrI: 0.20 to 1.50) respectively. The lower triangle of the SFP matrix allows the reader to easily identify pairs of interventions for which there were no head-to-head trials. In the example presented in Figure 3.6, the drug class of prostaglandin inhibitors is compared directly, in head-to-head trials, with all interventions except for the drug classes: others, oxytocin receptor

blockers (ORB) and nitrates (as indicated by the lack of PWMA estimate below the NMA estimate). This graph allows readers/clinicians to interpret key results of a NMA using a single plot, to compare the NMA and PWMA results and to identify which pairs of interventions could not be compared in a PWMA (due to a lack of head-to-head trials). These functionalities, together with the predictive intervals presented in the graphs could be used to guide potential areas of future clinical trials/epidemiological research studies.

Advantages and Limitations

Traditional forest plots display individual study effects together with the summary estimates to enable readers to assess the effects of each study, how different they are from one another and from the summary estimates, as well as their influence on the summary estimates. As much as it is desirable to display individual studies used in a NMA, it is cumbersome as the number of studies included in a NMA can often be large. Instead the graphical tools developed aimed to use the traditional forest plots, familiar to a great number of audiences in the medical area, to display the summary estimates from both NMA and PWMA and placing them side-by-side. The deliberate placement of the PWMA alongside the NMA results is to allow clinicians to directly address the question that naturally arises with NMA, that is, how different the results of NMA (that uses a network of trials) are compared to the results of traditional PWMA of head-to-head trials.

Along the diagonal, key NMA summary statistics (such as the median ranks together with rankograms which provide the full probability distribution of rankings for each intervention) which are commonly reported separately are included in the graph. By including these statistics on the same plot as the relative effects, it enables the reader to instantly identify which intervention is most likely to be the best and read its comparative effects with all the other remaining interventions (ordering on the median rank statistics further facilitates this by ensuring the “best” interventions are placed at the top of the plot).

Due to the matrix square design of the SFP Matrix, it works best for networks that are of moderate size (< 10 interventions). Displaying NMA results of larger network will

evidently require the SFP matrix to be separated into pages in the multiples of 2, hence reducing the readability and ease of interpretation of the NMA results. As such, options are created to sort in terms of key NMA summary statistics and print a user-specified range of interventions which will be discussed in greater details in the display options section.

Display Options

The SFP Matrix shown in Figure 3.6 is one variant of the many that can be displayed. In its simplest form, the SFP Matrix contains only the NMA and PWMA summary forest plots and estimates (with 95% CrI), with only the treatment names displayed along the diagonal and the heterogeneity estimates presented. Predictive intervals as shown in Figure 3.6 can be optionally included in the graph. Further NMA results components such as the ranking statistics and probability a treatment is best can be optionally included in the graph and displayed along the diagonal as shown in Figure 3.6 where the rankograms were displayed. Ranking statistics can be displayed in the form of (i) rankogram with median rank, (ii) bar chart with mean rank or (iii) the SUCRA estimates with cumulative ranking probability plots. Probability a treatment is best will be displayed with a pie chart with the probability estimates.

Apart from the display of key NMA results components, options to sort or reduce the number of interventions displayed in the graphs are available (with caution notes displayed as footnotes in the graphs to remind readers of the actual number of interventions used in the NMA to produce the displayed results). Although the presentation of all pairwise comparisons in the network is highly recommended, it is also acknowledged that there can be situation where it is necessary to display a reduced set of interventions, especially in the case of large networks. It may be helpful to clinicians and decision makers to restrict presentation of the NMA results to that of the top 5 or 10 ranking interventions when the network contains say 20 interventions or more. The NMA components that can be used for sorting the results are (i) median rank; (ii) mean rank; (iii) SUCRA percentages; (iv) probability a treatment is best and (v) relative treatment effect compared to the treatment coded as 1 (which is commonly placebo or standard of care) in the analysis. Footnotes in the graphs can also be

removed where it is necessary. Illustrative examples of SFP Matrix plots in its simplest format and with SUCRA percentages are shown in Appendix D.

3.3.6.2 Graph 2: Summary Forest Plot Table (SFP Table)

Description

The second graph, referred to as the Summary Forest Plot Table, is shown in Figure 3.7. This plot uses the presentational style of the traditional forest plot where the numerical estimates are reported alongside the summary forest plots. The SFP Table presents results for all possible combinations of intervention comparisons with each intervention in the second column compared to the intervention listed in the first column. Similar to the SFP Matrix above, the interventions have been ordered by their median rank. The third column reports the number of head-to-head (H-H) trials that compare the two interventions listed in columns 1 and 2 (a feature not incorporated in the SFP Matrix). In column 4 the numerical estimates of the relative effects with corresponding 95% CrI are presented for the NMA with the PWMA results directly below in grey. Finally, column 5 presents the summary forest plots; again the PWMA results are presented below the NMA results to allow visual assessment of consistency between the two analyses. Similar to the SFP Matrix, the display of the predictive intervals alongside the CrIs on the summary forest plot is optional. An estimate of heterogeneity across the trials included in the network is also presented. Median ranks of all interventions are also reported in this graph, numerically and graphically using a slider bar format (full rankograms, as presented on the SFP Matrix, were problematic for the SFP Table and difficult to read).

Advantages and Limitations

One advantage of this plot over the Matrix format is that the reference line of the summary forest plots for all pairwise comparisons is drawn on the same vertical line, hence facilitating the assessment of differences in comparative estimates and their precision between treatment pairs. Another advantage of this reporting style is that the

NMA results from large networks can be reported more easily with the SFP Table extending to multiple pages, where necessary.

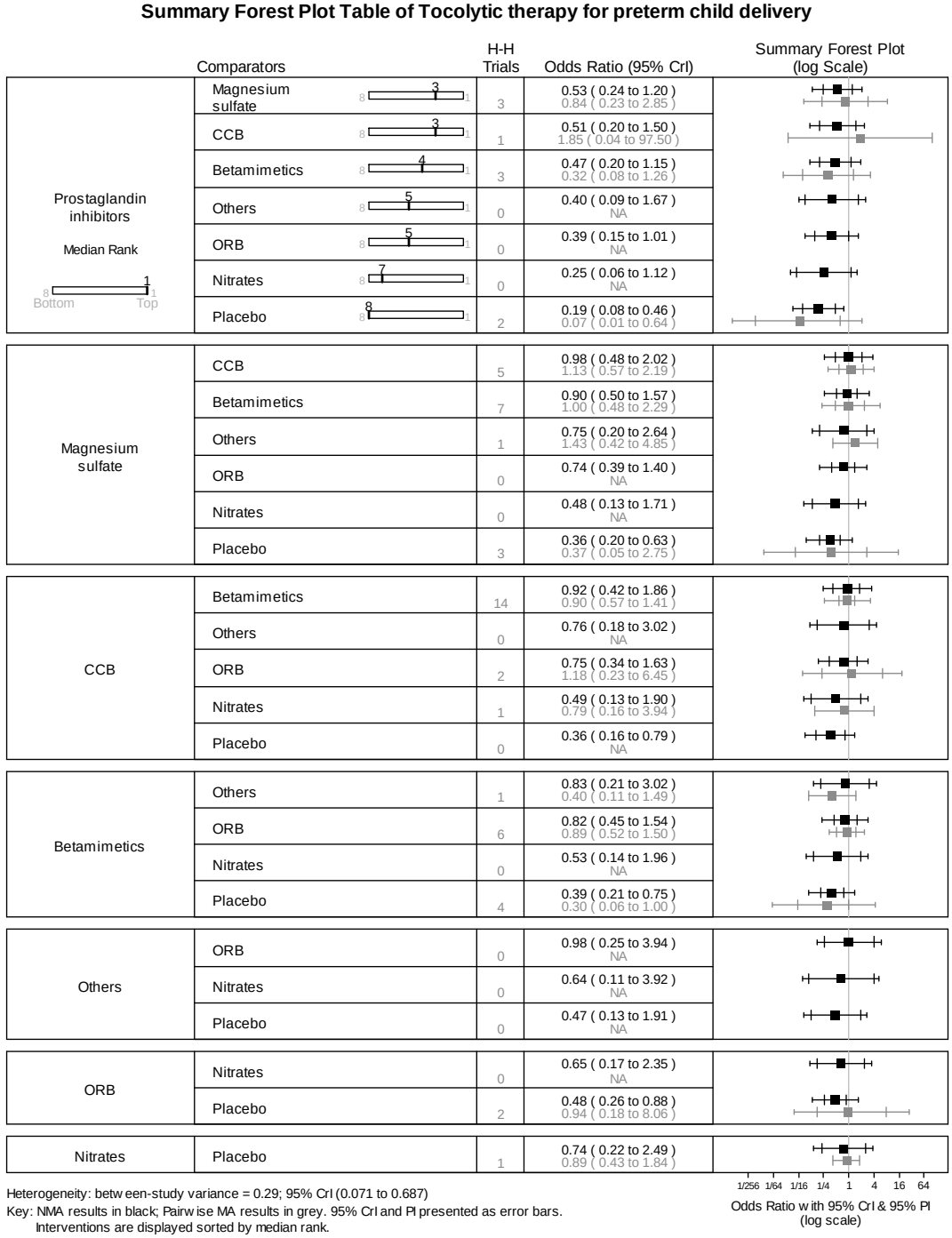

Figure 3.7: Summary forest plot table

Key NMA summary statistics such as the median rank, mean rank, SUCRA percentages and probability best statistics are presented in the top first box of NMA results. This is a result of the reduction in comparative pairs by one as the primary

comparator moves to the next intervention in the NMA. As such for a NMA with eight interventions (like in our example data), there will only be seven boxes on the SFP Table because the last intervention (in our example – Placebo) would have been compared to all preceding interventions and will not be listed in column 1 (the primary comparator column). Besides not having the last intervention in column 1, the size of the boxes decreases as it moves towards the next intervention, making it challenging to include rankograms on the graph. This is a limitation of this graph design but is also an advantage of this graph as the key NMA summary statistics had to be placed in the top box and this, in turn, allows readers to compare the interventions without having to flip through pages of the table when the network is large.

Display Options

The SFP Table shown in Figure 3.7 is one variant of the many that can be displayed. In its simplest form, the SFP Table contains only the NMA and PWMA summary forest plots and estimates (with 95% credible interval), and the number of head-to-head trials for each pair of intervention comparisons. Predictive intervals as shown in Figure 3.7 can be optionally included in the graph. Other NMA results components such as the ranking statistics and probability best can be optionally included in the graph and displayed in the first set of intervention comparisons as shown in Figure 3.7 where the median ranks were displayed. Choice of display of the ranking statistics are (i) median rank presented using slider bar, (ii) mean rank presented using slider bar and (iii) SUCRA percentages. Probability a treatment is best is displayed with a pie chart alongside the probability estimates.

Similar to SFP Matrix, options to sort or reduce the number of interventions displayed in the graphs are available (with caution notes displayed). NMA results components that can be used for sorting the results are (i) median rank; (ii) mean rank; (iii) SUCRA percentages; (iv) probability a treatment is best and (v) relative treatment effect compared to the treatment coded as 1 (which is commonly placebo or standard of care) in the analysis. Illustrative examples of SFP Table plots in its simplest format and with SUCRA percentages are shown in Appendix D.

3.3.6.3 Graph 3: Median Rank Chart

Description

The third graph, shown in Figure 3.8, presents the median ranks of all interventions included in the NMA with the aim to 'simplify' the presentation of rankings in NMA. A colour intensity scheme is employed in this graph to help draw attention to the best treatment(s) (using black ink in the lightest zone at the top of the chart) while simultaneously highlight the worst treatment(s) (in the darkest zone at the bottom of the chart). In our example (Figure 3.8), prostaglandin inhibitors are most likely to be the best with a median rank of 1, while nitrates and placebo are the worst, and the five other interventions have similar rankings between these extremes.

Median Rank Chart of Tocolytic therapy for preterm child delivery

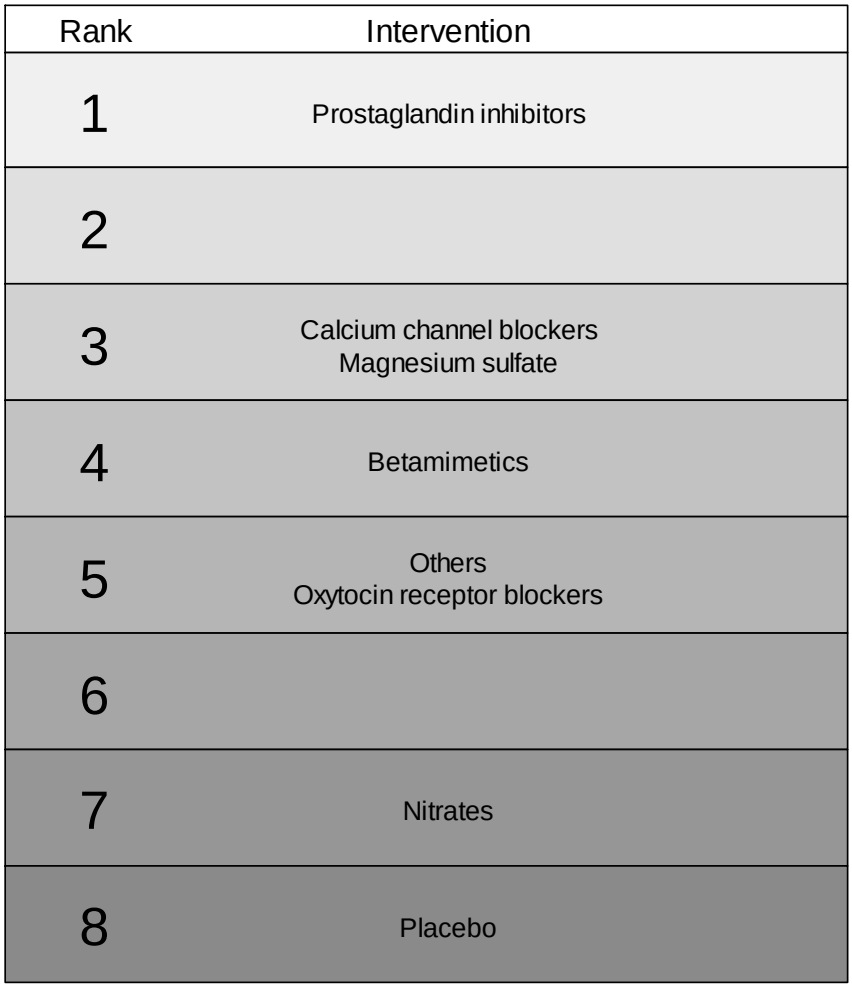

Figure 3.8: Median rank chart

Advantages and Limitations

As this graph allows all interventions included in the NMA to be presented in a single graph that can be printed on a single page, it is a particularly useful graphical tool when the network contains a large number of interventions. This graph provides readers with only the median ranking of the interventions; hence unlike the SFP Matrix and SFP Table, this does not provide quantitative information on the differences in efficacy estimates between the interventions included in the analysis.

3.3.7 Other graphical tools evaluated by experts

In the preliminary development of the graphical tools, other graphical tools were also developed. However, following the evaluation of the graphical tools by experts, two of these graphs did not contribute to the recommended graphical tools package. Descriptions and limitations of these two graphs are discussed in Section 3.3.7.1 and Section 3.3.7.2.

3.3.7.1 Summary Forest Plot Pie 1

Summary Forest Plot (SFP) Pie 1 is designed to present the probability best statistics estimated in NMA analysis and is presented in Figure 3.9. As the probability best statistics for all interventions sum to one, the thicknesses of the pie slices are proportional to the probability each intervention is the best (with the added benefit that the numerical estimates of the probability best are also displayed). Interventions are ordered by the probability best with the interventions with probability best less than 0.015 grouped together and presented as a single intervention named "Other interventions" (although this threshold for grouping low probability best interventions can be user-specified).

The circumference of the pie represents the line of no relative effect, analogous to the reference line (at OR=1 for binary outcomes, WMD (Weighted Mean Difference) = 0 for continuous outcomes) commonly seen in (summary) forest plots (albeit curved into a circle in this context). The summary forest plots for each of the interventions, relative to a chosen reference intervention, are plotted at the mid-angle of that intervention's pie

slice (note that the relative effect estimates for interventions with probability best statistics less than 0.015 are not displayed on the plot). The further an intervention's effect size lies outside the circumference of the pie, the better the intervention is relative to the chosen reference intervention. In the example, presented in Figure 3.9, it can be observed that the drug class of prostaglandin inhibitors has the largest relative effect with reference to placebo as well as the highest probability of being best (indicated by the widest pie slice). As with the SFP Matrix and the SFP Table, both credible and predictive intervals can be displayed. An estimate of heterogeneity is also included on the plot for completeness. To make this novel format of plot accessible, a clear legend is given explaining the various components. The key states the number of interventions included in the NMA, the reference treatment and the threshold of probability best used to group treatments into 'Other interventions' category. It also lists the treatments included in this 'Other interventions' category.

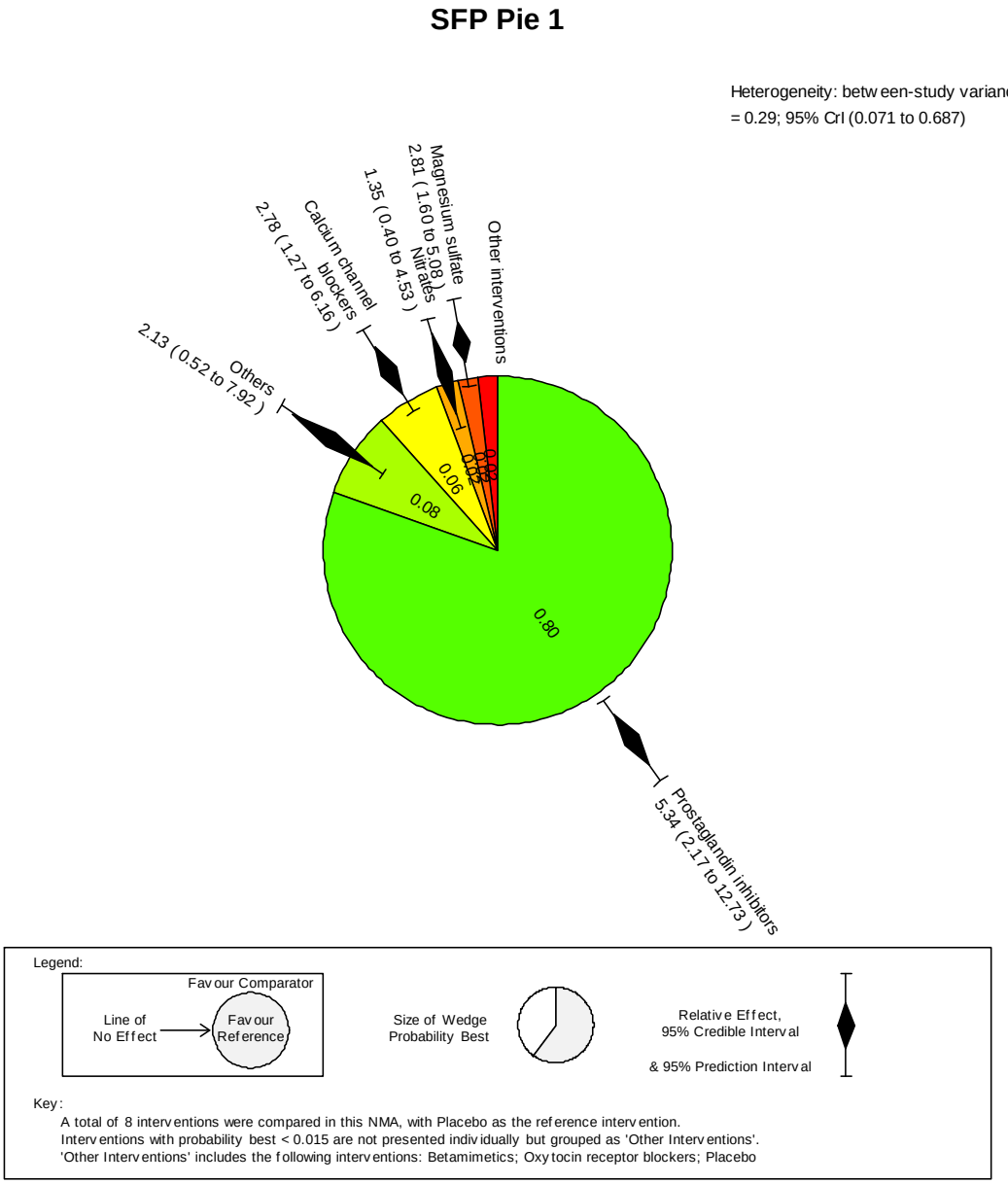

Figure 3.9: Summary forest plot Pie 1

A major limitation of this graph is that the graph can only present NMA results relative to only a chosen reference intervention, results for other reference interventions need to be plotted separately. Differences between the pairwise results are not immediately apparent without the help of the numerical summary estimates printed on the graph as it is more difficult to visualize differences along a circle and having circular grids on the graph makes the graph looks very busy.

3.3.7.2 Summary Forest Plot Pie 2

The Summary Forest Plot Pie 2 is a variant of the SFP Pie 1 graph that looks into the presentation of the probability best statistics estimated in NMA and is presented in Figure 3.10.

Interventions are ordered by the probability best with the interventions with the highest probability best starting at the 12 o'clock position clockwise to the interventions with the lowest probability best statistics. Multiple interventions with low probability best can be set to be grouped together as in the SFP Pie 1 graph. The circumference of the innermost pie has the same meaning as in the pie in the SFP Pie1 graph. Similarly, the summary forest plots for each of the interventions, relative to a chosen reference intervention, are plotted at the mid-angle of that intervention's pie slice. A legend describing various components of the graph is also provided.

One major difference between the SFP Pie 1 and SFP Pie 2 graphs is that the coloured area in SFP Pie 1 only takes into account the magnitude of the probability best for each intervention, while the coloured area in SFP Pie 2 represents not only the probability best for each intervention but also the precision of the summary estimate. The limitations of this graph are similar to those of SFP Pie 1.

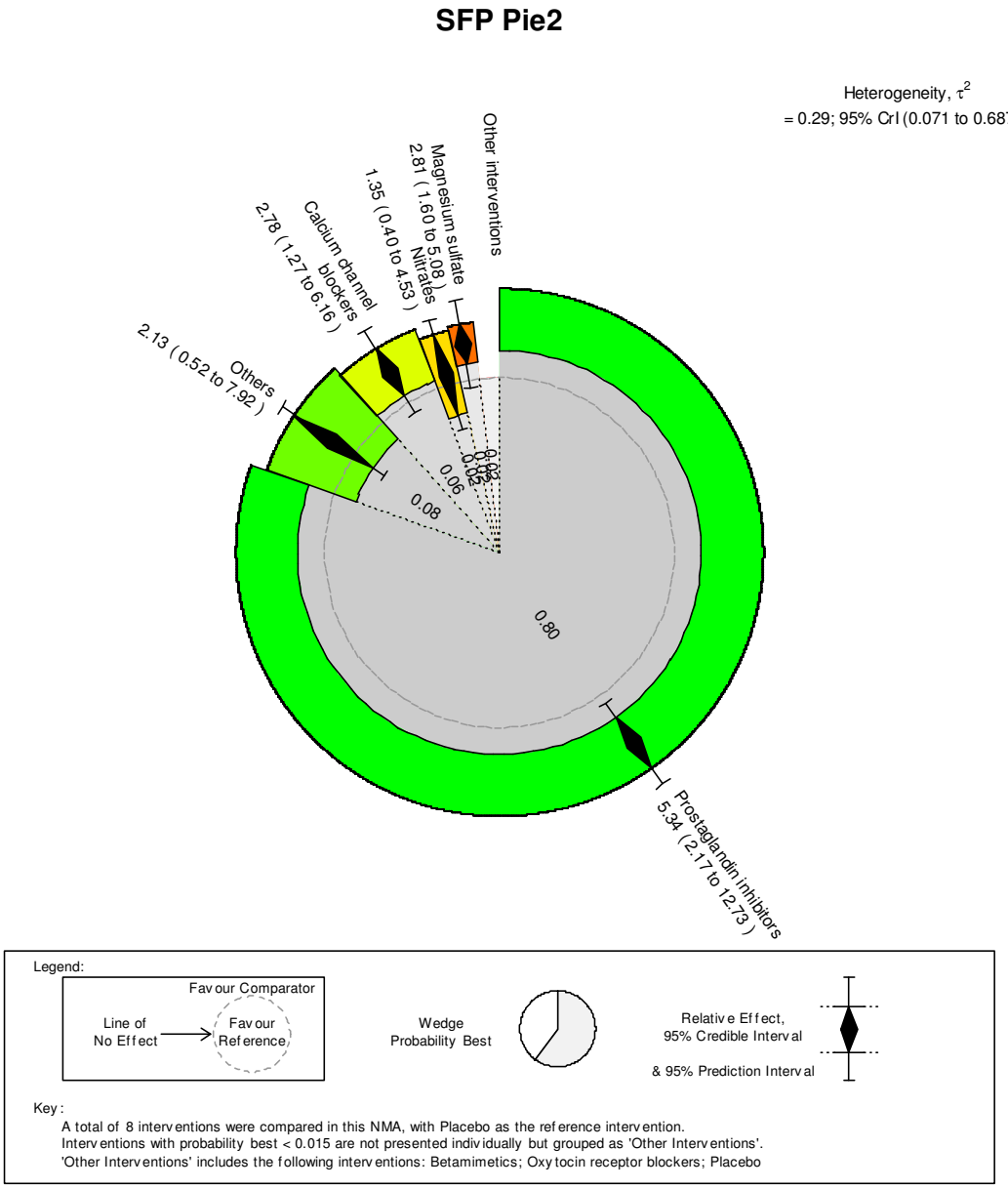

Figure 3.10: Summary forest plot Pie 2

One major issue of the design of this graph is that the area of the coloured area is inversely proportional to the precision of the summary statistics estimate of the NMA. As a result, an intervention having a summary estimate with low precision (wide 95% CrI/CI) will have a larger coloured area than another intervention with the same probability best statistics but a summary estimate with high precision (narrow 95% CrI/CI). This is inconsistent with graphical design standards (Cleveland, 1994, Few,

2004, Tufte, 2001) where important results should be given more data ink, which in this graph design is the area of coloured slices.

3.4 Discussion

The review summarises how NMA data, analysis methods and results are presented in UK HTA reports. It has shown that although the use of NMA in HTA has increased since the 2004 NICE guidance first advocated its use, there appears to be no standard tabular and/or graphical format for the presentation of the evidence structure or results. About a third of the reports reviewed did not present a network diagram or network table making it difficult to ascertain which studies and comparative evidence were used in the NMA synthesis and thus making the analysis neither transparent nor reproducible. Results were mainly presented in tabular format with a few also reporting summary forest plots. Although the latter is a well-established graphical tool in PWMA, its use is restricted in NMA as it is limited to the presentation of pooled results for pairwise comparisons (i.e. presentation of individual studies are not included). Only five HTA reports presented the probability that each treatment included in the network was the most effective, however, little consideration was given to other rankings; i.e. probability second most effective, third most effective and so on. The majority of reviewed HTAs implemented the NMA analysis in WinBUGS which has limited graphical functionality and therefore may require results to be exported to other statistical packages such as R for clearer presentation.

After the completion of the review, two further systematic reviews of the international literature of 42 NMA (Coleman et al., 2012) and 121 published NMA (Bafeta et al., 2014) were in broad agreement with the findings of this review. Coleman and colleagues found that the presentation of results was much more likely to be tabular (89.5%) rather than graphical (21.1%) formats. Bafeta and colleagues concluded that there were great disparities in the reporting of NMA results and essential components of reporting NMA results were missing in 98% of the articles. Thus, while the sample of NMAs in my review was limited to UK HTA reports, it is reassuring that a review of reports from a wider area obtained similar findings. This further strengthens the need

for additional guidance and presentational tools for reporting NMA results to aid ease of interpretability.

An initial attempt to provide recommendations regarding the presentation of NMA analyses was made. However, it should be acknowledged that no recommendation was developed for the reporting of advanced issues such as variable study quality, subgroup analysis or inconsistency analysis, due to the infancy of this methodology. After the publication of the review and recommendations (Tan et al., 2013), Hutton and colleagues conducted an overview of published reviews that discussed the quality of reporting NMA and provided suggestions to enhance future reporting quality of NMA (Hutton et al., 2014). They identified eight reports (including the review conducted in this thesis), which provided a comprehensive view of the reporting of NMA data network (such as study selection and risk of bias evaluations), analysis methods (including the use of Bayesian or Frequentist approaches, details on prior distribution when using Bayesian approach, inconsistency analysis and assumptions underlying the statistical model) and results (including presentational approaches and graphical tools). By utilising the findings derived from the reviews, Hutton and colleagues developed guidance for reporting NMA in the format of an extension of the PRISMA Statement (Hutton et al., 2015). The PRISMA statement (formerly known as QUOROM) is an internationally recognised guideline for good reporting of systematic reviews incorporating standard PWMA. Although much of this guideline for PWMA is highly relevant for NMA, the inherently greater complexity of the latter approaches presents further challenges for reporting such analyses. The review conducted in this thesis and the other published reviews collectively contributed to inform the development of the PRISMA NMA extension statement, an internationally recognised guideline that will be used by systematic reviewers and meta-analysts conducting and reporting NMA.

Also, a number of tutorial articles that focus on educating clinicians and methodologists alike on the fundamentals of NMA and how to interpret NMA results presented in journal articles were published after the publication of the review and recommendations (Tan et al., 2013). For example, Salanti (Salanti, 2012) summarises what the principles of NMA are, and its benefits and concerns as a next generation evidence synthesis tool. Articles by Dias and colleagues (Dias et al., 2013a, Dias et al., 2011d) provided technical guidance on the conduct of NMA through the use of tutorial examples, which

included useful program codes to enhance the understanding and facilitate the analysis. Other tutorial articles with greater relevance to clinicians on understanding the core concepts of NMA, interpreting results from published NMA and hence applying it to real-life clinical situation were published recently in medical journals, for example, by Mills and colleagues (Mills et al., 2012, Mills et al., 2013) and Cipriani and colleagues (Cipriani et al., 2013). Another systematic review article also looked at what guidance were available to researchers to present NMA results to end-users such as policymarkers and clinicians (Sullivan et al., 2014).

Given the increased understanding of NMA and accessibility to analysis tools by the publication of the tutorials, the popularity and use of NMA have increased. However, there is no standardised presentational tool for reporting NMA. In this thesis, three graphical tools to aid clear presentation and facilitate interpretation of NMA results were developed. SFP Matrix and SFP Table provide a comprehensive presentation of the important NMA and PWMA results displayed on a single plot. These plots not only enable easy comparison of NMA and PWMA results but also assist to reduce the number of tables and/or figures required for all relevant results to be presented in the main text of a journal article where space is often limited. The Median Rank Chart complements the SFP Matrix or the SFP Table by providing a visual summary of each intervention's median ranking within the network of interest; thus enabling decision makers to easily identify the "top-ranking" intervention(s) in terms of effectiveness.

Visual design principles were applied in the development of the graphs. NMA results presented in the SFP Matrix and SFP Table combine three main groups of results, namely (i) the summary forest plot graphs; (ii) the numerical estimates corresponding to the summary forest plots; and (iii) the ranking or probability best statistics. In the SFP Matrix, the most important intervention (e.g. Usual Care/Placebo or the top-ranking intervention when sorted by ranking) is usually at the top left-hand corner and hence the numerical estimates were strategically placed in the upper triangle of the matrix plot. This allows readers to read the relative effectiveness of interventions compared to the most important intervention easily as reading from left to right is generally the order that readers will scan a page. This therefore allows the summary forest plots to be placed in the lower triangle where the x-axis for the plots can be placed at the bottom of the matrix which is conventional with the usual placement of

the x-axis on graphs. The ranking or probability best statistics are placed along the diagonal with the intervention names in an enclosure so that readers can readily know what intervention the statistics correspond to. NMA and PWMA results are grouped and placed in an enclosure to allow the assessment of consistency of the results.

In the SFP Table, the three main groups of data are presented from left to right. Firstly, the intervention names together with the ranking or probability best statistics; secondly, the numerical estimates of the relative effectiveness and lastly the summary forest plots. In this design, the texts help to complement and enhance the summary forest plots that follow. Enclosures in the form of boxes present NMA results grouped by the reference intervention, allowing readers to easily recognise that all summary forest plots and numerical estimates within an enclosure are compared to the same reference intervention.

As both the graphs are developed for the presentation of NMA results, the NMA summary forest plots and numerical estimates are presented in stronger (black) ink to highlight the main results while the PWMA results in lighter (grey) ink, displayed for comparison. The intensity of the colours of the enclosures and axes, that do not represent the key results, are reduced to a minimum while light intermittent shading of enclosures in the SFP Matrix is employed to improve readability.

The Median Rank Chart presents the top-ranking intervention at the top, utilising the concept that readers will read from top to bottom, so attention is drawn to the top-ranking intervention first. Also, the top-ranking intervention is written in black ink in the lightest background shading compared to the worse intervention in the darkest background shading, utilising the visual perception concept of contrast to highlight the most important result (Few, 2004).

There has been an evolution of reporting standards initially for PWMA (Liberati et al., 2009a) and more recently for NMA (Hoaglin et al., 2011, Jansen et al., 2011). Further, Technical Support Documents (Dias et al., 2011c, Dias et al., 2011d, Dias et al., 2011a, Dias et al., 2011f, Dias et al., 2011e, Dias et al., 2011b, Ades et al., 2012) (commissioned by NICE), and a series of evidence synthesis for medical decision making tutorial articles (Dias et al., 2013d, Dias et al., 2013a, Dias et al., 2013b, Dias et al., 2013f, Dias et al., 2013e, Dias et al., 2013c, Ades et al., 2013) have recently been

published providing technical details of the implications and implementation of NMA methodology as well as guidance on reporting. These all highlight the need for a clear description of the NMA statistical model, and its assumptions, together with model fit statistics, including checks for inconsistency. Additionally, presentation of the evidence structure, in the form of a network diagram (Salanti et al., 2008), is also recommended. The graphical tools developed in this project aim to improve existing methods to report the results of a NMA and as such complement the aforementioned guidance documents. The graphs proposed focussed mainly on the presentation of single outcome but can potentially be adapted to present multiple outcomes in the future.

4 Bivariate and indirect comparison meta-analysis of overall survival and progression-free survival endpoints to improve economic evaluation of cancer treatments

4.1 Introduction

Economic evaluation of cancer treatments is based on information on the course of disease experienced by patients with cancer. It also requires collection of data on monetary spending by healthcare providers over the course of the patients' treatment to follow up and even till death for end of life treatment evaluations. However, evidence for the full set of parameters required to implement an economic model for cost-effectiveness evaluation is not always available.

Overall survival (OS) is the universally recognised primary endpoint of interest for evaluating the effectiveness of treatment regimens for cancer in Phase III randomised controlled trials (RCTs). Overall survival results (such as hazard ratios (HRs) or Kaplan-Meier survival curves) are reported in most (if not all) Phase III oncology RCTs. This is however, not the case for other outcome measures, for example progression-free survival (PFS), time to progression (TTP), tumour response and change in biomarker readings. This can limit the construction of health economic models such as multi-state Markov models which are designed to explain the course of the disease experienced by cancer patients and hence offer a more appropriate estimation of the cost-effectiveness of cancer regimens under evaluation.

For example, in 2007 the National Institute for Health and Care Excellence (NICE) carried out a technology appraisal of treatments for metastatic hormone-refractory prostate cancer (mHRPC) (Collins et al., 2007) which relied on a two-state Markov model (stable disease (StD) and death) due to the availability of sufficient data only on OS. A three-state Markov model (StD, progression (PD) and death) for cost-effectiveness analysis of cancer treatments would be better but requires information on both TTP (or PFS) and OS. Data on both these outcomes are not always reported in RCTs that evaluate the cancer treatment of interest and this was the case in one of the RCTs used in the above technology appraisal. As a result of limited data on PFS, a two-

state Markov model (StD and death) was specified using data on OS from a single RCT. In this chapter, the technology appraisal is used as an example to illustrate how bivariate evidence synthesis technique can be applied to predict PFS that was not reported and to enable the specification of a three-state model.

The objective of this project is to perform a Bayesian bivariate random-effects meta-analysis (BRMA) to obtain clinical effectiveness estimates of both PFS and OS for use in a three-state Markov model when data on OS and PFS in individual RCTs are incomplete.

4.1.1 Motivating example

As mentioned in the previous section, the UK NICE performed a technology appraisal of treatments for mHRPC in 2007. The Health Technology Assessment (HTA) report (Collins et al., 2007), evaluated the clinical- and cost-effectiveness of docetaxel in combination with either prednisone or prednisolone (D+P) for the treatment of mHRPC. In the report, a scoping search for studies evaluating the clinical- and cost-effectiveness of D+P was conducted. As only one RCT was identified to have compared D+P with mitoxantrone plus prednisone (M+P) and no other RCT compared D+P with any other possible interventions, RCTs that assessed mitoxantrone in combination with a corticosteroid compared with any chemotherapy regimen or best supportive care or placebo were also included in the scoping search. Extension of the studies selection to include studies that evaluated mitoxantrone in combination with a corticosteroid was to allow for the comparison between D+P and other relevant interventions using mitoxantrone in combination with a corticosteroid as a common comparator in indirect comparison analysis. In total, seven RCTs were identified based on the inclusion criteria, of which three RCTs used docetaxel compared with M+P, three RCTs used mitoxantrone plus a corticosteroid (M+C) compared with a corticosteroid and one RCT used M+P compared with mitoxantrone plus prednisone plus clodronate (M+P+Clo). The three RCTs that included docetaxel had docetaxel in the following combination: D+P, docetaxel with estramustine (D+E) and docetaxel with estramustine and prednisone (D+E+P).

Network table and network diagram showing the network structure of the seven RCTs are shown in Table 4.1 and Figure 4.1 respectively. As shown in Table 4.1, two dosage

schedules for D+P and D+E+P were administrated in trials TAX 327 and Oudard respectively. Docetaxel given 3-weekly (75mg) and 1-weekly (30mg) with prednisone was denoted as D+P and D1+P respectively and D+P is the docetaxel treatment regimen of interest in the HTA report. The RCT by Oudard investigated docetaxel given once in a 3 weeks cycle (i.e. 3-weekly) and twice in a 3 weeks cycle (on day 2 and 9) with estramustine and prednisone and these treatment schedules were defined as D70+P+E and D35+P+E respectively.

Table 4.1: Network Table of the seven identified RCTs in HTA report

	Treatment comparisons for OS endpoint					
Trial	D^ + P	D^ + P + E	D + E	M + P	M + P + Clo	P
CCI-NOV22 (Tannock et al., 1996)				80		81
CALGB 9182 (Kantoff et al., 1999)				119		123
Berry et al. (Berry et al., 2002)				56		63
TAX 327 (Tannock et al., 2004)	D+P: 335 D1+P: 334			337		
Ernst (Ernst et al., 2003)				105	104	
SWOG (Petrylak et al., 2004)			338	336		
Oudard (Oudard et al., 2005)		D70+P+E: 43 D35+P+E: 42		42		

Numbers in the network table represent the number of patients in the treatment arm of the trial
^ Note: Different schedules/dosages for Docetaxel drug

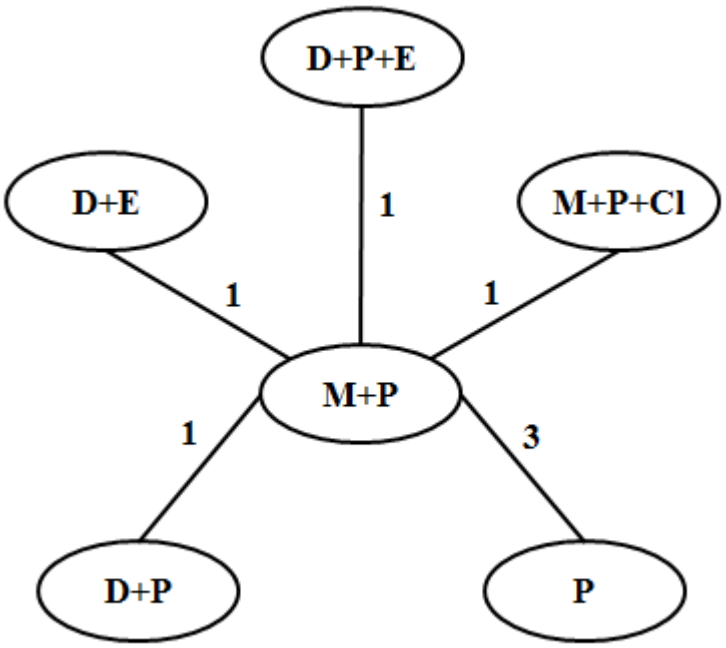

Figure 4.1: Network Diagram of the seven identified RCTs in HTA report. Numbers in the network diagram represents the number of RCTs that compared the treatments connected by the lines.

Clinical effectiveness assessed in the HTA report was performed using solely OS endpoint. Evidence synthesis using fixed-effect and random-effects meta-analysis were conducted for evaluating the clinical effectiveness comparing M+P to P using the three head-to-head RCTs that compared them. It was reported that the HR for M+P versus P was 0.99 (0.82 to 1.20) for both fixed-effect and random-effects meta-analysis. Indirect comparison meta-analysis (ICMA) was conducted for comparing D+P and P using M+P as the common comparator (see Figure 4.1 for network structure). A HR of 0.75 (0.57 to 0.99) for D+P versus P was reported using random-effects ICMA method by Bucher and colleagues (Bucher et al., 1997), which used the meta-analysis summary estimate of M+P versus P and the trial results of TAX 327 comparing M+P versus D+P (with a published HR of 0.76 (0.62 to 0.94)). No other indirect comparisons were conducted between the other interventions shown in the network diagram (See Figure 4.1) as it was not in the interest of the HTA to look at them.

For the cost-effectiveness assessment of D+P in the HTA report, two separate analyses were performed. This is due to the unlicensed status of some of the treatment regimens. The first analysis looked at three interventions (that are licensed at the time of the HTA report submission), namely D+P, M+P and P. The second analysis looked at eight

interventions, including the three in the first analysis and the following five: D1+P, D+E, D70+E+P, D35+E+P, M+P+Clo. Due to the unlicensed status of the interventions in the second analysis, the economic decision making in this HTA report was based on the results of the first analysis. In the research that follows in this chapter, the focus will be on the cost-effectiveness assessment of the interventions compared in the first analysis.

Four RCTs (CCI-NOV22, CALGB 9182, Berry et.al. and TAX 327) made up the network of trials for the first analysis that looked at interventions: D+P, M+P and P. Patient-level OS data from trial TAX 327 (Tannock et al., 2004) were used to explain the survival distribution for interventions M+P and D+P. Evidence synthesis results from the clinical effectiveness analysis for OS, comparing P to M+P, was used to estimate the survival distribution for P. Data on resource use and cost related to drug acquisition, treatment administration costs and subsequent follow-up costs (including the management of side-effects, hospitalisations, further chemotherapies and palliative care) for interventions D+P and M+P were based on patient-level cost data reported by Sanofi-Aventis (Sanofi-Aventis, 2005). Cost data for P were obtained based on patient-level data from cost-effectiveness study by Bloomfield and colleagues (Bloomfield et al., 1998). Utilities data used in the economic model were obtained from a study by Sandblom et al. (Sandblom et al., 2004). This study was one of seven studies identified in a separate systematic search of relevant databases to have reported potentially suitable health-related quality of life (HRQoL) utility values. Results from the first analysis showed that D+P was cost-effective compared to M+P with an Incremental Cost Effectiveness Ratio (ICER) of £32,706 per additional Quality-adjusted life-years (QALY), while P was dominated by M+P.

4.2 Methods

For clarity, OS is defined as the time from the start date of randomisation or study entry to the date of death or censored at the date when the patient was last known to be alive (also known as date of last follow-up). PFS is defined as the time from the start date of randomisation or study entry to the date of progression or date of death, whichever occurs first. Patients who were alive without progression were censored at the date of last follow-up. TTP is defined as the time from the start date of randomisation or study

entry to the date of progression or date of cancer-related death, whichever occurs first. Patients who were alive without progression were censored at the date of last follow-up and patients who died of other causes unrelated to cancer were censored at the date of death.

Although TTP is a more appropriate outcome variable for the specification of the three-state Markov model (StD, PD and death) for cost-effectiveness analysis of cancer treatments, it is not as commonly reported in published journal as PFS. The difference between PFS and TTP is that PFS considers death unrelated to cancer as an event of interest in the survival analysis while TTP does not.

Considering the advanced metastatic stage of the prostate cancer patient population in this project, the proportion of patients who would have died of other causes unrelated to prostate cancer is expected to be small. Hence, PFS is used in place of TTP in this project as an approximation to make it possible to present the methodological approach of using a BRMA model to jointly estimate the hazard ratios of PFS and OS for use in a three-state Markov model when data on OS and PFS in individual RCTs are incomplete.

Network of trials for OS and PFS and the Markov model reported in the HTA report are presented in Figure 4.2 (top). Interventions in the network include: D+P, M+P and P. As shown in the OS and PFS network diagrams in Figure 4.2 (top), there were no head-to-head RCTs that compared D+P to M+P for PFS.

In order to estimate the PFS HR of D+P compared with M+P, a Bayesian BRMA model that jointly estimates treatment effect on OS and PFS, is proposed. It allows the prediction of PFS for comparing D+P with M+P using the information on OS comparing D+P with M+P and the correlation between OS and PFS (obtained from studies reporting treatment effects for both outcomes). The predicted effect of PFS for D+P versus M+P in turn allows for the specification of a three-state economic Markov model incorporating a PD state as an intermediate state between the StD and death states as shown in Figure 4.2 (bottom).

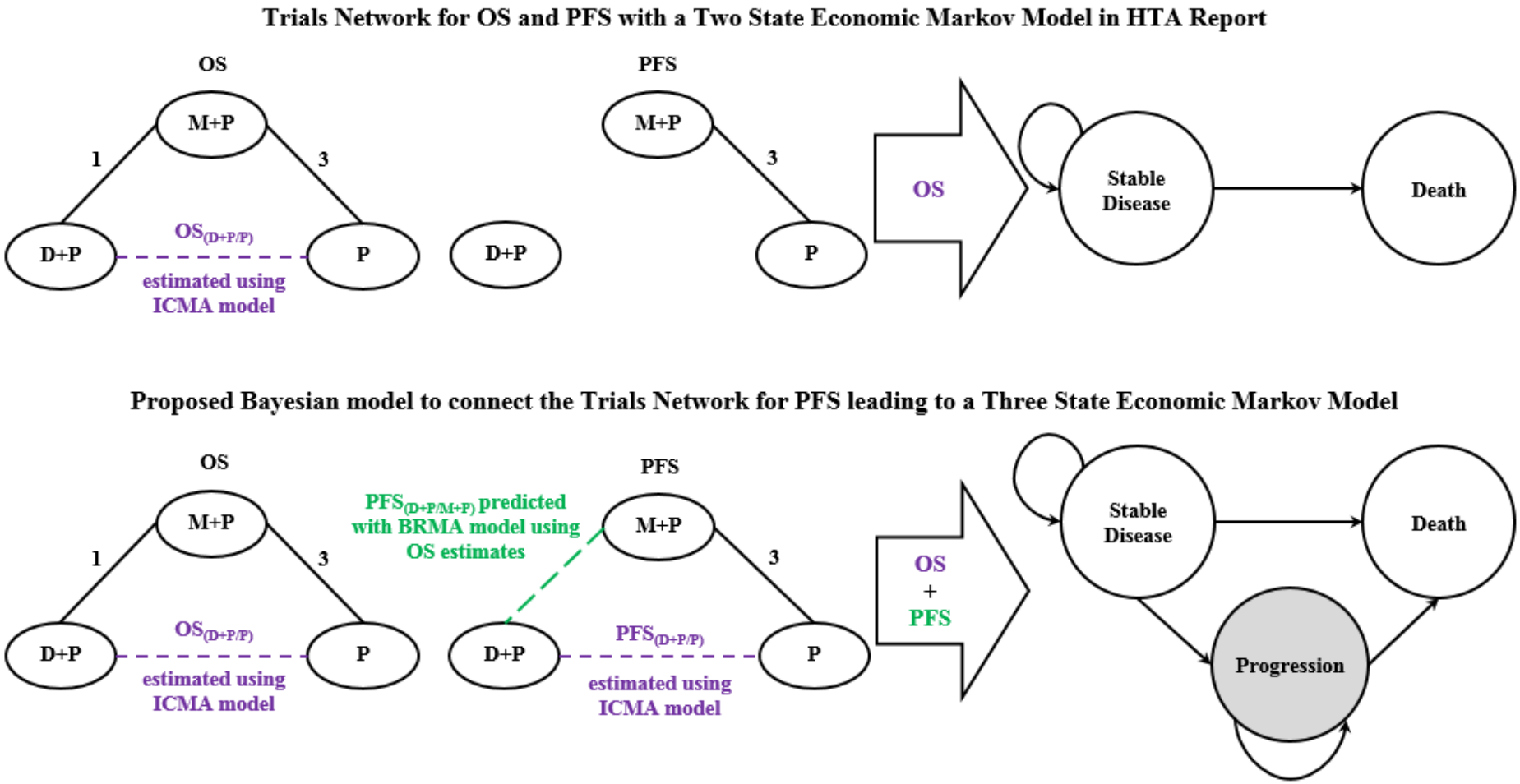

Figure 4.2: Original HTA model (top) and proposed Bayesian BRMA to predict PFS for the specification of a three-state economic Markov model (bottom)

Details of the four RCTs presented in the trial network in Figure 4.2 are presented in Table 4.2 (defined as HTA Set 1 RCTs). The remaining three RCTs that were identified in the clinical effectiveness scoping search in the HTA report are also presented in Table 4.2 (defined as HTA Set 2 RCTs). The set of seven RCTs in the HTA report (comprising HTA Set 1 and HTA Set 2 in Table 4.2) is termed "HTA Full Set" for the rest of this chapter.

Table 4.2: RCTs identified in the HTA report

HTA Set	Trial	No. of arms	Reference Treatment	Comparative Treatment(s)	Total no. of patients	OS data	PFS data
Set 1	CCI-NOV22 (Tannock et al., 1996)	2	M+P	P	161	Yes	Yes
	CALGB 9182 (Kantoff et al., 1999)	2	M+H	H	242	Yes	Yes
	Berry et al. (Berry et al., 2002)	2	M+P	P	120	Yes	Yes
	TAX 327 (Tannock et al., 2004)	3	M+P	D+P D1+P	1006	Yes	No
Set 2	Ernst (Ernst et al., 2003)	2	M+P	M+P+Cl	209	Yes	Yes
	SWOG (Petrylak et al., 2004)	2	M+P	D+E	674	Yes	Yes
	Oudard (Oudard et al., 2005)	3	M+P	D70+E+P D35+E+P	127	Yes	Yes

The Bayesian BRMA model for estimating OS and PFS for all treatment comparisons in the trial network shown in Figure 4.2 (bottom) is described in Section 4.2.1. Specifications of the three-state Markov model presented in Figure 4.2 (bottom) are described in Section 4.2.2.

4.2.1 Clinical effectiveness

Published articles for the seven RCTs in HTA Full Set identified in the scoping search of the HTA report were reviewed to assess the definition of the survival endpoints and quality of the summary data for use in the evidence synthesis. The results of this

assessment, presented in the results section (Section 4.3.1), revealed insufficient details and inconsistent scale of reporting of the survival estimates. Methods of obtaining the effectiveness estimates on the same scale required for the meta-analysis are presented in Section 4.2.1.1.

Evidence syntheses of the survival endpoints, OS and PFS, were performed for the endpoints independently where appropriate using fixed-effect and random-effects meta-analyses and ICMA. As illustrated in Figure 4.2 (top), no RCT compared D+P with either M+P or P for the PFS endpoint, hence, a Bayesian BRMA that simultaneously models treatment effects on OS and PFS was performed. This allows the prediction of the PFS HR of M+P versus D+P from OS by taking into account the correlation between the endpoints and assumption of exchangeability between the studies through "borrowing of strength" across studies and outcomes (which is discussed in Chapter 2). Details of Bayesian BRMA model and methods for constructing informative prior distributions for Bayesian models are also described in Chapter 2. Procedures for using Bayesian BRMA to predict PFS HR for comparing D+P to M+P are presented in Section 4.2.1.2. Sensitivity analyses using different prior distributions for the within-study and between-study correlations are described in Section 4.2.1.3.

4.2.1.1 Procedure of obtaining effectiveness estimates for meta-analysis

For the proposed Bayesian BRMA model, OS and PFS were analysed jointly using individual RCT's summary data on the Log HR (LHR) scale. The following analysis was conducted to obtain the estimates on the LHR scale. Articles in HTA Full Set were reviewed to assess if Kaplan-Meier survival curves for OS and PFS were reported. Where they were reported, OS and PFS individual patient data (IPD) for each of the RCTs were reconstructed from their respective Kaplan-Meier survival curves, using the method proposed by Guyot and colleagues (Guyot et al., 2012) using the DigitizeIt software (Bormann, 2013) and R software (R Core Team, 2012).

Survival analyses using the reconstructed IPD of the seven RCTs in HTA Full Set were performed to estimate the LHRs for the meta-analysis. Reconstructed IPD allow LHRs and corresponding standard errors to be estimated using survival analysis instead of crude estimation using median survival times and log-rank test p-values reported in the

RCTs. For trials where the survival curves and HRs were not presented in the published articles, LHRs and corresponding standard errors were calculated based on the summary statistics extraction methods, using reported median survival time, number of events and p-values, described by Parmar and colleagues (Parmar et al., 1998). Progression-free survival LHRs for trial CCI-NOV22 was obtained from the HTA report as it was not reported in the published article.

4.2.1.2 Procedures for predicting PFS HR for M+P versus D+P

This section first describes the procedures for predicting PFS HR using non-informative prior distributions for both the within-study and between-study correlations between OS and PFS in the BRMA model. Next, it describes the additional information required for constructing informative prior distributions for both the within-study and between-study correlations between OS and PFS and hence, the procedures for predicting PFS HR using the informative prior distributions in the BRMA model.

Using non-informative prior distributions for both within- and between-study correlations

Trial data for this analysis include the four RCTs in HTA Set 1, which were used in the main clinical effectiveness analysis of the HTA report. Procedures for predicting PFS are as follows:

1) Kaplan-Meier curves for OS and PFS were used to reconstruct the IPD for the four RCTs.
2) Two RCTs did not have PFS Kaplan-Meier curves. They are trial TAX 327 which did not record PFS in the trial and trial CCI-NOV22 which did not present a Kaplan-Meier curve for PFS in the trial journal article.
3) LHRs were estimated using the reconstructed IPD. PFS LHR for Trial CCI-NOV22 was obtained from the HTA report.
4) Non-informative prior distributions placed on the within-study and between-study correlation between OS and PFS LHRs are defined as follows:

   Within-study: $\rho_{w,i} \sim Uniform(-1, 1)$

   Between-study: $\rho_b \sim Uniform(-1, 1)$

Non-informative prior distributions in the form of half-normal distributions were used for the between-study standard deviations and are defined as follows:

$$\tau_{OS,i} \sim Normal(0,100)I(0,)$$

$$\tau_{PFS,i} \sim Normal(0,100)I(0,)$$

5) A BRMA model with non-informative prior distributions placed on the within-study and between-study correlation between OS and PFS LHRs was used to predict the PFS LHR for trial TAX 327 (that compared interventions: D+P to M+P) which was not reported. This was achieved by coding the LHR and corresponding standard error (SE) as missing values (NA in WinBUGS) and obtaining the predicted values from the model by MCMC simulation, as described in Chapter 2 Section 2.4.3.4.
6) The predicted PFS LHR summary statistics for trial TAX 327 was then used in a random-effect ICMA (for consistency with the random-effect model used in the OS analysis in the HTA report) for the estimation of the indirect PFS LHR of D+P versus P.

Using informative prior distributions for both within- and between-study correlations

Trial data used in this analysis also include the four RCTs in HTA Set 1 used in the previous section when using non-informative prior distributions for both the within-study and between-study correlations. Additional external trial data used for the construction of the informative prior distributions for the within-study and between-study correlations are: RCTs in HTA Set 2 (presented in Table 4.2) and the Cancer and Leukemia Group B (CALGB) prostate cancer trials (presented in article by Halabi and colleagues (Halabi et al., 2009)).

As discussed in Section 4.1.1, the RCTs in HTA Set 2 were selected in an extension to the main scoping search for studies using D+P because only one RCT was identified to have compared D+P to another intervention (M+P in trial TAX 327). Hence, the studies search selection was extended to include RCTs that were compared to M+P. However, among the interventions reviewed in the extended search, interventions in the RCTs in HTA Set 2 were not used in the main HTA clinical effectiveness analysis because it was not in the interest of the HTA to compare them with D+P. Therefore,

these RCTs were used to form part of the external trial evidence (together with the CALGB trials) for constructing informative prior distributions in this section.

A review of studies examining the relationship between PFS and OS in advanced or metastatic cancer was performed by Davis and colleagues (Davis et al., 2012). For mHRPC, an article by Halabi and colleagues (Halabi et al., 2009), which reported summary statistics on the correlation between OS and PFS using IPD data of nine CALGB prostate cancer trials, was identified in the review; a total of 1201 men with mHRPC were enrolled into the CALGB trials included in the analysis. Details of the trials included in the article by Halabi and colleagues are presented in Table 4.3. For the rest of this chapter, this set of nine CALGB trials is called "Halabi Set" for clarity of presentation.

The trials reported in Halabi Set were used to form part of the external trial evidence for the construction of an informative prior distribution for the between-study correlation between OS LHR and PFS LHR. The construction of an informative prior distribution for the within-study correlation between OS LHR and PFS LHR presents additional complexity as it could only be estimated using IPD of both OS and PFS from a RCT (containing docetaxel versus not containing docetaxel) using bootstrapping (Efron and Tibshirani, 1986) as described by Daniels and Hughes (Daniels and Hughes, 1997) and Bujkiewicz and colleagues (Bujkiewicz et al., 2013). Although the within-study correlation between OS and PFS reported in the article by Halabi and colleagues might not be the same as the within-study correlation between OS LHR and PFS LHR, the former was used as a crude approximation to the latter. This made it possible to illustrate how external evidence could be used to inform the BRMA model parameters and potentially improve prediction of PFS HR by modelling jointly the treatment effects on OS and PFS.

Table 4.3: Summary data of the nine trials in the article by Halabi and colleagues

Trial	Year	Number of arms	Treatment 1	Comparative Treatment(s)	Total number of patients	OS data	PFS data	Included for construction of prior?
CALGB 9182	1999	2	Mitoxantrone (M) +Hydrocortisone	Hydrocortisone	242	Yes	Yes	No
CALGB 9181	2000	2	Megestrol acetate (MA) Low dose	MA High dose	149	Yes	Yes	Yes
CALGB 9780	2001	1	Docetaxel (D) +Estramustine (E) +Hydrocortisone	None	46	Yes	Yes	Yes
CALGB 9480	2002	3	Suramin (SU) Low dose	SU Intermediate dose SU High dose	390	Yes	Yes	Yes
CALGB 9680	2002	2	M Low dose	M High dose	45	Yes	No	No
CALGB 99813	2003	1	D+E +Carboplatin with G-CSF support	None	40	Yes	Yes	Yes
CALGB 9583	2004	2	Antiandrogen withdrawal alone	Antiandrogen withdrawal +Ketoconazole	260	Yes	No	No
CALGB 90004	2008	1	D+E +Exisulind	None	75	Yes	Yes	No
CALGB 90006	2010	1	D+E +Bevacizumab	None	77	Yes	No	No

A schematic diagram presenting the procedures for the prediction of PFS HR of D+P versus M+P, including the construction of informative prior distributions for the within-study and between-study correlations, is shown in Figure 4.3. Procedures for the prediction of PFS HR of D+P versus M+P are as follows:

1) Kaplan-Meier curves for OS and PFS were used to reconstruct the IPD for the four RCTs.
2) Two trials did not have PFS Kaplan-Meier curves. They are trial TAX 327 which did not record PFS in the trial and trial CCI-NOV22 which did not present a Kaplan-Meier curve for PFS in the trial journal article.
3) Informative prior distribution on the within-study correlation between PFS LHR and OS LHR was constructed based on the correlation estimates (for PFS with OS) reported by Halabi and colleagues (Halabi et al., 2009). In order to sample from this prior distribution for the correlation, Fisher transformation method (discussed in Chapter 2 Section 2.4.3.6) was used to convert the correlation to a

corresponding Fisher correlation parameter, z, which follows a normal distribution for sampling.

4) To construct the informative prior distribution for the between-study correlation between OS and PFS LHRs, trials from the external trial evidences were reviewed for suitability to be used for prior distribution construction. The inclusion criteria for identifying the trial set for constructing the informative prior distribution are:
    i. the trial must be reported before year 2007 (this is the year of publication of the HTA report);
    ii. both OS and PFS must be reported in the trial and
    iii. RCT that cannot be collapsed into single-arm trial must have docetaxel (or a combination of it) in one of the treatment arms.

    Based on the inclusion criteria, a set of trials suitable for constructing the informative prior distribution for the between-study correlation between OS and PFS LHRs were identified and presented in the results (Section 4.3.1). For the rest of this chapter, this identified set of trials is termed "external trial data set (ETD Set)".
5) For the construction of the informative prior distribution using the ETD Set, a "complete" set of OS and PFS estimates for all trials in the ETD Set was required. A "complete" set means that OS and PFS log hazard rates for all treatment arms in all trials must be available. To achieve that, OS and PFS estimates for the trials in the ETD Set were estimated using either reconstructed IPD as described in Section 4.2.1.1, reported median survival times or prediction using a network meta-analysis (NMA). Similar to the RCTs in HTA Set 1, Kaplan-Meier curves for OS and PFS, if reported in the published articles, were used to reconstruct the IPD. Parametric survival models that best fit the data were used for the estimation of log hazard rates of each individual treatment in the trials. The log hazard rates for trials that did not present the Kaplan-Meier curves were estimated using the reported median survival time and corresponding 95% confidence intervals (CIs) and/or p-values using the methods by Parmar and colleagues (Parmar et al., 1998). As some of the included trials were single-arm trials, log hazard rate for the "missing" comparative arm did not exist. Therefore, NMAs were conducted for OS and

PFS independently to predict the log hazard rates for the arms that did not exist. This, in turn, allowed the corresponding LHRs (comparing treatments containing Docetaxel versus treatments not containing Docetaxel) to be predicted. This resulted in a set of bivariate outcome data of LHRs on OS and PFS that were "complete" with no "missing" data, which could be entered into a Bayesian BRMA model to obtain a posterior distribution for the between-study correlation between LHRs on OS and PFS

6) A Bayesian BRMA model was used to model the OS and PFS data simultaneously to obtain a posterior distribution for the between-study correlation between LHRs on OS and PFS. This posterior distribution was subsequently used as an informative prior distribution for the BRMA analysis of the main HTA trials. Similarly, a Fisher correlation parameter, which follows a normal distribution, was required for sampling. The Fisher correlation distribution of the posterior distribution for the between-study correlation was used for this purpose.
7) A Bayesian BRMA model, with informative prior distributions placed on both the within-study (mentioned in point 3) and between-study (obtained in point 6) correlations between LHRs on OS and PFS, was used to predict the PFS LHR for trial TAX 327 (that compared interventions: D+P to M+P) which was not reported.
8) The predicted PFS LHR summary statistics for trial TAX 327 was subsequently used in a random-effects ICMA for estimating the indirect PFS LHR of D+P versus P.

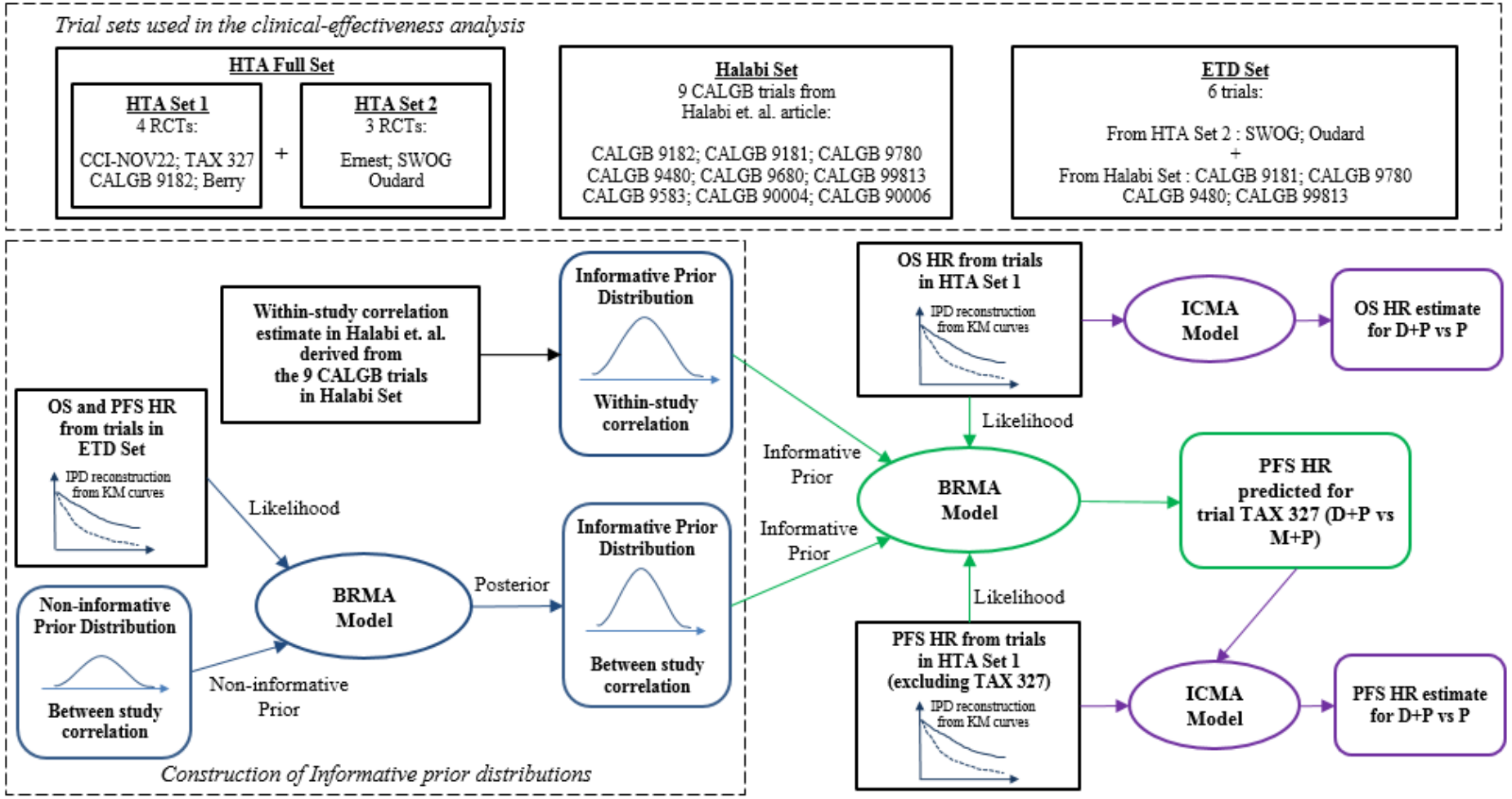

Figure 4.3: Diagram for the clinical effectiveness analysis using BRMA and ICMA models with corresponding trial evidence sets

4.2.1.3 Sensitivity analysis

A sensitivity analysis was performed to investigate the effect of the choice of prior distributions on the predicted PFS HR for D+P versus M+P. Combination pairs of non-informative, optimistic and pessimistic prior distributions for the within-study and between-study correlations were investigated. Non-informative prior distributions take the form of uniform distributions, confined between the range of -1 and 1 inclusive, for both the within-study and between-study correlations as follows:

Non-informative prior distributions:

Within-study: $\rho_{w,i} \sim Uniform(-1, 1)$

Between-study: $\rho_b \sim Uniform(-1, 1)$

In this sensitivity analysis, optimistic prior distribution for the between-study correlation was defined using a positive high correlation while pessimistic prior distribution for the between-study correlation was defined using a positive low correlation. The prior distributions take the following parameter form (before using Fisher transformation):

optimistic prior parameter values $\rho_b = 0.8$ (95% CI: $0.7, 0.9$)

pessimistic prior parameter values $\rho_b = 0.2$ (95% CI: $0.0, 0.4$)

Correspondingly, PFS HRs for D+P versus P were estimated using fixed-effect and random-effects ICMA.

4.2.2 Cost-effectiveness

A two-state Markov model was specified in the UK HTA report that assessed the cost-effectiveness of docetaxel with either prednisone or prednisolone for the treatment of mHRPC and was developed for the HTA report using Microsoft Excel software. A diagram showing the two-state model specified in the HTA report and the proposed three-state model in this thesis is shown in Figure 4.2. The proposed three-state model incorporated a PD state not included in the original two-state model in the HTA report and was developed using the WinBUGS software (Lunn et al., 2000, Spiegelhalter et al., 2003). For the comparison of the results of the proposed three-state model with the results of the two-state model reported in the HTA, the two-state model was reproduced using WinBUGS. For clarity, the re-produced two-state model will be called the "WinBUGS two-state model" and the original two-state model in the HTA report will be called the "HTA two-state model" in the rest of this chapter.

Similar to the HTA two-state model, both the WinBUGS two-state model and WinBUGS three-state model were run for 180 cycles, where one cycle represented one month. This allowed the models to be run for a time horizon of 15 years. Based on the transition probabilities that were calculated using trial TAX 327 (with median age reported to be 68 years), a robust estimate of the mean survival was obtained as it is expected that most of the patients in the cohort would have died in the model. A cohort size of 10,000 was used in each of the models.

To develop the economic models, data for the construction of transition probabilities, definition of costs and utilities for each of the interventions need to be extracted. These data can be extracted from reviews, single RCT or evidence synthesis from a number of trials/studies. Specifications of the transition probabilities, cost and utilities are described in sections 4.2.2.1, 4.2.2.2 and 4.2.2.3 respectively.

4.2.2.1 Transition probabilities

For the WinBUGS two-state model, the transition probabilities were estimated using the Weibull parameters reported in the HTA report, which used IPD from trial TAX 327. This was for consistency with the parameters used in the HTA model. As for the three-state model, which incorporated a PD state, the transition probabilities for

transition from StD state to PD state were estimated using parametric Weibull survival modelling on reconstructed PFS IPD from one of the six RCTs (excluding trial TAX 327) in the HTA report.

The selection criteria for the RCT to be used for estimating the transition probabilities for treatment arm M+P are: (i) comparable OS profile of the selected trial and trial TAX 327; (ii) selected trial having a mean time of progression closest to the reported mean cycle of M+P administrated in trial TAX 327 (that is: 5.9 cycles as reported in the HTA report). As PFS was not recorded for TAX 327, selection criteria (ii) is based on the assumption that the patients in trial TAX 327 M+P arm were administrated M+P till progression. The mean number of cycles of M+P administrated was then used as a crude approximation of the 'potential' mean time to progression for patients administrated M+P in trial TAX 327. Transition probability for transition from the StD state to death state was obtained from an article on cost-effectiveness analysis for advance hormone-dependent prostate cancer (Lu et al., 2012). In the absence of patient-level data on both OS and PFS for any of the interventions, the transition probability for transition from PD state to death state could not be estimated. Methods for the estimation of (i) transition probabilities using parametric Weibull survival model and (ii) transition probabilities for transition from PD state to death state in the three-state model are described in the next two subsections.

Transition probabilities estimated using parametric Weibull survival model

Survival analysis using the parametric Weibull model was used to implement time-dependency in the transition probabilities in the economic models (transition probabilities for transition from StD state to death state in WinBUGS two-state model; and transition probabilities for transition from StD state to PD state in three-state model). The Weibull distribution takes the following probability density function:

$$f(t) = \lambda\gamma t^{\gamma-1} exp(-\lambda t\gamma)$$

where λ gives the scale of the distribution and γ defines the shape. The hazard function for this distribution is therefore:

$$h(t) = \lambda\gamma t^{\gamma-1}$$

with a cumulative hazard function of:

$$H(t) = \lambda t^{\gamma}$$

where the survival function is related to the cumulative hazard function in the following form:

$$S(t) = exp[-H(t)]$$

Since hazards are instantaneous, these need to be converted to a transition probability for a given period, such as a Markov cycle. Using the survival function, transition probability between time-points $(t-u)$ and t, denoted as $TP(t_u)$ where u is the cycle length, was defined as one minus the ratio of the survival function at the end of the interval to the survival function at the beginning of the interval. This function defined as:

$$TP(t_u) = 1 - S(t)/S(t-u)$$

was re-written using the cumulative hazard function as:

$$\begin{aligned} TP(t_u) &= 1 - exp[-H(t)]/exp[-H(t-u)] \\ &= 1 - exp[H(t-u) - H(t)] \end{aligned}$$

Therefore, transition probability was defined using the Weibull parameters as follows:

$$TP(t_u) = 1 - exp[\lambda(t-u)^{\gamma} - \lambda t^{\gamma}]$$

In the HTA report, results of the Weibull survival analysis model were presented in the form of the regression coefficients of the intercept and scale parameters. These two parameters are expressed in terms of the Weibull parameters, λ and γ, as follows:

$$\lambda = exp(-\beta/\alpha)$$

$$\gamma = \frac{1}{\alpha}$$

where β is the intercept and α is the scale regression coefficient parameters from the Weibull survival analysis.

When performing the probabilistic analysis, the covariance between the intercept and scale regression parameter from the Weibull survival analysis were also incorporated in the WinBUGS two-state model. This was achieved by using the Cholesky decomposition matrix derived from the covariance matrix obtained from the Weibull survival regression model. Given a covariance matrix of the form:

$$Covariance, C = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

the Cholesky decomposition matrix takes the form:

$$D = \begin{pmatrix} \sqrt{a} & 0 \\ \dfrac{b}{\sqrt{a}} & \sqrt{c - \dfrac{b^2}{a}} \end{pmatrix}$$

such that $C = D\,D^*$ where D^* denotes the conjugate transpose of D. Cholesky decomposition matrices of the covariance matrices for the interventions, D+P and M+P, were calculated independently and applied to the transition probabilities of D+P and M+P respectively in the WinBUGS two-state model to allow for the correlation between the intercept and scale parameters when sampling the random normal draws for the two parameters. Assuming that the Cholesky decomposition matrix of the covariance matrix for M+P is:

$$D_{M+P} = \begin{pmatrix} u_{D,M+P} & 0 \\ v_{D,M+P} & w_{D,M+P} \end{pmatrix}$$

the transition probability incorporating parameter uncertainties for transition from StD state to death state for M+P is defined as:

$$TP_{D,M+P}(t_u) = 1 - exp[H_D(t-u) - H_D(t)]$$

$$TP_{D,M+P}(t_u) = 1 - exp\left[\lambda_{D,M+P}(t-u)^{\gamma_{D,M+P}} - \lambda_{D,M+P}t^{\gamma_{D,M+P}}\right]$$

where:

$$\lambda_{D,M+P} = \exp\left(\frac{-\beta_{D,M+P}}{\alpha_{D,M+P}}\right)$$

$$\gamma_{D,M+P} = \frac{1}{\alpha_{D,M+P}}$$

and

$$\beta_{D,M+P} = \beta_{M+P} + u_{D,M+P} Z_{\beta,D,M+P}$$

$$\alpha_{D,M+P} = \alpha_{M+P} + v_{D,M+P} Z_{\beta,D,M+P} + w_{D,M+P} Z_{\alpha,D,M+P}$$

where β_{M+P} and α_{M+P} are the intercept and scale regression coefficients for M+P presented in the HTA report; and $Z_{\beta,\mathrm{D},M+P} \sim Normal(0,1)$ and $Z_{\alpha,\mathrm{D},M+P} \sim Normal(0,1)$.

Transition probabilities of interventions D+P and M+P for the WinBUGS two-state model were calculated using the regression coefficients and Cholesky decomposition matrix from the HTA report (Table 28 and 29 respectively (Collins et al., 2007)). Transition probabilities for P were calculated by applying the HR of P versus M+P or HR of P versus D+P to the hazard rates of M+P and D+P in the transition probabilities respectively. Therefore, assuming that the transition probability for M+P is given by:

$$TP_{M+P}(t_u) = 1 - exp[H(t-u) - H(t)]$$

and with a HR for P versus M+P, denoted as $HR_{P/M+P}$, the transition probability for P is given by:

$$\begin{aligned} TP_P(t_u) &= 1 - exp\{HR_{P/M+P}[H(t-u) - H(t)]\} \\ &= 1 - exp[H(t-u) - H(t)]^{HR_{P/M+P}} \end{aligned}$$

Uncertainty associated with the HR was incorporated in the model by assigning a normal distribution to the logarithm of the HR as follows:

$$LHR \sim Normal(\bar{\mu}, \sigma^2)$$

where $\bar{\mu}$ and σ^2 are the mean and variance estimate of the LHR from random-effects meta-analysis.

For the three-state model, the set of transition probabilities for intervention M+P was calculated using regression coefficients of the parameters of a Weibull survival model for PFS using re-constructed IPD from one of the RCTs in HTA Full Set selected based on the criteria outlined above. As no PFS patient-level data was available for the interventions D+P and P, transition probabilities for each of the interventions were

calculated by applying their HR with respect to M+P to the transition probabilities of M+P. Similarly, uncertainty associated with each of the HRs was included in the respective models by assigning normal distribution to the LHRs.

Transition probabilities from PD state to death state (three-state model)

Although IPD were reconstructed for PFS (together with OS) for the trial selected for estimating the transition probability from StD state to PD state, the reconstructed IPD for PFS and OS were not paired by patient. Hence, it would not be possible to estimate the transition probabilities from PD state to death state using parametric survival analysis performed using reconstructed IPD as described in the previous section. To overcome this issue, transition probabilities were estimated by assuming the mean total survival time was equal to the weighted sum of combined survival time from stable disease to progression and then to death and the survival time when death occurred from other causes:

$$mean(Total\ Time) = W_{StDtoPDtoDeath}[mean(Time_{StDtoPD}) + mean(Time_{PDtoDeath})] + W_{StDtoDeath}mean(Time_{StDtoDeath})$$

where $mean(Time_{StDtoPD})$ defines the mean time that patients stayed in the StD state before transition to the PD state; $mean(Time_{PDtoDeath})$ and $mean(Time_{StDtoDeath})$ define the mean time for PD state to death state and StD state to death state respectively; W defines the weight assigned to the mean time and is related to the number of patients who transition through the two potential pathways in the economic model as shown in Figure 4.2 (bottom), from stable disease to death either with or without disease progression.

As the proportion of patients who died of causes unrelated to prostate cancer was expected to be small (<1%), we assumed that $W_{StDtoDeath} \rightarrow 0$, therefore $W_{StDtoPDtoDeath} \rightarrow 1$. Hence,

$$mean(Total\ Time) = W_{StDtoPDtoDeath}[mean(Time_{StDtoPD}) + mean(Time_{PDtoDeath})] = mean(Time_{StDtoPD}) + mean(Time_{PDtoDeath})$$

and therefore,

$$mean(Time_{PDtoDeath}) = mean(Total\ Time) - mean(Time_{StDtoPD})$$

Assuming that the survival data for patients from PD to death follows an exponential survival distribution, the transition probability between time-points $(t-u)$ and t, denoted as $TP(t_u)$ where u is the cycle length, is defined as follows:

$$\begin{aligned} TP(t_u) &= 1 - exp\{\lambda(t-u)^\gamma - \lambda t^\gamma\} \\ &= 1 - exp(-\lambda u) \end{aligned}$$

where $\gamma = 1$ for the exponential survival model.

As the hazard rate, $\lambda = \frac{1}{mean(Time)}$,

$$TP(t_u) = 1 - exp\left(\frac{-u}{mean(Time)}\right)$$

For M+P and D+P, the $mean(Total\ Time)$ for each of the interventions were estimated using the mean survival time calculated from the reconstructed OS IPD of trial TAX 327. For P, the mean survival time was estimated by a random-effect meta-analysis of the log hazard rate of the three RCTs that had a P treatment regimen arm.

As PFS endpoint were not recorded for trial TAX 327, the mean number of cycles of drug reported in the HTA report were used to represent the mean time from stable disease to progression, $mean(Time_{StDtoPD})$, based on the assumption that patients stopped drug treatment on the onset of disease progression. Mean number of cycles of drug P was not reported in the HTA report. Therefore, the mean time to progression was also estimated using meta-analysis of the log hazard rate of the two RCTs that reported PFS data for P.

Transition probabilities for transition from PD state to death state for each intervention were therefore calculated using the equation:

$$TP(t_u) = 1 - exp\left(\frac{-u}{[mean(Total\ Time) - mean(Time_{StDtoPD})]}\right)$$

Uncertainty associated with the mean survival time or log hazard rate were also incorporated using normal distributions and propagated in the economic model. As the

exponential survival model is a single parameter model, Cholesky decomposition was not required for defining the uncertainty.

4.2.2.2 Cost

Cost data comprises drug acquisition and administration cost for each interventions, cost of the management of adverse side effects and subsequent follow up cost that included cost of further chemotherapy after disease progression, management of side-effects and palliative cost. Cost for each of the interventions to be used in the WinBUGS 2-state and 3-state model were extracted from cost data presented in the HTA report. In the report, costs were categorised into three components: namely, (i) the drug cost, (ii) the follow up cost and (iii) the terminal care cost. Drug cost included cost of acquisition and administration of each intervention.

Follow up cost included the cost of managing side-effects, subsequent chemotherapies and hospitalisation for palliative care. Terminal care costs were one-off costs used to incorporate the cost of caring for patients in the last month of life. As stated in the HTA report, terminal care cost data were not recorded in the trial (TAX 327), hence these costs were estimated from patients who died in the first six months after entering the trial. In the absence of costs per cycle for follow up cost, these costs were assigned and calculated as one-off cost, in a similar fashion as terminal care cost, as patient died. Cost data for interventions D+P and M+P were estimated using patient level data from trial TAX 327 while cost data for P were estimated using published cost-effectiveness analyses by Bloomfield and colleagues (Bloomfield et al., 1998).

Gamma distribution was used to represent uncertainty in the follow up costs and terminal care costs. Drug costs for each of the interventions were calculated based on the mean number of cycles of drugs administrated. Normal distribution was used to describe the number of cycles of drugs administrated to reflect uncertainty in the drug costs.

For the WinBUGS two-state model, the total costs were calculated as the summation of all three categories of costs. For the three-state model, drug costs and terminal care costs were calculated in a similar way to that calculated in the WinBUGS two-state

model while follow-up costs were calculated by dividing the follow-up costs into two unequal parts (using a parameter defined as ψ in the equations that follow).

Costs of subsequent chemotherapy and hospitalisations accounted for between 70% and 80% of follow-up costs which most likely occurred post-progression and the remaining follow-up cost (20% to 30%) were related to side effects likely to occur prior to progression (but may also be associated with the subsequent chemotherapy post-progression). Therefore the follow up costs were divided into portions corresponding to StD state and PD state.

As the base case analysis for the three-state model, 75% $(\psi = 0.75)$ of the follow-up costs were assigned to the PD state to account for the cost of subsequent chemotherapy, managing side-effects and hospitalisations post-progression. Computation of these costs were based on the number of patients who died per cycle while the remaining 25% of the costs that were assigned to the StD state were computed based on the number of patients who progressed per cycle. Follow-up costs were assigned as one-off cost in a similar way as the WinBUGS two-state model.

In the WinBUGS two-state model, follow-up costs were calculated based on patients died per cycle as:

$$Total\ Cost_{FU} = \sum_{i=1}^{180} \left(Cost_{FU,i} \times N_{Died,i}\right)$$

where $Cost_{FU,i}$ represents follow-up cost data for cycle i, and $N_{Died,i}$ represents the number of patients who died in cycle i.

For the three-state model, the follow-up costs were calculated as:

$$Total\ Cost_{FU} = Total\ Cost_{StDFU} + Total\ Cost_{PDFU}$$

where:

$$Total\ Cost_{StDFU} = \sum_{i=1}^{180} \left[(1-\psi) \times Cost_{FU,i} \times N_{Progressed,i}\right]$$

$$Total\ Cost_{PDFU} = \sum_{i=1}^{180} \left(\psi \times Cost_{FU,i} \times N_{Died,i}\right)$$

ψ represents the proportion of follow-up costs associated with the PD state (termed "division factor" for the rest of this chapter) and $N_{Progressed,i}$ represents the number of patients who progressed in cycle *i*.

Sensitivity analyses for assessing different proportions of follow-up costs associated with the PD state are described in Section 4.2.2.5. An annual discount rate of 3.5% was used for discounting the cost after the first year.

4.2.2.3 Quality-adjusted life-years

Quality-adjusted life-years were used as a measure of effect in this cost-effectiveness analysis. It is a single measure that explains an individual's health-related quality of life and length of life gained due to the implementation of an intervention. To estimate the QALY, utility data in the form of health-related quality of life (HRQoL) were required to quantify the potential health status of patients with mHRPC, as well as the impact the interventions (in terms of disease progression and serious adverse effects) had on their HRQoL.

Quality of life data used in the HTA two-state model were extracted from a study conducted by Sandblom and colleagues (Sandblom et al., 2004). This study was selected (out of seven studies that were identified to have reported suitable HRQoL utility values) because (i) it reported HRQoL values using a generic HRQoL instrument that met the requirement specified by NICE; (ii) the population under assessment was representative of the target population of the HTA; and (iii) it provided end-of-life HRQoL values of prostate cancer patients in their last year before death. However, no suitable utility data are identified from the literature to estimate the impact of each individual intervention on the HRQoL. The HRQoL instrument used in the study by Sandblom and colleagues that met the requirement specified by NICE is the EuroQoL five-dimensional (EQ-5D) questionnaire. Health state utility values obtained from the preference-based measure, EQ-5D questionnaire values were used in the HTA two-state model.

The EQ-5D values used in the HTA two-state model are used in the WinBUGS two-state model. Mean and corresponding 95% CI of the EQ-5D values reported for all

patients in the 12 months prior to death were used to define the distribution for the utilities data for a probabilistic analysis. Beta-distribution was utilised to describe the uncertainty pertaining to utility data for the StD state. A utility value of zero is assigned for the death state.

Similarly, utility data reported in the study by Sandblom et al. were used to define the utility distributions in the three-state model. With the inclusion of a PD state in the three-state model, additional utility data for the PD state were required. EQ-5D value used in the two-state models for the StD state were used to describe the HRQoL of patients in the PD state (denoted as U_{PD}). This was based on the argument that this utility reflects the patients' HRQoL prior to prostate cancer death and the state prior to death would be the PD state. As the EQ-5D value for the StD state in the two-state model was used to describe the utility for the PD state, utility for the StD state need to be estimated. This is achieved by splitting the patients in the StD state into three groups and using the following EQ-5D values from the Sandblom et al. study: (i) EQ-5D values of all patients who died of other causes (denoted as $U_{Other\ causes}$) (ii) EQ-5D values of all patients who were still surviving with prostate cancer (denoted as $U_{Surviving}$) and lastly (iii) U_{PD} described earlier. These EQ-5D values, together with transition probabilities are used to formulate the utility for the StD state as follows:

$$U_{StD} = TP_{StD}U_{Surviving} + TP_{StDtoPD}U_{PD} + TP_{StDtoDeath}U_{Other\ causes}$$

where TP_{StD} defines the probability of remaining in StD state, $TP_{StDtoPD}$ defines the probability of transition from StD state to PD state and $TP_{StDtoDeath}$ defines the probability of transition from StD state to death state without passing through the PD state. A utility value of zero ($U_{Death} = 0$) was assigned for the death state. Similarly, an annual discount rate of 3.5% was used for discounting the utilities after the first year.

4.2.2.4 Cost-effectiveness analysis using WinBUGS

For the assessment of the cost-effectiveness of the interventions in each model, the mean costs and mean QALYs gained for the interventions and ICERs for the comparison of the two interventions of interest (M+P and D+P) were estimated using WinBUGS. For each model, the simulation was conducted using Markov Chain Monte

Carlo (MCMC) implemented in WinBUGS with 50,000 iterations and 30,000 burn-in iterations that were discarded. The annual discount rate was applied from cycle 13 onwards).

Mean costs and QALYs gained for each intervention were obtained from the WinBUGS output. ICER was calculated by taking the difference between the mean values of the cost of interventions over the difference between the mean values of the QALYs gained of interventions; as:

$$ICER = \frac{\overline{Cost_{D+P}} - \overline{Cost_{M+P}}}{\overline{QALY_{D+P}} - \overline{QALY_{M+P}}}$$

where $\overline{Cost_{D+P}}$ and $\overline{Cost_{M+P}}$ defines the mean cost of D+P and M+P respectively and $\overline{QALY_{D+P}}$ and $\overline{QALY_{M+P}}$ defines the mean QALY gained per patient for D+P and M+P respectively.

Output from the simulations for the WinBUGS two-state and three-state models were exported from WinBUGS to the R software to generate CEACs for comparing the three interventions, as well as CEAC and cost-effectiveness plane for comparing the proposed three-state model with the WinBUGS two-state model (which is expected to be comparable to the HTA two-state model) when evaluating the difference between D+P and M+P.

4.2.2.5 Sensitivity Analysis

For the base case analysis of the WinBUGS two-state model, transition probability for P was calculated using random-effects meta-analysis OS HR for M+P versus P. Sensitivity analysis for the WinBUGS two-state model was performed using the OS HR for D+P versus P obtained from ICMA. The HR obtained from the indirect comparison was placed on the transition probability of the D+P arm to obtain the transition probability for the P arm. Results for this WinBUGS two-state (Indirect) were compared to the results from the WinBUGS two-state (Direct) where the transition probability for the P arm was calculated using random-effects meta-analysis OS HR of M+P versus P.

Several prior distributions were placed on the between-study and within-study correlation in the BRMA model to assess the influence of the choice of the prior distributions on the predicted PFS HR as described in Section 4.2.1. The alternative predicted PFS HR values, together with their associated uncertainties, were propagated in the three-state model to assess the impact of the choice of prior distributions for the correlation between OS and PFS LHRs on the cost-effectiveness estimates.

Sensitivity analysis was also performed for the cost-effectiveness analysis using alternative values of the division factor ψ for the follow-up costs in the three-state model (defining proportion of this cost assigned to StD and PD states). The ratio of 25:75 ($\psi = 0.75$) was proposed for the base-case analysis, whilst alternative scenarios considered ratios of 20:80 ($\psi = 0.80$) and 30:70 ($\psi = 0.70$) to assess the influence of the division factor ψ on the cost-effectiveness estimates. Additionally, analysis using a probabilistic distribution for the division factor ψ was carried out; a beta distribution with mean 0.75 and standard deviation of 0.075 [assuming that the 95% CI is 0.6 to 0.9, this gives a standard deviation of (0.9-0.6)/4 = 0.075] was placed on the division factor ψ to assess its influence on the cost-effectiveness results.

4.3 Results

4.3.1 Clinical effectiveness

Definitions of OS were consistent for the seven RCTs in HTA Full Set except for trial Berry, where OS was not explicitly defined in the published article (Berry et al., 2002). Overall survival was defined as the time from the start date of randomisation to the date of death or censored at the date when the patient was last known to be alive. There were inconsistencies in the definition of progression endpoint across the six RCTs in HTA Full Set (excluding trial TAX327 which did not record PFS endpoint). Specifically, PFS was reported for three RCTs (Kantoff et al., 1999, Ernst et al., 2003, Petrylak et al., 2004) and defined as the time from the start date of randomisation to the date of progression or date of death, whichever occurred first. Neither PFS nor TTP was reported in the published article for trial CCI-NOV22 (Tannock et al., 1996); however, the HTA report presented TTP estimates for this trial. Time to progression was reported by Berry and colleagues in the published article (Berry et al., 2002) but no

explicit definition for TTP was provided. Oudard and colleagues reported time to PSA progression and defined this endpoint as the time from the date of randomisation to the date of PSA progression (Oudard et al., 2005). Due to the variability in the progression endpoints reported, standardising the definition of the progression endpoint estimates for all the RCTs would require IPD from each of the trials, which is not achievable within the resources of this project.

Kaplan-Meier curves for OS were reported in the original articles of the seven trials in the HTA report (see Table 4.2). Progression-free survival was not reported for trial TAX 327 and PFS Kaplan-Meier curve were not reported for Trial CCI-NOV22. Hazard ratios on OS and PFS reported in the original articles and those obtained from the survival analysis of reconstructed IPD are presented in Table 4.4 and Table 4.5 respectively. Hazard ratios calculated using the reconstructed IPD were comparable to the results reported in the original trials' publication. For trials CALGB 9182, Berry and Oudard, although the OS HR point estimates were in the reverse direction to the published point estimates, the 95% CIs were consistent with those reported in the trials.

To obtain summary estimates for the HR of OS comparing M+P with P, both fixed-effect and random-effects meta-analysis were used, combining estimates obtained by reconstructing IPD from the three RCTs that directly compared M+P and P (trials in HTA Set 1 excluding trial TAX 327). The HRs were 0.903 (0.751 to 1.084) and 0.901 (0.405 to 2.023) respectively, which were different from 0.99 (0.82 to 1.20) for both fixed-effect and random-effects results in the HTA report. The difference in the point estimate of the HRs was largely due to the lower HRs obtained using the reconstructed IPD for trials: CALGB 9182, CCI-NOV22 and Berry, compared to the HRs reported in the HTA report. However, the 95% credible interval (CrI) estimated using fixed-effect meta-analysis was comparable to the 95% CI reported in the HTA report.

Overall survival HR for treatment comparison D+P versus P (as shown in Figure 4.2) using reconstructed IPD from the three RCTs and trial TAX 327 was computed using both fixed-effect and random-effects ICMA. Hazard ratios of 0.688 (0.523, 0.907) and 0.688 (0.300, 1.604) were estimated for fixed-effect and random-effects ICMA respectively as compared to the HR of 0.75 (0.57, 0.99) for random-effects ICMA published in the HTA report. Similarly, the HRs estimated were lower than that reported in the HTA report due to the HRs estimated for trials: CALGB 9182, CCI-

NOV22 and Berry using reconstructed IPD. Overall survival results from meta-analysis and ICMA using reconstructed IPD are presented in Table 4.6.

Fixed-effect and random-effects meta-analysis results for PFS comparing M+P versus P using reconstructed IPD were 0.641 (0.532 to 0.772) and 0.619 (0.170 to 2.048) respectively. No summary estimates for this comparison was reported in the HTA report.

As trial TAX 327 did not investigate the PFS endpoint for the treatment comparison D+P versus M+P, the Bayesian BRMA model described in Chapter 2 Section 2.4.3 was used to jointly model the correlated outcomes of OS and PFS to enable the prediction of PFS for treatment comparison D+P versus M+P of trial TAX 327. To model the correlated outcomes, prior distributions for the within-study and between-study correlations were specified in the model. Both non-informative and informative prior distributions were used in the analysis.

Table 4.4: Individual trial's OS results using IPD reconstructed from Kaplan-Meier survival curves

Trial	Comparison	HR (95% CI) reported in journal article	HR (95% CI) reported in HTA report	HR (95% CrI) from reconstructed IPD
Overall Survival				
TAX 327	D+P / M+P	0.76 (0.62-0.94)	0.76 (0.62-0.94)	0.76 (0.620, 0.936)
CALGB 9182	M+H / H	Not reported but median survival reported as: M+H 12.3 months and; H 12.6 months (p=0.77)	1.05 (0.74, 1.49)	0.96 (0.732, 1.251)
CCI-NOV 22	M+P / P	Not reported but a total of 140 deaths reported at time of analysis (p=0.27)	0.91 (0.69, 1.19)	0.81 (0.590, 1.110)
Berry	M+P / P	Not reported but median survival reported as: M+P 23 months and; P 19 months (p=0.569)	1.13 (0.75, 1.70)	0.95 (0.628, 1.432)
Ernst	M+P+Cl/M+P	1.05 (0.78, 1.42)	1.05 (0.78, 1.42)	1.08 (0.799, 1.452)
SWOG	D+E / M+P	0.8 (0.67, 0.97)	0.8 (0.67, 0.97)	0.79 (0.659, 0.955)
Oudard	D70+E+P / M+P	Not reported but median survival reported as: D70+E+P 18.6 months,	0.94 (0.29, 1.02)	1.08 (0.675, 1.715)
	D35+E+P / M+P	D35+E+P 18.4 months and; M+P 13.4 months	0.86 (0.68, 1.08)	0.75 (0.448, 1.245)

Table 4.5: Individual trial's PFS results using IPD reconstructed from Kaplan-Meier survival curves

Trial	Comparison	HR (95% CI) reported in journal article	HR (95% CI) reported in HTA report	HR (95% CrI) from reconstructed IPD
Progression-free Survival				
TAX 327	D+P / M+P	Endpoint not collected	Not possible	Not possible
CALGB 9182	M+H / H	Not reported but median survival reported as: M+H 3.7 months and; H 2.3 months (p=0.0218)	Time to progression (calculated from number of events and p-value presented in the trial publication) HR= 1.50 (1.06, 2.13); p = 0.0218	0.74 (0.574, 0.954)
CCI-NOV 22*	M+P / P	Not reported	0.47 (0.32, 0.68)	Not possible+
Berry*	M+P / P	Not reported but median survival reported as: M+P 8.1 months and; P 4.1 months (p=0.018)	Estimated from the Kaplan-Meier curve for PFS presented in the trial publication. HR= 0.64 (0.48, 0.86)	0.63 (0.432, 0.927)
Ernst	M+P+Cl/M+P	0.81 (0.61, 1.07)	0.81 (0.61, 1.07)	0.84 (0.63, 1.112)
SWOG	D+E / M+P	Not reported but median survival reported as: D+E 6.3 months and; M+P 3.2 months (p<0.001)	time to disease progression observed for the docetaxel group compared with the mitoxantrone group: HR=1.30 (1.11, 1.52); p < 0.001	0.73 (0.627, 0.860)
Oudard*	D70+E+P / M+P	Not reported but median survival for time to PSA progression is reported as: D70+E+P 8.8 months,	Not reported	Not possible
	D735+E+P / M+P	D35+E+P 9.3 months and; M+P 1.7 months	Not reported	Not possible

*Trials where TTP was reported in the journal article or HTA report instead of PFS
+No Kaplan-Meier survival curve in published article

Table 4.6: OS and PFS HRs estimated from traditional and indirect comparison meta-analysis using reconstructed IPD

	Hazard Ratio (95% Confidence/CrI)			
	Overall Survival		Progression-free Survival	
Evidence Synthesis	Reported in HTA report	Estimated using reconstructed IPD (with non-informative priors)	Estimated using reconstructed IPD (with non-informative priors)	Estimated using reconstructed IPD (with informative priors)
Meta-analysis (M+P/P)				
Fixed Effect Analysis	0.99 (0.82, 1.20)	0.903 (0.751, 1.084)	0.641 (0.532, 0.772)	0.641 (0.532, 0.772)
Random Effects Analysis	0.99 (0.82, 1.20)	0.901 (0.405, 2.023)	0.619 (0.170, 2.048)	0.619 (0.170, 2.048)
Relative estimate (D+P/M+P)	0.76 (0.62, 0.94)	0.76 (0.620, 0.936)	0.618 (0.383, 0.941)*	0.608 (0.416, 0.861)*
Indirect comparison (D+P/P)				
Fixed Effect Analysis	Not performed	0.688 (0.523, 0.907)	0.396 (0.307, 0.512)	0.388 (0.299, 0.504)
Random Effects Analysis	0.75 (0.57, 0.99)	0.688 (0.300, 1.604)	0.381 (0.107, 1.280)	0.374 (0.105, 1.266)

*HR predicted using BRMA model

Informative prior distribution on the within-study correlation between PFS and OS was constructed based on the correlation estimates ρ_w = 0.3 (0.26, 0.32) reported by Halabi and colleagues (Halabi et al., 2009). For the construction of the informative prior distribution for the between-study correlation between OS and PFS, external trial evidence described in Section 4.2.1.1 are reviewed. Based on the inclusion criteria for identifying the trials to be used for constructing the informative prior distribution outlined in Section 4.2.1.1, six trials were identified to form the ETD Set. These included 2 RCTs from the HTA Set 2 trials and four trials from the article by Halabi and colleagues, which was used for the construction of the informative prior distribution for the within-study correlation between PFS and OS.

All three RCTs (in HTA set 2 shown in Table 4.2) presented both OS and PFS results, including Kaplan-Meier curves. However, only two of the RCTs contained docetaxel as one of its intervention and were included in the ETD Set. One of the RCTs from HTA set 2 was a three-arm RCT with two arms having docetaxel; as one of these arms administrated docetaxel once every 21-day cycle (similar to how it was administrated in Trial TAX327), only this arm and the reference arm administrating M+P were used for the prior distribution construction.

There were nine CALGB trials in the article by Halabi and colleagues, of which four trials were selected to be included in the ETD Set. One of the trials (CALGB 9182) was excluded because it was included in the main HTA clinical effectiveness analysis, two trials were excluded because they were trials reported post year 2007 and the remaining two trials were excluded because they did not report useful PFS results. Overall survival was defined as the time from randomisation or study entry to date of death from any cause. Progression-free survival was defined uniformly in all CALGB trial protocols as the time from randomisation or study entry to progression or death, whichever occurs first. Time to progression was not reported for the nine CALGB trials. Of the four trials included for the construction of the informative prior distribution for the between-study correlation between OS and PFS, two trials were RCTs and two trials were single-arm trials. As the RCTs were comparing different doses of the same drug, the arms in each of the RCTs were collapsed into a single arm and the trials were included in the prior distribution construction as single-arm trials, resulting in four single-arm trials. These trials are classified as either intervention with

or without docetaxel. All four trials presented OS Kaplan-Meier curves while only one trial presented a PFS Kaplan-Meier curve.

Treatment contrast OS and PFS estimates for the four single-arm trials were predicted using a NMA that included all six trials in the ETD Set. Details of the trials and their treatment contrast results estimated or predicted in the NMA are presented in Table 4.7. Bivariate random-effects meta-analysis of the treatment contrasts (predicted or estimated in the NMA) for OS and PFS using non-informative prior distributions on the correlations ($\rho_{w,i}$ and ρ_b) produced a posterior distribution of ρ_b = 0.07 (95% CrI: -0.94 to 0.96) for the between-study correlation of OS and PFS.

Table 4.7: NMA summary estimates for the six trials in ETD Set

Trial	Year	Number of arms	Collapsed to single-arm?	Total number of patients	NMA results for treatment contrast (Treatment with docetaxel/Treatment without docetaxel)			
					Overall Survival		Progression-free Survival	
					Log(HR)	Var(Log(HR))	Log(HR)	Var(Log(HR))
CALGB 9181	2000	2	Yes	149	-0.586	1253.006	-0.524	949.407
CALGB 9780	2001	1	NA	46	-0.535	382.202	-0.476	299.374
CALGB 9480	2002	3	Yes	390	-0.585	3225.069	-0.542	2524.280
CALGB 99813	2003	1	NA	40	-0.580	335.081	-0.552	257.394
SWOG	2004	2	No	674	-0.189	5.946	-0.302	4.383
Oudard	2005	2*	No	85	-1.027	68.919	-0.797	4.311

*Only two treatment arms of this three-armed RCT were used

The posterior distribution for the between-study correlation estimated using the ETD Set in turn forms an informative prior distribution for the between-study correlation, ρ_b, in the BRMA model for the prediction of PFS HR for treatment comparison D+P versus M+P (for trial TAX 327). Informative prior distribution for the within-study correlation, $\rho_{w,i}$, was defined using the summary statistics reported by Halabi and colleagues. In order to sample from the prior distribution for the within-study and between-study correlations, ρ_b and $\rho_{w,i}$, Fisher transformation method (discussed in Chapter 2 Section 2.4.3.5) was used to convert the correlations to corresponding Fisher correlations, $z_{w,i}$ and z_b respectively, which follow normal distributions for sampling as follows:

Within-study: $z_{w,i} \sim Normal(0.310, 0.017^2)$

Between-study: $z_b \sim Normal(0.072, 0.930^2)$

Plots of the empirical posterior distribution synthesised for the between-study correlation (on the Fisher transformation scale) and the prior distribution for the between-study correlation (specified using summary statistics of the posterior distribution) used in the BRMA model are presented in Figure 4.4.

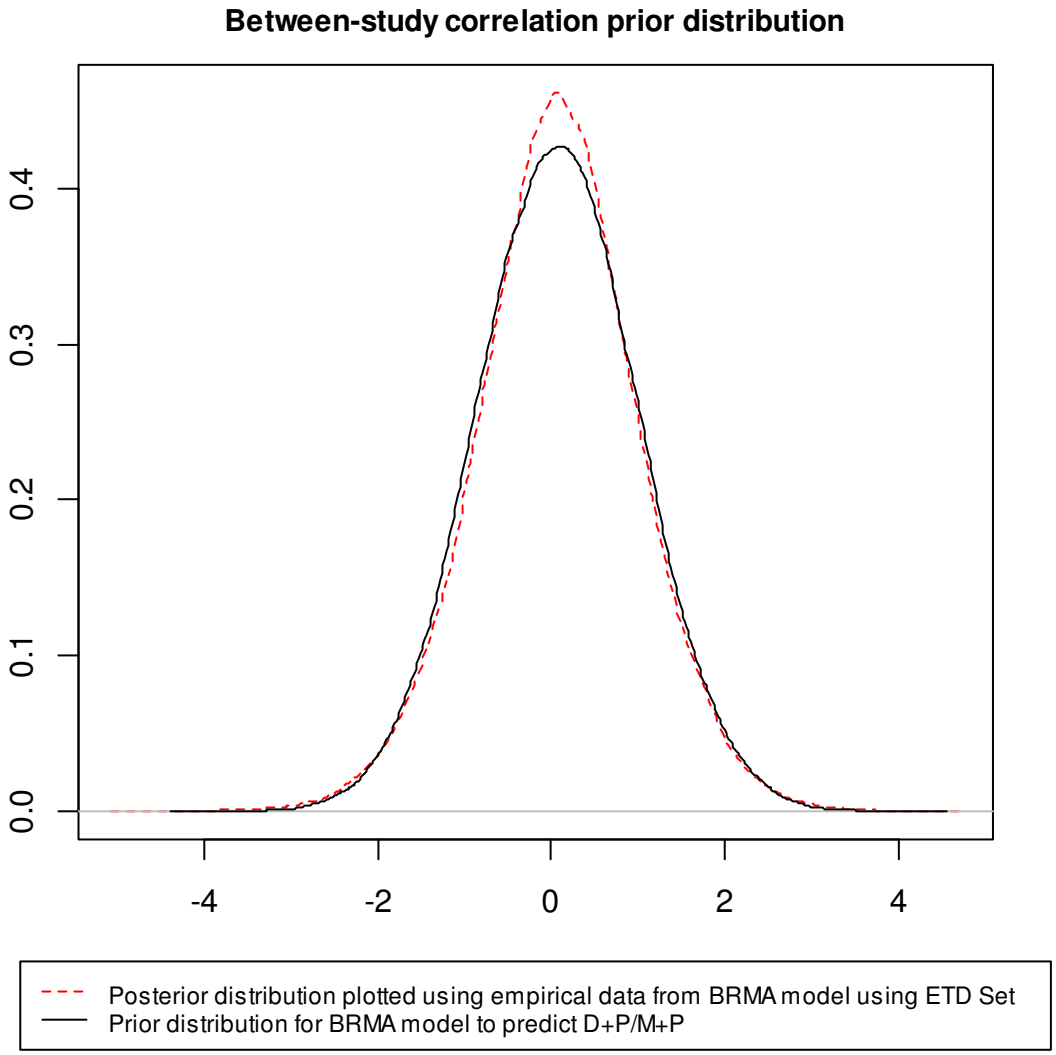

Figure 4.4: Posterior distribution and prior distribution for the between-study correlation (z_b on Fisher transformation scale) between OS and PFS

Using the informative prior distributions for the within-study and between-study correlations constructed, the predicted PFS HR for D+P versus M+P (for trial TAX

327) was 0.608 (95% CrI: 0.416 to 0.861). The predicted PFS HR for trial TAX 327 thus allow an indirect comparison to be performed to estimate treatment comparison D+P versus P. Progression-free survival HRs for D+P versus P estimated using fixed-effect and random-effects ICMA were 0.388 (95% CrI: 0.299 to 0.504) and 0.374 (95% CrI: 0.105 to 1.266) respectively.

Non-informative prior distributions for both within-study and between-study correlations between OS and PFS were employed in the BRMA model. The non-informative prior distributions take the form of uniform distributions between -1 and 1 inclusive as follows:

Within-study: $\rho_{w,i} \sim Uniform(-1, 1)$

Between-study: $\rho_b \sim Uniform(-1, 1)$

When using non-informative prior distributions for both within-study and between-study correlations between LHRs on OS and PFS, the predicted PFS HR for the comparison of D+P versus M+P (for trial TAX 327) was estimated to be 0.618 (95% CrI: 0.383 to 0.941). Progression-free survival HRs for intervention D+P versus P using non-informative prior distributions on both correlations, $\rho_{w,i}$ and ρ_b, were 0.396 (95% CrI: 0.307 to 0.512) and 0.381 (95% CrI: 0.107 to 1.280) for fixed-effect and random-effects ICMA respectively. Results of the predicted PFS HRs (for D+P versus M+P) and corresponding ICMA PFS HRs (for D+P versus P) are presented in Table 4.6.

The informative prior distributions on the within-study and between-study correlation on the LHRs between OS and PFS constructed using the Halabi Set and ETD Set provided a predicted PFS HR of D+P versus M+P that was comparable to that using non-informative prior distributions but with higher precision (width of 95% CrI: 0.445 versus 0.558). Similarly, the same was observed for the random-effects ICMA results for PFS HR M+P versus P between these two combinations of prior distributions.

4.3.1.1 Sensitivity analysis

Sensitivity analyses were conducted using combinations of prior distributions (with non-informative and/or informative) for the within-study and between-study

correlations between OS and PFS. Besides informative prior distributions constructed using ETD Set, optimistic and pessimistic (informative) prior distributions for the between-study correlation were also employed in the BRMA model. Combination pairs of prior distributions used in the BRMA model for predicting the PFS HRs of D+P versus M+P are shown in Table 4.8, together with the predicted PFS HR for D+P versus M+P and corresponding PFS HR for D+P versus P estimated using fixed-effect and random-effects ICMA.

When using informative prior distribution on the between-study correlation only, the predicted PFS HR for D+P versus M+P had higher precision (case B1: 95% CrI width=0.476) compared to the estimate using non-informative prior distributions for both the within-study and between-study correlation (case A2: 95% CrI width=0.558). The increase in precision when using informative prior distribution on only the within-study correlation was however not as great (case B2: 95% CrI width=0.520) as the former case.

However, the PFS HRs for M+P versus P estimated from ICMA in case B1 had lower precision (95% CrI width: fixed-effect=0.209, random-effects=1.187) than the estimates from ICMA in case A2 (95% CrI width: : fixed-effect=0.205, random-effects=1.173). For case B2, the PFS HRs for M+P versus P estimated from ICMA had higher precision (95% CrI width: fixed-effect=0.202, random-effects=1.160) than case A2.

Informative prior distributions in the form of pessimistic (case B3) and optimistic (case B4) prior distributions on the between-study correlation were used in this sensitivity analysis to evaluate the PFS HRs estimated considering low and high positive correlation between OS and PFS when used together with informative prior distribution on the within-study correlation constructed using Halabi Set. Both case B3 and case B4 gave PFS HRs (for D+P versus M+P and M+P versus P) with higher precision than PFS HRs estimated using the base case analysis.

Table 4.8: PFS HR estimated from the proposed Bayesian BRMA and ICMA models using non-informative and informative prior distributions

	Case	Prior distributions on the correlation between OS and PFS		Progression-free Survival Hazard Ratio (95% CrI)		
				Predicted PFS HR (D+P/M+P) using BRMA Model	Indirect Comparison Meta-Analysis (D+P/P)	
		Within-study Correlation, ρ_w	Between-study Correlation, ρ_b		Fixed Effect	Random Effects
Base Case Analysis	A1	Informative using Halabi Set ρ_w = 0.3 (0.26, 0.32)	Informative using ETD Set ρ_b = 0.07 (-0.94, 0.96)	0.61 (0.416, 0.861)	0.39 (0.299, 0.504)	0.37 (0.105, 1.266)
	A2	Non-informative ρ_w ~ Uniform (-1,1)	Non-informative ρ_b ~ Uniform (-1,1)	0.62 (0.383, 0.941)	0.40 (0.307, 0.512)	0.38 (0.107, 1.280)
Sensitivity Analysis	B1	Non-informative ρ_w ~ Uniform (-1,1)	Informative using ETD Set ρ_b = 0.07 (-0.94, 0.96)	0.62 (0.422, 0.898)	0.40 (0.306, 0.515)	0.38 (0.108, 1.295)
	B2	Informative using Halabi Set ρ_w = 0.3 (0.26, 0.32)	Non-informative ρ_b ~ Uniform (-1,1)	0.61 (0.387, 0.907)	0.39 (0.301, 0.503)	0.38 (0.106, 1.266)
	B3	Informative using Halabi Set ρ_w = 0.3 (0.26, 0.32)	Informative (Pessimistic) ρ_b = 0.2 (0.0, 0.4)	0.61 (0.414, 0.856)	0.39 (0.298, 0.503)	0.37 (0.105, 1.262)
	B4	Informative using Halabi Set ρ_w = 0.3 (0.26, 0.32)	Informative (Optimistic) ρ_b = 0.8 (0.7, 0.9)	0.59 (0.406, 0.815)	0.38 (0.291, 0.490)	0.36 (0.102, 1.230)

4.3.2 Cost-effectiveness

4.3.2.1 Transition probabilities

Transition probabilities for the WinBUGS two-state and three-state model estimated using the methods described in Section 4.2.2.1 are presented in this section. For the two-state model, only one set of transition probabilities for each of the interventions was estimated. As a result of the addition of a PD state in the three-state model, two additional sets of transition probabilities for each intervention as compared to the two-state model need to be defined. The three sets of transition probabilities that need to be defined were: (i) the transition probabilities from StD state to PD state, (ii) the transition probabilities from StD state to death state (due to other causes without progression) and (iii) the transition probabilities from PD state to death state.

The sets of transition probabilities estimated for the two models are presented in the next two sub-sections.

WinBUGS two-state model

Transition probabilities for the interventions in the HTA two-state model were estimated using the Weibull parameter values reported in the HTA report, which were obtained from an economic review submitted by Sanofi-Aventis. The parameter values were estimated from survival analysis using IPD from trial TAX 327.

Similarly, Weibull survival regression analysis was performed using the IPD reconstructed for trial TAX 327 in this thesis. Comparison of the Weibull survival analysis regression coefficients using the reconstructed OS IPD with the coefficients reported in the HTA report are presented in Table 4.9. The comparison of the corresponding covariance and Cholesky decomposition matrices are also presented in Table 4.9. The Weibull regression coefficients, covariance and Cholesky decomposition matrices using the reconstructed OS IPD were comparable to the estimates reported in HTA report.

Considering the accuracy of the parameter values estimated in the HTA report, which used patient-level data from the trial TAX 327 by Sanofi-Aventis, the Weibull parameters estimated using reconstructed OS IPD were not used for the WinBUGS

two-state model. This also enables the comparison of the cost-effectiveness analysis results of the Bayesian WinBUGS two-state model with the HTA two-state model. Transition probabilities for the interventions M+P and D+P were calculated using the Weibull regression coefficients and Cholesky decomposition matrix for each of the interventions reported in the HTA report.

Transition probabilities for intervention P were calculated by applying the (direct) HR of P versus M+P (HR=1.01 (95% CI: 0.833 to 1.220)) to the transition probabilities for M+P. A sensitivity analysis was performed to assess how the transition probabilities for P would differ if (indirect) HR of P versus D+P (HR=1.33 (95% CrI: 1.010 to 1.754)) was used by applying the indirect HR to the transition probabilities of D+P. Uncertainties around the HR parameters were also incorporated in the model as described in the methods section (Section 4.2.2). The (mean) transition probabilities for the first 12 cycles and distributions for the LHRs for the two methods of estimating transition probabilities for P (using direct and indirect HR) are presented in Table 4.10.

Table 4.9: Regression coefficients, Covariance and Cholesky decomposition matrices estimated using Weibull survival analyses

		Reported in HTA report				Using reconstructed IPD for Trial Tax 327			
Treatment		D+P		M+P		D+P		M+P	
		Mean	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
Results from Weibull regression analysis	*Intercept*	3.214	0.0546	3.036	0.0447	3.260	0.0548	3.052	0.0440
	Scale	0.6482	0.0438	0.6184	0.0371	0.6718	0.0437	0.6182	0.0360
		Intercept	*Scale*	*Intercept*	*Scale*	*Intercept*	*Scale*	*Intercept*	*Scale*
Covariance matrix	*Intercept*	0.002981		0.001998		0.003003		0.001936	
	Scale	0.000925	0.001918	0.000356	0.001376	0.000732	0.001910	0.000218	0.001296
		Intercept	*Scale*	*Intercept*	*Scale*	*Intercept*	*Scale*	*Intercept*	*Scale*
Cholesky decomposition	*Intercept*	0.054599		0.044699		0.054800		0.044000	
	Scale	0.016942	0.040385	0.007964	0.036229	0.013358	0.041612	0.004955	0.035657

Table 4.10: Transition probabilities (mean) for the first 12 cycles and distributions of LHRs for the estimation of transition probabilities for P (WinBUGS two-state model)

Cycle	D+P	M+P	P (Direct)^	P (Indirect)#
1	0.0072	0.0075	0.0076	0.0097
2	0.0134	0.0152	0.0154	0.0181
3	0.0177	0.0208	0.0211	0.0237
4	0.0211	0.0255	0.0258	0.0283
5	0.0242	0.0297	0.0301	0.0324
6	0.0269	0.0335	0.0340	0.0361
7	0.0294	0.0371	0.0376	0.0394
8	0.0318	0.0405	0.0411	0.0426
9	0.0340	0.0437	0.0443	0.0455
10	0.0361	0.0467	0.0474	0.0483
11	0.0381	0.0497	0.0504	0.0510
12	0.0401	0.0525	0.0533	0.0535

^ Distribution used was: $LHR_{P/(M+P)} \sim Normal\ (0.0101, 0.0971^2)$

Distribution used was: $LHR_{P/(D+P)} \sim Normal\ (0.2877, 0.1408^2)$

WinBUGS three-state model

Transition probabilities in the HTA and WinBUGS two-state models were calculated using Weibull parameter values estimated using OS IPD from trial TAX 327. For consistency and comparability of the three-state model with the two-state models, it is desirable to use the PFS IPD from trial TAX 327; however, this is not possible as PFS data were not recorded for trial TAX 327. Hence, trial SWOG included in the HTA report was used to calculate the transition probabilities for the three-state model based on the selection criteria described in Section 4.2.2.1 to identify the trial that has a PFS profile closest to the 'potential' PFS profile of trial TAX 327. Justifications for the choice of trial SWOG are discussed in this section.

Overall survival curves for four RCTs in HTA Full Set (excluding CALGB 9182 which used hydrocortisone instead of prednisone and CCI-NOV22 which did not report PFS) were compared to the OS curve of trial TAX 327 and presented in Figure 4.5. The OS Kaplan-Meier curves suggested that trial SWOG has an OS profile closest to trial TAX 327.

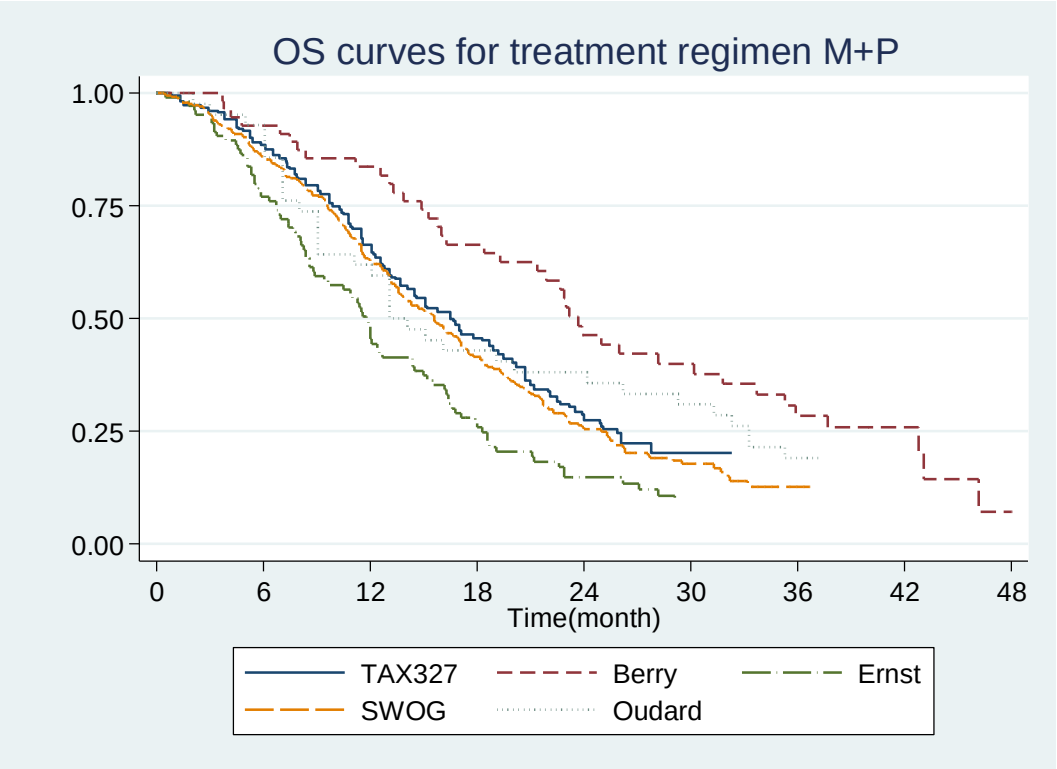

Figure 4.5: Overall survival Kaplan-Meier curves for RCTs in the HTA report

Mean time to progression for patients administrated M+P for the four RCTs were estimated using the IPD reconstructed from PFS Kaplan-Meier curves. As described in Section 4.2.2.1, as PFS was not recorded in trial TAX 327, mean number of cycles of M+P administrated in trial TAX 327 was used for comparing with the mean time to progression in the four RCTs. The mean time to progression for the four RCTs and the mean number of cycles of M+P administrated in trial TAX 327 are shown in Table 4.11. Trials Ernst and SWOG have mean time to progression closest to the assumed mean time to progression of TAX 327. The results suggested that trial SWOG has an OS profile closest to trial TAX 327 and therefore potentially a PFS profile closest to TAX 327 if the PFS endpoint had been recorded. Hence, reconstructed IPD of trial SWOG were used to estimate the transition probabilities from StD state to PD state for M+P.

Table 4.11: Mean time to progression for RCTs in the HTA report

Trial	Mean Time to Progression (SE)
Berry et al.	12.8 (1.63)
Ernst	5.9 (0.53)
SWOG	5.9 (0.33)
Oudard	4.2 (0.88)
	Mean no.of cycles (SE)
TAX 327	5.9 (0.17)

Transition probabilities from StD state to PD state for M+P and D+P need to be defined using reconstructed PFS IPD. However, this was only estimated for M+P as there was no PFS data for D+P in any of the seven RCTs in the HTA report. This transition probability from StD state to PD state for M+P was estimated from reconstructed PFS IPD of trial SWOG. Transition probability for D+P was in turn calculated by applying the predicted PFS HR of D+P versus M+P (HR=0.61 (95% CrI: 0.416 to 0.861)) to the transition probability of M+P.

Transition probability for transition from the StD state to death state was extracted from an article on cost-effectiveness analysis for advance hormone-dependent prostate cancer (Lu et al., 2012) and was applied to the model with no uncertainty as 0.005. This value was applied as the probability of transition from the StD state to death state for all interventions in the model as it defines the probability of a patient dying due to other causes (other than prostate cancer) and was expected to be the same regardless of the intervention administrated.

In the absence of patient-level data on correlated progression and death for trial SWOG, probability of transition from PD state to death state cannot be estimated. Therefore, transition probabilities for each of the interventions are estimated by assuming that the mean survival time equals the sum of the mean time from stable disease to progression and the mean time from progression to death, as described in Section 4.2.2.1. The (mean) transition probabilities estimated are presented in Table 4.12.

Table 4.12: Transition probabilities for transition from the progression state to death state for interventions D+P, M+P and P (WinBUGS three-state model)

Cycle	M+P	D+P^	P#
1	0.1485	0.0931	0.2571
2	0.1459	0.0914	0.2533
3	0.1451	0.0909	0.2521
4	0.1446	0.0905	0.2513
5	0.1442	0.0903	0.2508
6	0.1440	0.0902	0.2504
7	0.1437	0.0900	0.2500
8	0.1436	0.0899	0.2497
9	0.1434	0.0898	0.2495
10	0.1433	0.0897	0.2493
11	0.1432	0.0896	0.2491
12	0.1431	0.0896	0.2489

^ Distribution used was: $LHR_{(D+P)/(M+P)} \sim$ Normal(-0.502,0.093^2)

Distribution used was: $LHR_{P/(M+P)} \sim$ Normal(0.493,0.662^2)

4.3.2.2 Cost

Costs of drug for interventions M+P and D+P were estimated using the mean cycle of chemotherapy administrated. The reported mean number of cycles for the interventions M+P and D+P were 5.9 (SE=0.17) and 7.3 (SE=0.18) respectively, Normal distributions were assigned to the number of treatment cycles to incorporate uncertainty around these values in the two economic models. Cost of drug for M+P was £347.73/cycle and £1253.92/cycle for D+P, including £177.46/cycle for outpatient attendance fees for both drug regimens. Cost of drug for P was calculated at £1.48 per patient per cycle. In the WinBUGS two-state model, the drug costs for a patient taking P was calculated for the number of cycles that the patient remains in the StD state before transition to the death state. In the three-state model, the cost drug for a patient taking P was calculated for the number of cycles that the patient remains in the StD state before transition to the PD state. It was assumed that the patient would stop taking P after progression and hence, no drug cost would be calculated for the cycles post-progression before transition to the death state.

Follow-up and terminal care costs for interventions M+P and D+P were estimated from trial TAX 327 as reported in the HTA report (Table 36 and Table 37 in (Collins et al.,

2007)). Uncertainties for the costs were applied in the two economic models using Gamma distributions. Follow-up and terminal care costs for drug P were not available and were estimated from the costs of intervention M+P from trial TAX 327. In order to estimate the costs for drug P, a cost ratio of drug P with reference to drug M+P was estimated using costing data of P and M+P from a review article (Bloomfield et al., 1998). The mean cost ratio estimated in the WinBUGS two-state model was 1.278 (95% CrI: 0.946 to 1.691) which suggested that the mean cost of P was higher than the mean cost of M+P. This cost ratio was calculated by assigning Gamma distributions [Gamma(α, β)] of Gamma(105, 276) and Gamma(81, 285) to the cost data of P and M+P respectively. The mean cost (drug, follow-up and terminal care) per patient at each state in the economic model for each of the interventions are presented in Table 4.13.

Table 4.13: Mean cost, mean QALY and mean time spent per patient at each state in the economic model

Economic Model		Drug	Mean Cost (£) (95% CrI)	Drug cost (£)	Mean QALYs (95% CrI)	Mean Time (95% CrI)
WinBUGS two-state model						
Stable & Progression Disease State		P (direct)	11772 (6127, 20280)	26 (23, 31)	0.809 (0.5590, 1.0760)	18.1 (15.54, 21.03)
		P (indirect)	11772 (6128, 20290)	27 (22, 32)	0.812 (0.5476, 1.1100)	18.2 (14.68, 22.41)
		M+P	11237 (6855, 17030)	2057 (427, 3679)	0.813 (0.5718, 1.0580)	18.2 (16.55, 19.93)
		D+P	15862 (9066, 23020)	9152 (3261, 15050)	0.967 (0.6746, 1.2690)	21.9 (19.50, 24.58)
WinBUGS three-state model						
Stable Disease State		P (direct)	NA	6 (1, 18)	0.276 (0.0614, 0.7646)	4.1 (0.68, 12.03)
Progression Disease State		P (direct)	NA	mean cycles = 4.12 (SE:1.47)	0.573 (0.2667, 1.0080)	13.4 (6.63, 22.81)
					0.849 (0.4389, 1.4400)	17.5 (9.32, 28.90)
			10152 (5160.0, 17760.0)#			
Stable Disease State		M+P	371 (211.6, 589.9)	2047 (400, 3678)	0.377 (0.3330, 0.4247)	5.7 (5.05, 6.40)
Progression Disease State		M+P	3804 (2088.0, 6345.0)	mean cycles = 5.34 (SE:0.17)	0.512 (0.3596, 0.6674)	11.9 (10.78, 13.12)
	Terminal care	M+P	3756 (1026.0, 8239.0)		0.889 (0.7200, 1.0600)	17.6 (16.32, 18.99)
			9977 (5995.0, 15250.0)			
Stable Disease State		D+P	373 (218.1, 578.7)	9164 (3268, 15000)	0.619 (0.4967, 0.7602)	9.6 (7.67, 11.93)
Progression Disease State		D+P	2474 (1370.0, 4009.0)	mean cycles = 6.62 (SE:0.26)	0.522 (0.3664, 0.6796)	12.3 (11.15, 13.55)
	Terminal care	D+P	3326 (916.2, 7266.0)		1.141 (0.9369, 1.3520)	22.0 (19.75, 24.40)
			15337 (8594.9, 22440.0)			

#Calculated using mean cost ratio of 1.278 (95% CrI: 0.946 to 1.691) for P compared to M+P

4.3.2.3 Quality-adjusted life years

Mean EQ-5D HRQoL for all patients in the 12 months prior to death was 0.538 (95% CI: 0.461 to 0.615) as reported in the study by Sandblom and colleagues (Sandblom et al., 2004). Using this EQ-5D data, the utility for StD state in the WinBUGS two-state model was defined using a beta distribution with parameter values: Beta(21.1, 18.1). For the three-state model, the utility distribution, Beta(21.1, 18.1), for the StD in the WinBUGS two-state model was assigned as the utility distribution for the PD state. Two additional EQ-5D values as discussed in Section 4.2.2.3 were extracted from the Sandblom study (Sandblom et al., 2004), they were the EQ-5D for patients who were still surviving at the time of analysis of the study, EQ-5D = 0.770 (95% CI: 0.755 to 0.785), and EQ-5D of patients who died of other non-prostate cancer related death, EQ-5D = 0.564 (95% CI: 0.497 to 0.631). These two utility data were defined in the economic model using the following beta distributions, Beta(581.3, 173.6) and Beta(29.1, 22.5) respectively. Utility for the StD state was calculated based on the method described in Section 4.2.2.3.

As the utilities defined were selected to reflect the general HRQoL of patients with advanced prostate cancer and were independent of the interventions administrated by the patients, the utilities were used in the model for all three interventions. Mean QALY per patient for each of the interventions in the economic models are presented in Table 4.13.

4.3.2.4 Cost-effectiveness analysis results

Incremental cost-effectiveness ratios of D+P to M+P, which indicate the additional cost per extra unit of QALY from D+P compared to M+P for the HTA two-state, WinBUGS two-state and three-state models are presented in Table 4.14.

Table 4.14: Summary cost-effectiveness results for all interventions

	Cost (£)	QALY	ICER (£)	Probability cost-effectiveness (Probability of Error)		Net Benefit (95% CrI)	
Intervention	Mean (SE)	Mean (SE)	Mean	£20,000	£30,000	£20,000	£30,000
HTA two-state model (using direct MA HR for M+P/P)							
P (using direct HR)	11227	0.81001	Dominated	0.39	0.33	Not Reported	Not Reported
M+P	10834	0.81364		0.39	0.29	Not Reported	Not Reported
D+P (3 weekly)	15883	0.96801	32706	0.22	0.38	Not Reported	Not Reported
WinBUGS two-state model							
P (using direct HR)	11772 (3642.9)	0.809 (0.1324)		0.32 (0.68)	0.28 (0.72)	4417 (-5119.0, 12360.0)	12512 (1647.9, 22500.3)
M+P	11237 (2615.1)	0.813 (0.1246)		0.36 (0.64)	0.28 (0.72)	5023 (-2327.0, 11680.3)	13153 (4116.0, 21800.0)
D+P (3 weekly)	15862 (3556.8)	0.967 (0.1517)	30026	0.32 (0.68)	0.44 (0.56)	3479 (-5749.1, 12560.0)	13149 (1859.0, 24430.0)
P (using indirect HR)	11772 (3642.8)	0.812 (0.1451)		0.33 (0.67)	0.29 (0.71)	4475 (-5383.2, 12990.0)	12599 (1178.0, 23520.3)
M+P	11237 (2615.1)	0.813 (0.1246)		0.37 (0.63)	0.30 (0.70)	5023 (-2327.0, 11680.3)	13153 (4116.0, 21800.0)
D+P (3 weekly)	15862 (3556.8)	0.967 (0.1517)	30026	0.30 (0.70)	0.41 (0.59)	3479 (-5749.1, 12560.0)	13149 (1859.0, 24430.0)
WinBUGS three-state model							
P (using direct HR)	10152 (3255.6)	0.849 (0.2584)		0.30 (0.70)	0.27 (0.74)	6829 (-4364.0, 19720.0)	15320 (829.8, 33820.0)
M+P	9977 (2365.9)	0.889 (0.0871)		0.33 (0.67)	0.20 (0.81)	7810 (1724.0, 13090.0)	16704 (9635.0, 23240.3)
D+P (3 weekly)	15337 (3514.3)	1.141 (0.1061)	21340	0.37 (0.63)	0.54 (0.47)	7474 (-647.4, 15390.0)	18879 (9606.0, 28120.0)

The results of the analyses showed that using our WinBUGS two-state model, the ICER calculated was £30026 as compared to £32706 reported in the HTA report. As for the three-state model, the ICER calculated using a predicted PFS HR of 0.61 (95% CrI: 0.416 to 0.861) for D+P versus M+P, which was estimated using informative prior distributions for both the within-study and between-study correlations, was £21340. Our results showed that by specifying a three-state model and taking into account the cost and QALY at the PD state, the ICER estimated was lower than that of the two-state model.

The net benefit for each intervention and the probability that each intervention was cost-effective at £20000 and £30000 are presented in Table 4.14. At £20000, for the two-state models, M+P had the highest probability of being cost-effective amongst all three interventions while D+P has the highest probability of being cost-effective in the three-state model. However, at £30000, D+P had the highest probability of being cost-effective amongst all three interventions for all the models, with the probability being as high as 0.54 in the three-state model compared to the two-state models (at 0.44 and 0.41 for the WinBUGS direct and indirect HR models). This is also presented graphically in cost-effectiveness acceptability curves over a range of cost values. Cost-effectiveness acceptability curves comparing all the interventions for the WinBUGS two-state model using direct head-to-head meta-analysis HR of P versus M+P, WinBUGS two-state model using ICMA HR of P versus M+P and WinBUGS three-state model using direct head-to-head meta-analysis HR of P versus M+P are presented in Figure 4.6, Figure 4.7 and Figure 4.8 respectively. In terms of net benefit, at £20000 M+P had the highest net benefit amongst all three interventions in all models. However, at £30000, M+P had the highest net benefit in the two-state model while D+P had the highest net benefit in the three-state model.

Cost-effectiveness acceptability curves and cost-effectiveness plane looking into the difference between intervention D+P and M+P (with M+P as the reference intervention) were also plotted for comparing the WinBUGS two-state model with the three-state model and are presented in Figure 4.9 and Figure 4.10 respectively. The cost-effectiveness acceptability curves showed that after approximately £10000, the probability that D+P was more cost-effective than M+P was higher in the three-state model than the two-state model. For example, at £30000, the probability was around

0.7 for the three-state model compared to 0.5 for the two-state model as shown in Figure 4.9. The cost-effectiveness plane showed that D+P was more effective (in terms of utility) than M+P for both the two-state and three-state models, although the certainty of its effectiveness was greater in the three-state model than the two-state model as all the 50000 Monte Carlo simulations points of the three-state model (in green colour) were in the positive utility difference area. Besides being more effective, D+P was also more costly than M+P. However, the degrees of uncertainty for the difference in cost of the two interventions were comparable for both models.

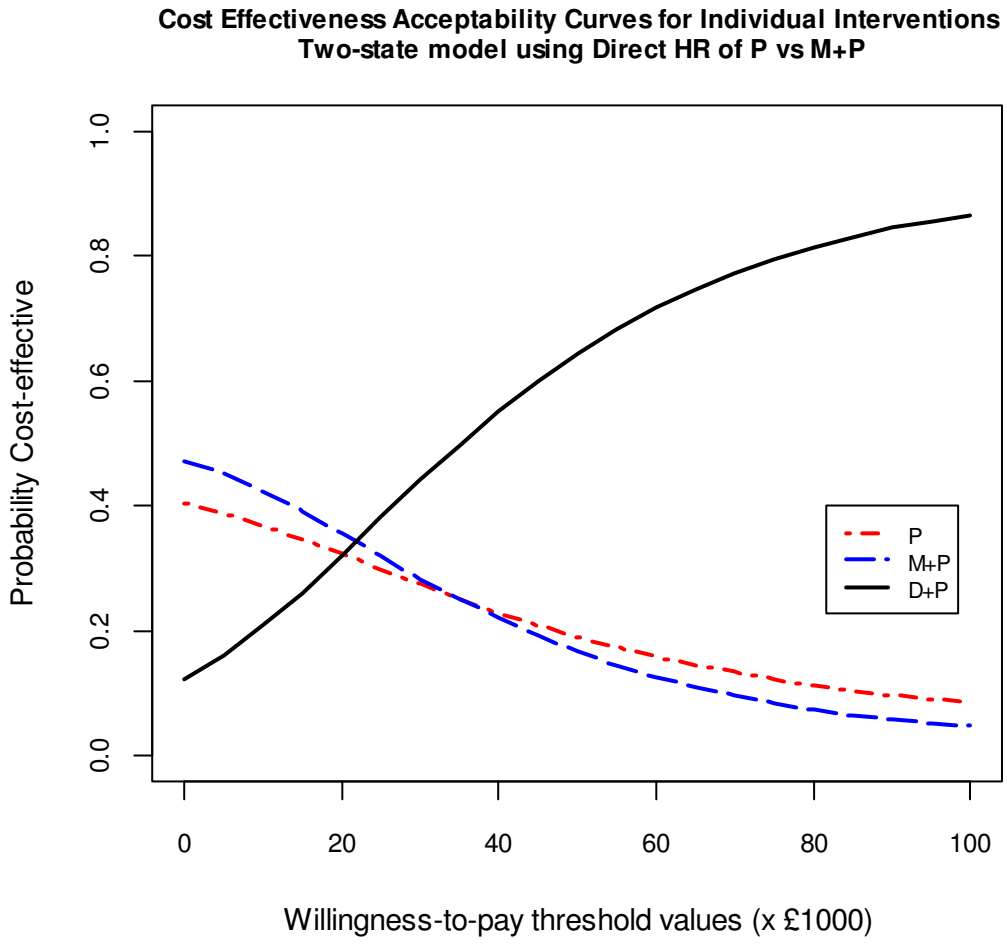

Figure 4.6: Cost-effectiveness acceptability curves for all three interventions – WinBUGS two-state model using direct HR of P versus M+P

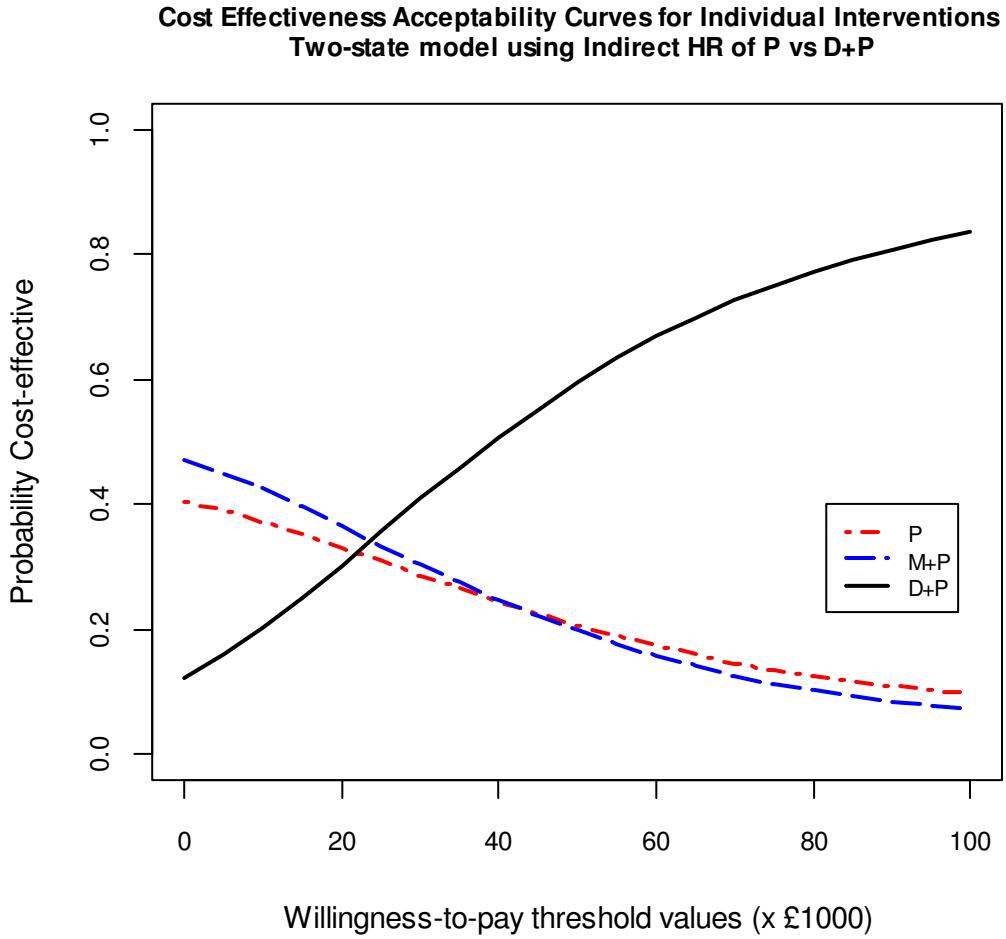

Figure 4.7: Cost-effectiveness acceptability curves for all three interventions – WinBUGS two-state model using indirect HR of P versus D+P

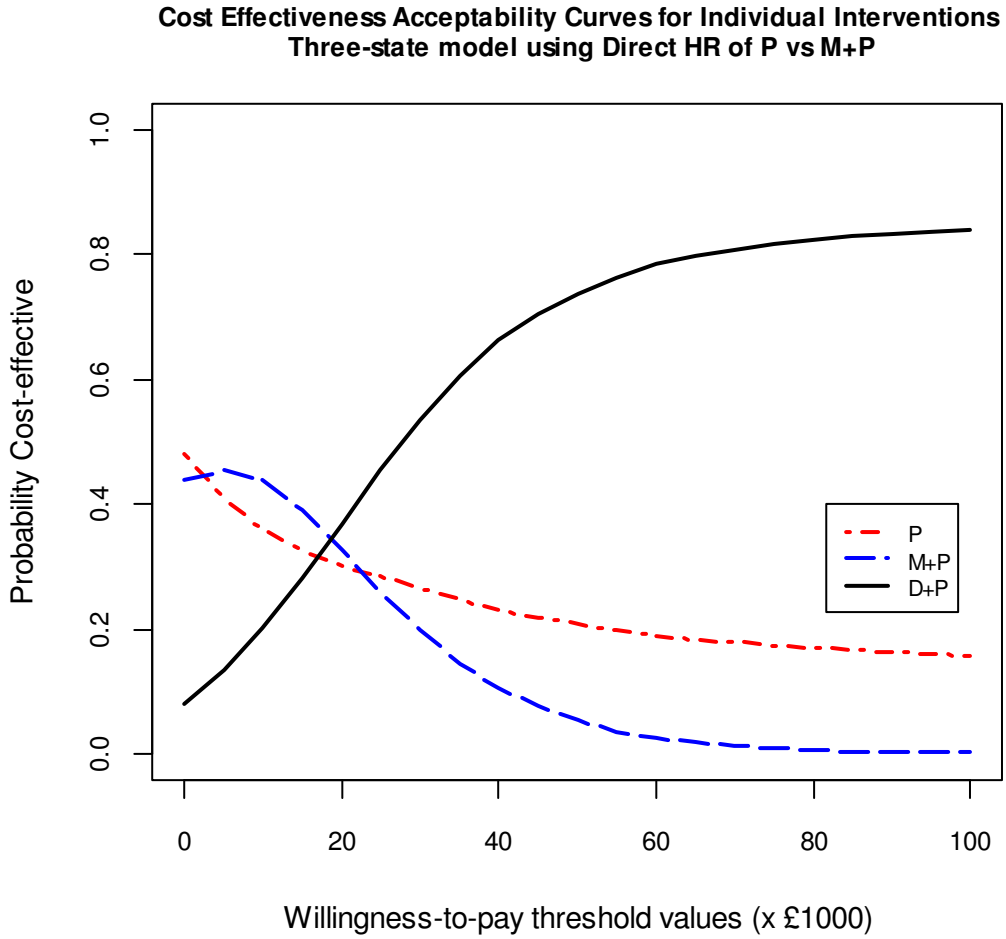

Figure 4.8: Cost-effectiveness acceptability curves for all three interventions – three-state model using direct HR of P versus M+P

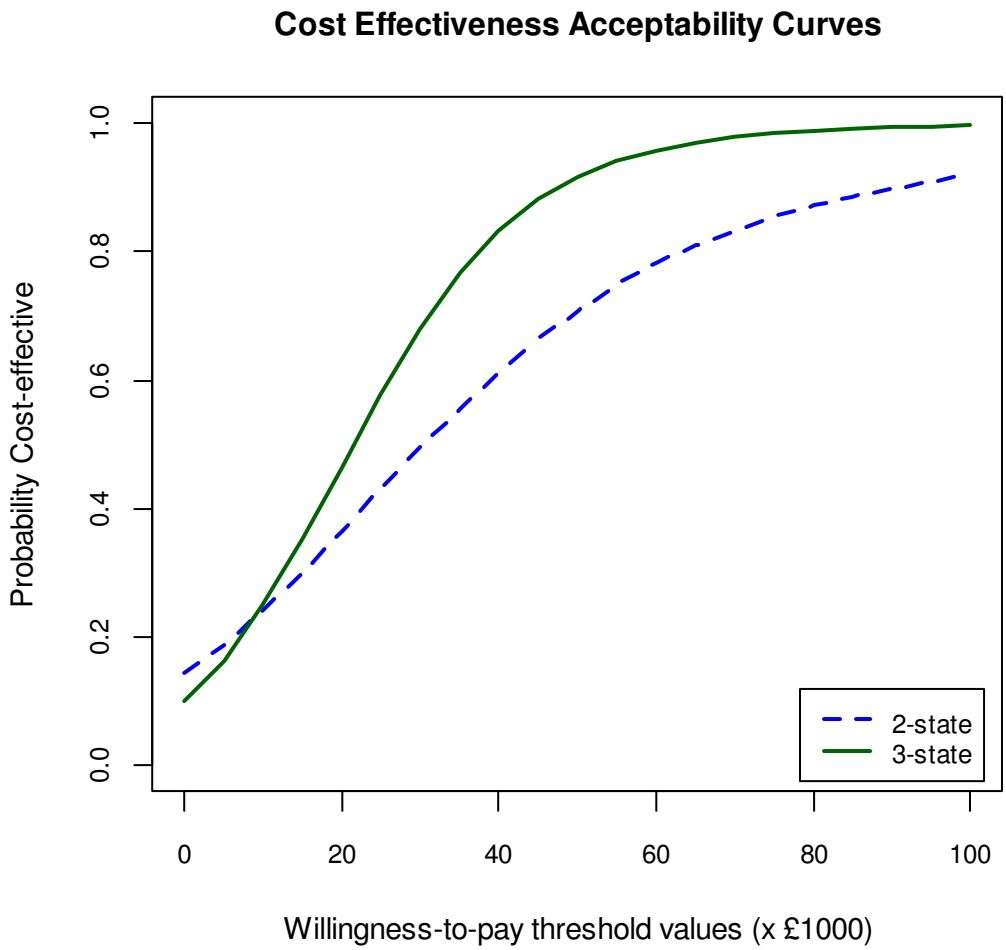

Figure 4.9: Cost-effectiveness acceptability curves for WinBUGS two-state and three-state economic models

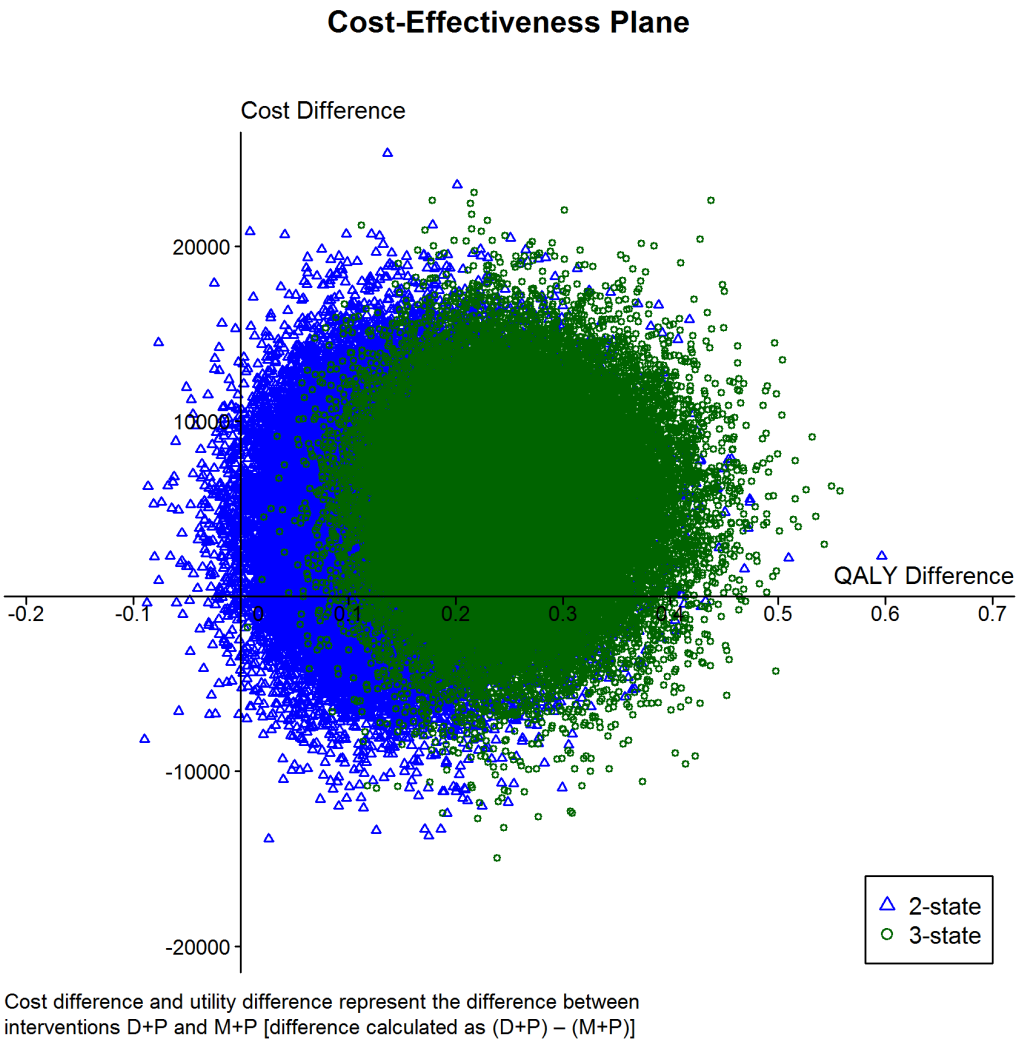

Figure 4.10: Cost-effectiveness plane for WinBUGS two-state and three-state economic models

4.3.2.5 Sensitivity Analysis

Sensitivity analyses were conducted using combinations of non-informative and informative prior distribution for the within- and between-study correlation in the BRMA model and various follow up cost division factors to assess their influence on the estimation of the ICER values. The analyses showed that if the between-study correlation between OS and PFS was relatively high (ρ_b=0.8), the ICER calculated would be lower than the ICER calculated in the base case (with $\rho_b = 0.07$). Varying the follow up cost division factors from 0.7 to 0.8 changed the ICER slightly from £21694 to £20986. Results of the sensitivity analyses are presented in Table 4.15.

Table 4.15: Sensitivity Analyses using combinations of non-informative and informative prior distribution for the within- and between-study correlation in the BRMA model and various follow up cost division factors

Clinical Effectiveness BRMA Model			Cost-Effectiveness Markov Three-state Model			
Prior distributions on the correlation between OS and PFS		Predicted		Difference [(D+P) minus (M+P)]		
Within-study Correlation, ρ_w	Between-study Correlation, ρ_b	PFS LHR (D+P/M+P) using BRMA Model	Follow up cost Division Factor	Cost Mean (SE)	QALY Mean (SE)	ICER
Base Case for cost-effectiveness analysis						
Informative* ρ_w = 0.3 (0.26, 0.32)	Informative^ ρ_b = 0.07 (-0.94, 0.96)	LHR ~ Normal (-0.50, 0.0934^2)	ψ = 0.75	5360 (4244.60)	0.251 (0.0646)	21340
Sensitivity Analyses						
Non-informative ρ_w ~ Uniform (-1,1)	Non-informative ρ_b ~ Uniform (-1,1)	LHR ~ Normal (-0.49, 0.0930^2)	ψ = 0.75	5350 (4243.95)	0.244 (0.0638)	21966
Non-informative ρ_w ~ Uniform (-1,1)	Informative^ ρ_b = 0.07 (-0.94, 0.96)	LHR ~ Normal (-0.48, 0.0932^2)	ψ = 0.75	5344 (4243.56)	0.239 (0.0636)	22360
Informative* ρ_w = 0.3 (0.26, 0.32)	Non-informative ρ_b ~ Uniform (-1,1)	LHR ~ Normal (-0.51, 0.0931^2)	ψ = 0.75	5363 (4244.79)	0.254 (0.0647)	21146
Informative* ρ_w = 0.3 (0.26, 0.32)	Informative (Pessimistic) ρ_b = 0.2 (0.0, 0.4)	LHR ~ Normal (-0.50, 0.0933^2)	ψ = 0.75	5362 (4244.72)	0.253 (0.0647)	21216
Informative* ρ_w = 0.3 (0.26, 0.32)	Informative (Optimistic) ρ_b = 0.8 (0.7, 0.9)	LHR ~ Normal (-0.53, 0.0931^2)	ψ = 0.75	5380 (4245.94)	0.267 (0.0658)	20145
Informative* ρ_w = 0.3 (0.26, 0.32)	Informative^ ρ_b = 0.07 (-0.94, 0.96)	LHR ~ Normal (-0.50, 0.0934^2)	ψ ~ Beta(24.25, 8.08)	5418 (4198.72)	0.251 (0.0644)	21594
Informative* ρ_w = 0.3 (0.26, 0.32)	Informative^ ρ_b = 0.07 (-0.94, 0.96)	LHR ~ Normal (-0.50, 0.0934^2)	ψ = 0.70	5449 (4222.42)	0.251 (0.0646)	21694
Informative* ρ_w = 0.3 (0.26, 0.32)	Informative^ ρ_b = 0.07 (-0.94, 0.96)	LHR ~ Normal (-0.50, 0.0934^2)	ψ = 0.80	5271 (4267.87)	0.251 (0.0646)	20986

*Informative prior distribution using Halabi Set; ^Informative prior distribution using ETD Set

4.4 Discussions

The proposed Bayesian BRMA model illustrated how OS and PFS endpoints could be modelled jointly for the prediction of PFS HR for D+P versus M+P. The PFS HRs estimated by this model using informative prior distributions for either or both of the within-study and between-study correlations between the OS and PFS endpoints were comparable to the PFS HR estimated using solely non-informative prior distributions for the correlations although higher precision was achieved when informative priors were used. Sensitivity analysis case B1 and B2 (scenarios in Table 4.8) demonstrated that the informative prior distribution for the between-study correlation contributed more to the increase in precision of the predicted PFS HR than the informative prior distribution for the within-study correlation. Sensitivity case B2 (solely informative within-study correlation) when compared to case A2 (non-informative between- and within-study correlations) showed that the predicted PFS HRs were comparable and hence the use of the correlation between OS and PFS as a crude approximation for the correlation between OS LHR and PFS LHR has minimal impact of the predicted HR for D+P versus M+P. When keeping the informative prior distribution for the within-study correlation the same, the relative effectiveness of the interventions on the PFS increases as the magnitude of the between-study correlation increases (as seen in cases A1, B3 and B4).

Similarly, there is an increase in precision for the indirect comparison results for PFS HR of D+P versus P estimated using the predicted PFS HR in the ICMA, except for case B1 (informative prior for between-study correlation using ETD Set). Although the ICMA PFS HR point estimate for case B1 was comparable to that reported in case A2 (non-informative priors), the PFS HR had lower precision than case A2. This highlights that although the use of external information in the form of informative priors generally results in increase precision for the analysis estimates, the use of informative priors can sometimes results in increased uncertainty instead as illustrated in this project. Hence, the estimates from using informative priors in Bayesian meta-analysis also depend upon whether there is an associated decrease or increase in heterogeneity especially for random-effects models.

The aim of this project is to predict the PFS HR of D+P versus M+P to allow for the specification of a three-state model for mHRPC, which was not possible in the previous

analysis in the HTA report due to the absence of PFS data for trial TAX 327; this is achieved with the use of the BRMA model. The intervention D+P was accepted by NICE technology appraisal committee to be introduced for the treatment of mHRPC at an ICER of £32706. This intervention (D+P) with this high ICER was accepted as the disease under consideration is considered under the end of life treatment classification of NICE. The ICER computed using the proposed three-state model of £21340 therefore would not have changed the decision as it is lower than the ICER that the treatment regimen was accepted. However, the proposed three-state model is a more realistic representation of the disease course of patients with mHRPC. It also allows more detailed estimation of the cost and QALY experience of the cancer patients, hence achieving a more appropriate and representative estimate of the cost-effectiveness of the interventions under evaluation.

Cost data used in the three-state were estimated using the data reported in the HTA report. As the economic model specified in the HTA report is a two-state model, cost data were available for only two states: the StD state and the death state. Assumptions on the possible cost at the PD state have to be made when specifying the three-state model. One assumption is that the follow-up cost at PD state was included in the StD state of the two-state model. Another assumption is that the follow-up cost for the PD state made up of 75% of the follow-up cost in the StD state of the two-state model (based on the fact that the cost was associated with subsequent chemotherapy and palliative care most likely to occur post progression). The lack of detailed cost data is a limitation in the calculation of cost data for the three-state economic model; however, the sensitivity analyses show that the ICER did not vary greatly using different proportions (70% and 80%) for calculating the follow-up cost at PD state in the three-state model.

Introducing more parameters to the three-state model, such as for the QALY, may result in increased uncertainty around the resulting cost-effectiveness estimates. In this project, however, introducing more parameters as a more realistic approach has led to reduced uncertainty. This is because the additional parameters are estimated with high precision whilst the parameters in the two-state model may carry a lot of uncertainty.

The reconstructed IPD for OS and PFS endpoints in the trials allowed time-dependency to be implemented in the transition probabilities using parametric Weibull survival

analysis. Hence, transition probability for transition from StD state to death state in the WinBUGS two-state model and transition probability for transition from StD state to PD state in the three state model were appropriately calculated using reconstructed IPD for OS and PFS respectively. Although IPD for OS and PFS were reconstructed, the data are not paired, meaning we do not have both the death and progression event data of each individual subject in the trial. As such, it is not possible to estimate the survival function from progression to death to estimate the transition probabilities for transition from PD state to death state in the WinBUGS three-state model. Transition probabilities for each of the interventions are estimated by assuming that the mean survival time equals the sum of the mean time from stable disease to progression and the mean time from progression to death, assuming that the contribution to mean survival time by patients who died of causes other than prostate cancer is negligible as the proportion of such patients is small. Estimation of this transition probability without making such assumption would have been possible if patient-level IPD are available.

Another limitation of this project is the use of PFS as an approximation to TTP because PFS is more commonly reported in published articles than TTP. This was the case for the trials in this project. As PFS considers death unrelated to cancer progression as an event in the survival analysis compared to TTP, the rate of transition from StD state to PD state using PFS would be faster than that using TTP. However, based on the assumptions that the proportion of patients who died of other causes unrelated to prostate cancer disease progression is expected to be small (due to the advanced metastatic stage of the disease) and comparable between the treatment arms (non-cancer deaths independent of cancer treatments), the difference in rates of transit from StD state to PD state between using PFS and using TTP is expected to be minimal and the PFS HR and TTP HR between the treatment arms are expected to be comparable. Hence, although the cost and QALY for the StD state is expected to be (minimally) lower when using PFS as compared to using TTP, the ICER for D+P compared to M+P computed using PFS HR is expected to be comparable to the ICER computed using TTP HR. Nevertheless, it would have been possible to predict TTP HR for trial TAX 327 if IPD for the derivation of TTP from the trials utilised in the BRMA and ICMA were available or obtainable within the resources of this project.

As mentioned earlier, patients who started treatment on M+P were likely to be given D+P as second-line chemotherapy treatment post-progression. The OS HR reported for trial TAX 327 using the intention-to-treat principle for clinical trials is likely to be an underestimation of the "true" OS HR where no treatment switching has occurred, giving a predicted PFS HR that is also an underestimation of the "true" PFS HR should the PFS data had been recorded for trial TAX327. This implies that for the economic analysis, when applying the predicted PFS HR to the transition probabilities for M+P to estimate the transition probabilities for D+P, the cohort of patients taking D+P was transiting to the PD state at a faster rate than if the original PFS data had been recorded. Equivalently, this will have an impact on the cost and QALY estimates of the cost-effectiveness analysis.

In summary, the inclusion of the PD state to the original two-state model enables a more realistic representation of the natural disease pathway of mHRPC patients. However, transition probability has to be appropriately specified and it is not easy and can be a challenge especially when PFS data are not available. Where PFS data was not recorded, the Bayesian BRMA model proposed in this project serves to predict the "missing" PFS HR to enable the specification of a three-state model incorporating the PD state. It also allows for the inclusion of competing risk of morbidity to be incorporated into the model for cost-effectiveness analysis to better inform health policy decision making.

5 Use of Bayesian evidence synthesis to inform estimation of utility in multiple sclerosis

5.1 Introduction

In the UK, Health Technology Assessment (HTA) agencies such as NICE appraise new health technologies on the basis of cost-effectiveness analyses. In such analyses, the estimation of health related quality of life (HRQoL) used to estimate the gain in quality adjusted life years (QALY) under new treatment compared to the standard care is required. Typically, HRQoL is measured using the EuroQol 5-Dimensions (EQ-5D) Questionnaire (Rabin and de Charro, 2001) across most diseases and is the recommended HRQoL score for cost-effectiveness analyses by NICE. However, this score is not always reported in randomised controlled trials (RCTs) and often mapping techniques are used to estimate EQ-5D from disease specific measures.

For example, in rheumatoid arthritis, a common measure of functional status is Health Assessment Questionnaire (HAQ) and regression-based methods for mapping of HAQ onto EQ-5D have been developed (Pennington and Davis, 2014). In cancer, the Functional Assessment of Cancer Therapy-General (FACT-G) and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30 (EORTC QLQ-C30) are the most widely used cancer-specific HRQoL instruments and regression methods for mapping these cancer-specific HRQoL measures onto EQ-5D have been developed (Cheung et al., 2009, Teckle et al., 2013, Crott and Briggs, 2010). Other cancer-specific HRQoL, such as the functional assessment of cancer therapy-prostate (FACT-P) may be used for metastatic castrate-resistant prostate cancer and similar regression-based models have also been constructed for mapping FACT-P onto EQ-5D (Diels et al., 2015).

In relapsing remitting multiple sclerosis (MS), a common long-term outcome measure in RCTs is disability progression measured by Expanded Disability Status Scale (EDSS). Orme and colleagues (Orme et al., 2007) developed a mapping model to estimate EQ-5D in patients with MS depending on their EDSS score. This mapping algorithm has been used in a technology appraisal (Asaria et al., 2011, Novartis

Pharmaceuticals UK Ltd, 2011) conducted by NICE. The algorithm, however is developed based on a single cohort of patients and applied to studies even when data on EQ-5D was collected as part of the RCT of the treatments under consideration. This was the case of the FREEDOMS (Kappos et al., 2010) and TRANSFORMS (Cohen et al., 2010) trials of Fingolimod. This approach may lead to biased estimates by not taking into account the variability between populations. However, individual patient data (IPD) is not always available to the analysts conducting the technology appraisal and they need to rely on evidence from publicly accessible resources such as the publication by Orme and colleagues. The aim of the work presented in this chapter was to investigate the use of advanced meta-analytic methods to include all available evidence in estimating EQ-5D from EDSS beyond the data from a single cohort. The technology appraisal of Fingolimod is used in this chapter as illustrative example of extending the evidence on HRQoL in MS beyond the Orme data.

5.1.1 Motivating example

In 2011, NICE conducted a single technology appraisal (STA) of Fingolimod for the treatment of relapsing remitting MS in adults (Asaria et al., 2011, Novartis Pharmaceuticals UK Ltd, 2011). Two RCTs (FREEDOMS (Kappos et al., 2010) and TRANSFORMS (Cohen et al., 2010)) were used for the cost-effectiveness analysis. Although utility data for the patients in the trials were collected, these data were not used in the cost-effectiveness analysis. In place of the EQ-5D data collected as part of the RCTs, EQ-5D utility values were estimated from data collected on the EDSS scores by using the mapping model developed by Orme and colleagues (Orme et al., 2007). The relationship between EQ-5D and EDSS reported in the article was estimated using data from a questionnaire conducted by the MS Trust UK in 2005.

The objective of this project is to make use of Bayesian evidence synthesis methods to incorporate available evidence about the relationship between EDSS and EQ-5D from published articles in the development of mapping from EDSS onto EQ-5D, such as the relationship developed by Orme and colleagues. EQ-5D utility values estimated for different EDSS scores have important application in cost-effectiveness analysis. The economic evaluation of new interventions in MS is based on multi-state models, where the health states relate to different levels of disease severity characterised by the EDSS

scores. In order to attach utility values to all health states, such values need to be associated with EDSS scores of the states.

A request to obtain IPD of the MS cohort in Orme's study was made to the first author of the article. However, the author no longer has access to the IPD and hence the IPD data was not available for analysis in this project. In the absence of IPD from Orme's study, data were simulated to serve the purpose of methodological development of mapping with the use of external summary data. Methods of data simulation, search strategy for the identification of external evidence on the relationship between EQ-5D and EDSS scores and modelling techniques of both the simulated IPD and the external summary data are presented in Section 5.2. Results of the Bayesian evidence synthesis and regression analyses are presented in Section 5.3. Strengths and limitations of the use of Bayesian evidence synthesis for informing the estimates of EQ-5D utility in MS as illustrated in this chapter are discussed in Section 5.4.

5.2 Methods

Methods used for the estimation of health utility values (EQ-5D) from EDSS scores in patients with MS using Bayesian evidence synthesis techniques are described in this section. A schematic diagram depicting the procedures involved in the estimation of EQ-5D for patients with MS is shown in Figure 5.1.

Firstly, patient-level data on EQ-5D, EDSS, gender, recent relapse status, type of MS, education and years since diagnosis were simulated based on the model developed by Orme and colleagues. Methods of simulation for these data are presented in Section 5.2.1. The simulated data are then used as the likelihood data in the Bayesian linear regression analysis as shown in Figure 5.1 (procedures depicted in black).

Secondly, Bayesian linear regression models to estimate EQ-5D values corresponding to different EDSS scores are described in Section 5.2.2. Non-informative prior distributions (shown in orange in Figure 5.1) for the regression coefficients of the Bayesian linear regression model are also presented.

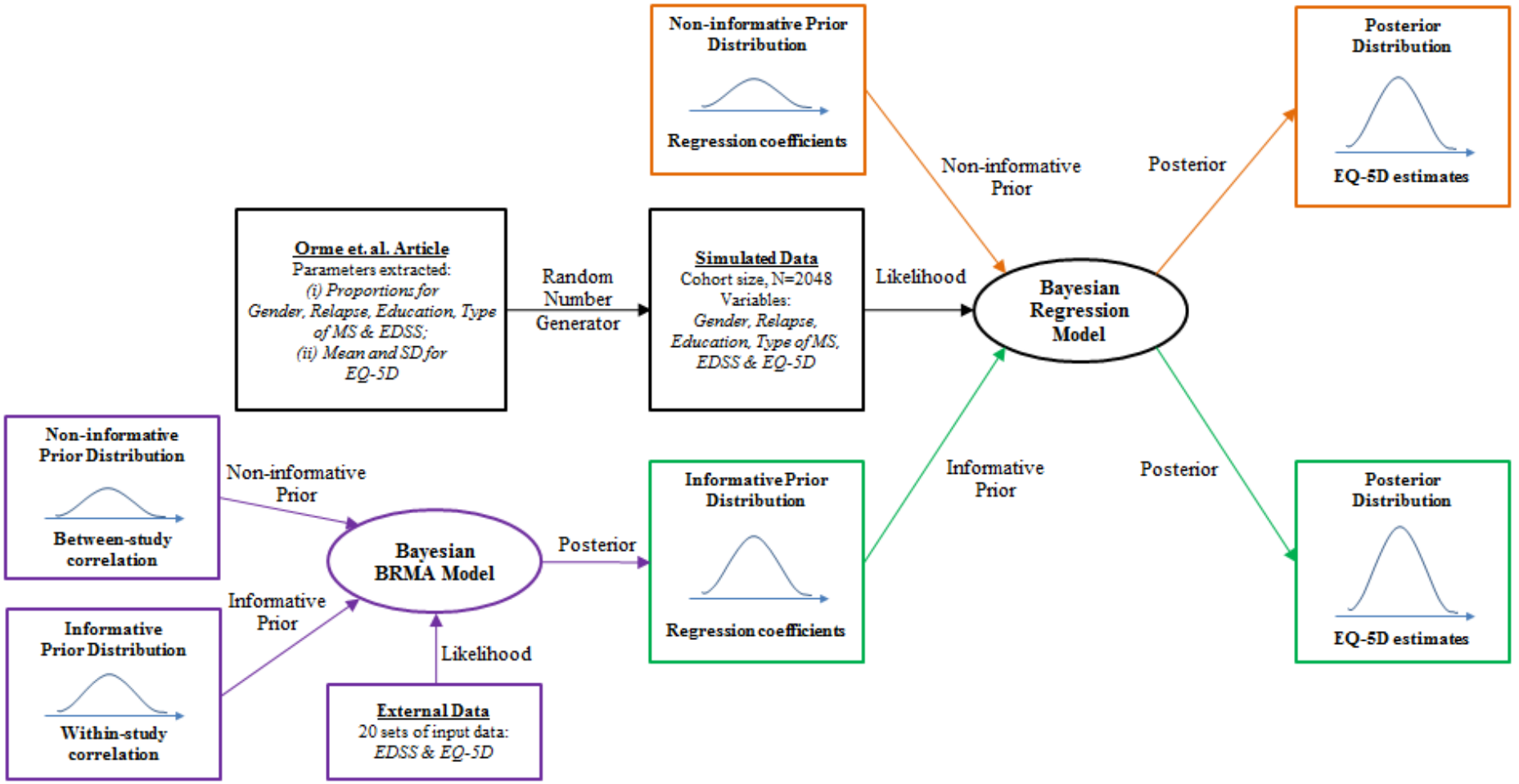

Figure 5.1: Schematic Diagram showing analysis plan for the estimation of EQ-5D using Bayesian techniques.

Thirdly, external summary data on EQ-5D and EDSS obtained from published articles are used to construct informative prior distributions (the informative prior distribution in green in Figure 5.1) for the regression coefficients in the Bayesian linear regression models. Search strategy and selection criteria of studies for constructing the prior distributions using Bayesian bivariate random-effects meta-analysis (BRMA) are discussed in Section 5.2.3.1. The BRMA model for constructing the informative prior distributions (in green in Figure 5.1) using external data is described in Section 5.2.3.2. Informative prior distribution on the within-study correlation and non-informative prior distribution on the between-study correlation between EQ-5D and EDSS in the BRMA model (all in purple in Figure 5.1) are also described in Section 5.2.3.2.

Lastly, data transformation considerations when using sequential Bayesian models (Bayesian BRMA and Bayesian linear regression models) are discussed in Section 5.2.4. Sensitivity analyses utilising alternative methods of simulating patient-level data on EQ-5D and modelling assumptions in the regression model for estimating EQ-5D are described in Section 5.2.5.

5.2.1 Simulation of patient-level data

Patient-level data were simulated to mimic the patient cohort in Orme's study reported in the published article (Orme et al., 2007). The cohort of 2048 patients was simulated using the model developed by Orme and colleagues. Variables identified to be associated with the independent variable, EQ-5D, in a multivariable linear regression model reported in the article were simulated using summary statistics reported for all the variables. The variables included EDSS, gender, recent relapse status, type of MS, education and years since diagnosis. As the summary statistics for the variable: years since diagnosis was not reported in the article, IPD for this variable could not be simulated, however the effect of this variable was small compared to the other covariates.

Mean EQ-5D for each EDSS score was not reported in the article; however, regression coefficients for EQ-5D at different EDSS scores from a multivariable linear regression analysis were reported. Mean EQ-5D values corresponding to each EDSS value were estimated using the regression coefficients (and applying them to the baseline value).

Estimation of the standard deviations for EQ-5D at each EDSS score was not straight-forward as the model by Orme reported predicted values from a regression model assuming normality of the data. Using the estimates from the model for data simulation (using normal distribution) resulted in larger variability of the data (which was observed when applying the same model to the simulated data). In order to simulate the EQ-5D values for this cohort of patients, assumptions about the standard deviations of the EQ-5D values were made as described in Section 5.2.1.2. All variables are simulated independently of the other variables as the relationships between the variables in this study population are not known from the article. The R software (R Core Team, 2012) was used for the simulation of the data.

5.2.1.1 Data simulation of categorical variables

Data for binary variables (gender and recent relapse status) for the cohort were simulated using binomial distribution with probabilities reported in Table 3 of the article (Orme et al., 2007). Categorical variables with more than two groups (EDSS, education and type of MS) were simulated using multinomial distribution with probabilities reported in Table 3 (for education and type of MS) and Figure 1 (for EDSS) of the article (Orme et al., 2007). Summary data of patient demographics reported in Table 3 and Figure 1 of the article published by Orme and colleagues were reproduced in Table 5.3 (in Section 5.3.1) for comparisons with the simulated data. Data for categories defined as "missing" (for gender) or "not answered" (for education level) in Table 3 of Orme's article were not simulated. The proportions for each of the other categories (in variables: gender and education level) were adjusted using the sum of proportions in each category as the denominator.

5.2.1.2 Data simulation of continuous variable

EQ-5D is a quality of life utility measure that is bounded by a range from -0.594 to 1 (Dolan and Roberts, 2002), where value of 1 indicates a high quality of life while lower values indicate worse quality of life. As a result of the bounded range of values for EQ-5D, EQ-5D data are usually not normally distributed. This rendered simulating EQ-5D data using normal distribution inefficient. However, a number of probability

distributions (including the normal distribution) are explored in this thesis to evaluate which is the best distribution for simulating EQ-5D data for this cohort of patients.

Statistical distributions explored included the normal distribution, lognormal distribution and beta distribution, where the lognormal distribution and beta distribution were discussed as possible distributions for utilities by Briggs and colleagues (Briggs et al., 2006) (page 92). The sub-sections that follow describe the methods for simulating EQ-5D using the three distributions. Methods for assessing the simulated data are discussed in Section 5.2.1.3.

Normal distribution

As mentioned above, EQ-5D data are generally not normally distributed and simulating such data from a normal distribution does not seem appropriate. However, the mean values and standard deviations for EQ-5D at each EDSS score were not reported by Orme and colleagues, and hence results of the model fitted by the authors were used for data simulation. Since the authors of the paper conducted multivariable linear regression analysis (with EQ-5D as dependent variable) under assumption of normality, it was reasonable in this situation to use the normal distribution to simulate the EQ-5D in order to reproduce IPD for this study.

The 95% confidence intervals (CIs) reported for the mean values of EQ-5D at each EDSS score are used to compute the standard deviations using the equation below:

$$\widehat{\sigma_i} = \frac{\sqrt{N_i} \times W_{95\%CI,i}}{2 \times 1.96}$$

where $\widehat{\sigma_i}$ is the estimated standard deviation for EQ-5D at a given EDSS score i ($i = 0,1 \dots ,6,6.5,7, \dots ,9$); $W_{95\%CI,i}$ is the width of the 95% CI for mean EQ-5D at EDSS score i; N_i is the number of patients with EDSS score i. This standard deviation, which is termed the "original SD" for the rest of this chapter, is mathematically an overestimation of the true standard deviation for each of the EDSS scores. This is because the above equation did not take into account the covariance between the regression coefficients in the multivariable linear regression model. Besides that, the 95% CIs were estimated using a linear regression model that assumes normality of data where the data can include any real values, which was not the case for EQ-5D values.

Drawing from a normal distribution, some of the simulated EQ-5D values are expected to be outside the plausible range of [-0.594, 1] for EQ-5D. To reduce this problem, the "original SD" is reduced by truncating the simulated EQ-5D data to exclude values outside of the range [-0.594, 1]. Using this reduced set of data, "new" standard deviations for each EDSS score are calculated and termed the "new SD" for the rest of this chapter. This set of standard deviations is in turn used to simulate EQ-5D data from the normal distribution. Although this approach does not fully resolve the issue of having simulated EQ-5D data outside the plausible range for EQ-5D values, it reduces the proportion of EQ-5D data that fall outside the range.

An intuitive approach to circumvent the issue of having implausible EQ-5D data (resulting from a normal distribution) is to use a truncated normal distribution. This distribution restricts all data simulated to be constrained within a pre-specified range, which is [-0.594, 1] in this case. However, it is not a desirable method for simulating the required data because the resulting simulated EQ-5D data are not normally distributed, which would be required to model the data by the use of linear regression, and also have mean values at each EDSS score that are shifted due to the fact that the truncation was not symmetric about the mean.

In this thesis, all three method of simulating EQ-5D data using normal distribution are conducted to test the performance of each method. The set of mean values of EQ-5D at each EDSS score reported in the published article is used for the data simulation. The data simulated using (i) the normal distribution with "original SD" is named "Normal (original SD)"; (ii) the normal distribution with "new SD" is named "Normal"; (iii) the truncated normal distribution is named "Truncated Normal" for the rest of this thesis.

Log-normal distribution

The log-normal distribution has also been proposed for modelling utility values in economic modelling as utility values are generally constrained within a range of real values that can include negative values and are not normally distributed (Briggs et al., 2006).

An advantage of using the log-normal distribution to simulate EQ-5D data is that the data generated follow a normal distribution and hence the normality assumption of the

linear regression analysis applied to the simulated data is satisfied. As the log-normal distribution only takes values from zero to positive infinity, a transformation of the EQ-5D variable onto an interval from 0 to positive infinity was applied. This in turn allowed the data to be simulated using a log-normal distribution on the transformed EQ-5D variable. The transformation that was applied is as follows:

$$T_{EQ5D} = \frac{X_{EQ5D} + 0.594}{1 - X_{EQ5D}}$$

where X_{EQ5D} represents the EQ-5D values and T_{EQ5D} is the transformed value of EQ-5D; and $X_{EQ5D} \in [-0.594, 1]$ and $T_{EQ5D} \in [0, \infty]$. Considering that $T_{EQ5D} \sim LogNormal(\mu, \sigma^2)$, it can be noted that $Log(T_{EQ5D}) \sim Normal(\mu, \sigma^2)$. Denoting $Z_{EQ5D} = Log(T_{EQ5D})$, Z_{EQ5D} was simulated using the normal distribution. Mean and variance of Z_{EQ5D}, were estimated using the Taylor expansions for the moments of functions of random variables. The mean and variance of Z_{EQ5D}were approximated using the first moment and second moment equations as follow:

First moment (for Mean):

$$E[f(X)] \approx f(\mu_X) + \frac{f''(\mu_X)}{2}\sigma_X^2$$

Second moment (for Variance):

$$Var[f(X)] \approx [f'(E[X])]^2 var[X] = [f'(\mu_X)]^2 \sigma_X^2$$

Taking

$$f(X_{EQ5D}) = Z_{EQ5D} = Log\left(\frac{X_{EQ5D} + 0.594}{1 - X_{EQ5D}}\right)$$

$$f'(X_{EQ5D}) = \frac{1}{X_{EQ5D} + 0.594} + \frac{1}{1 - X_{EQ5D}}$$

$$f''^{(X_{EQ5D})} = \frac{1}{(1 - X_{EQ5D})^2} - \frac{1}{(X_{EQ5D} + 0.594)^2}$$

the mean of Z_{EQ5D} is:

$$Mean(Z_{EQ5D}) = \mu_{Z_{EQ5D}} = log\left(\frac{\mu_{X_{EQ5D}} + 0.594}{1 - \mu_{X_{EQ5D}}}\right) + \frac{\sigma^2_{X_{EQ5D}}}{2}\left(\frac{1}{\left(1 - \mu_{X_{EQ5D}}\right)^2} - \frac{1}{\left(\mu_{X_{EQ5D}} + 0.594\right)^2}\right) \quad (5.1)$$

and the variance of Z_{EQ5D} is:

$$Variance(Z_{EQ5D}) = \sigma^2_{Z_{EQ5D}} = \left(\frac{1}{\mu_{X_{EQ5D}} + 0.594} + \frac{1}{1 - \mu_{X_{EQ5D}}}\right)^2 \sigma^2_{X_{EQ5D}} \quad (5.2)$$

where $\mu_{X_{EQ5D}}$ is the mean EQ-5D and $\sigma^2_{X_{EQ5D}}$ is the variance for EQ-5D.

Taking the reported mean EQ-5D at each EDSS score, $\mu_{X_{EQ5D,i}}$, from the article by Orme and colleagues and the corresponding standard deviation at each EDSS scores, $\sigma_{X_{EQ5D},i}$, mean and variance of $Z_{EQ5D,i}$ were computed. Using the mean and variance of $Z_{EQ5D,i}$ at each EDSS score, data of $Z_{EQ5D,i}$ at each EDSS score are simulated from a normal distribution. Simulated data of Z_{EQ5D} ($\sum_{i=0}^{9} Z_{EQ5D,i}$; $i = 0, 1, 2, \ldots, 6, 6.5, 7, \ldots, 9$) for the cohort of 2048 patients were used in the Bayesian linear regression model. To summarise the simulated EQ-5D data and regression analysis results on the EQ-5D scale, back-transformation of Z_{EQ5D} data/values to X_{EQ5D} data/values was performed using the following equation:

$$X_{EQ5D} = \frac{\exp(Z_{EQ5D}) - 0.594}{1 + \exp(Z_{EQ5D})} \quad (5.3)$$

Two sets of Z_{EQ5D} data using lognormal distribution are simulated using the two sets of standard deviations explained in the previous section on Normal distribution. Data simulated using (i) the log-normal distribution with "original SD" is named "LogNormal (original SD)"; and (ii) the log-normal distribution with the "new SD" is named "LogNormal" for the rest of this thesis.

Beta distribution

Another possible approach for simulating EQ-5D data is to use a distribution for quantities that are constrained by a specific range of values, such as the beta distribution which takes values within a range of 0 to 1. To allow for the full range of EQ-5D values, a transformation of EQ-5D is applied as follows:

$$B_{EQ5D} = \frac{X_{EQ5D} + 0.594}{1.594}$$

where X_{EQ5D} represents the EQ-5D and B_{EQ5D} is the transformed EQ-5D that follows a Beta distribution $B_{EQ5D} \sim Beta(\alpha, \beta)$ with shape parameters α and β. Denoting the mean and variance for X_{EQ5D} as $\mu_{X_{EQ5D}}$ and $\sigma^2_{X_{EQ5D}}$ respectively, the mean of B_{EQ5D} is:

$$Mean(B_{EQ5D}) = \mu_{B_{EQ5D}} = \frac{\mu_{X_{EQ5D}} + 0.594}{1.594}$$

and the variance of B_{EQ5D} is:

$$Variance(B_{EQ5D}) = \sigma^2_{B_{EQ5D}} = \frac{\sigma^2_{X_{EQ5D}}}{(1.594)^2}$$

Using the mean values and corresponding variances of X_{EQ5D} at each EDSS score, a set of corresponding mean values and variances of the transformed EQ-5D $\left(\left(\mu_{B_{EQ5D,0}}, \sigma^2_{B_{EQ5D,0}}\right), \left(\mu_{B_{EQ5D,1}}, \sigma^2_{B_{EQ5D,1}}\right), \ldots \left(\mu_{B_{EQ5D,9}}, \sigma^2_{B_{EQ5D,9}}\right)\right)$ are calculated. This set of means and variances is in turn used to calculate the set of corresponding shape parameters for generating B_{EQ5D} from the beta distribution using the following equations:

$$\alpha_i = \frac{\mu^2_{B_{EQ5D,i}}\left(1 - \mu_{B_{EQ5D,i}}\right)}{\sigma^2_{B_{EQ5D,i}}} - \mu_{B_{EQ5D,i}}$$

$$\beta_i = \frac{\alpha_i\left(1 - \mu_{B_{EQ5D,i}}\right)}{\mu_{B_{EQ5D,i}}}$$

where $i = (0, 1, 2, \ldots, 6, 6.5, \ 7, \ldots, 9)$ denotes an EDSS score.

A cohort of 2048 patients was generated using the beta distribution with the set of shape parameters calculated. EQ-5D data for the cohort of 2048 patients were calculated by back-transforming the simulated data, B_{EQ5D} using the following equation:

$$X_{EQ5D} = 1.594B_{EQ5D} - 0.594$$

The two sets of standard deviations ("original SD" and "new SD") discussed in the previous sections are used to simulate the EQ-5D data. However, as explained in the previous sections, the set of "original SD" is large, resulting in the implausible negative values for the shape parameters rendering EQ-5D data simulation using a beta distribution impossible. This issue does not occur when using the set of "new SD". Another method explored for correcting the overestimation of the "original SD" is by reducing the standard deviations such that the 95% reference intervals for EQ-5D at each EDSS score are constrained to be within the range of -0.594 to 1. Hence, the 95% reference interval for EQ-5D at each EDSS score is truncated at the minimum and maximum interval values of -0.594 and 1 respectively and the standard deviation recomputed using this truncated 95% reference interval.

These two sets of standard deviations were used to compute the corresponding sets of shape parameters for beta distribution from which the transformed values of EQ-5D, B_{EQ5D}, were simulated. These two sets of B_{EQ5D} were back-transformed into two sets of X_{EQ5D}. The X_{EQ5D} data simulated using (i) the beta distribution with the "new SD" is named "Beta"; (ii) the beta distribution with standard deviations recomputed using truncated 95% reference intervals is named "Beta Truncated" for the rest of this thesis.

5.2.1.3 Assessment of simulated data

Simulated data for the categorical variables: gender, recent relapse status, EDSS, education and type of MS, are tabulated using frequency and proportions. These summary statistics are compared to the summary statistics reported in Table 3 of the article by Orme and colleagues (Orme et al., 2007) to assess if the simulated data are comparable to the study cohort data.

EQ-5D data for the cohort of patients reported in the article were presented only as regression coefficients (at each EDSS score categories) of a multivariable linear regression analysis, and no summary statistics for EQ-5D were reported in the article. Hence, to assess how comparable the simulated EQ-5D data are with the Orme's data, multivariable linear regression analyses using EDSS as a categorical variable (similar to the regression model in the article) were performed on all sets of simulated EQ-5D data. The degree of bias each set of simulated EQ-5D data has from the original data is computed using the following equation:

$$Bias = \sum_{i=0}^{9} \left(\left(\mu_{Sim,i} - \mu_{Orme,i} \right)^2 + \left(\sigma_{Sim,i} - \sigma_{Orme,i} \right)^2 \right)$$

where $i = (0, 1, 2, \dots, 6, 6.5, 7, \dots, 9)$ denotes the EDSS scores, $\mu_{Orme,i}$ and $\sigma_{Orme,i}$ represent the mean and standard deviation of EQ-5D at EDSS score i estimated from the regression coefficients reported by Orme and colleagues respectively and $\mu_{Sim,i}$ and $\sigma_{Sim,i}$ represent the mean and standard deviation of EQ-5D at EDSS score i estimated from the regression coefficients performed using the simulated EQ-5D data respectively. This bias factor was computed for each of the seven sets of simulated EQ-5D data. The set of simulated EQ-5D (except the Normal (original SD) set of data) that gave the smallest bias was used as the base-case EQ-5D data in this project. The reason for excluding the Normal (original SD) set of EQ-5D data was that although it was expected to give the smallest bias, it was a highly implausible set of EQ-5D data as a large proportion (>10%) of the data were expected to be outside the range of [-0.594,1].

5.2.2 Bayesian linear regression models for estimating EQ-5D in MS

The objective of this project was to obtain estimates of EQ-5D utility values for EDSS scores between 0 and 9 using simulated IPD and incorporating information about the correlation between EQ-5D and EDSS from published articles that reported both EQ-5D and EDSS. This section describes the linear regression models that are used for estimating the EQ-5D utility values.

In order to estimate EQ-5D utility values for values of EDSS, a multivariable linear regression model, similar to the one defined in the article by Orme and colleagues (Orme et al., 2007) was utilised. Besides the multivariable linear regression model,

analyses using univariable linear regression models were also explored. The univariable linear regression equation for the association between EQ-5D and EDSS takes the form:

$$X_{EQ5D,i} = \alpha + \beta_1 X_{EDSS,i} + \varepsilon_i \tag{5.4}$$

with EQ-5D and EDSS treated as continuous data in the regression model and i representing the i-th subject in the simulated study cohort. The multivariable linear regression equation adjusting for other covariates takes the form:

$$\begin{aligned} X_{EQ5D,i} = \alpha + \beta_1 X_{EDSS,i} + \beta_G X_{Gender,i} + \beta_R X_{Relapse,i} \\ + \beta_M X_{MS,i} + \beta_E X_{Edu,i} + \varepsilon_i \end{aligned} \tag{5.5}$$

where $X_{Gender,i}$, $X_{Relapse,i}$, $X_{MS,i}$ and $X_{Edu,i}$ are the covariate terms for variables gender, recent relapse status, type of MS and education respectively. Gender and recent relapse status are defined as binary data and type of MS and education are defined as categorical data in the regression model.

In the article by Orme and colleagues (Orme et al., 2007), the EDSS variable, X_{EDSS}, was used as categorical data in the linear regression model (similar to the one specified in Equation (5.5)). In this thesis, the linear regression analyses were performed under the Bayesian framework where EDSS was defined as a continuous variable in the linear regression model. This was to allow for the inclusion of external summary data on EDSS which were generally reported using mean and standard deviation in published articles instead of proportion or frequency count at each EDSS score. To make such inclusion possible, the EDSS data in both analyses, the regression model and the external model (used to construct the prior distributions as described in Section 5.2.3.2), had to be modelled on the same scale; hence the specification of EDSS could only take the form of continuous variable.

In order to take into account the non-linearity of the relationship between EQ-5D and EDSS, polynomials of various degrees for the regression term, X_{EDSS}, were explored. Step-wise linear regression models fitted to the simulated data discussed in the previous section (Section 5.2.1) were used to decide the degrees of polynomial that best fit the simulated data to explain the relationship between EQ-5D and EDSS. The term, EDSS (X_{EDSS}), was centred at its mean value (rounded to the nearest integer) in all the

models. Model fitting were performed using STATA software (StataCorp, 2011) and the likelihood ratio test was used to assess the model fit of nested polynomial models.

The polynomial model which was found to be the best fit to the simulated data was in turn used in the Bayesian linear regression model for estimating EQ-5D at each EDSS score and also in the external analysis model (described in Section 5.2.3) for constructing informative prior distributions for the regression coefficients of the Bayesian linear regression model. Non-informative prior distributions for the regression coefficients of the Bayesian linear regression models are also applied in the analyses. The non-informative prior distributions for the regression coefficients of the polynomial models explored are presented in Table 5.1. Methods for the construction of informative prior distributions for the regression coefficients of the polynomial models are described in Section 5.2.3.

Table 5.1: Regression models used in Frequentist model selection and subsequently in Bayesian regression analyses.

Degree of Polynomial	Regression Model	Prior Distribution for Regression Coefficients
1st	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS}$	$\alpha \sim Normal(0.0, 10^6)$
		$\beta_1 \sim Normal(0.0, 10^6)$
	With adjustment for covariates	$\beta_G \sim Normal(0.0, 10^6)$
	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS}$	$\beta_R \sim Normal(0.0, 10^6)$
	$+ \beta_G X_{Gender} + \beta_R X_{Relapse} + \beta_M X_{MS} + \beta_E X_{Edu}$	$\beta_M \sim Normal(0.0, 10^6)$
		$\beta_E \sim Normal(0.0, 10^6)$
2nd	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS}$	$\alpha \sim Normal(0.0, 10^6)$
		$\beta_1 \sim Normal(0.0, 10^6)$
	With adjustment for covariates	$\beta_2 \sim Normal(0.0, 10^6)$
	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS}$	$\beta_G \sim Normal(0.0, 10^6)$
	$+ \beta_G X_{Gender} + \beta_R X_{Relapse} + \beta_M X_{MS} + \beta_E X_{Edu}$	$\beta_R \sim Normal(0.0, 10^6)$
		$\beta_M \sim Normal(0.0, 10^6)$
		$\beta_E \sim Normal(0.0, 10^6)$
3rd	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS} + \beta_3 X^3_{EDSS}$	$\alpha \sim Normal(0.0, 10^6)$
		$\beta_1 \sim Normal(0.0, 10^6)$
	With adjustment for covariates	$\beta_2 \sim Normal(0.0, 10^6)$
	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS} + \beta_3 X^3_{EDSS}$	$\beta_3 \sim Normal(0.0, 10^6)$
	$+ \beta_G X_{Gender} + \beta_R X_{Relapse} + \beta_M X_{MS} + \beta_E X_{Edu}$	$\beta_G \sim Normal(0.0, 10^6)$
		$\beta_R \sim Normal(0.0, 10^6)$
		$\beta_M \sim Normal(0.0, 10^6)$
		$\beta_E \sim Normal(0.0, 10^6)$
4th	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS} + \beta_3 X^3_{EDSS} + \beta_4 X^4_{EDSS}$	$\alpha \sim Normal(0.0, 10^6)$
		$\beta_1 \sim Normal(0.0, 10^6)$
	With adjustment for covariates	$\beta_2 \sim Normal(0.0, 10^6)$
	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS} + \beta_3 X^3_{EDSS} + \beta_4 X^4_{EDSS}$	$\beta_3 \sim Normal(0.0, 10^6)$
	$+ \beta_G X_{Gender} + \beta_R X_{Relapse} + \beta_M X_{MS} + \beta_E X_{Edu}$	$\beta_4 \sim Normal(0.0, 10^6)$
		$\beta_G \sim Normal(0.0, 10^6)$
		$\beta_R \sim Normal(0.0, 10^6)$
		$\beta_M \sim Normal(0.0, 10^6)$
		$\beta_E \sim Normal(0.0, 10^6)$
5th	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS} + \beta_3 X^3_{EDSS} + \beta_4 X^4_{EDSS} + \beta_5 X^5_{EDSS}$	$\alpha \sim Normal(0.0, 10^6)$
		$\beta_1 \sim Normal(0.0, 10^6)$
	With adjustment for covariates	$\beta_2 \sim Normal(0.0, 10^6)$
	$X_{EQ5D} = \alpha + \beta_1 X_{EDSS} + \beta_2 X^2_{EDSS} + \beta_3 X^3_{EDSS} + \beta_4 X^4_{EDSS} + \beta_5 X^5_{EDSS}$	$\beta_3 \sim Normal(0.0, 10^6)$
	$+ \beta_G X_{Gender} + \beta_R X_{Relapse} + \beta_M X_{MS} + \beta_E X_{Edu}$	$\beta_4 \sim Normal(0.0, 10^6)$
		$\beta_5 \sim Normal(0.0, 10^6)$
		$\beta_G \sim Normal(0.0, 10^6)$
		$\beta_R \sim Normal(0.0, 10^6)$
		$\beta_M \sim Normal(0.0, 10^6)$
		$\beta_E \sim Normal(0.0, 10^6)$

5.2.3 Construction of prior distributions

To construct informative prior distributions for the linear regression coefficients that define the association between EQ-5D and EDSS, external summary data (from publicly available sources) on both outcomes is used. A Bayesian BRMA model that jointly synthesises both outcomes while accounting for the correlation between the two outcomes was proposed to model the external data.

In this section, the methods and procedures for constructing the prior distributions are discussed. Firstly, the search strategy and selection criteria of studies to form the likelihood data to be used in the Bayesian BRMA are described. Next, the specifications of the BRMA model for constructing prior distributions are described.

5.2.3.1 Search strategy and selection criteria of studies

To identify studies for the construction of the prior distributions, a search in PubMed for studies in multiple sclerosis published between the year 1999 and June 2014 inclusive was performed. The search was conducted using the search terms: (i) "multiple sclerosis" and "EDSS" and "EQ-5D", (ii) "multiple sclerosis" and "EuroQoL" and (iii) "multiple sclerosis" and "EQ-5D". Published studies from the search were retrieved and reviewed. Studies that reported both EDSS and EQ-5D summary statistics at the same time point in the study (such as at baseline or end of follow-up period) were identified to form the dataset for the construction of the prior distributions using the proposed Bayesian BRMA described in the next section. Summary statistics in terms of frequency, mean and standard deviation for EDSS and EQ-5D were extracted to form the likelihood data for the Bayesian BRMA. For the rest of this Chapter, this dataset of studies is referred to as the "external dataset".

5.2.3.2 Bayesian BRMA model

The purpose for using a Bayesian BRMA model in this project was to construct prior distributions for the regression coefficients described in Section 5.2.2. The Bayesian BRMA model described in Chapter 2 (Section 2.4.3.5) was adapted to jointly model the aggregate data on EDSS and EQ-5D outcomes in such a way to allow the construction

of prior distributions for the regression coefficient for the polynomial terms in the linear regression model. In order to do this, both EDSS and EQ-5D were treated as continuous outcomes and the BRMA model in the product normal formulation (first degree polynomial case was described in Chapter 2 Equation (2.5)) was extended to contain the polynomial terms to match those in the linear regression model (of the simulated IPD as discussed in Section 5.2.2).

This model takes the following hierarchical form with the within-study model of the BRMA:

$$\begin{pmatrix} Y_{EDSS,i} \\ Y_{EQ5D,i} \end{pmatrix} \sim Normal\left(\begin{pmatrix} \mu_{EDSS,i} \\ \mu_{EQ5D,i} \end{pmatrix}, \Sigma_i\right) \qquad (5.6)$$

with

$$\Sigma_i = \begin{pmatrix} \sigma^2_{EDSS,i} & \sigma_{EDSS,i}\sigma_{EQ5D,i}\rho_{w,i} \\ \sigma_{EQ5D,i}\sigma_{EDSS,i}\rho_{w,i} & \sigma^2_{EQ5D,i} \end{pmatrix}$$

and the between-study model parameterised using a product normal formulation with n-th degree polynomial:

$$\begin{gathered} \mu_{EDSS,i} \sim Normal(\eta_{EDSS}, \psi^2_{EDSS}) \\ \mu_{EQ5D,i} | \mu_{EDSS,i} \sim Normal(\eta_{EQ5D,i}, \psi^2_{EQ5D}) \\ \eta_{EQ5D,i} = \lambda_0 + \lambda_1 \mu_{EDSS,i} + \lambda_2 \mu^2_{EDSS,i} + \cdots + \lambda_n \mu^n_{EDSS,i} \end{gathered} \qquad (5.7)$$

where $Y_{EDSS,i}$ and $Y_{EQ5D,i}$ are correlated estimates of true mean values $\mu_{EDSS,i}$ and $\mu_{EQ5D,i}$ of EDSS and EQ-5D in each study i ($i = 1,2,3, \ldots$) identified in the search and selection criteria described in Section 5.2.3.1, with corresponding within-study covariance matrix Σ_i. Here, the outcome variables $\mu_{EDSS,i}$ and $\mu_{EQ5D,i}$ were associated in the form of a linear regression model with n-th degree polynomials (Equation (5.7)) in the product normal formulation of the BRMA model. As the association of EQ-5D with EDSS is not linear in practice, their relationship using different degrees of polynomial as shown in Table 5.2 were investigated. Similar to the Bayesian linear regression model, the EDSS term in Equation (5.7), $\mu_{EDSS,i}$, was also centred at the

same mean EDSS value (rounded to the nearest integer) defined in Section 5.2.2 for all models in Table 5.2.

The purpose for using the BRMA model in this project was to obtain posterior distributions for the regression coefficients (λ_0, λ_1, λ_2, etc) of the linear regression model (of the between-study component of the BRMA model) to be used as prior distributions for the corresponding regression coefficients (α, β_1, β_2, etc) in the linear regression models defined in Table 5.1 in Section 5.2.2.

Non-informative prior distributions were placed on the regression coefficients of the linear regression model (λ_0, λ_1, λ_2, … , λ_n, in Equation (5.7)) in the between-study component of the Bayesian BRMA model. In the case of the first degree polynomial relationship between $\mu_{EDSS,i}$ and $\mu_{EQ5D,i}$, two ways of specifying non-informative priors were used. In Case 1, non-informative prior was placed on the regression coefficient λ_0, between-study correlation ρ_b and the between-study standard deviations (ψ^2_{EDSS} and ψ^2_{EQ5D}) as described in Chapter 2 Equation (2.8); and in Case 2, it is placed on the regression coefficients (λ_0 and λ_1).

For the within-study correlation in the within-study component of the Bayesian BRMA model, informative prior distribution was defined using information from a published study (Fisk et al., 2005). Using the correlation coefficient of -0.66 and study sample size of 187 reported in the article (Fisk et al., 2005), the informative prior distribution for the within-study correlation between $Y_{EDSS,i}$ and $Y_{EQ5D,i}$ was constructed using the method discussed in Chapter 2 Section 2.4.3.6. Table 5.2 presents the prior distributions specified for the parameters in the between-study and within-study components of the BRMA model with polynomial of degree one to five.

The summary statistics (mean and standard errors) from the posterior distributions of the regression coefficients in the product normal models of BRMA presented in Equation (5.7) were subsequently used to construct prior distributions for its corresponding regression coefficients in the linear regression models discussed in Section 5.2.2. The posterior distributions and corresponding prior distributions are presented in Section 5.3.3.2.

Table 5.2: Polynomials used in BRMA model and prior distributions specified for the within- and between-study correlation parameters

	Between-study model of BRMA		Within-study model of BRMA
Polynomial Type	Linear regression model	Non-informative prior distribution	Informative prior distribution
1st-degree polynomial (Case 1)	$\eta_{EQ5D,i} = \lambda_0+\lambda_1\mu_{EDSS,i}$	$\lambda_0 \sim$ Normal (0, 1000) $\lambda_1 = (\sigma_{EQ5D}\rho_b)/(\sigma_{EDSS}\sqrt{(1-\rho_b^2)})$ $\rho_b \sim$ Uniform (-1, 1)	$z_{wi} \sim$ Normal (-0.7928, 0.0737^2)
1st-degree polynomial (Case 2)	$\eta_{EQ5D,i} = \lambda_0+\lambda_1\mu_{EDSS,i}$	$\lambda_0 \sim$ Normal (0, 1000) $\lambda_1 \sim$ Normal (0, 1000)	$z_{wi} \sim$ Normal (-0.7928, 0.0737^2)
2nd-degree polynomial	$\eta_{EQ5D,i} = \lambda_0+\lambda_1\mu_{EDSS,i}+\lambda_2\mu^2_{EDSS,i}$	$\lambda_0 \sim$ Normal (0, 1000) $\lambda_1 \sim$ Normal (0, 1000) $\lambda_2 \sim$ Normal (0, 1000)	$z_{wi} \sim$ Normal (-0.7928, 0.0737^2)
3rd-degree polynomial	$\eta_{EQ5D,i} = \lambda_0+\lambda_1\mu_{EDSS,i}+\lambda_2\mu^2_{EDSS,i}+\lambda_3\mu^3_{EDSS,i}$	$\lambda_0 \sim$ Normal (0, 1000) $\lambda_1 \sim$ Normal (0, 1000) $\lambda_2 \sim$ Normal (0, 1000) $\lambda_3 \sim$ Normal (0, 1000)	$z_{wi} \sim$ Normal (-0.7928, 0.0737^2)
4th-degree polynomial	$\eta_{EQ5D,i} = \lambda_0+\lambda_1\mu_{EDSS,i}+\lambda_2\mu^2_{EDSS,i}+\lambda_3\mu^3_{EDSS,i}+\lambda_4\mu^4_{EDSS,i}$	$\lambda_0 \sim$ Normal (0, 1000) $\lambda_1 \sim$ Normal (0, 1000) $\lambda_2 \sim$ Normal (0, 1000) $\lambda_3 \sim$ Normal (0, 1000) $\lambda_4 \sim$ Normal (0, 1000)	$z_{wi} \sim$ Normal (-0.7928, 0.0737^2)
5th-degree polynomial	$\eta_{EQ5D,i} = \lambda_0+\lambda_1\mu_{EDSS,i}+\lambda_2\mu^2_{EDSS,i}+\lambda_3\mu^3_{EDSS,i}+\lambda_4\mu^4_{EDSS,i}+\lambda_5\mu^5_{EDSS,i}$	$\lambda_0 \sim$ Normal (0, 100) $\lambda_1 \sim$ Normal (0, 100) $\lambda_2 \sim$ Normal (0, 100) $\lambda_3 \sim$ Normal (0, 100) $\lambda_4 \sim$ Normal (0, 100) $\lambda_5 \sim$ Normal (0, 100)	$z_{wi} \sim$ Normal (-0.7928, 0.0737^2)

5.2.4 Data transformation considerations in sequential Bayesian models

The Bayesian linear regression models for the estimation of EQ-5D in MS are described in Section 5.2.2 and the EQ-5D IPD simulated using the methods presented in Section 5.2.1 are used for the analyses. The summary data of studies identified using the search strategy and selection criteria presented in Section 5.2.3.1 were used as the likelihood data for the Bayesian BRMA model discussed in Section 5.2.3.2. Here, there are two Bayesian analysis models using two different sets of likelihood data. However, the scale of the two sets of data used in the two Bayesian analysis models had to be the same due to the sequential use of the two Bayesian models. The posterior distributions of the regression coefficients derived from the Bayesian BRMA analysis were subsequently used as prior distributions in the Bayesian linear regression analysis.

When the EQ-5D data were simulated using the normal distribution, the data inherently satisfied the normality assumption of the Bayesian linear regression model. The simulated data were directly utilised in the Bayesian linear regression model and EQ-5D summary data of studies identified for the Bayesian BRMA model were also directly applied in the BRMA analysis.

In Section 5.2.1.2, besides the normal distribution, the log-normal distribution was also used for the simulation of EQ-5D data. When the EQ-5D data were simulated using the log-normal distribution, the back-transformed EQ-5D data (where transformation were applied for the purpose of data simulation) were not normally distributed. However, the simulated data before performing the back-transformation satisfy the normality assumption for linear regression analysis and were used for the analysis. As the analysis was performed on the log-normal scale of the EQ-5D data, similar transformation of the likelihood data (EQ-5D summary data of studies identified) for the Bayesian BRMA was required and are described in the next section.

5.2.4.1 Log-normal distribution

The Bayesian linear regression analyses were performed using the transformed log-normal EQ-5D IPD, Z_{EQ5D} described in Section 5.2.1.2, which satisfy the normality assumption for linear regression analysis.

Methods used for the construction of prior distributions for the regression coefficients of the linear regression model were the same as those described in Section 5.2.3.2. However, the likelihood data for the BRMA model had to be transformed onto the log-normal scale for the BRMA analysis (to match the scale of the EQ-5D IPD and hence to construct the prior distributions for the regression coefficients on the same scale). The mean and variance of EQ-5D ($\mu_{X_{EQ5D}}$ and $\sigma^2_{X_{EQ5D}}$respectively) were transformed onto the log-normal scale (as $\mu_{Z_{EQ5D}}$ and $\sigma^2_{Z_{EQ5D}}$respectively) for all the studies in the external dataset. The mean and the variance of the transformed EQ-5D for the studies in the external data are computed using Equations (5.1) and (5.2) respectively derived in Section 5.2.1.2.

The posterior distribution for the regression coefficients in the BRMA model were then used as prior distribution for the regression coefficients in the Bayesian linear regression model using the simulated IPD on the log-normal scale. Final estimates of EQ-5D at each EDSS score from the Bayesian linear regression model were computed by back-transforming the log-normal EQ-5D estimate (Z_{EQ5D}) to EQ-5D estimates (X_{EQ5D}) using equation (5.3). The 95% credible interval (CrI) for the mean EQ-5D estimate at each EDSS score was calculated using the same equation.

5.2.5 Sensitivity analyses

5.2.5.1 Non-constant variance models for EQ-5D

In the base case model where EDSS was treated as a continuous variable in the linear regression model (instead of as a categorical variable in the case of the Orme's model), the variance of EQ-5D was assumed constant across all EDSS scores. However, the distribution of the data in Orme's paper suggested that the variance of EQ-5D varied across EDSS states. As a sensitivity analysis, the variances of EQ-5D across all EDSS were allowed to vary with different EDSS scores using either polynomials or exponential functions that best fit the structure of the variance data. Further variability in the variance was introduced by adding an error term, which follows a chi-square distribution, to the polynomial and exponential models. Details of the choice of polynomial and exponential model for the base-case (simulated) EQ-5D data as mentioned in Section 5.2.1.3 are described in Section 5.3.5.1.

5.2.5.2 Estimation of EQ-5D using a different set of simulated data

Seven sets of EQ-5D data were simulated using three different statistical distributions (normal, log-normal and beta distributions) as described in Section 5.2.1.2 but only the set of simulated EQ-5D data with the smallest bias as discussed in Section 5.2.1.3 was chosen as the base-case EQ-5D data for analysis in this project.

To explore how sensitive the estimates for EQ-5D were to the choice of statistical distribution used for the data simulation, the set of EQ-5D simulated using the log-normal distribution was used for the sensitivity analysis. This set of simulated data was selected for the sensitivity analysis as it does not require further transformation to satisfy the normality assumption of the linear regression model.

5.3 Results

In this section, results of the analyses are presented. Firstly, in the absence of IPD for the patient population in Orme's study, simulation techniques described in Section 5.2.1 were utilised to generate the IPD. Summary statistics of the patient demographics and disease status estimated using the simulated IPD are presented in Section 5.3.1. After assessing the simulated IPD and deciding on the set of EQ-5D data to be used as the base-case EQ-5D data for analyses in Section 5.3.1, the choice of regression model for the estimation of EQ-5D in MS is discussed in Section 5.3.2.

Next, informative prior distributions for the regression coefficients of the regression model were constructed and are presented in Section 5.3.3. Using the IPD data, the prior distributions (both non-informative and informative) and the regression model, Bayesian linear regression analyses were performed to estimate EQ-5D at each EDSS score. Results of the analyses are presented in Section 5.3.4.

As the regression analyses were performed using EDSS as a continuous variable instead of as a categorical data, sensitivity analysis allowing the variance model of the regression model to be non-constant were performed and reported in Section 5.3.5.1. Further sensitivity analysis using an alternative set of simulated EQ-5D (using log-normal distribution) in place of the base-case EQ-5D data was conducted and reported in Section 5.3.5.2.

5.3.1 Simulated individual patient data

Individual patient data for the study cohort published by Orme and colleagues (Orme et al., 2007) were simulated using the methods described in Section 5.2.1. Summary statistics of patient demographics and disease characteristics estimated using the simulated IPD are presented in Table 5.3, alongside those reported for Orme's study. The results showed that the simulated data for variables: gender, education, type of MS, recent relapse and EDSS were comparable to that of the study cohort in Orme's study.

Table 5.3: Patient Demographics and disease information in Orme article (Orme et al., 2007) and simulated data

	Orme Data^	Simulated Data
Demographics	Proportions in % (N)	Proportions in % (N)
Gender		
Male	24.7	24.7 (505)
Female	74.5	75.3 (1543)
Missing	0.8	NS
Education		
Secondary school	32.2	32.8 (672)
College or sixth form	26.5	26.1 (534)
University or polytechnic degree	29.7	31.0 (634)
Postgraduate degree	10.1	10.2 (208)
No answer	1.6	NS
Type of MS		
RRMS	35.5	36.0 (737)
SPMS	37.2	38.0 (778)
PPMS	27.3	26.0 (533)
Relapse during last 3 months		
Yes	28.9	28.1 (575)
No	71.1	71.9 (1473)
EDSS scores*		
0	1.4 (28)	1.4 (29)
1	7.4 (151)	7.6 (155)
2	8.8 (180)	8.7 (178)
3	3.8 (77)	3.8 (78)
4	9.4 (193)	9.2 (188)
5	15.8 (323)	15.4 (316)
6	19.3 (396)	19.5 (399)
6.5	15.1 (309)	15.4 (315)
7	10.3 (210)	10.3 (210)
8	8.1 (165)	8.3 (169)
9	0.8 (16)	0.5 (11)

^ Data extracted from Orme et. al. Table 3; * Data extracted from Orme et. al. Figure 1
NS: Not Simulated

As discussed in Section 5.2.1.2, there were no EQ-5D summary statistics reported in the published article. As a result, mean and standard deviation for EQ-5D at each EDSS score was approximated using the multivariable linear regression results in the article. Means and standard deviations approximated using the methods discussed in Section 5.2.1.2 are presented in Table 5.4.

Table 5.4: Means and standard deviations used for simulating EQ-5D data

EDSS score	Mean EQ-5D	Original SD	new SD	Truncated standard deviation^
0	0.870	0.238	0.169	0.066
1	0.799	0.589	0.327	0.103
2	0.705	0.640	0.396	0.151
3	0.574	0.454	0.346	0.217
4	0.610	0.663	0.394	0.199
5	0.518	0.844	0.433	0.246
6	0.458	0.944	0.442	0.277
6.5	0.462	0.843	0.420	0.274
7	0.297	0.713	0.438	0.359
8	-0.049	0.646	0.404	0.278
9	-0.195	0.297	0.238	0.204

^ Only used for data simulation using the beta distribution

Using the various sets of standard deviations for EQ-5D, together with the mean EQ-5D, seven sets of EQ-5D data were simulated using either normal, log-normal or beta distribution as described in Section 5.2.1.2. EQ-5D data simulated from normal distribution using original SD and new SD are called “Normal (original SD)” and “Normal” respectively. One set of EQ-5D data, named “Truncated Normal”, was simulated using a truncated normal distribution. Those simulated from log-normal distribution with the original SD and new SD are called “LogNormal (original SD)” and “LogNormal” respectively. Beta distribution was used for simulating two sets of EQ-5D data, named “Beta” and “Beta Truncated” which used the new SD and truncated standard deviations respectively. Bar charts and summary statistics of the seven sets of EQ-5D data are presented in Figure 5.2 and Table 5.5 respectively. Error bars in Figure 5.2 represent the 95% CIs.

Ideally, the summary statistics of the seven sets of EQ-5D data should be compared to the summary statistics for EQ-5D in Orme study, however it is not possible as EQ-5D data in the article were presented only as regression coefficients (at each EDSS score categories) of a multivariable linear regression analysis. It therefore makes more sense

to conduct similar regression analysis on the simulated data and compare their results to the regression results in Orme's article. Hence, multivariable linear regression analyses assessing the association between EQ-5D and EDSS (where EDSS was treated as a categorical variable), adjusting for gender, recent relapse, education and type of MS were performed for all sets of simulated EQ-5D data. Results of the regression analyses are presented in Figure 5.3 and Table 5.6 respectively. Error bars in Figure 5.3 represent the 95% CrIs from Bayesian multivariable linear regression analyses, apart from the Orme data which represent 95% CIs.

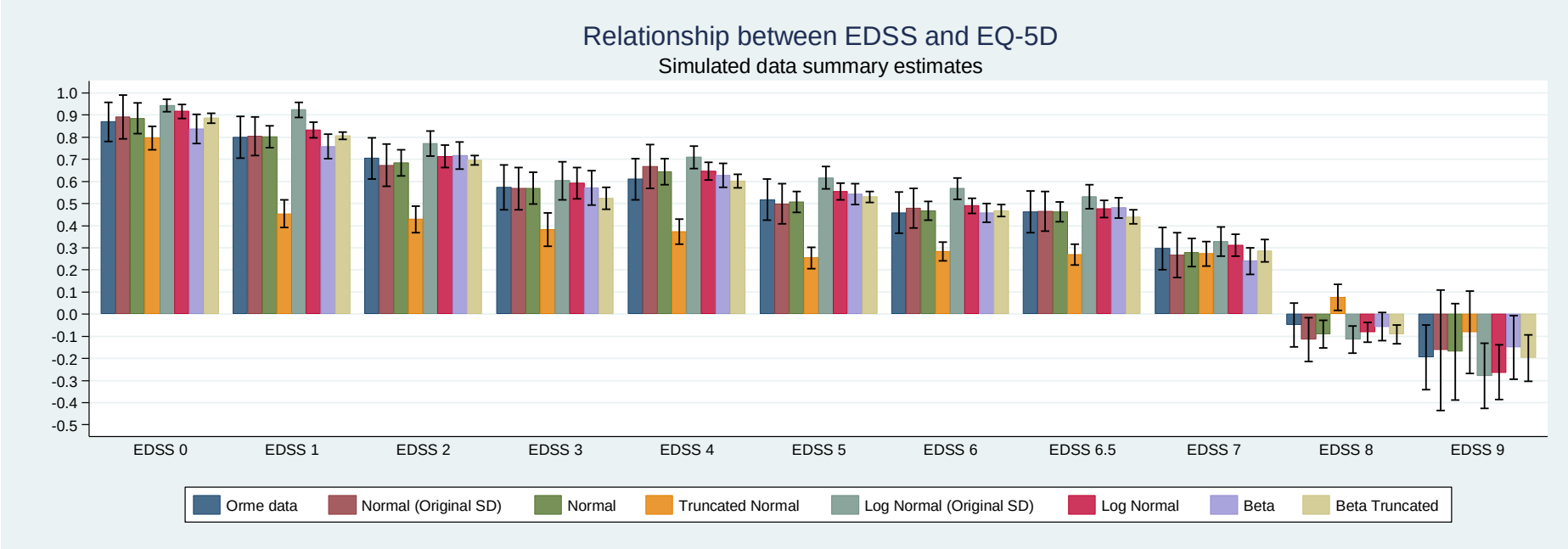

Figure 5.2: Relationship between EDSS and EQ-5D for the seven sets of simulated EQ-5D data

Table 5.5: EQ-5D summary statistics for the seven sets of simulated EQ-5D data

EQ-5D	Orme's Data	Normal (original SD)	Normal	Truncated Normal	LogNormal (original SD)	LogNormal	Beta	Beta Truncated
for	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
EDSS=0	0.870 (0.782, 0.958)	0.891 (0.793, 0.990)	0.885 (0.815, 0.955)	0.797 (0.744, 0.850)	0.944 (0.915, 0.973)	0.917 (0.885, 0.948)	0.837 (0.771, 0.903)	0.886 (0.864, 0.909)
EDSS=1	0.799 (0.705, 0.893)	0.804 (0.717, 0.892)	0.802 (0.754, 0.850)	0.454 (0.391, 0.516)	0.924 (0.890, 0.958)	0.833 (0.797, 0.869)	0.758 (0.703, 0.813)	0.807 (0.791, 0.824)
EDSS=2	0.705 (0.611, 0.798)	0.673 (0.578, 0.768)	0.685 (0.626, 0.744)	0.429 (0.369, 0.488)	0.772 (0.714, 0.829)	0.714 (0.664, 0.763)	0.716 (0.655, 0.778)	0.696 (0.675, 0.717)
EDSS=3	0.574 (0.472, 0.675)	0.568 (0.473, 0.664)	0.570 (0.497, 0.642)	0.382 (0.306, 0.458)	0.603 (0.518, 0.689)	0.592 (0.521, 0.664)	0.570 (0.492, 0.648)	0.524 (0.475, 0.573)
EDSS=4	0.610 (0.516, 0.703)	0.668 (0.568, 0.767)	0.644 (0.585, 0.703)	0.373 (0.318, 0.429)	0.710 (0.659, 0.761)	0.646 (0.606, 0.687)	0.628 (0.573, 0.683)	0.602 (0.572, 0.632)
EDSS=5	0.518 (0.426, 0.610)	0.499 (0.407, 0.590)	0.508 (0.461, 0.555)	0.255 (0.206, 0.303)	0.617 (0.566, 0.668)	0.554 (0.517, 0.592)	0.543 (0.496, 0.589)	0.530 (0.504, 0.555)
EDSS=6	0.458 (0.365, 0.551)	0.479 (0.390, 0.568)	0.468 (0.426, 0.510)	0.284 (0.242, 0.326)	0.568 (0.520, 0.615)	0.490 (0.455, 0.525)	0.458 (0.416, 0.500)	0.469 (0.441, 0.496)
EDSS=6.5	0.462 (0.368, 0.556)	0.464 (0.375, 0.554)	0.463 (0.418, 0.508)	0.270 (0.223, 0.317)	0.531 (0.477, 0.586)	0.476 (0.437, 0.515)	0.481 (0.435, 0.527)	0.439 (0.407, 0.471)
EDSS=7	0.297 (0.200, 0.393)	0.267 (0.165, 0.369)	0.279 (0.216, 0.341)	0.273 (0.217, 0.329)	0.328 (0.262, 0.394)	0.312 (0.263, 0.361)	0.240 (0.180, 0.300)	0.286 (0.236, 0.337)
EDSS=8	-0.049 (-0.147, 0.050)	-0.115 (-0.215, -0.015)	-0.090 (-0.153, -0.028)	0.076 (0.016, 0.135)	-0.115 (-0.175, -0.055)	-0.082 (-0.127, -0.036)	-0.056 (-0.120, 0.007)	-0.091 (-0.133, -0.050)
EDSS=9	-0.195 (-0.340, -0.049)	-0.163 (-0.435, 0.109)	-0.169 (-0.387, 0.049)	-0.082 (-0.269, 0.105)	-0.280 (-0.427, -0.132)	-0.263 (-0.387, -0.138)	-0.150 (-0.293, -0.006)	-0.198 (-0.303, -0.092)

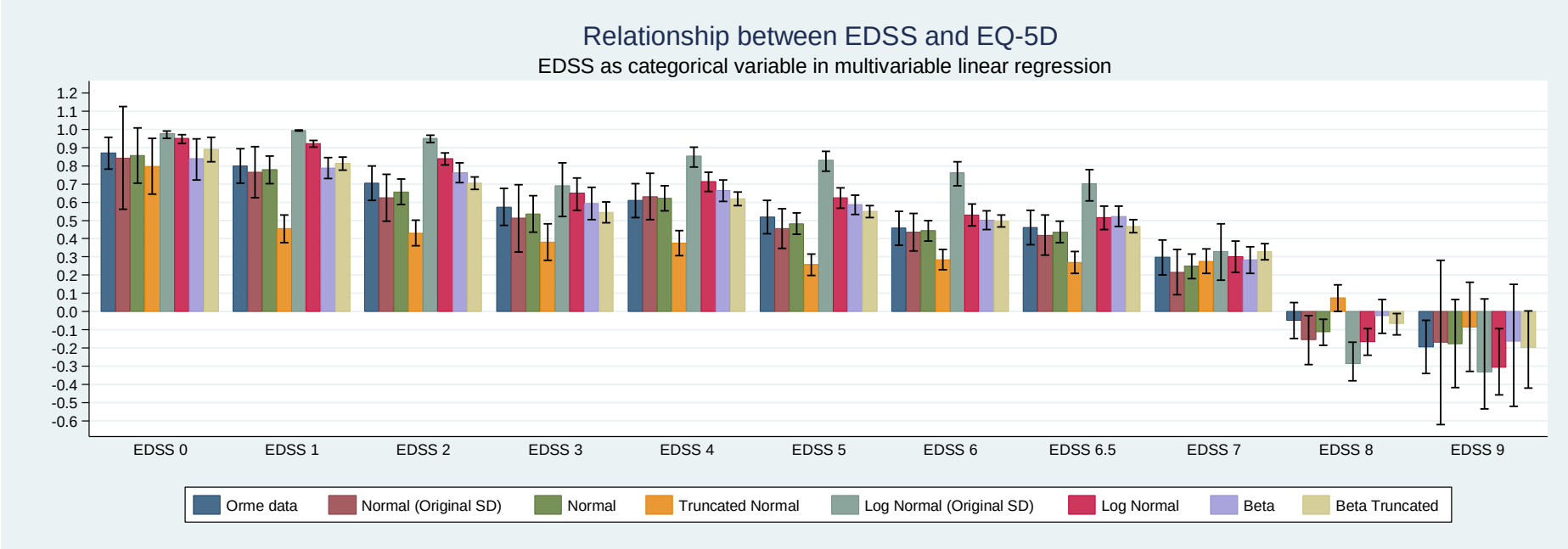

Figure 5.3: Association between EDSS and EQ-5D estimated using multivariable linear regression model for the seven sets of simulated EQ-5D data

Table 5.6: EQ-5D summary statistics reported in Orme paper and estimated using multivariable linear regression model (with EDSS as categorical data) for the seven sets of simulated EQ-5D data

EQ-5D	Orme's Data	Normal (original SD)	Normal	Truncated Normal	LogNormal (original SD)	LogNormal	Beta	Beta Truncated
for	Mean (95% CI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)
EDSS=0	0.870 (0.782, 0.958)	0.841 (0.560, 1.125)	0.856 (0.705, 1.009)	0.798 (0.645, 0.951)	0.977 (0.950, 0.992)	0.950 (0.923, 0.970)	0.840 (0.723, 0.948)	0.891 (0.822, 0.956)
EDSS=1	0.799 (0.705, 0.893)	0.765 (0.624, 0.905)	0.778 (0.703, 0.854)	0.455 (0.379, 0.531)	0.994 (0.992, 0.997)	0.922 (0.902, 0.939)	0.789 (0.729, 0.846)	0.814 (0.778, 0.849)
EDSS=2	0.705 (0.611, 0.798)	0.625 (0.497, 0.754)	0.657 (0.588, 0.727)	0.431 (0.361, 0.501)	0.951 (0.928, 0.968)	0.839 (0.804, 0.870)	0.762 (0.707, 0.816)	0.705 (0.670, 0.741)
EDSS=3	0.574 (0.472, 0.675)	0.512 (0.326, 0.695)	0.537 (0.437, 0.636)	0.382 (0.281, 0.481)	0.690 (0.522, 0.817)	0.651 (0.556, 0.732)	0.594 (0.504, 0.682)	0.544 (0.486, 0.601)
EDSS=4	0.610 (0.516, 0.703)	0.632 (0.504, 0.759)	0.623 (0.554, 0.691)	0.375 (0.306, 0.444)	0.855 (0.794, 0.902)	0.714 (0.659, 0.764)	0.664 (0.605, 0.722)	0.619 (0.581, 0.656)
EDSS=5	0.518 (0.426, 0.610)	0.455 (0.347, 0.565)	0.482 (0.424, 0.541)	0.257 (0.199, 0.317)	0.830 (0.771, 0.878)	0.625 (0.569, 0.678)	0.588 (0.534, 0.639)	0.550 (0.516, 0.583)
EDSS=6	0.458 (0.365, 0.551)	0.437 (0.334, 0.539)	0.443 (0.388, 0.498)	0.285 (0.229, 0.340)	0.762 (0.689, 0.824)	0.531 (0.470, 0.589)	0.502 (0.449, 0.555)	0.496 (0.463, 0.529)
EDSS=6.5	0.462 (0.368, 0.556)	0.419 (0.309, 0.529)	0.437 (0.378, 0.496)	0.271 (0.211, 0.330)	0.701 (0.608, 0.780)	0.517 (0.450, 0.579)	0.523 (0.467, 0.578)	0.468 (0.432, 0.504)
EDSS=7	0.297 (0.200, 0.393)	0.216 (0.092, 0.342)	0.249 (0.182, 0.316)	0.276 (0.209, 0.344)	0.331 (0.173, 0.483)	0.302 (0.215, 0.387)	0.284 (0.209, 0.356)	0.328 (0.283, 0.372)
EDSS=8	-0.049 (-0.147, 0.050)	-0.156 (-0.290, -0.022)	-0.114 (-0.186, -0.042)	0.075 (0.002, 0.147)	-0.285 (-0.382, -0.168)	-0.169 (-0.240, -0.094)	-0.026 (-0.120, 0.066)	-0.069 (-0.129, -0.011)
EDSS=9	-0.195 (-0.340, -0.049)	-0.172 (-0.622, 0.281)	-0.177 (-0.418, 0.067)	-0.084 (-0.328, 0.161)	-0.332 (-0.535, 0.070)	-0.309 (-0.457, -0.093)	-0.165 (-0.521, 0.150)	-0.200 (-0.420, 0.004)

Bias between the multivariable regression analysis results (with EDSS as a categorical variable) and the results published in Orme's study were computed and presented in Table 5.7.

Table 5.7: Bias between published regression results and results from simulated EQ-5D data

EDSS Score	Normal (original SD)	Normal	Truncated Normal	LogNormal (original SD)	LogNormal	Beta	Beta Truncated
EDSS=0	0.0096	0.0259	0.0307	0.0629	0.0572	0.0334	0.0419
EDSS=1	0.2695	0.3040	0.4216	0.3840	0.3515	0.3134	0.3266
EDSS=2	0.3360	0.3679	0.4403	0.4572	0.4064	0.3779	0.3867
EDSS=3	0.1336	0.1645	0.1998	0.1572	0.1736	0.1678	0.1817
EDSS=4	0.3580	0.3945	0.4490	0.4631	0.4154	0.4037	0.4144
EDSS=5	0.6248	0.6634	0.7296	0.7634	0.6767	0.6718	0.6839
EDSS=6	0.7957	0.8394	0.8687	0.9206	0.8403	0.8434	0.8615
EDSS=6.5	0.6214	0.6616	0.6971	0.6959	0.6591	0.6675	0.6803
EDSS=7	0.4287	0.4637	0.4614	0.4035	0.4485	0.4571	0.4781
EDSS=8	0.3448	0.3750	0.3857	0.4047	0.3846	0.3584	0.3792
EDSS=9	0.0048	0.0302	0.0417	0.0378	0.0539	0.0165	0.0355
Total Bias	3.9268	4.2900	4.7256	4.7502	4.4673	4.3107	4.4699

Excluding "Normal (original SD)", "Normal" was the set of simulated EQ-5D data closest to the EQ-5D data in Orme's study with a bias of 4.29, followed by the "Beta" (bias=4.31) and "LogNormal" (bias=4.47). Overall, EQ-5D data simulated using the new SD (described in Section 5.2.1.2 and presented in Table 5.4) produced the best sets of simulated EQ-5D data from each of the three distributions used. Based on the results, "Normal" was used as the base-case EQ-5D data for analysis in this thesis. "LogNormal" was used for the sensitivity analysis.

5.3.2 Choice of regression model with EDSS as continuous data

Using the base-case "Normal" EQ-5D data, linear regression models with different degrees of polynomial (on the EDSS score) were fitted in a stepwise fashion by increasing the degree of the polynomial at each step. Likelihood ratio test was performed at each step to determine the polynomial that best fits the EQ-5D data. When the next higher degree polynomial was not statistically a better fit for the data than the current degree polynomial, the current degree polynomial was selected as the polynomial that best fit the data. The polynomial with the best fit for the data was the

cubic model. Similar model selection approach was applied to the "LogNormal" EQ-5D data and the cubic model was also the model that best fit the data.

For the purpose of further exploration, the next higher degree polynomial (quintic model) was also fitted. It was statistically a better fit for the data than both the quartic and cubic models. Further exploration of higher degrees of polynomial above five was not performed as there were only 20 sets of EQ-5D and EDSS in the external dataset for the Bayesian BRMA model. Fitting higher degree polynomial using this set of data is therefore not recommended due to insufficient power. Hence, polynomial models of up to the fifth degree were analysed for both "Normal" and "LogNormal" EQ-5D data.

5.3.3 Construction of the prior distributions

5.3.3.1 List of studies identified for the construction of prior distributions

Using the search strategy and selection criteria described in Section 5.2.3.1, a total of 64 studies (including RCTs and cohort studies) were identified. Twelve of the 64 studies reviewed reported baseline summary statistics for EDSS and EQ-5D that were suitable for the BRMA of EDSS and EQ-5D. Among the 12 studies, two reported summary statistics for three groups of patients or treatment arms, four were two-arm studies and the remaining six were single cohort studies. In total, 20 sets of baseline summary statistics for EDSS and EQ-5D were extracted to be used as data for the construction of prior distribution for the regression coefficients in the regression model associating EQ-5D with EDSS. Summary statistics of the studies identified are presented in Table 5.8.

Table 5.8: Summary statistics of EDSS and EQ-5D reported in the studies identified.

				EDSS			EQ-5D		
Ref	Study	Year	Groups	N	Mean	SD	N	Mean	SD
1	Kobelt (Kobelt et al., 2001)	2001	Unknown	737	4.40	1.000	717	0.55	0.331
2	Putzki (Putzki et al., 2009)	2009	IFNb-1a 30 mcg once weekly	1025	2.00	1.500	1137	0.75	0.172
3	Vaney (Vaney et al., 2012)	2012	Walking Group	23	5.72	1.060	23	0.59	0.170
			Lokomat	26	5.88	0.900	26	0.65	0.150
4	DEFINE	2013	Placebo	408	2.48	1.240	395	0.71	0.250
	(Gold et al., 2012, Kappos et al., 2014)		BG-12 (240 mg BID)	410	2.40	1.290	402	0.72	0.230
			BG-12 (240 mg TID)	416	2.36	1.190	404	0.71	0.240
5	Zettl (Zettl et al., 2013)	2013	Unknown	414	5.30	1.700	414	0.50	0.300
6	Limone (Limone et al., 2013)	2013	Placebo	72	5.80	1.103	72	0.70	0.068
			Dalfampridine-ER 10 mg BID	224	5.80	1.048	221	0.69	0.058
7	Moss-Morris (Moss-Morris et al., 2013)	2013	CBT	48	4.90	1.350	47	0.66	0.220
			SL	46	5.10	1.010	46	0.60	0.260
8	Forgarty (Fogarty et al., 2013)	2013	Mixed	214	3.59	2.640	213	0.59	0.330
9	Reese (Reese et al., 2013)	2013	RRMS	92	2.60	1.500	92	0.83	0.180
			SPMS	38	5.40	1.500	38	0.64	0.278
			PPMS	7	5.90	1.200	7	0.64	0.260
10	Kuspinar (Kuspinar and Mayo, 2013)	2014	Unknown	189	2.00	1.852	189	0.69	0.180
11	Garopoulou (Garopoulou et al., 2014)	2014	MS-I	5	1.40	0.700	5	0.64	0.229
			MS-C	5	1.40	0.200	5	0.74	0.109
12	Mitosek-Szewcvyk (Mitosek-Szewczyk et al., 2014)	2014	Unknown	3521	3.34	2.200	3521	0.80	0.270

5.3.3.2 Prior distributions for the Bayesian linear regression models

Prior distributions for the regression coefficients of the Bayesian linear regression model were constructed using the posterior distributions of the regression coefficients from its corresponding Bayesian BRMA model. For example, in the case of the first degree polynomial linear regression model, the prior distribution for α and β_1 in the Bayesian linear regression model was constructed using the posterior distribution of λ_0 and λ_1 respectively from the Bayesian BRMA model. For the higher degree polynomials, the prior distributions for β_k in the Bayesian linear regression models were constructed using the posterior distributions of λ_k from the Bayesian BRMA model where k =2 to 5. Summary statistics for the posterior distribution of the lambdas (λ) in the BRMA models estimated using the external dataset (identified in Section 5.3.3.1 as the likelihood data) are presented in Table 5.9. Non-informative prior distributions for the between-study model and informative prior distribution for the within-study correlation of the within-study model of the BRMA were also used as discussed in Section 5.2.3.2 (Table 5.2).

Table 5.9: Summary statistics for the posterior distributions of the regression coefficients in the BRMA models

Regression Coefficients	Mean (SD)					
	1st	1st	2nd	3rd	4th	5th
$\lambda_0 \rightarrow \alpha$	0.642 (0.0225)	0.641 (0.0230)	0.633 (0.0268)	0.585 (0.0320)	0.569 (0.0370)	0.565 (0.0363)
$\lambda_1 \rightarrow \beta_1$	-0.028 (0.0117)	-0.029 (0.0121)	-0.013 (0.0332)	0.019 (0.0340)	-0.005 (0.0467)	0.070 (0.0892)
$\lambda_2 \rightarrow \beta_2$			0.007 (0.0132)	0.083 (0.0371)	0.128 (0.0647)	0.141 (0.0630)
$\lambda_3 \rightarrow \beta_3$				0.020 (0.0093)	0.059 (0.0462)	-0.032 (0.1046)
$\lambda_4 \rightarrow \beta_4$					0.007 (0.0083)	-0.046 (0.0554)
$\lambda_5 \rightarrow \beta_5$						-0.008 (0.0083)

Density plots of the posterior distributions of the regression coefficients appeared reasonably normally distributed. Therefore, prior distributions for the corresponding regression coefficients were specified using normal distribution with means and standard deviations presented in Table 5.9. Graphs showing the posterior distributions plotted using empirical data from the BRMA analyses and the prior distributions plotted using normal distribution with parameter values shown in Table 5.9 are presented in Figure 5.4.

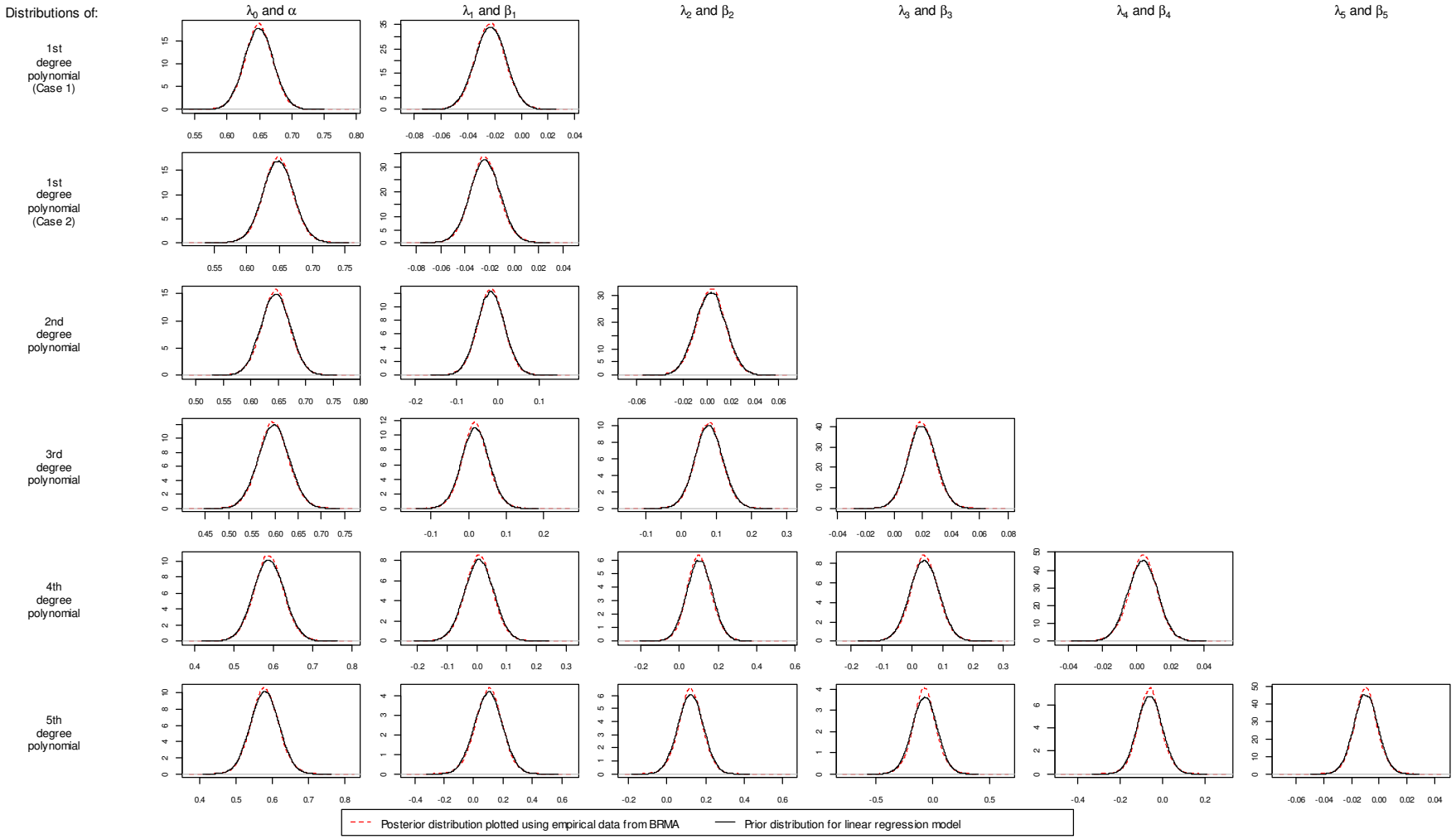

Figure 5.4: Distributions of posterior distributions from BRMA and prior distributions plotted using normal distribution

5.3.4 Estimation of EQ-5D for each EDSS score using linear regression

Bayesian univariable linear regression model was used to investigate the association of EQ-5D with EDSS, where EDSS was treated as a continuous variable. As mentioned in Section 5.3.2, using stepwise likelihood ratio test, the cubic model best explained the association between EQ-5D and EDSS. However, with further exploration, the quintic model fitted the data better than the cubic model and hence it was also included in the analyses. Bayesian linear regression analyses with polynomials from the first degree to the fifth degree using "Normal" EQ-5D data were performed. The Bayesian model fit was assessed using Deviance Information Criterion (DIC), which are presented in Table 5.10.

Table 5.10: Bayesian DIC for the 1st degree to 5th degree polynomial models

	Polynomial					
Prior	1st (Case 1)	1st (Case 2)	2nd	3rd	4th	5th
Univariable						
Non-informative	2270.57		2201.1	2155.27	2156.49	2151.55
Informative	2280.62	2279.21	2201.78	2155.01	2156.19	2151.26
Multivariable						
Non-informative	2269.86		2200.89	2155.77	2157.21	2151.46
Informative	2290.91	2288.87	2204.04	2155.69	2156.96	2150.99

The Bayesian DIC were consistent with the results from the likelihood ratio test, which indicated that the quintic model was better than the cubic model in explaining the association between EQ-5D and EDSS. The DIC revealed that for the higher degree polynomial models (from cubic onwards), the models with informative priors fitted the data better than the models with non-informative priors; the reverse was true for the lower degree polynomials. These results were consistent for both the univariable and multivariable linear regression analyses.

Results of the Bayesian univariable linear regression analyses with the cubic and quintic models, using non-informative and informative prior distributions (presented in Table 5.2 and Table 5.9), are presented in Figure 5.5 and Figure 5.6 respectively. Summary statistics of the EQ-5D estimates at each EDSS score for the models are presented in Table 5.11.

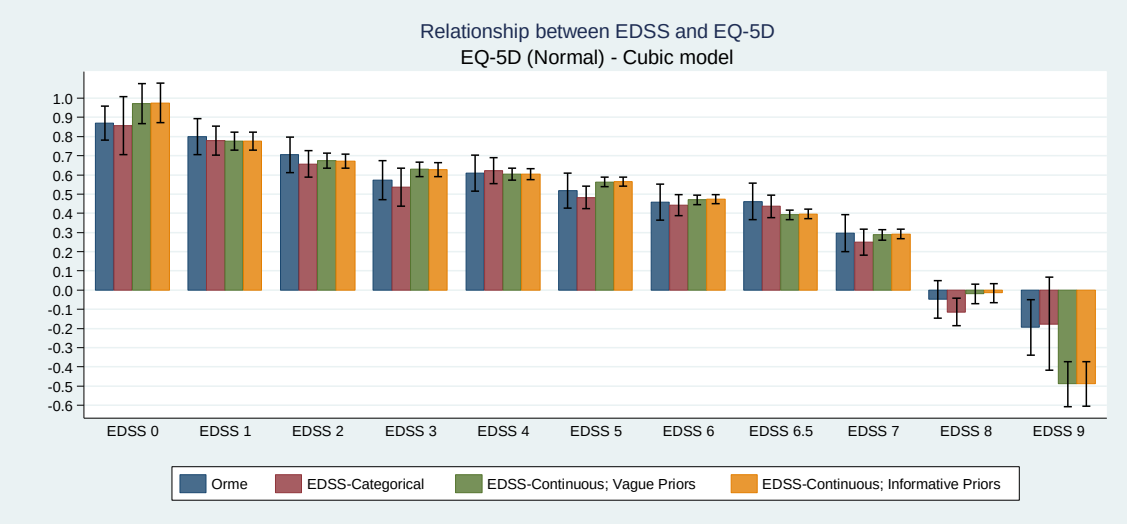

Figure 5.5: Association between EDSS and EQ-5D estimated using univariable linear regression with cubic model

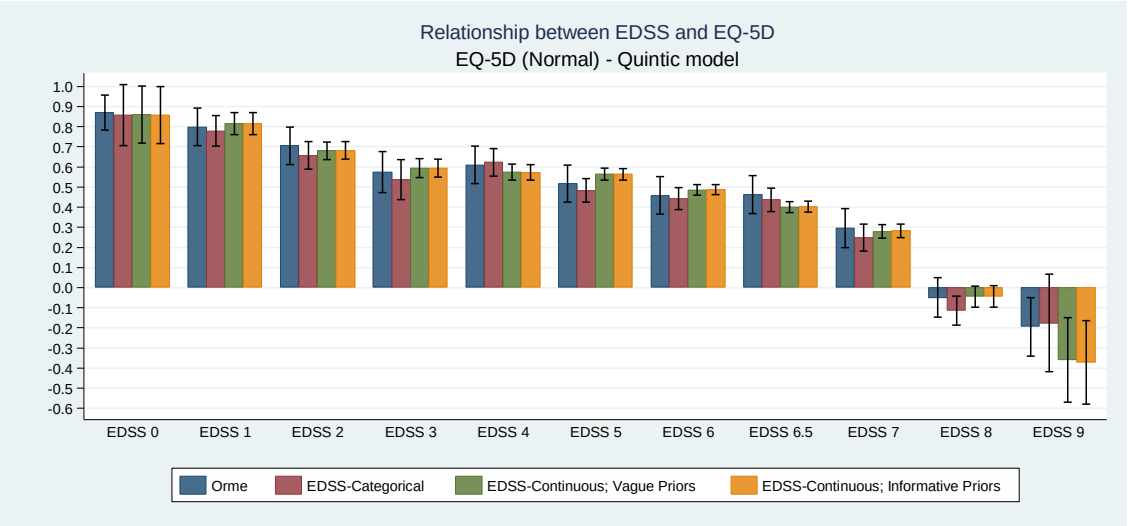

Figure 5.6: Association between EDSS and EQ-5D estimated using univariable linear regression with quintic model

The EQ-5D estimates from the cubic model extended over a wider range of values compared to the estimates from both the quantic model and linear regression model with EDSS as a categorical data. The EQ-5D estimates from the quintic model were comparable to the estimates from the categorical analysis except for estimates at EDSS 9.

Using the cubic linear regression model, the 95% CrI for EQ-5D estimates had upper limit above one at EDSS 0 and lower limit below -0.594 at EDSS 9, which was beyond the range of plausible EQ-5D values (of [-0.594,1]). This was the case when using both

non-informative and informative priors. Precision of the estimates for EQ-5D were improved when using informative prior distributions as expected. Widths of the 95% CrIs for the EQ-5D estimates when using informative priors were reduced by 0% to 7% compared to the widths of the intervals when using non-informative priors. For the quintic linear regression model, only the 95% CrI for EQ-5D at EDSS 0 using non-informative prior extended beyond the plausible value for EQ-5D. The precision of the estimates when using informative priors were also higher than that when using non-informative priors. The reductions in the width of the 95% CrIs were between 0.2% and 7.4% inclusive.

Multivariable linear regression analyses (using non-informative and informative prior distributions) adjusting for gender, recent relapse status, education level and type of MS were performed and results of the analyses are presented in Table 5.12. Graphs showing the relationship between EDSS and EQ-5D using the cubic and quintic linear regression models, adjusting for covariates, are shown in Figure 5.7 and Figure 5.8 respectively.

Table 5.11: EQ-5D estimates from univariable linear regression analyses

"Normal" EQ-5D for	Simulated data	Categorical Analysis	Cubic model (3rd-degree polynomial)			Quintic model (5th-degree polynomial)		
			Non-informative	Informative	Reduction in	Non-informative	Informative	Reduction in
	Mean (95% CI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)	width of CrIs	Mean (95% CrI)	Mean (95% CrI)	width of CrIs
EDSS=0	0.885 (0.815, 0.955)	0.856 (0.705, 1.009)	0.971 (0.867, 1.076)	0.975 (0.872, 1.078)	1.4%	0.860 (0.718, 1.002)	0.857 (0.715, 0.999)	0.3%
EDSS=1	0.802 (0.754, 0.850)	0.778 (0.703, 0.854)	0.776 (0.730, 0.823)	0.775 (0.729, 0.822)	0.0%	0.815 (0.761, 0.869)	0.816 (0.762, 0.870)	0.3%
EDSS=2	0.685 (0.626, 0.744)	0.657 (0.588, 0.727)	0.674 (0.636, 0.713)	0.672 (0.635, 0.709)	2.6%	0.681 (0.637, 0.725)	0.682 (0.638, 0.726)	0.2%
EDSS=3	0.570 (0.497, 0.642)	0.537 (0.437, 0.636)	0.629 (0.591, 0.667)	0.627 (0.591, 0.663)	4.7%	0.594 (0.548, 0.640)	0.594 (0.548, 0.640)	1.2%
EDSS=4	0.644 (0.585, 0.703)	0.623 (0.554, 0.691)	0.604 (0.573, 0.636)	0.604 (0.574, 0.634)	6.0%	0.574 (0.534, 0.613)	0.573 (0.534, 0.610)	3.9%
EDSS=5	0.508 (0.461, 0.555)	0.482 (0.424, 0.541)	0.564 (0.538, 0.589)	0.565 (0.542, 0.589)	7.0%	0.564 (0.534, 0.595)	0.563 (0.536, 0.592)	7.4%
EDSS=6	0.468 (0.426, 0.510)	0.443 (0.388, 0.498)	0.470 (0.446, 0.494)	0.474 (0.451, 0.497)	4.9%	0.486 (0.459, 0.512)	0.487 (0.461, 0.513)	3.4%
EDSS=6.5	0.463 (0.418, 0.508)	0.437 (0.378, 0.496)	0.392 (0.368, 0.417)	0.397 (0.373, 0.421)	3.3%	0.400 (0.373, 0.427)	0.402 (0.376, 0.429)	0.8%
EDSS=7	0.279 (0.216, 0.341)	0.249 (0.182, 0.316)	0.288 (0.261, 0.314)	0.293 (0.267, 0.319)	1.5%	0.279 (0.246, 0.313)	0.283 (0.250, 0.316)	0.5%
EDSS=8	-0.090 (-0.153, -0.028)	-0.114 (-0.186, -0.042)	-0.020 (-0.071, 0.030)	-0.016 (-0.066, 0.034)	0.7%	-0.045 (-0.098, 0.008)	-0.044 (-0.096, 0.009)	0.7%
EDSS=9	-0.169 (-0.387, 0.049)	-0.177 (-0.418, 0.067)	-0.490 (-0.609, -0.373)	-0.490 (-0.606, -0.374)	1.4%	-0.360 (-0.570, -0.149)	-0.372 (-0.580, -0.163)	0.9%

Table 5.12: EQ-5D estimates from multivariable linear regression analyses, adjusting for gender, recent relapse status, education level and type of MS

"Normal" EQ-5D for	Simulated data	Categorical Analysis	Cubic model (3rd-degree polynomial)			Quintic model (5th-degree polynomial)		
			Non-informative	Informative	Reduction in	Non-informative	Informative	Reduction in
	Mean (95% CI)	Mean (95% CrI)	Mean (95% CrI)	Mean (95% CrI)	width of CrIs	Mean (95% CrI)	Mean (95% CrI)	width of CrIs
EDSS=0	0.885 (0.815, 0.955)	0.856 (0.705, 1.009)	0.943 (0.832, 1.054)	0.960 (0.850, 1.068)	1.9%	0.835 (0.690, 0.980)	0.840 (0.696, 0.984)	0.7%
EDSS=1	0.802 (0.754, 0.850)	0.778 (0.703, 0.854)	0.750 (0.690, 0.811)	0.763 (0.705, 0.821)	4.0%	0.792 (0.725, 0.859)	0.800 (0.735, 0.865)	2.5%
EDSS=2	0.685 (0.626, 0.744)	0.657 (0.588, 0.727)	0.649 (0.596, 0.704)	0.661 (0.613, 0.710)	9.6%	0.656 (0.598, 0.714)	0.664 (0.608, 0.720)	3.5%
EDSS=3	0.570 (0.497, 0.642)	0.537 (0.437, 0.636)	0.604 (0.551, 0.658)	0.617 (0.570, 0.665)	12.3%	0.569 (0.510, 0.629)	0.577 (0.521, 0.632)	6.2%
EDSS=4	0.644 (0.585, 0.703)	0.623 (0.554, 0.691)	0.579 (0.530, 0.629)	0.594 (0.553, 0.636)	15.8%	0.550 (0.496, 0.605)	0.557 (0.509, 0.606)	11.2%
EDSS=5	0.508 (0.461, 0.555)	0.482 (0.424, 0.541)	0.538 (0.492, 0.584)	0.555 (0.518, 0.593)	17.9%	0.542 (0.493, 0.590)	0.550 (0.509, 0.591)	16.0%
EDSS=6	0.468 (0.426, 0.510)	0.443 (0.388, 0.498)	0.444 (0.399, 0.490)	0.463 (0.425, 0.502)	15.7%	0.462 (0.415, 0.509)	0.472 (0.430, 0.513)	11.8%
EDSS=6.5	0.463 (0.418, 0.508)	0.437 (0.378, 0.496)	0.366 (0.321, 0.412)	0.386 (0.347, 0.426)	13.4%	0.375 (0.328, 0.422)	0.385 (0.342, 0.428)	8.0%
EDSS=7	0.279 (0.216, 0.341)	0.249 (0.182, 0.316)	0.262 (0.215, 0.309)	0.281 (0.240, 0.323)	10.5%	0.253 (0.203, 0.304)	0.263 (0.215, 0.312)	4.1%
EDSS=8	-0.090 (-0.153, -0.028)	-0.114 (-0.186, -0.042)	-0.045 (-0.108, 0.018)	-0.027 (-0.089, 0.035)	2.5%	-0.068 (-0.133, -0.003)	-0.061 (-0.124, 0.003)	1.8%
EDSS=9	-0.169 (-0.387, 0.049)	-0.177 (-0.418, 0.067)	-0.512 (-0.634, -0.389)	-0.499 (-0.620, -0.378)	0.9%	-0.361 (-0.575, -0.146)	-0.363 (-0.573, -0.151)	1.7%

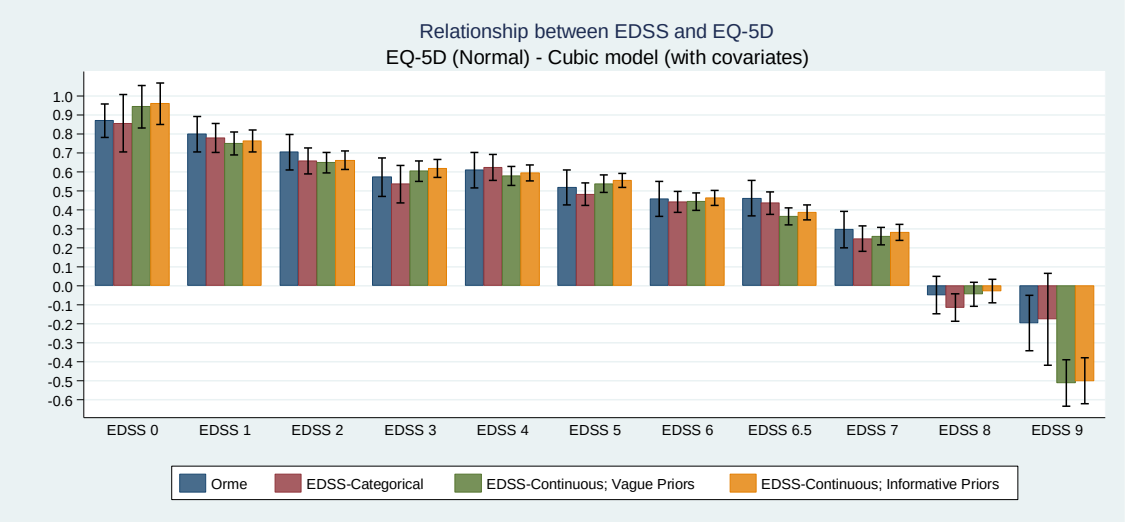

Figure 5.7: Association between EDSS and EQ-5D estimated using multivariable linear regression with cubic model

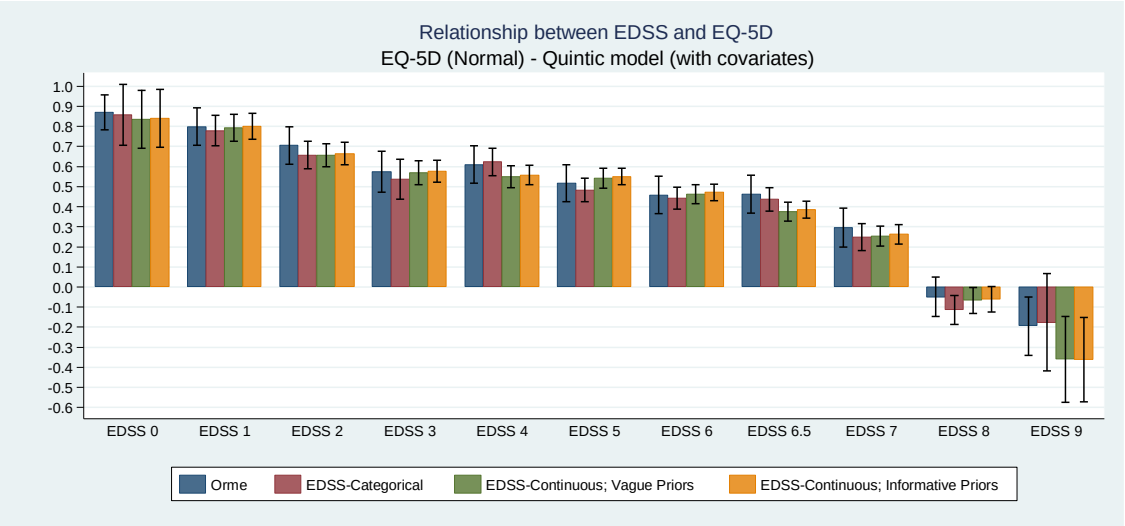

Figure 5.8: Association between EDSS and EQ-5D estimated using multivariable linear regression with quintic model

Similar to the univariable linear regression analyses results, the EQ-5D estimates when using the quintic model were comparable to the estimates from the categorical linear regression analysis. The 95% CrIs for EQ-5D at EDSS 0 and EDSS 9 in the cubic model also extended beyond the plausible range of EQ-5D values as in the case of the univariable analysis. This was however not the case for the quintic model. The 95% CrI for EQ-5D when using both non-informative and informative priors were within the range of [-0.594, 1]. Reduction in the width of the 95% CrIs when using informative priors compared to non-informative ranged from 0.9% to 17.9% for the cubic model and from 0.7% to 16.0% for the quintic model.

In the multivariable linear regression model where EDSS was analysed as a categorical variable, the variance for EQ-5D at each EDSS score was different, whereas in the case of the regression model where EDSS was treated as a continuous variable, the variance for EQ-5D each EDSS score was assumed constant across all EDSS scores from 0 to 9. As mentioned in Section 5.2.2, the reason why EDSS was analysed as a continuous variable in this thesis was due to the fact that the informative priors for the Bayesian linear regression analyses were constructed from the BRMA model where EDSS could only be analysed as a continuous variable. This was because most published articles (which formed the external evidence for the construction of the informative priors) reported mainly mean with standard deviation or 95% CIs for both EDSS and EQ-5D. To investigate the effect of allowing the variance of EQ-5D at each EDSS score to vary on the estimates of EQ-5D, sensitivity analysis were performed. Details on the modelling of the EQ-5D variance across all EDSS scores and the results are presented in the next section.

5.3.5 Sensitivity Analyses

5.3.5.1 Non-constant variance models for EQ-5D

In the results presented in Section 5.3.4, the variance of EQ-5D at each EDSS score was assumed the same in the linear regression model. Considering that the variances of EQ-5D were not expected to be constant across all EDSS scores, the model was modified to allow the variances of EQ-5D to vary. Variance of the base-case EQ-5D at each EDSS score i ($i = 1, 2, 3, \dots, 6, 6.5, 7, 8, 9$) was calculated for fitting polynomial and exponential functions that explain the structure of the variances across all EDSS scores. The polynomial model that best fit the structure of the variance of EQ-5D was the quadratic model and is defined as:

$$Var\left(X_{EQ5D,i}\right) = 0.1812 + 0.0051\left(X_{EDSS,i} - \overline{X_{EDSS,\iota}}\right) - 0.0047\left(X_{EDSS,i} - \overline{X_{EDSS,\iota}}\right)^2 + \varepsilon_i$$

and exponential function that best fit the EQ-5D variance is defined as:

$$Var(X_{EQ5D,i}) = 0.2015\ exp\left(-0.0555\left(X_{EDSS,i} - \overline{X_{EDSS,i}}\right)^2\right) + \varepsilon_i$$

where ε_i is the residual error term and follows a chi-square distribution as follows:

$$\varepsilon_i \sim \chi^2(n_{EDSS,i} - 1)$$

and $n_{EDSS,i}$ is the number of patients in each EDSS score i. Figure 5.9 shows the variance of EQ-5D at each EDSS score with the quadratic model and exponential function fitted to the variance data.

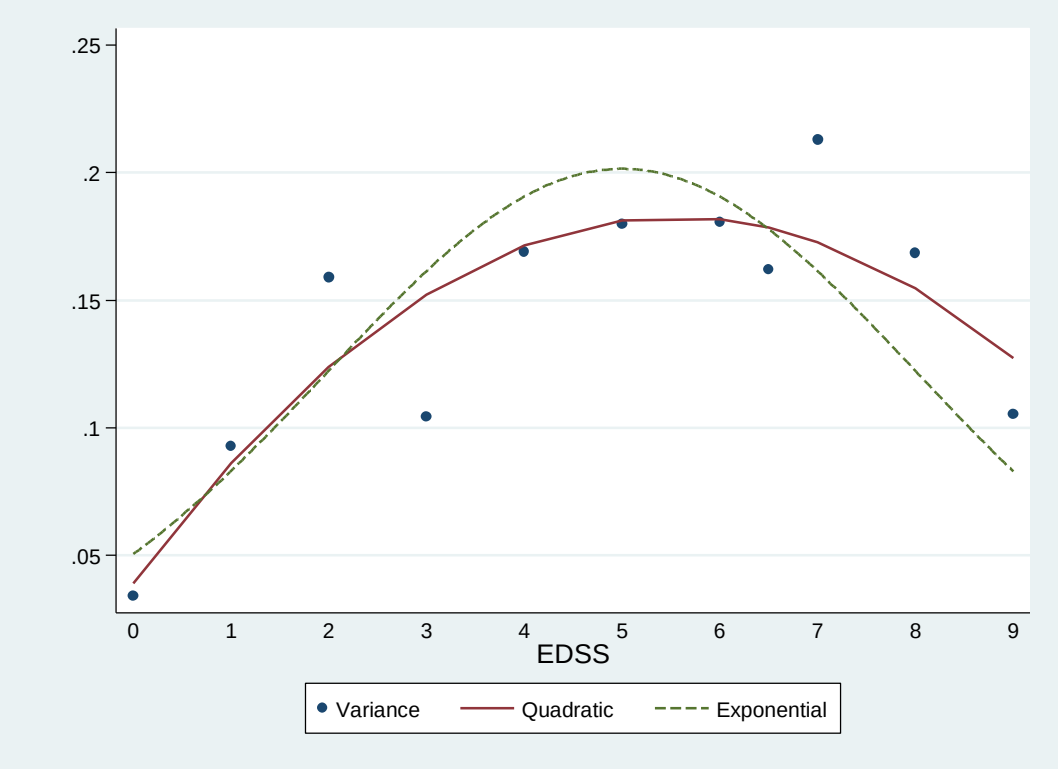

Figure 5.9: Variances of EQ-5D by EDSS scores

Results from the multivariable linear regression analysis (adjusting for gender, recent relapse status, education level and type of MS) using the quadratic variance model are presented in Table 5.13. Results from using the exponential function for the variances of EQ-5D are presented in Table 5.14.

Table 5.13: EQ-5D estimates using quadratic variance model in Bayesian multivariable linear regression

	Mean EQ-5D (95% CrI)			
	Cubic model (3rd-degree polynomial)		Quintic model (5th-degree polynomial)	
EDSS	Non-informative	Informative	Non-informative	Informative
0	0.872 (0.388, 1.364)	1.001 (0.586, 1.450)	0.845 (0.270, 1.418)	0.877 (0.299, 1.402)
1	0.737 (0.397, 1.074)	0.802 (0.547, 1.063)	0.789 (0.371, 1.204)	0.918 (0.559, 1.293)
2	0.658 (0.313, 1.005)	0.689 (0.509, 0.870)	0.639 (0.266, 1.008)	0.760 (0.463, 1.066)
3	0.611 (0.269, 0.955)	0.634 (0.502, 0.766)	0.555 (0.169, 0.940)	0.618 (0.385, 0.855)
4	0.572 (0.250, 0.895)	0.611 (0.524, 0.697)	0.549 (0.187, 0.912)	0.564 (0.415, 0.713)
5	0.516 (0.206, 0.826)	0.591 (0.527, 0.655)	0.549 (0.205, 0.888)	0.572 (0.497, 0.647)
6	0.420 (0.108, 0.731)	0.547 (0.462, 0.632)	0.460 (0.121, 0.791)	0.560 (0.424, 0.698)
6.5	0.349 (0.035, 0.660)	0.508 (0.397, 0.620)	0.364 (0.030, 0.687)	0.517 (0.327, 0.711)
7	0.258 (-0.058, 0.575)	0.452 (0.301, 0.609)	0.234 (-0.115, 0.580)	0.433 (0.174, 0.700)
8	0.006 (-0.378, 0.380)	0.278 (-0.022, 0.601)	-0.071 (-0.508, 0.369)	0.125 (-0.262, 0.532)
9	-0.360 (-1.033, 0.262)	-0.002 (-0.581, 0.619)	-0.227 (-1.113, 0.618)	-0.357 (-1.360, 0.408)
DIC	8135.71	8138.33	8153.47	8147.5

Table 5.14: EQ-5D estimates using exponential variance model in Bayesian multivariable linear regression

	Mean EQ-5D (95% CrI)			
	Cubic model (3rd-degree polynomial)		Quintic model (5th-degree polynomial)	
EDSS	Non-informative	Informative	Non-informative	Informative
0	0.870 (0.380, 1.365)	1.003 (0.580, 1.455)	0.841 (0.247, 1.425)	0.872 (0.291, 1.408)
1	0.736 (0.393, 1.075)	0.802 (0.544, 1.067)	0.789 (0.371, 1.201)	0.918 (0.558, 1.291)
2	0.657 (0.312, 1.000)	0.688 (0.508, 0.870)	0.638 (0.269, 1.008)	0.757 (0.459, 1.065)
3	0.610 (0.266, 0.953)	0.634 (0.503, 0.766)	0.554 (0.167, 0.938)	0.613 (0.378, 0.853)
4	0.569 (0.248, 0.892)	0.611 (0.524, 0.698)	0.547 (0.178, 0.912)	0.561 (0.411, 0.711)
5	0.513 (0.203, 0.827)	0.591 (0.527, 0.655)	0.547 (0.202, 0.893)	0.572 (0.497, 0.647)
6	0.416 (0.106, 0.731)	0.546 (0.462, 0.631)	0.459 (0.122, 0.794)	0.560 (0.423, 0.697)
6.5	0.345 (0.035, 0.660)	0.506 (0.396, 0.617)	0.363 (0.032, 0.690)	0.514 (0.324, 0.708)
7	0.255 (-0.059, 0.574)	0.449 (0.301, 0.603)	0.233 (-0.107, 0.575)	0.429 (0.174, 0.696)
8	0.006 (-0.372, 0.379)	0.270 (-0.023, 0.585)	-0.072 (-0.500, 0.359)	0.121 (-0.256, 0.522)
9	-0.354 (-1.011, 0.258)	-0.017 (-0.582, 0.589)	-0.225 (-1.088, 0.597)	-0.340 (-1.298, 0.393)
DIC	8142.09	8144.98	8161.17	8156.18

Comparing between the two models, based on the DIC, the quadratic model appeared to be slightly better than the exponential model in explaining the variability in the variance of EQ-5D across the different values of EDSS. However, allowing the variance of EQ-5D to vary at different EDSS scores introduced more variability to the estimates of EQ-5D as compared to the results presented in the base-case analyses in Section 5.3.4 where the variance of EQ-5D was constant across all EDSS scores. The 95% CrIs were markedly wider than those in the base-case analyses and the mean EQ-

5D at EDSS 0 for the cubic regression model using informative priors were above 1 for both the quadratic and exponential variance models. Using DIC for model comparison, the use of varying variance did not improve the estimates of EQ-5D when compared to the constant variance cubic model with DIC=2155.77 (non-informative prior) and DIC=2155.69 (informative prior) as shown in Table 5.10.

5.3.5.2 Estimates of EQ-5D using "LogNormal" EQ-5D data

Individual patient data of Orme's study were not available and were simulated using several standard statistical distributions. Assessment of the simulated data suggested that the "Normal" EQ-5D data simulated using the Normal distribution was the best (i.e. having the least bias when compared to Orme's data). In this section, sensitivity analysis to explore how the estimate of EQ-5D at each EDSS score when using the "LogNormal" EQ-5D data would compare with the results when using the "Normal" EQ-5D data.

Bayesian univariable and multivariable linear regression analyses were performed using the "LogNormal" EQ-5D data. Results of the analysis when using the cubic model in the regression model are presented in Table 5.15, alongside the results when using the "Normal" EQ-5D data.

Mean EQ-5D from the analyses using "LogNormal" data was consistently higher than those estimated using "Normal" except for mean EQ-5D at EDSS 8. The "Normal" data gave estimates that extended beyond the plausible range of EQ-5D value of [-0.594, 1] on both upper and lower limits while the estimates from the "LogNormal" data were within the range of plausible EQ-5D values. This was largely a result of the back-transformation that was applied at the end of the Bayesian linear regression analysis to map the log-normal EQ-5D estimates back to the EQ-5D score scale.

Results from the Bayesian linear regression analysis conducted using "LogNormal" EQ-5D data compared to "Normal" EQ-5D data when using the quintic model in the regression model are presented in Table 5.16. The results were similar to those reported in the cubic models.

Table 5.15: EQ-5D estimates from linear regression analyses using cubic model

	Mean EQ-5D (95% CrI)			
	"Normal"		"LogNormal"	
EDSS	Non-informative	Informative	Non-informative	Informative
Univariable Analysis				
0	0.971 (0.867, 1.076)	0.975 (0.872, 1.078)	0.973 (0.963, 0.981)	0.973 (0.963, 0.981)
1	0.776 (0.730, 0.823)	0.775 (0.729, 0.822)	0.917 (0.904, 0.928)	0.917 (0.904, 0.928)
2	0.674 (0.636, 0.713)	0.672 (0.635, 0.709)	0.840 (0.821, 0.857)	0.840 (0.822, 0.857)
3	0.629 (0.591, 0.667)	0.627 (0.591, 0.663)	0.773 (0.748, 0.796)	0.772 (0.748, 0.795)
4	0.604 (0.573, 0.636)	0.604 (0.574, 0.634)	0.724 (0.700, 0.747)	0.722 (0.699, 0.745)
5	0.564 (0.538, 0.589)	0.565 (0.542, 0.589)	0.671 (0.649, 0.692)	0.668 (0.647, 0.688)
6	0.470 (0.446, 0.494)	0.474 (0.451, 0.497)	0.569 (0.544, 0.594)	0.567 (0.542, 0.590)
6.5	0.392 (0.368, 0.417)	0.397 (0.373, 0.421)	0.478 (0.450, 0.506)	0.476 (0.449, 0.504)
7	0.288 (0.261, 0.314)	0.293 (0.267, 0.319)	0.344 (0.311, 0.377)	0.344 (0.311, 0.376)
8	-0.020 (-0.071, 0.030)	-0.016 (-0.066, 0.034)	-0.070 (-0.126, -0.010)	-0.063 (-0.120, -0.003)
9	-0.490 (-0.609, -0.373)	-0.490 (-0.606, -0.374)	-0.456 (-0.499, -0.402)	-0.449 (-0.494, -0.394)
Multivariable Analysis				
0	0.943 (0.832, 1.054)	0.960 (0.850, 1.068)	0.971 (0.959, 0.980)	0.970 (0.958, 0.979)
1	0.750 (0.690, 0.811)	0.763 (0.705, 0.821)	0.909 (0.892, 0.925)	0.910 (0.892, 0.925)
2	0.649 (0.596, 0.704)	0.661 (0.613, 0.710)	0.826 (0.798, 0.852)	0.827 (0.801, 0.852)
3	0.604 (0.551, 0.658)	0.617 (0.570, 0.665)	0.755 (0.718, 0.790)	0.757 (0.722, 0.788)
4	0.579 (0.530, 0.629)	0.594 (0.553, 0.636)	0.704 (0.663, 0.741)	0.704 (0.669, 0.738)
5	0.538 (0.492, 0.584)	0.555 (0.518, 0.593)	0.648 (0.606, 0.687)	0.647 (0.611, 0.682)
6	0.444 (0.399, 0.490)	0.463 (0.425, 0.502)	0.541 (0.492, 0.589)	0.541 (0.497, 0.583)
6.5	0.366 (0.321, 0.412)	0.386 (0.347, 0.426)	0.447 (0.392, 0.500)	0.446 (0.397, 0.495)
7	0.262 (0.215, 0.309)	0.281 (0.240, 0.323)	0.309 (0.249, 0.368)	0.310 (0.255, 0.365)
8	-0.045 (-0.108, 0.018)	-0.027 (-0.089, 0.035)	-0.103 (-0.171, -0.031)	-0.097 (-0.164, -0.026)
9	-0.512 (-0.634, -0.389)	-0.499 (-0.620, -0.378)	-0.468 (-0.509, -0.416)	-0.463 (-0.505, -0.410)

Table 5.16: EQ-5D estimates from linear regression analyses using quintic model

	Mean EQ-5D (95% CrI)			
	"Normal"		"LogNormal"	
EDSS	Non-informative	Informative	Non-informative	Informative
Univariable Analysis				
0	0.860 (0.718, 1.002)	0.857 (0.715, 0.999)	0.953 (0.928, 0.971)	0.952 (0.927, 0.970)
1	0.815 (0.761, 0.869)	0.816 (0.762, 0.870)	0.930 (0.918, 0.941)	0.931 (0.918, 0.942)
2	0.681 (0.637, 0.725)	0.682 (0.638, 0.726)	0.845 (0.824, 0.864)	0.846 (0.826, 0.865)
3	0.594 (0.548, 0.640)	0.594 (0.548, 0.640)	0.739 (0.704, 0.770)	0.737 (0.703, 0.768)
4	0.574 (0.534, 0.613)	0.573 (0.534, 0.610)	0.687 (0.654, 0.719)	0.680 (0.647, 0.711)
5	0.564 (0.534, 0.595)	0.563 (0.536, 0.592)	0.670 (0.644, 0.696)	0.660 (0.635, 0.685)
6	0.486 (0.459, 0.512)	0.487 (0.461, 0.513)	0.592 (0.566, 0.618)	0.587 (0.561, 0.613)
6.5	0.400 (0.373, 0.427)	0.402 (0.376, 0.429)	0.493 (0.462, 0.522)	0.493 (0.463, 0.522)
7	0.279 (0.246, 0.313)	0.283 (0.250, 0.316)	0.331 (0.290, 0.373)	0.338 (0.297, 0.379)
8	-0.045 (-0.098, 0.008)	-0.044 (-0.096, 0.009)	-0.109 (-0.166, -0.050)	-0.104 (-0.160, -0.044)
9	-0.360 (-0.570, -0.149)	-0.372 (-0.580, -0.163)	-0.359 (-0.471, -0.198)	-0.381 (-0.483, -0.233)
Multivariable Analysis				
0	0.835 (0.690, 0.980)	0.840 (0.696, 0.984)	0.949 (0.922, 0.969)	0.948 (0.920, 0.968)
1	0.792 (0.725, 0.859)	0.800 (0.735, 0.865)	0.925 (0.908, 0.939)	0.923 (0.907, 0.938)
2	0.656 (0.598, 0.714)	0.664 (0.608, 0.720)	0.834 (0.804, 0.860)	0.832 (0.803, 0.858)
3	0.569 (0.510, 0.629)	0.577 (0.521, 0.632)	0.722 (0.676, 0.764)	0.716 (0.672, 0.757)
4	0.550 (0.496, 0.605)	0.557 (0.509, 0.606)	0.668 (0.620, 0.713)	0.657 (0.613, 0.699)
5	0.542 (0.493, 0.590)	0.550 (0.509, 0.591)	0.648 (0.603, 0.690)	0.636 (0.596, 0.673)
6	0.462 (0.415, 0.509)	0.472 (0.430, 0.513)	0.567 (0.517, 0.614)	0.557 (0.512, 0.600)
6.5	0.375 (0.328, 0.422)	0.385 (0.342, 0.428)	0.464 (0.409, 0.517)	0.458 (0.406, 0.507)
7	0.253 (0.203, 0.304)	0.263 (0.215, 0.312)	0.300 (0.235, 0.364)	0.298 (0.235, 0.360)
8	-0.068 (-0.133, -0.003)	-0.061 (-0.124, 0.003)	-0.137 (-0.204, -0.066)	-0.140 (-0.205, -0.070)
9	-0.361 (-0.575, -0.146)	-0.363 (-0.573, -0.151)	-0.381 (-0.485, -0.228)	-0.401 (-0.495, -0.262)

5.4 Discussions

Evidence synthesis using meta-analysis is widely used in clinical research to understand the results of any study in the context of all other relevant and available evidence. Average values of outcomes of interest (or average treatment effects on those outcomes) have commonly been obtained by meta-analysing all relevant published evidence. In this project on the estimation of EQ-5D in MS, evidence synthesis takes on a different approach. Evidence synthesis technique was not applied to estimate the average endpoint of interest (EQ-5D), instead it was utilised for the construction of informative prior distributions for the subsequent estimation of EQ-5D (using regression models) in a Bayesian framework.

Results in this project show that instead of using EQ-5D estimates solely from a single study (Orme's study in this case) to inform the economic model for cost-effectiveness analysis, it is possible to incorporate external evidence on EQ-5D and EDSS to obtain the EQ-5D estimates. When using informative prior distributions in a quintic multivariable linear regression model for the estimation of EQ-5D at each EDSS, the precision of the estimates increased compared to those obtained from analyses using non-informative priors. For example, at EDSS 5, the width of the 95% CrI was narrower by 16% compared to estimates obtained from analysis using non-informative priors.

The joint analysis of EQ-5D and EDSS using a BRMA model allows the relationship between the two endpoints to be embedded in the prior distributions constructed and utilised to inform the EQ-5D estimates in the subsequent regression analyses. Estimates of EQ-5D when incorporating external evidence have higher precision and can in turn be used to construct utilities distribution with higher precision for use in a probabilistic cost-effectiveness analysis.

A recent study (Fogarty et al., 2013) in the Ireland also reported the relationship between EQ-5D and EDSS for the purpose of informing economic decision making. Instead of using a linear regression model with EDSS as a categorical variable (as in Orme's study) in their analysis, they modelled EQ-5D with EDSS as a continuous variable using a piecewise linear regression. Apart from that, no adjustment for other covariate was deemed necessary in Fogarty's study.

The distribution of EDSS scores was bimodal with the highest peak at EDSS 6 for both studies. However, Fogarty's study classified patients with EDSS 6 and EDSS 6.5 in the category EDSS 6 while Orme's study had them as two separate categories. It is therefore not appropriate to compare the results of the studies at EDSS 6. The second peak in Fogarty's study was at EDSS 1 with mean EQ-5D of 0.80 (95% CI = 0.75 to 0.85). The mean EQ-5D reported in Orme's study was 0.799 (95% CI = 0.705 to 0.893) and the estimates from the Bayesian analyses in this thesis were 0.800 (95% CrI = 0.735 to 0.865) and 0.792 (95% CrI = 0.725 to 0.859) when using informative and non-informative prior distributions respectively. The results were similar except that the 95% CIs for EQ-5D in Fogarty's study are narrower compared to that in Orme's study and also the 95% CrIs for EQ-5D in the Bayesian analyses. The 95% CIs of EQ-5D in Fogarty's study may have been narrower because it is a univariable analysis and uses a piecewise linear regression model instead of a simple linear regression model.

One major limitation of this project is the absence of IPD of the population in Orme's study, which was not obtainable from the author of Orme's study. Coupled with that, the mean and standard deviation of the patients' EQ-5D at each EDSS score and year since diagnosis were not reported in the article by Orme and colleagues. The mean EQ-5D at each EDSS score from the multivariable linear regression analysis results was used in place of the mean EQ-5D from the patient data for simulating the EQ-5D data. Strong assumptions had to be made to estimate the standard deviation of EQ-5D at each EDSS score. It was not possible to simulate the year since diagnosis data and this variable was not adjusted for in the Bayesian multivariable analyses. Apart from this, data for the other variables (gender, recent relapse status, type of MS and education) were simulated independently as the relationships of these variables with one another and with EDSS and EQ-5D were not known. When these variables were adjusted for in the Bayesian multivariable analyses, it is unlikely that any effects of these covariates on EQ-5D would be seen, unless purely by chance, since the correlations between the covariates and EQ-5D were not accounted for in the data simulation. Thus, if the IPD were available, simulation of the data would not be necessary and a better assessment of the use of external information to inform EQ-5D summary estimate at each EDSS score would be viable. Although EQ-5D utility at each EDSS score are estimated using a simulated dataset in this project, the simulated data were comparable to the data from

the study population. Hence, the increase in precision for the EQ-5D estimates would also be expected if the original data from the study were used.

Studies for the construction of prior distributions for the regression coefficients of the Bayesian linear regression models were identified using solely PubMed. As other databases such as EMBASE and Cochrane Library were not used to search for potential studies to form the external dataset, it is possible that some studies were not identified for the construction of the prior distributions. Hence using a single database could exacerbate publication bias. Besides publication bias, another source of potential bias from using PubMed is language bias as PubMed do not contain an extensive collection of non-English language journal articles. Additional databases such as LILACS (Latin American and Caribbean Health Sciences Literature) that contain collections of non-English language publications could potentially help to minimize language bias induced when using solely PubMed.

As EQ-5D and EDSS are reported using means and standard deviations in most articles on MS except for articles that investigate the relationship between them, both EQ-5D and EDSS extracted from these articles (that form the external evidence for a Bayesian analysis) had to be analysed on the continuous scale. Hence, a Bayesian linear regression analysis with EDSS as a categorical variable is not possible in this project. Looking ahead, it may be possible to perform Bayesian linear regression analysis with EDSS as a categorical variable (incorporating external evidence) if future articles in MS are to report patients' EQ-5D summary statistics for each category of EDSS.

Although the increase in precision of the EQ-5D estimate using Bayesian evidence synthesis (incorporating informative prior distributions) in this project is small, the same statistical methodology when applied to a different disease area may provide greater gain in precision of its estimates of interest, depending on the amount of evidence available as well as the heterogeneity of the evidence sources.

Although utility data were collected as part of the RCTs (TRANSFORMS and FREEDOMS) used in the manufacturer's submission for the health technology appraisal, these data were not used in the cost-effectiveness analysis. In place of these data, EQ-5D utility values were estimated using the mapping model of EQ-5D and EDSS developed by Orme and colleagues. The methodology described in this thesis

can potentially be applied to include data from Orme's study as part of the external source of evidence to construct the informative prior distributions when the utility data from the RCTs (TRANSFORMS and FREEDOMS) were made available in the future.

In conclusion, this project shows how Bayesian evidence synthesis can be employed to make use of all available evidence to inform utility estimates, which can then be used in cost-effectiveness analysis of disease-modifying therapies in MS to inform economic decision making. Similarly, the same statistical methodologies can be applied to other disease area to inform outcome estimates when external sources of evidence are available.

6 Discussions and Conclusions

6.1 Thesis Overview

This thesis considers three areas of methodological and presentational challenges related to clinical effectiveness and cost-effectiveness analysis in HTA and develops novel approaches to address these challenges. The three methodological and presentational challenges identified in this thesis are:

1. There is no established standardised presentational approaches for reporting NMA results in HTA
2. Data for the full set of parameters required to specify a multi-state model for economic modelling may not always be available
3. Mapping of a disease specific health effect to a generic measure of health effect in cost-effectiveness evaluation is often performed using estimates from a single study

The first chapter of this thesis outlined the aims and provided an introduction to the methodological and presentational challenges that are explored and answered in this thesis. General statistical methodologies and software are described in Chapter 2; with specific, and further statistical methods unique to addressing each of the challenges identified, are presented in their respective Chapters 3, 4 or 5.

Chapter 3, 4 and 5 are self-contained chapters addressing challenge 1, 2 and 3 listed above respectively. Each chapter starts with an introduction to the background of the research question, and where appropriate presents the motivating case study that forms the example for illustrating the methods developed. It then follows with a methods section that describes the data evidence and statistical methodologies applied for the analysis. Schematic diagrams presenting an overview of the data evidence, methods and procedures involved in the development of the applied methodologies for answering the identified issues in HTA are also included.

Results in the form of recommendations, quantitative estimates, graphical tools and software are presented. The chapters then conclude with a discussion section, which

interprets and explains the results to answer the research question specific to the chapter. It also justifies the data evidence and approaches used, discusses the advantages and limitations of the applied methodologies developed and provides suggestions for further work.

As detailed discussions of the findings, using different case studies to highlight and address the challenges identified, are presented in the discussion section of their respective chapters, this chapter seeks to provide a more general overview. In particular it highlights the strengths and limitations of the approaches proposed and implemented in this thesis. It also discusses potential further extensions to these methods from a broader perspective.

6.2 Strengths and limitations of this work

The strengths and limitations of the graphical tools, and evidence synthesis approaches, developed in this thesis to address the challenges identified are discussed in Sections 6.2.1 and 6.2.2 respectively. Strengths and limitations relating to other methodological work that is unique to the specific HTA projects of Chapters 4 and 5 are discussed in the discussion section of their respective chapters. These include problems with the reconstruction of individual patient data (IPD) for survival outcomes and specification of transition probabilities for cost-effectiveness analysis in Chapter 4, and simulation of IPD for Bayesian linear regression analysis in Chapter 5.

6.2.1 Graphical tools

A review that summarises how NMA data, analysis methods and results are presented in UK HTA reports revealed that there is no standard tabular and/or graphical format for the presentation of NMA evidence structure or results. Results were mainly presented in tabular format with only a few also reporting summary forest plots. This is probably due to the fact that most of the reviewed HTAs implemented the NMA in WinBUGS which has limited graphical functionality.

After the completion of my review, two published systematic reviews of the international literature (Coleman et al., 2012, Bafeta et al., 2014) were in broad

agreement with our findings. Thus, while the sample of NMAs in my review was limited to UK HTA reports, it is reassuring that reviews with a wider perspective obtained similar findings. This further strengthens the need for both additional guidance and improved presentational tools for reporting NMA results to aid their interpretability.

Recommendations regarding the presentation of NMA analyses are made. However, it should be acknowledged that no recommendation was developed for the reporting of issues such as variable study quality, subgroup analysis or inconsistency analysis, due to the infancy of this methodology.

Given the increasing popularity and use of NMA, there remains no standardised presentational tool for their reporting. In this thesis, graphical tools to aid clear presentation and thereby to facilitate interpretation of NMA results were developed to address this challenge. Three graphical tools, namely SFP Matrix, SFP Table and Median Rank Chart were developed. SFP Matrix and SFP Table provide a comprehensive presentation of the important NMA and PWMA results displayed on a single plot. These plots not only enable easy comparison of NMA and PWMA results but also assist in reducing the number of tables and/or figures required for all relevant results to be presented in the main text of a journal article where space is often limited. The Median Rank Chart complements the SFP Matrix or the SFP Table by providing a visual summary of each intervention's median ranking within the network of interest; thus enabling decision makers to easily identify the "top-ranking" intervention(s) in terms of effectiveness.

As the majority of HTAs reviewed implemented the NMA analysis by calling the WinBUGS software from the R2WinBUGS library, the graphical tools were also developed in R. Hence, users familiar with R can readily modify the NMA WinBUGS model to allow for further extensions such as the use of a hierarchical model whereby the similarity of treatment effects within the same class of treatments is assumed (Owen et al., 2015, Warren et al., 2014). Although the graphical tools were developed for the presentation of a single outcome, potentially they can be extended to describe multiple outcome situations.

6.2.2 Evidence synthesis

Evidence synthesis using meta-analysis is widely used in clinical research to understand the results of any study in the context of all other relevant and available evidence. Average values of outcomes of interest (or average treatment effects on those outcomes) have commonly been obtained by a meta-analysis of all relevant published evidence. However, evidence synthesis techniques are applied in a number of different ways to address methodological challenges in HTA identified in this thesis.

In Chapter 4, when data for the appropriate specification of an economic model for cost-effectiveness evaluation is not available, bivariate random-effects meta-analysis (BRMA) was used. This predicts the PFS HR, when PFS is not reported, for a specific randomised controlled trial (RCT) in mHRPC, by jointly modelling OS and PFS data in a Bayesian meta-analytic framework incorporating informative prior distributions from available published evidence on both the within- and between-study correlations between PFS and OS. Indirect comparison meta-analysis (ICMA) was also utilised for the estimation of relative clinical effectiveness between two interventions when no head-to-head RCT exists. In Chapter 5, Bayesian BRMA was used to jointly model EQ-5D and EDSS in multiple sclerosis for the construction of informative prior distributions for the regression coefficients to inform the subsequent estimation of EQ-5D using Bayesian regression models.

When using the BRMA to jointly model two correlated endpoints (such as the PFS and OS in mHRPC presented in Chapter 4; EDSS and EQ-5D for MS presented in Chapter 5), an important item of information for the bivariate evidence synthesis is the within-study correlation between the two endpoints. As is the case for our two projects, the within-study correlation for the studies included in the evidence synthesis is often not available. To overcome this limitation, where the within-study correlations cannot be directly obtained from the summary data and IPD are not available, informative prior distributions for the within-study correlations were constructed using correlation data from published external sources of evidence.

Informative prior distributions for the between-study correlations can also be constructed using the same (or different) set of external evidence although it is acknowledged that this correlation can be directly estimated using the summary data of the studies included in the BRMA. Hence, to perform a fully Bayesian analysis that

incorporate external information, informative prior distributions for both the within-study and between-study correlations between PFS and OS were constructed in the mHRPC project (Chapter 4). However, in the MS project (Chapter 5), BRMA in the form of product normal with polynomial terms was applied to external evidence to produce posterior distributions. These in turn served as informative prior distributions for the coefficients in a subsequent multi-polynomial linear regression model.

Results from the two projects showed that when using BRMA to jointly model two outcomes, the use of external evidence in the form of informative priors for the within-study and between-study correlations between the two outcomes results in increased precision of the estimates as compared to the case when non-informative priors were used. However, this is not always the case. In the mHRPC project, when the predicted PFS HR of D+P versus M+P, obtained from BRMA using informative priors for the between-study correlation and non-informative priors for the within-study correlation, were subsequently used in an ICMA to estimate the corresponding PFS HRs for D+P versus P, the PFS HR estimate had lower precision compared to that obtained using the predicted PFS from BRMA with the non-informative priors. Thus, although the use of informative priors generally results in increased precision of the estimates, they can sometimes result in increased uncertainty. This highlights that the estimates obtained encompassing informative prior distributions in a Bayesian meta-analysis also depend upon whether there is an associated decrease or increase in heterogeneity especially for random-effects models.

However, as with any evidence synthesis the precision of the estimate depends on the amount of evidence available as well as the heterogeneity of the evidence sources. Although the focus of this thesis has been on the use of evidence technique for estimating quantitative results and little has been done to evaluate the quality of the studies used in the meta-analysis performed in this thesis, it is acknowledged that the quality of the evidence supporting the analysis is equally important.

This thesis demonstrates how evidence synthesis methodologies can be applied for the purpose of outcome prediction or the construction of informative prior distributions for a Bayesian analysis, in addition to the traditional pooling of data for obtaining the average values of the outcomes of interest. Besides that, fully Bayesian analyses that utilise external information in the form of informative prior distributions are presented.

Within the Bayesian framework, the approaches proposed also allow sets of evidence to be synthesised sequentially using various evidence synthesis techniques to inform the final estimate of interest.

6.3 Further work

Potential extensions to the graphical tools and evidence synthesis approaches are outlined in sections 6.3.1 and 6.3.2 respectively.

6.3.1 Graphical tools

The presentational challenges of reporting quantitative NMA results in HTA is the first of the three challenges addressed in this thesis. Although the focus has been on quantitative evidence synthesis results and methods, the quality of the evidence supporting the analysis is equally important to enable appropriate reporting and interpretation of the results. However, there was no clear guidance for evaluating the quality of the evidence for NMA such as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group and PRISMA statement for traditional meta-analysis during the development of the graphical tools in this thesis.

Approaches to assess the quality of evidence from NMA based on the methodology developed by the GRADE Working Group was proposed by Salanti and colleagues (Salanti et al., 2014). Following that, the GRADE Working Group presented their four-step approach to rate the quality of evidence for NMA estimates (Puhan et al., 2014) and the PRISMA extension statement for reporting systematic reviews used for NMA (Hutton et al., 2015) was published in 2015.

Concurrent to the development of the graphical tools, Chaimani and colleagues developed graphical tools for frequentist NMA performed in STATA (Chaimani et al., 2013). Although the graphical tools in this thesis are developed using the R software and use the output from WinBUGS where a Bayesian NMA was carried out, the graphics can also be used for presenting NMA performed under the frequentist framework with appropriate modification of the R code.

With the increasing popularity of NMA, recent methodological work has also lead to the development of multivariate NMA for multiple correlated binary (Efthimiou et al., 2014, Achana et al., 2014) and continuous (Hong et al., 2013) outcomes. Efthimiou and colleagues further extended their work to allow for the joint synthesis of continuous, binary, time-to-event or mixed outcomes (Efthimiou et al., 2015). Multivariate NMA requires estimates of the within-study and between-study correlations of the multiple outcomes under evaluation. It is therefore imperative to present the estimates of the outcomes and correlations appropriately to ensure clear interpretation of the multivariate NMA results.

In summary, areas for further work include: Incorporating enhanced flexibility for printing graphics of NMA results from various software and analysis framework using the published R code; Packaging the R code into a R-library or stand-alone software to allow greater dissimilation of the graphical tools; Incorporating the presentational approaches for reporting the quality of NMA evidence; and adapt the graphics for the presentation of multiple outcomes NMA.

6.3.2 Evidence synthesis

The next two challenges, concerning economic model and utility for cost-effectiveness analysis in HTA, were addressed utilising Bayesian BRMA that jointly synthesises correlated outcomes. Other evidence synthesis methods, such as NMA and ICMA, were also used in conjunction with the BRMA to achieve the aims of this thesis. All these evidence syntheses were performed using aggregated trial data as no patient-level data was sourced from investigators of the studies. One area for further work would be to perform the BRMA using (where available) IPD from the studies to enable the direct incorporation of the within-study correlation between the multiple outcomes into the model.

Individual patient data for PFS and OS were reconstructed for the mHRPC RCTs in Chapter 4 using Kaplan-Meier curves reported in the corresponding publications. However, the OS and PFS data so reconstructed were regarded as two independent sets of time-to-event data that are not correlated. To address this limitation, informative prior distributions for the within-study correlation between OS and PFS LHRs were

constructed using external trial data. Had the IPD been available it would have enabled the within-study correlation between PFS and OS LHRs to be estimated using bootstrapping. This is particularly important when the aim is to predict, for example, the PFS HR from the OS HR of a trial by jointly modelling OS and PFS using a BRMA model. The correlation reported by Halabi and colleagues was between OS and PFS and was applied as a crude approximation to construct the informative prior distributions for the within-study correlation between OS and PFS LHRs used in the BRMA model. Further research could also be carried out to derive the approximate relationship between the correlation between OS and PFS LHRs (needed in the BRMA model) and the Kendall's statistics for the association between OS and PFS (available from Halabi et al.). Such derivation could be conducted in a similar manner as derived by Wei and Higgins (Wei and Higgins, 2013), for example, for the relationship between the correlation between log odds ratios (LORs) on two outcomes and correlation between probabilities of event on those two outcomes or as in Bujkiewicz et al. where correlation between LOR and log rate ratio was expressed in terms of the regression-based Prentice's criteria for association between surrogate endpoints (Bujkiewicz et al., 2015a). Recent methodologies proposed by Boucher and colleagues (Boucher et al., 2015) that use an illness-death model framework to simulate OS and PFS IPD when only summary data are available can also be applied.

Similarly, IPD would allow the within-study correlation between baseline quality of life as assessed by EDSS and EQ-5D to be directly estimated for the BRMA model used to construct the informative prior distributions to inform the utility estimates for the multiple sclerosis study considered in Chapter 5. Methods for multivariate meta-analysis using IPD, both in the frequentist and Bayesian framework, have been proposed by Riley and colleagues (Riley et al., 2015) for multiple outcome data types and can be applied to extend the analysis in this thesis.

Another potential area for further work is to extend the BRMA model used in the prediction of PFS HR for trial TAX 327 to a bivariate random-effects network meta-analysis model, including all RCTs containing docetaxel (D) for the first-line chemotherapy treatment of men with mHRPC. Although in the last few years, new drugs such as cabazitaxel, abiraterone and enzalutamide have been developed, they are primarily trialled in men who had previously been treated with a D-containing

chemotherapy regimen. Hence, to date, D with either prednisone or prednisolone (P) in the combination (D+P) remain the NICE recommended treatment options for men with mHRPC based on the health technology appraisal published in 2007. Since then, a number of trials containing D+P or D with other combination drugs have been reported. These can be included to form a network of trials for the prediction of PFS HR using bivariate NMA for trial TAX 327, which compared D+P versus mitoxantrone plus prednisone, M+P. More recent methodologies for multivariate NMA developed by Ades and Efthimiou and their colleagues (Ades et al., 2010, Efthimiou et al., 2014, Efthimiou et al., 2015) can potentially be applied to the trial TAX 327 situation. Given the possible network of trials that may be identified, it is also acknowledged that univariable NMA for PFS would also be possible. Nevertheless, a multivariate NMA incorporating the correlations between PFS and OS will not only enable joint inference to be made between the correlated outcomes but also provide improved value over separate univariable NMA as evidence synthesis is performed in one coherent analysis.

6.4 Conclusions

Methodologies for evidence synthesis have evolved over the years, from the traditional meta-analysis that compares a single outcome between two interventions to multivariate meta-analysis and NMA that allow the evaluation of multiple outcomes and interventions. Although the BRMA model was used in the mHRPC project for prediction of PFS and in the MS project for the construction of prior distributions for the regression coefficients defining the association between EDSS and EQ-5D, it can also be applied in the context of any correlated outcomes and in particular surrogate endpoints (Bujkiewicz et al., 2015b). With its potential statistical advantages and application in medical research, this thesis has demonstrated, with the use of case examples from HTA reports, approaches where BRMA can be applied in a number of different ways to address methodological challenges in HTA.

Due to the inherent complexity of these advanced evidence synthesis methods, it is imperative that there is consistency and clarity in the presentation of results from these analyses to ease their ready interpretability. This is especially important for NMA where multiple interventions are evaluated. The review conducted in this thesis of existing methods of presenting NMA results in HTA revealed that there was great

variability in reporting styles and highlighted the need for additional guidance and presentational tools for reporting NMA results. Novel graphical tools were developed in this thesis with the aim to improve existing reporting methods. Ultimately, the hope is that the graphical tools developed as part of this thesis will be disseminated widely and recommended in updated guidance setting the standards for future HTA reports.

Appendices

Appendix A : Data extraction form for reviewing HTA reports

DATA EXTRACTION FORM

A. General

A1 HTA report appraisal type
□ NICE □ Non-NICE

A2 Method of evidence synthesis
□ Indirect Comparison □ Mixed Treatment Comparison

A3 Section where IC/MTC results were reported
□ Clinical Effectiveness □ Cost Effectiveness
□ Separate IC/MTC □ Other _____

A4 Was IC/MTC also performed for non-clinical effectiveness endpoints such as adverse events?
□ Yes, endpoint _____
□ No

A5 Where an economic decision model is developed as part of the report, are the MTC results used in the decision model?
□ Yes □ No

B. Presentation of IC/MTC data

B1 Number of interventions evaluated in the HTA report using IC/MTC
_____

B2 Trial studies presented using a Network Table?
□ Yes □ No

B3 Intervention relationships presented using a Network Diagram
□ Yes □ No

C. Presentation of IC/MTC synthesis model and its implementation

C1 Is a Bayesian evidence synthesis approach used? If Yes, answer all following questions
□ Yes □ No

C2 Was Prior used for the Bayesian evidence synthesis presented?
□ Yes □ No

C3 Has sensitivity analysis of Prior used for the Bayesian evidence synthesis been performed and presented?
□ Yes □ No

C4 Was check for inconsistency in the MTC analysis performed and presented?
□ Yes □ No

C5 Was convergence check performed?
□ Yes □ No

C6 Was WinBUGS used for the evidence synthesis analyses?
□ Yes □ No

D. Presentation of IC/MTC results

D1 Was the IC/MTC results presented using any of the following tools?
□ Matrix Table □ Table (ratio) □ Table (rate)
□ Summary forest plot □ Text

D2 Was Probability Best statistics presented?
□ Yes □ No

How was it presented?
□ Table □ Figure □ Text

D3 Was Ranking statistics presented?
□ Yes □ No

How was it presented?
□ Table □ Figure □ Text

END

Appendix B : Evaluation form for graphical tools

Evaluation of Graphical Presentation Displays for MTC – Feedback Form

1. Have you ever been involved in a Mixed Treatment Comparison (MTC)? Yes / No

2. Overall, in terms of (i) Content & (ii) Clarity, how would you rate the 4 Summary Forest Plots (SFPs)?

	CONTENT										CLARITY									
	Inappropriate content									Appropriate Content	Difficult to Understand									Excellent Clarity
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
SFP Matrix	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
SFP Table	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
SFP Pie (version 1)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
SFP Pie (version 2)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

3. Which would you consider using in your MTC? Please select as many as you wish.
   ☐ SFP Matrix ☐ SFP Table ☐ SFP Pie (v1) ☐ SFP Pie (v2)

4. Compared to current practice, do these plots improve / add value to the presentation of MTC? Yes / No

5. Which other important information would you like to be added in either these or future plots?
   ☐ Absolute Effect
   ☐ Ranks
   ☐ Inconsistency
   ☐ Others, please state: _____

6. Any Other Comments

Please return the form or email it back to Sze Huey at sht10@leicester.ac.uk.
Thank you!

Appendix C : Presentation for NMA with outcome on the continuous scale

The graphical tools proposed can be extended to other outcome measurements such as mean difference and hazard ratio. Examples of its use for continuous outcome are shown in Figure C1 to Figure C3. The dataset used for illustrating can be found in the NICE TSD 2 example 5 on Parkinson's Disease.

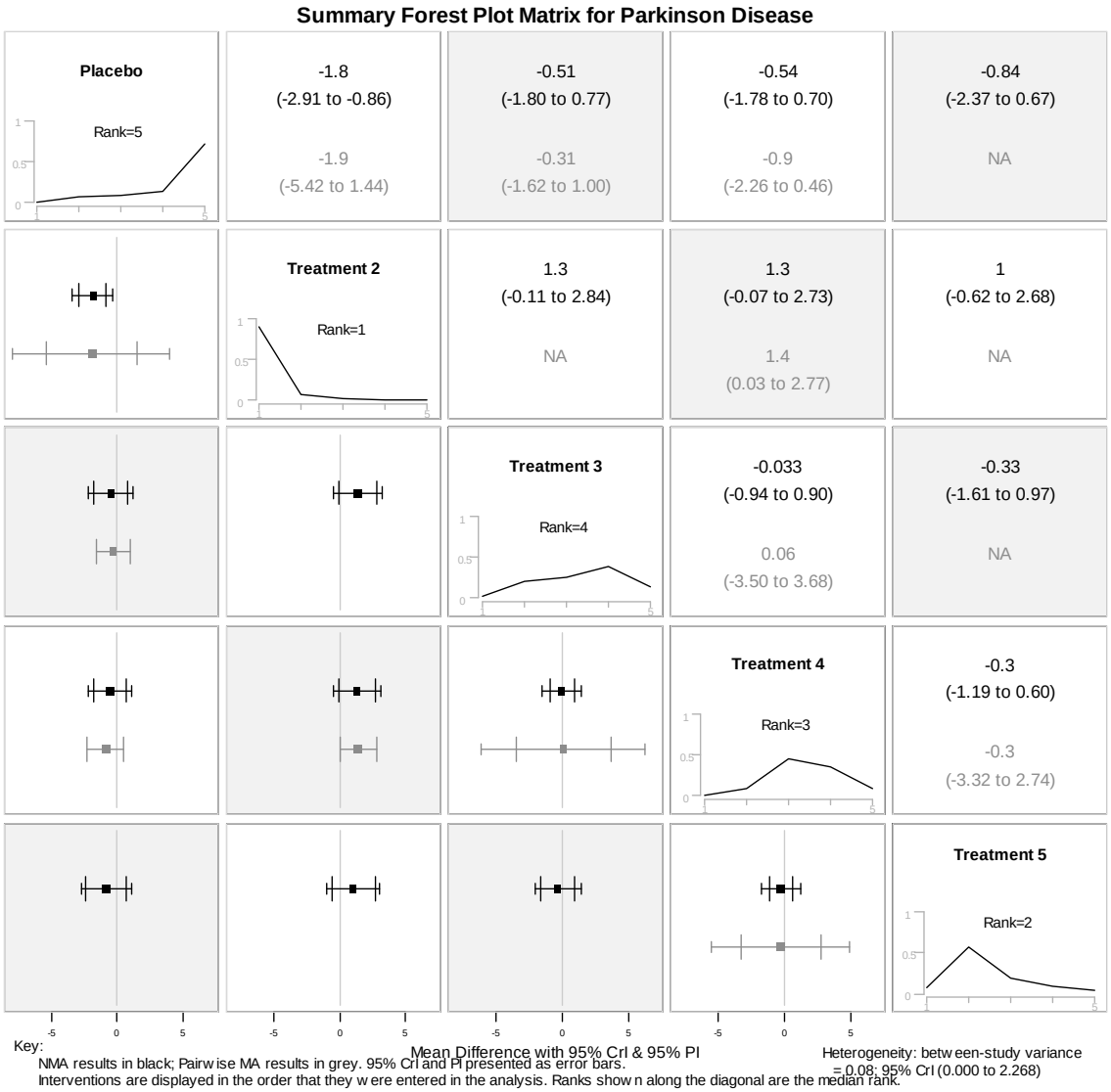

Figure C1: Summary Forest Plot Matrix presenting median rank of all interventions

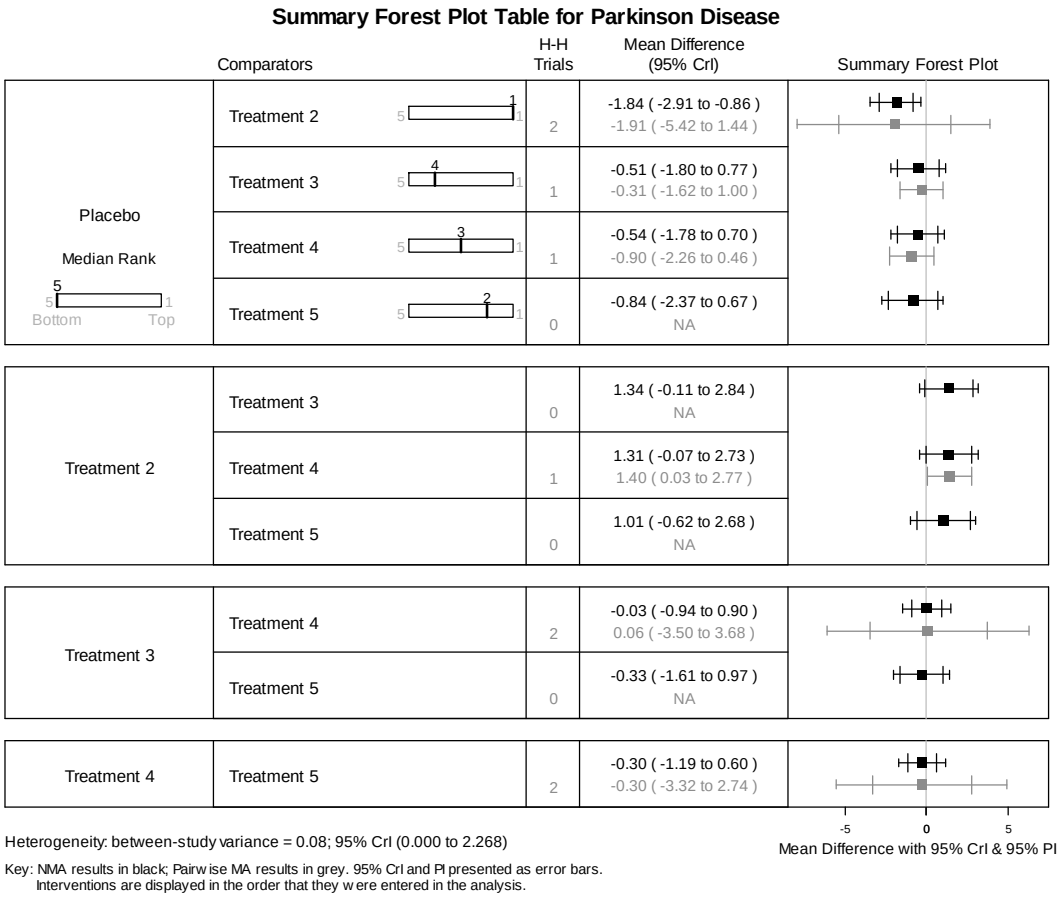

Figure C2: Summary Forest Plot Table presenting median rank of all interventions

Median Rank Chart for Parkinson Disease

Rank	Intervention
1	Treatment 2
2	Treatment 5
3	Treatment 4
4	Treatment 3
5	Placebo

Figure C3: Median Rank Chart showing the ranking of all interventions in terms of efficacy

Appendix D : Display options for SFP Matrix and SFP Table

SFP Matrix and SFP Table in its simplest format and without footnotes are shown in Figure D1.1 and D2.1. Additional display options are available for the SFP Matrix and SFP Table that allows the reporting of mean rank, SUCRA percentage and probability best statistics instead of median rank, as well as the exclusion of 95% prediction intervals. Figure D1.2 shows the SFP Matrix with the presentation of the SUCRA percentage using cumulative rankograms while Appendix Figure D2.2 presents the same information using the SFP Table graph.

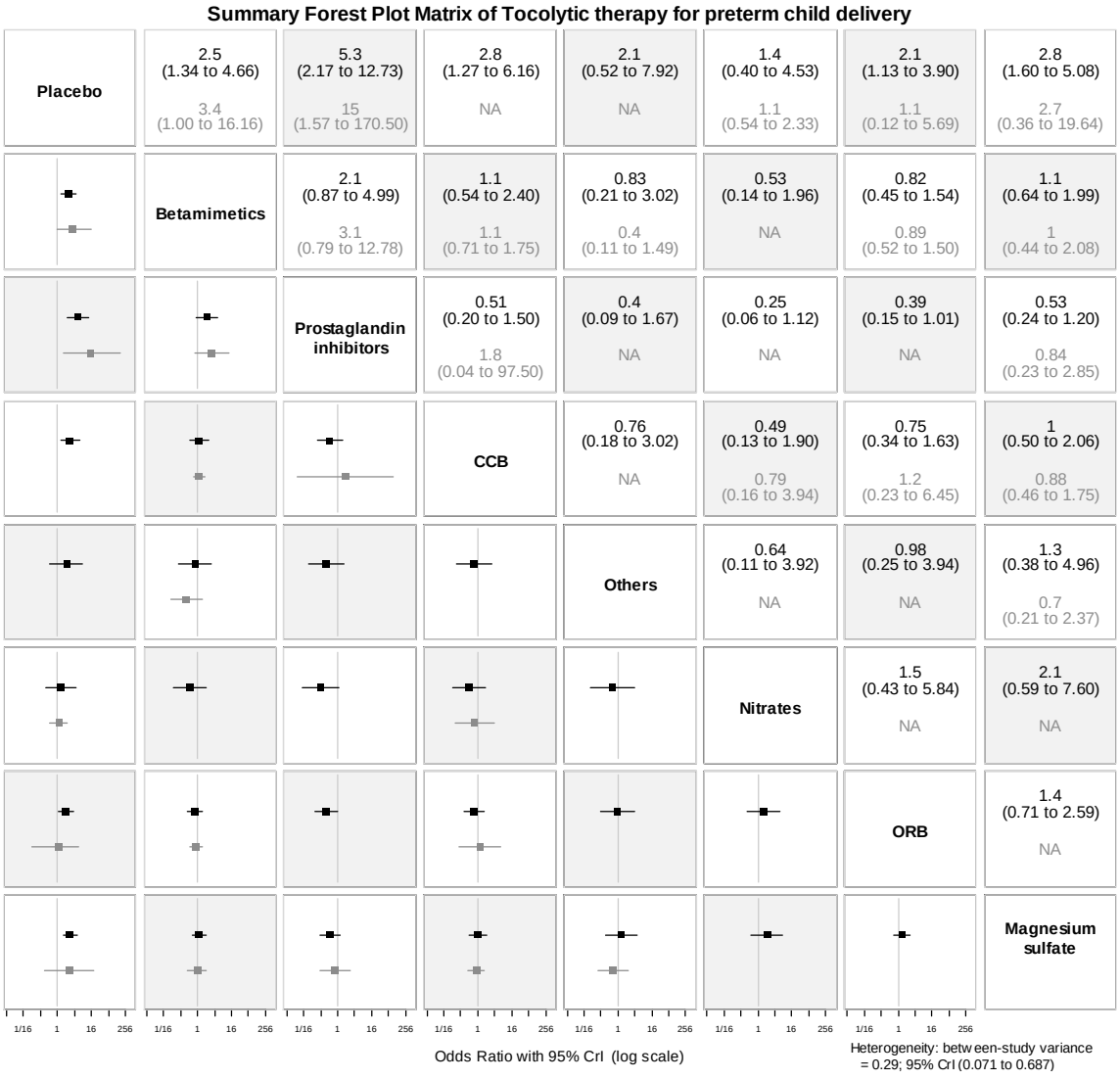

Figure D1.1: Summary Forest Plot Matrix in its simplest format with no sorting and footnote

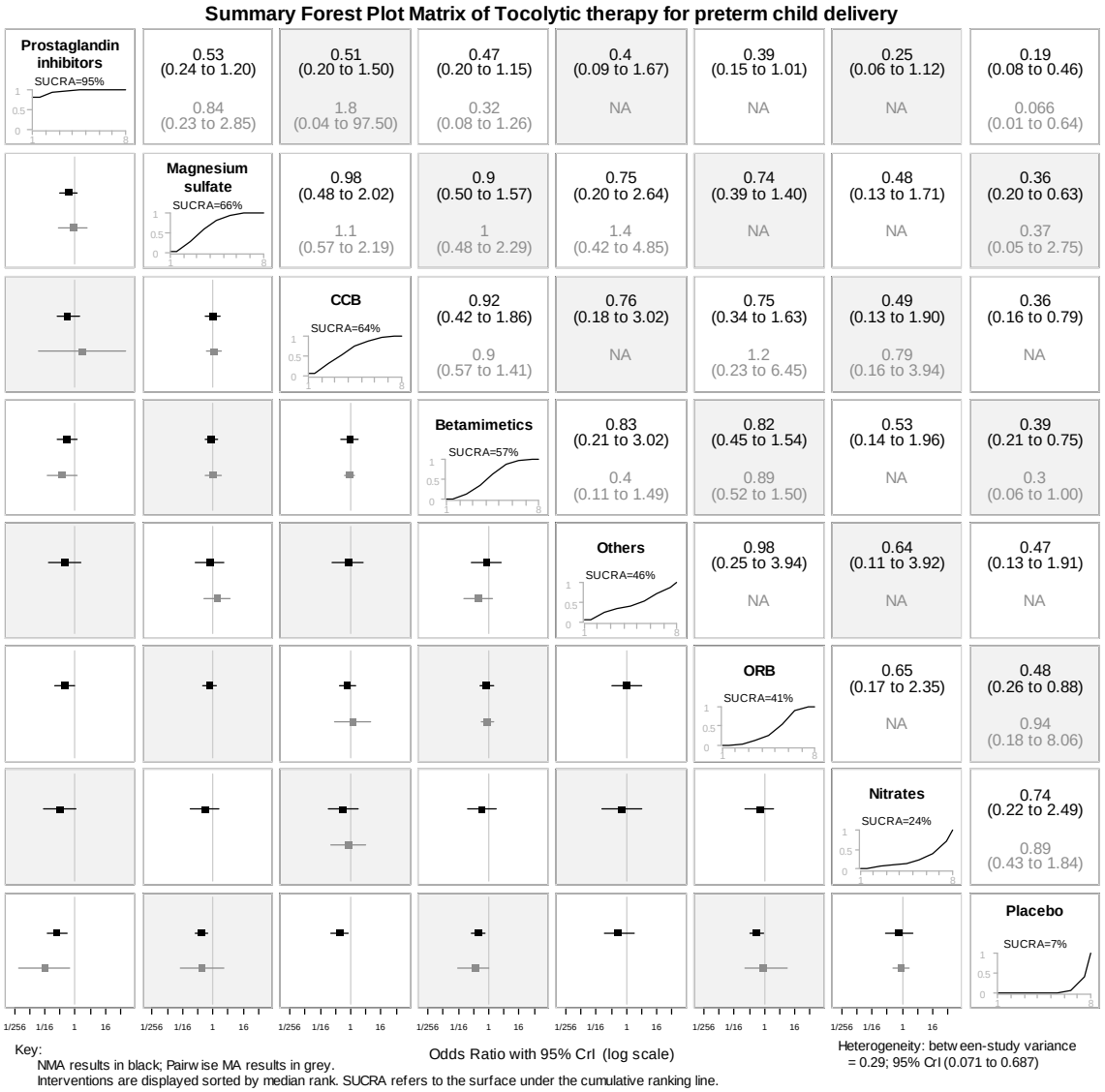

Figure D1.2: Summary Forest Plot Matrix presenting SUCRA percentage of all interventions along the diagonal

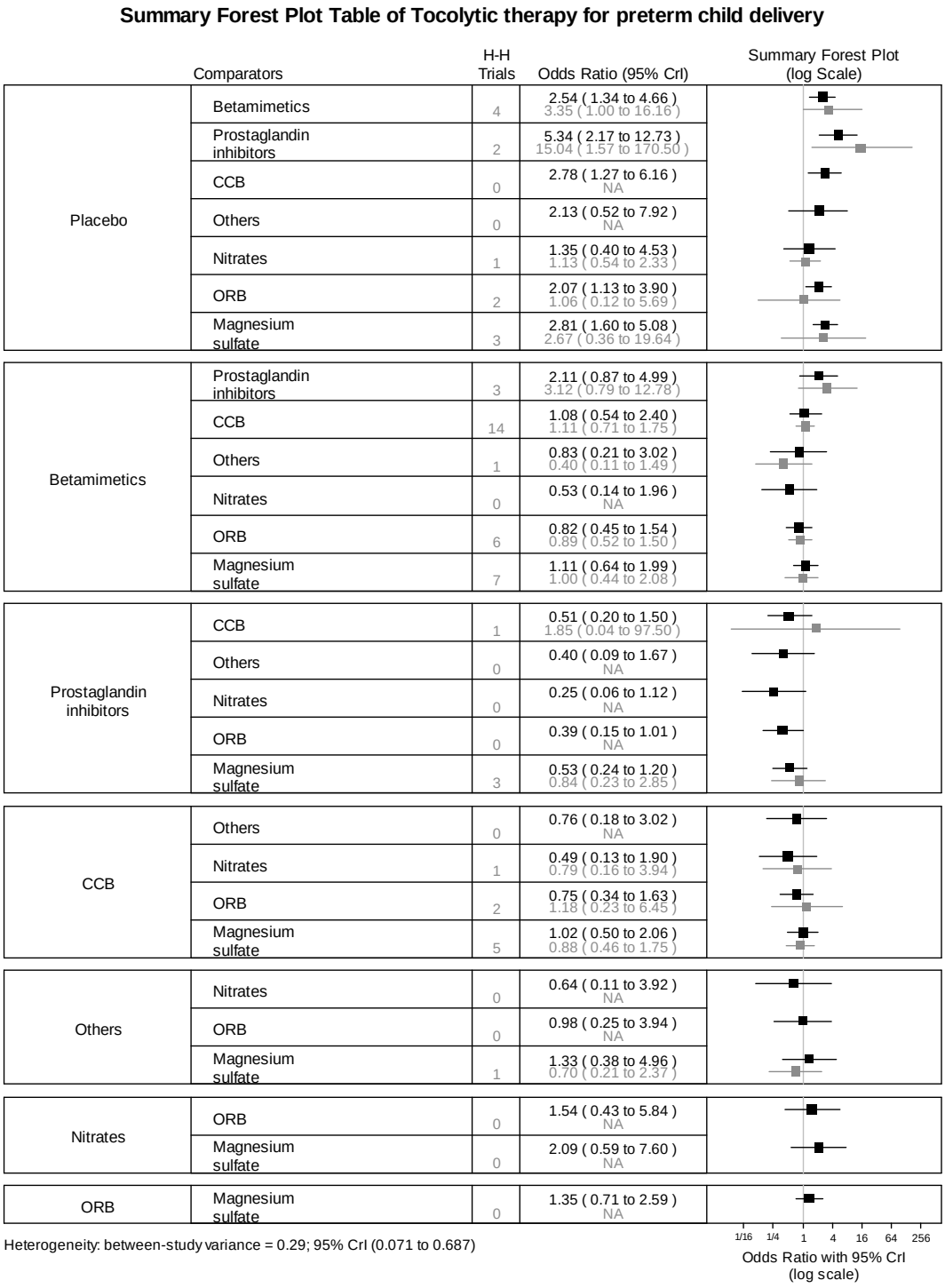

Figure D2.1: Summary Forest Plot Table in its simplest format with no sorting and footnote

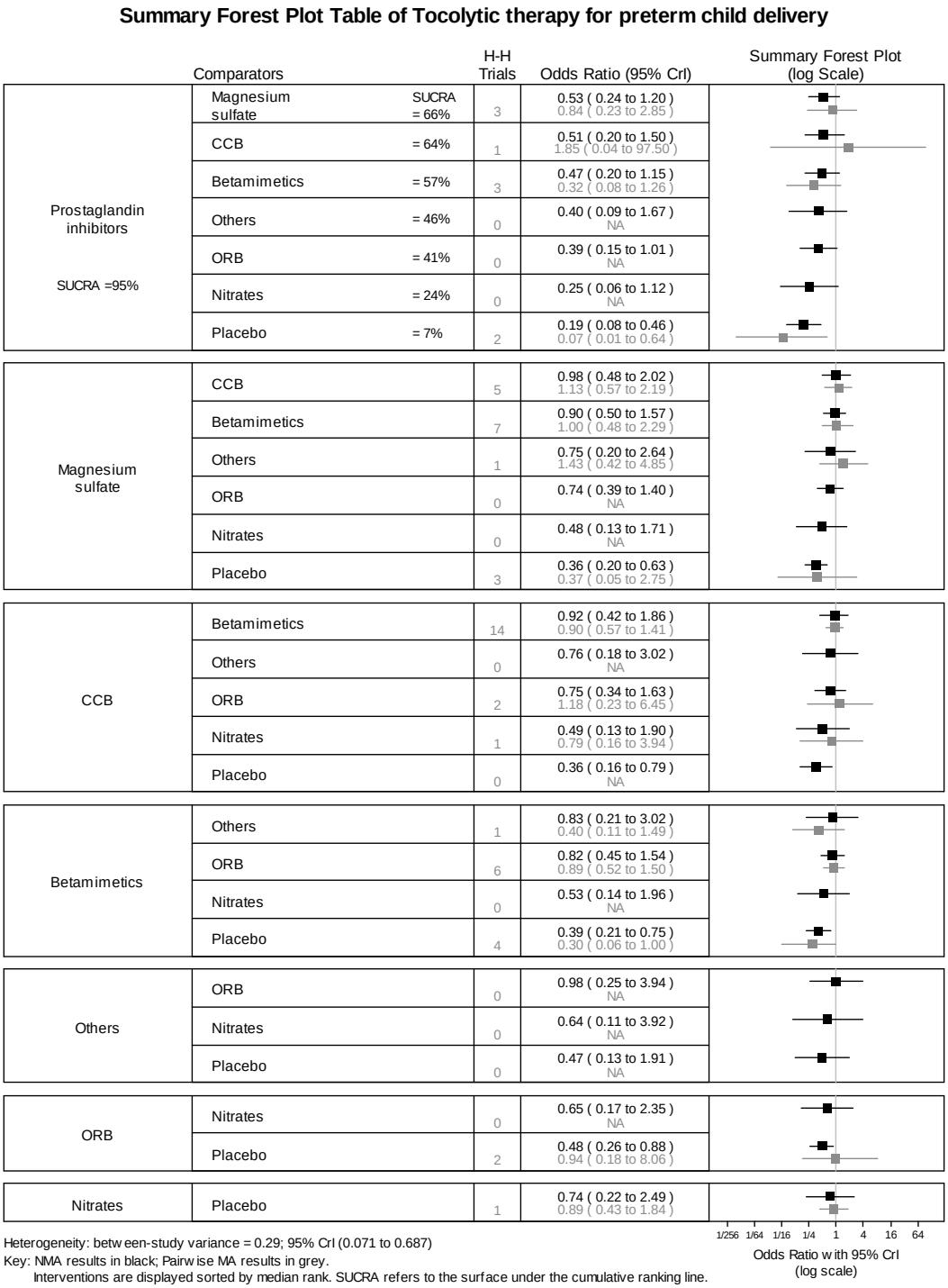

Figure D2.2: Summary Forest Plot Table presenting SUCRA percentage of all interventions

Appendix E : Program codes for Chapter 4

Box E1: WinBUGS codes for BRMA

Case 1: both ρ_{BS} and ρ_{WI} non-informative
Case 2: ρ_{BS} non-informative; ρ_{WI} informative
Case 3: ρ_{BS} informative; ρ_{WI} non-informative
Case 4: both ρ_{BS} and ρ_{WI} informative

```
model{

for (k in 1:num){
        var[k,1]~dnorm(0,h1)I(0,)
        var[k,2]~dnorm(0,h2)I(0,)
        }
        h1~dgamma(1.0,0.01)
        h2~dgamma(1.0,0.01)

#within study precision matrix

  # Case 2 and Case 4; ρ_WI informative
  fisher.corr_w~dnorm(zw,prec.zw)
  prec.zw<-1/pow(sd.zw,2)

for (i in 1:num) {

        # Case 1 and 3; ρ_WI non-informative
        rho_w[i]~dunif(-1,1)
        # Case 2 and Case 4; ρ_WI informative
        rho_w[i]<-(exp(2*fisher.corr_w)-1)/(exp(2*fisher.corr_w)+1)
        prec_w[i,1:2,1:2]<-inverse(delta[i,1:2,1:2])
#covariance matrix for the j-th study
        delta[i,1,1]<-var[i,1]/n[i,1]
        delta[i,2,2]<-var[i,2]/n[i,2]
        delta[i,1,2]<-sqrt(delta[i,1,1])*sqrt(delta[i,2,2])*rho_w[i]
        delta[i,2,1]<-sqrt(delta[i,1,1])*sqrt(delta[i,2,2])*rho_w[i]
}

# Random-effects model
for (i in 1:num) {
        Y[i,1:2]~dmnorm(mu[i,1:2], prec_w[i,1:2,1:2])
#  product normal formulation for the between study part:
        mu[i,1]~dnorm(los,prec_los)
        mu[i,2]~dnorm(lpfs[i],prec_lpfs)
        lpfs[i]<-lambda0+lambda1*(mu[i,1] - mean(mu[,1]))
        }
los~dnorm(0.0, 0.01)

  # Case 1 and Case 2; ρ_BS non-informative

  gam_los~dnorm(0,100)I(0,)
  gam_lpfs~dnorm(0,100)I(0,)
  gam_los.sq<-gam_los*gam_los
  gam_lpfs.sq<-gam_lpfs*gam_lpfs
  prec_los<-1/gam_los.sq
  prec_lpfs<-1/gam_lpfs.sq

  # prior between study correlations:
  corr.lpfs.los~dunif(-1,1)
```

```
#priors for lambda coefficients
lambda0~dnorm(0.0, 1.0E-3)
# implied prior for lambda coefficients
lambda1<-(gam_lpfs/gam_los)*(corr.lpfs.los/sqrt(1-corr.lpfs.los*corr.lpfs.los))

# Fisher Transformation of correlations:
z.fisher<-(1/2) * log((1+corr.lpfs.los)/(1-corr.lpfs.los))
```

```
# Case 3 and Case 4; ρBS informative

fisher.corr~dnorm(z, prec.z)
prec.z<-1/pow(sd.z,2)
corr.lpfs.los<-(exp(2*fisher.corr)-1)/(exp(2*fisher.corr)+1)

# informative between study heterogeneity:
sd.los~dnorm(0,100)I(0,)
sd.lpfs~dnorm(0,100)I(0,)
sd.los.sq<-sd.los*sd.los
sd.lpfs.sq<-sd.lpfs*sd.lpfs
prec.los<-1/sd.los.sq
prec_lpfs<-1/(sd.lpfs.sq - sd.los.sq*pow(lambda1,2))

#priors for lambda coefficients
lambda0~dnorm(0.0, 0.01)
#lambda1<-corr.lpfs.los*tau.lpfs/tau.los==>
lambda1<-corr.lpfs.los*sd.lpfs/sd.los
```

```
# estimates:
mean.log.os<-los
mean.log.pfs<-lambda0

sd.log.os<-gam_los
sd.log.pfs<-sqrt(gam_lpfs.sq+gam_los.sq*pow(lambda1,2))

mean.os<-exp(mean.log.os)
mean.pfs<-exp(mean.log.pfs)

#predicted PFS for TAX327; trial index =4
new.log.pfs<-Y[4,2]
new.pfs<-exp(new.log.pfs)
sd.new.log.pfs<-sqrt(delta[4,2,2])
}
```

Box E2: WinBUGS codes for two-state Markov model

```
model{

#lambda == transition probabilities; piX==patient count still in StD,
#dX==patient count WHO <just entered> death state.
#Note that cycle starts at 2 and ends at 181;
#this is to allow patient count matrix to be N at [1] index which represent cycle 0.

#Index for lambda, pi,d,a,hcpi : first is for (1:C, 2:T, 3:PM, 4:PD)
#Transition probabiliities are from TAX327 trial

for(t in 2:Cycles) {

  lambda[1, t,1,2]<-1-exp(w.lambdaC*(pow(((t-1)-1), w.gammaC) - pow((t-1), w.gammaC)))
  lambda[2, t,1,2]<-1-exp(w.lambdaT*(pow(((t-1)-1), w.gammaT) - pow( (t-1), w.gammaT)))
  lambda[3, t,1,2]<-1-exp(hr.os.pm*w.lambdaC*(pow(((t-1)-1), w.gammaC) - pow((t-1), w.gammaC)))
  lambda[4, t,1,2]<-1-exp(hr.os.pd*w.lambdaT*(pow(((t-1)-1), w.gammaT) - pow((t-1), w.gammaT)))

        for(d in 1:4){
                lambda[d,t,1,1]<-1-lambda[d,t,1,2]
                lambda[d,t,2,1]<-0
                lambda[d,t,2,2]<-1

                for(s in 1:S){
                        pi[d,t,s]<-inprod( pi[d,(t-1),], lambda[d,t, ,s])
                }

        #No of deaths that occured in the cycle, not the cumulative number of death in the cycle.
                dpi[d,t]<-pi[d,t,2] - pi[d,(t-1), 2]

                #patient that moved state in each cycle
                api[d,t]<-(pi[d,(t-1), 1] - pi[d,t,1])

                #0.5 to half the paitent count to get half cycle patient counts
                hcpi[d,t,1]<-pi[d,t,1] + 0.5*api[d,t]
                hcpi[d,t,2]<-pi[d,t,2]
        } # end of d=1:4
}# end of t=2:Cycles


for(d in 1:4){
        pi[d,1,1]<-N
        pi[d,1,2]<-0

        dpi[d,1]<-0
        hcpi[d,1,1]<-pi[d,1,1]
        hcpi[d,1,2]<-pi[d,1,2]

} # end of d=1:4

        #(i)Cholesky Decomposition
        Z.inteC~dnorm(0, 1)
        Z.scaleC~dnorm(0, 1)
        Z.inteT~dnorm(0, 1)
        Z.scaleT~dnorm(0, 1)

        #Values taken from Table 29 in HTA report Page 55
        w.inteC<-3.036 + 0.0447*Z.inteC
```

```
w.scaleC<-0.6184 + 0.0079620581655481*Z.inteC + 0.0362350855623601*Z.scaleC

w.inteT<-3.214 + 0.0546*Z.inteT
w.scaleT<-0.6482 + 0.0169413919413919*Z.inteT + 0.040390954916765*Z.scaleT

w.lambdaC<-exp(-w.inteC/w.scaleC)          #Equations from Page 54
w.gammaC<-1/w.scaleC
w.lambdaT<-exp(-w.inteT/w.scaleT)
w.gammaT<-1/w.scaleT



# Direct MA OS HR (P vs M+P) : from HTA report  1/0.99 (1/1.20, 1/0.82)
se.pm<-(log(1/0.82)-log(1/1.20))/(2*1.959964)
prec_lhr.pm<-1/pow(se.pm, 2)
mean.lhr.pm<-log(1/0.99)
lhr.pm~dnorm(mean.lhr.pm, prec_lhr.pm)
hr.os.pm<-exp(lhr.pm)

# Indirect MA OS HR (P vs D+P) : from HTA report  1/0.75 (1/0.99, 1/0.57)
se.pd<-(log(1/0.57)-log(1/0.99))/(2*1.959964)
prec_lhr.pd<-1/pow(se.pd, 2)
mean.lhr.pd<-log(1/0.75)
lhr.pd~dnorm(mean.lhr.pd, prec_lhr.pd)
hr.os.pd<-exp(lhr.pd)


#Costing data
for(ind in 2:7){
    betaC[ind]<-1/betaC.par[ind-1]
        cC.dis[ind]~dgamma(4.0, betaC[ind])
    betaT[ind]<-1/betaT.par[ind-1]
        cT.dis[ind]~dgamma(4.0, betaT[ind])
}
cC.dis[1]<-0.0
cT.dis[1]<-0.0

c.ter[1]~dgamma(4.0, 0.00101467)
c.ter[2]~dgamma(4.0, 0.0011338)

#Data for relative difference factor for costing of the P arm
rd.mparm~dgamma(80.62938, 0.00350856)
rd.parm~dgamma(105.09062, 0.00361896)
rd.factor<-rd.parm/rd.mparm

#(ii)Distribution placed on Drug cost
#d.dose.cycle.C<-22.8
d.ncycle.C~dnorm(5.9, 0.174835)
d.cost.C[1]<-(169.25 + 1.02)*d.ncycle.C          #drug cost of M+P
d.cost.C[2]<-177.46*d.ncycle.C                   #clinic outpatient cost
cDrug.tot[1]<-sum(d.cost.C[])

#d.dose.cycle.T<-142.5
d.ncycle.T~dnorm(7.3, 0.17562282)
d.cost.T[1]<-(1069.50 + 1.02+5.94)*d.ncycle.T          #drug cost of D+P+Dex
d.cost.T[2]<-177.46*d.ncycle.T                   #clinic outpatient cost
cDrug.tot[2]<-sum(d.cost.T[])

#First 5 cycles, no follow up cost
for(t in 2:Cycles){
```

```
		indx[t]<-min(7, max(1,1+round((t-5)/4)))
		c[1,t,1]<-cC.dis[indx[t]]
		c[1,t,2]<-cDead

		c[2,t,1]<-cT.dis[indx[t]]
		c[2,t,2]<-cDead

		cP[t,1]<-1.48	#fixed cost of £1.48 /cycle/patient
		cP[t,2]<-cDead
	}

	#Utilities
	#Data from Sandblom
	#0.538 ±0.077
	se_SD<-(2*hw_SD)/1.96
	u.alpha<-mn_SD*(mn_SD*(1-mn_SD)/pow(se_SD,2)-1)
	u.beta<-mn_SD*(1-mn_SD)/pow(se_SD,2)-1-u.alpha
	uSD~dbeta(u.alpha, u.beta)

	for(t in 1:Cycles){
		u[t,1]<-uSD
		u[t,2]<-uDead
	}

	for(t in 2:13){
		for(d in 1:2){				#note:difference in index max count
			ct[d,t]<-dpi[d,t]*c[d,t,1]		#Follow up cost
			cd[d,t]<-dpi[d,t] *c.ter[d]		#Termination cost
			ut[d,t]<-inprod(hcpi[d,t,], u[t,])
		}
		for(d in 3:4){
			ct.drug[d,t]<-pi[d,t,1]*cP[t,1]	#Using pts at each cycle; only for d=3,4
			ut[d,t]<-inprod(hcpi[d,t,], u[t,])
		}
	}
	for(t in 14:26){
		for(d in 1:2){
			#midpt discount adjustment!
			mc.c[d,t]<-(c[d,t,1]) / pow((1+0.00291667), 4*(round((t-1)/4)))
			ct[d,t]<-dpi[d,t]*mc.c[d,t]
			cd[d,t]<-(dpi[d,t] * c.ter[d]) / pow((1+0.00291667), (t-1))
			ut[d,t]<-inprod(hcpi[d,t,], u[t,]) / pow((1+0.00291667), (t-1))

		}
		for(d in 3:4){
			ct.drug[d,t]<-(pi[d,t,1]*cP[t,1]) / pow((1+0.00291667), (t-13))
			ut[d,t]<-inprod(hcpi[d,t,], u[t,]) / pow((1+0.00291667), (t-1))
		}
	}
	for(t in 27:Cycles){
		for(d in 1:2){
			ct[d,t]<-(dpi[d,t]*c[d,t,1]) / pow((1+0.00291667), (t-1))

			cd[d,t]<-(dpi[d,t] * c.ter[d]) / pow((1+0.00291667), (t-1))
			ut[d,t]<-inprod(hcpi[d,t,], u[t,]) / pow((1+0.00291667), (t-1))
		}
		for(d in 3:4){
			ct.drug[d,t]<-(pi[d,t,1]*cP[t,1]) / pow((1+0.00291667), (t-13))
			ut[d,t]<-inprod(hcpi[d,t,], u[t,]) / pow((1+0.00291667), (t-1))
		}
```

```
        }

        for(d in 1:2){
        mean.2C[d,1]<-sum(ct[d,2:Cycles])/N
        mean.2C[d,2]<-sum(cd[d,2:Cycles])/N
        }

        cDrug.tot[3]<-(sum(ct.drug[3, 2:Cycles]))/N
        cDrug.tot[4]<-(sum(ct.drug[4, 2:Cycles]))/N
        mean.C[1]<-sum(mean.2C[1,1:2]) + cDrug.tot[1]
        mean.C[2]<-sum(mean.2C[2,1:2]) + cDrug.tot[2]
        mean.C[3]<-rd.factor*(sum(mean.2C[1,1:2])) + cDrug.tot[3]
        mean.C[4]<-rd.factor*(sum(mean.2C[1,1:2])) + cDrug.tot[4]

        for(d in 1:4){
                mean.Tm[d]<-sum(pi[d, 2:Cycles, 1]) / N
                mean.U[d]<-sum(ut[d, 2:Cycles])/N/12          #mean utilities for each Tx
        }

        Cost.diff<-mean.C[2] - mean.C[1]
        Util.diff<-mean.U[2] - mean.U[1]
        ICER<-Cost.diff/Util.diff

 for (g in 1:21)  {
        Rc[g]<-(g-1)*5000

        IncNetbenefit[g]<-Rc[g]*Util.diff- Cost.diff
        ProbCE[g]<-step(IncNetbenefit[g])

        NB[3,1,g]<-Rc[g]*mean.U[3]- mean.C[3]
        NB[3,2,g]<-Rc[g]*mean.U[4]- mean.C[4]

        for(k in 1:2){
          NB[1,k,g]<-Rc[g]*mean.U[1]- mean.C[1]
          NB[2,k,g]<-Rc[g]*mean.U[2]- mean.C[2]

        for(j in 1:3){
            pCE[j,k,g]<-equals(rank(NB[,k,g],j),3)
    }
    Prob.C[k,g]<-pCE[1,k,g]
    Prob.T[k,g]<-pCE[2,k,g]
    Prob.P[k,g]<-pCE[3,k,g]
        }

 }

}
```

Box E3: WinBUGS codes for three-state Markov model

```
model{

prec_ost1<-1/(0.5872765*0.5872765)
ost1~dnorm(17.75561, prec_ost1)
prec_dct1<-1/(0.17*0.17)
dct1~dnorm(5.9, prec_dct1)

prec_ost2<-1/(0.6170532*0.6170532)
ost2~dnorm(19.83514, prec_ost2)
prec_dct2<-1/(0.18*0.18)
dct2~dnorm(7.3, prec_dct2)

prec_oslh3<-1/(0.223078192*0.223078192)
oslh3~dnorm(-2.905034262, prec_oslh3)
prec_dclh3<-1/(0.074743477*0.074743477)
dclh3~dnorm(-1.699704398, prec_dclh3)

ost3<-1/exp(oslh3)
dct3<-1/exp(dclh3)

for(t in 2:Cycles) {

 lambda[1,t,1,2]<-1-exp(w.lambdaC*(pow(((t-1)-1), w.gammaC) -  pow( (t-1), w.gammaC)))
 lambda[2,t,1,2]<-1-exp(hr.pfs*w.lambdaC*(pow(((t-1)-1), w.gammaC) -  pow( (t-1), w.gammaC)))
 lambda[3,t,1,2]<-1-exp(hr.pfs.pm*w.lambdaC*(pow(((t-1)-1), w.gammaC) - pow((t-1), w.gammaC)))


        lambda[1,t,2,3]<-1-exp(-1/(ost1-dct1))
        lambda[2,t,2,3]<-1-exp(-1/(ost2-dct2))
        lambda[3,t,2,3]<-1-exp(-1/(ost3-dct3))

        for(d in 1:3){
                lambda[d,t,1,1]<-1 - (lambda[d,t,1,2] +lambda[d,t,1,3])
                lambda[d,t,1,3]<-0.005

                lambda[d,t,2,1]<-0.0
                lambda[d,t,2,2]<-1-(lambda[d,t,2,1] +lambda[d,t,2,3])

                lambda[d,t,3,1]<-0
                lambda[d,t,3,2]<-0
                lambda[d,t,3,3]<-1

                for(s in 1:S){
                        pi[d,t,s]<-inprod( pi[d,(t-1),], lambda[d,t, ,s])
                }
                #No of deaths that occured, not the cumulative no of death in the cycle.
                dpi[d,t]<-pi[d,t,3] - pi[d,(t-1), 3]

                #No of progression that occured, not the cumulative no of progression in the cycle.
                pg[d,t]<-pi[d,(t-1), 1]*lambda[d,t,1,2]   #no. of pts who moved from state 1 to state 2

                # Variables for calculation mean no of drug cycles taken:
                # to compare with Table 35 in HTA report
                #no. of pts who moved from state 1 to state 3 without state 2
                s13[d,t]<-pi[d,(t-1), 1]*lambda[d,t,1,3]
                tot.cyc[d,t]<-(t-1)*pg[d,t] + (t-1)*s13[d,t]
```

```
                    #Half cycle patient alive count -- to add
                    api[d,t,1]<-0.5*(pi[d,(t-1), 1] - pi[d,t,1])
                    api[d,t,2]<-0.5*(pi[d,(t-1), 2] - pi[d,t,2])
                    #half cycle patient counts
                    hcpi[d,t,1]<-pi[d,t,1] + api[d,t,1]
                    hcpi[d,t,2]<-pi[d,t,2] + api[d,t,2]
                    hcpi[d,t,3]<-pi[d,t,3]

        }  #end for d=1:2

}  # end of t=2:Cycles

        for(d in 1:3){
        #total no of patients who progressed. Is <N as some pts went straight to death state.
        #tot.pg[d]<-sum(pg[d,2:Cycles])
        #pts who completed 10 cycles of treatment
        diff.pg[d]<-sum(pg[d,11:Cycles]) + sum(s13[d,11:Cycles])
        cyc[d]<-(sum(tot.cyc[d,2:10]) + 10*diff.pg[d]) /N

                    pi[d,1,1]<-N
                    pi[d,1,2]<-0
                    pi[d,1,3]<-0
                    dpi[d,1]<-0
    }

        #(i)Cholesky Decomposition from SWOG trial
        Z.inteC~dnorm(0, 1)
        Z.scaleC~dnorm(0, 1)
        w.inteC<-1.85249 + 0.0595041*Z.inteC
        w.scaleC<-1.012566 + (-0.0113286311363419)*Z.inteC + 0.0426494760007284*Z.scaleC

        w.lambdaC<-exp(-w.inteC/w.scaleC)
        w.gammaC<-1/w.scaleC


    #Predicted PFS HR here.....
        prec_lhr.pfs<-1/(se.lhr.pfs*se.lhr.pfs)
        lhr.pfs~dnorm(mn.lhr.pfs, prec_lhr.pfs)                    #predicted PFS HR
        hr.pfs<-exp(lhr.pfs)


#    Direct MA PFS HR (P vs M+P) : from MA using HTA trials
        se.pm<-0.622
        prec_lhr.pm<-1/pow(se.pm, 2)
        mean.lhr.pm<-0.493                #note inversed to be positive!!
        lhr.pm~dnorm(mean.lhr.pm, prec_lhr.pm)
        hr.pfs.pm<-exp(lhr.pm)


        #Costing data
        for(ind in 2:7){
            betaC[ind]<-1/betaC.par[ind-1]
                cC.dis[ind]~dgamma(4.0, betaC[ind])
            betaT[ind]<-1/betaT.par[ind-1]
                cT.dis[ind]~dgamma(4.0, betaT[ind])
        }
        cC.dis[1]<-0.0
        cT.dis[1]<-0.0

        c.ter[1]~dgamma(4.0, 0.00101467)
```

```
        c.ter[2]~dgamma(4.0, 0.0011338)


        #Data for relative difference factor for costing of the P arm
        rd.mparm~dgamma(80.62938, 0.00350856)
        rd.parm~dgamma(105.09062, 0.00361896)
        rd.factor<-rd.parm/rd.mparm


        #(ii)Distribution placed on Drug cost
        d.ncycle.C~dnorm(5.9, 0.174835)
        d.cost.C[1]<-(169.25 + 1.02)*d.ncycle.C                    #drug cost of M+P
        d.cost.C[2]<-177.46*d.ncycle.C                             #clinic outpatient cost
        cDrug.tot[1]<-sum(d.cost.C[])

        d.ncycle.T~dnorm(7.3, 0.17562282)
        d.cost.T[1]<-(1069.50 + 1.02+5.94)*d.ncycle.T              #drug cost of D+P+Dex
        d.cost.T[2]<-177.46*d.ncycle.T                             #clinic outpatient cost
        cDrug.tot[2]<-sum(d.cost.T[])


        #Distribution placed on division factor for follow-up cost
#       mn.fu<-0.7
#       se.fu<-(0.9-0.6)/4
#       fu.alpha<-mn.fu*(mn.fu*(1-mn.fu)/pow(se.fu,2)-1)
#       fu.beta<-mn.fu*(1-mn.fu)/pow(se.fu,2)-1-fu.alpha
#       fu.cp~dbeta(fu.alpha, fu.beta)

        #First 5 cycles, no follow up cost
        for(t in 2:Cycles){
                indx[t]<-min(7, max(1,1+round((t-5)/4)))
                c[1,t,1]<-(1-fu.cp)*cC.dis[indx[t]]
                c[1,t,2]<-fu.cp*cC.dis[indx[t]]
                c[1,t,3]<-cDead

                c[2,t,1]<-(1-fu.cp)*cT.dis[indx[t]]
                c[2,t,2]<-fu.cp*cT.dis[indx[t]]
                c[2,t,3]<-cDead

                cP[t,1]<-1.48                    # fixed cost of £1.48 /cycle/patient
                cP[t,2]<-0.0
                cP[t,3]<-cDead

        }

        #Utilities
        #Data from Sandblom  paper
        #Asymp : 0.770 ±0.015; Progressive : 0.538 ±0.077; DnoPG : 0.564±0.067;
        mn_Asymp<-0.770
        se_Asymp<-(2*0.015)/1.96
        mn_Prog<-0.538
        se_Prog<-(2*0.077)/1.96
        mn_DnoPG<-0.564
        se_DnoPG<-(2*0.067)/1.96
        u.alpha_Asymp<-mn_Asymp*(mn_Asymp*(1-mn_Asymp)/pow(se_Asymp,2)-1)
        u.beta_Asymp<-mn_Asymp*(1-mn_Asymp)/pow(se_Asymp,2)-1-u.alpha_Asymp
        u.alpha_Prog<-mn_Prog*(mn_Prog*(1-mn_Prog)/pow(se_Prog,2)-1)
        u.beta_Prog<-mn_Prog*(1-mn_Prog)/pow(se_Prog,2)-1-u.alpha_Prog
        u.alpha_DnoPG<-mn_DnoPG*(mn_DnoPG*(1-mn_DnoPG)/pow(se_DnoPG,2)-1)
        u.beta_DnoPG<-mn_DnoPG*(1-mn_DnoPG)/pow(se_DnoPG,2)-1-u.alpha_DnoPG
```

```
uAsymp~dbeta(u.alpha_Asymp, u.beta_Asymp)
uProg~dbeta(u.alpha_Prog, u.beta_Prog)
uDnoPG~dbeta(u.alpha_DnoPG, u.beta_DnoPG)

for(t in 2:Cycles){
        for(d in 1:3){
u[d,t,1]<-lambda[d,t,1,1]*uAsymp + lambda[d,t,1,2]*uProg + lambda[d,t,1,3]*uDnoPG

                u[d,t,2]<-uProg
                u[d,t,3]<-uDead
        }
}

for(t in 2:13){
        for(d in 1:2){
                f1.ct[d,t]<-inprod(pg[d,t] , c[d,t,1])          #Progressed
                ct[d,t]<-inprod(dpi[d,t] , c[d,t,2])                    #Death
                cd[d,t]<-(dpi[d,t] * c.ter[d])
        }
        ct.drug[3,t]<-pi[3,t,1]*cP[t,1]

        for(d in 1:3){
                for(s in 1:S){
                        ut[d,t,s]<-inprod(hcpi[d,t,s], u[d,t,s])
                }
        }
}


for(t in 14:Cycles){
        for(d in 1:2){
                f1.ct[d,t]<-inprod(pg[d,t] , c[d,t,1]) / pow((1+0.00291667), (t-1))#Progressed
                ct[d,t]<-inprod(dpi[d,t] , c[d,t,2]) / pow((1+0.00291667), (t-1))#Death
                cd[d,t]<-(dpi[d,t] * c.ter[d]) / pow((1+0.00291667), (t-1))
        }
        ct.drug[3,t]<-(pi[3,t,1]*cP[t,1]) / pow((1+0.00291667), (t-13))

        for(d in 1:3){
                for(s in 1:S){
                        ut[d,t,s]<-inprod(hcpi[d,t,s], u[d,t,s]) / pow((1+0.00291667), (t-1))
                }
        }
}


for(d in 1:2){
mean.3C[d,1]<-sum(f1.ct[d,2:Cycles]) /N
mean.3C[d,2]<-sum(ct[d,2:Cycles]) /N
mean.3C[d,3]<-sum(cd[d,2:Cycles]) /N
        mean.C[d]<-sum(mean.3C[d,1:3]) + cDrug.tot[d]
}
cDrug.tot[3]<-(sum(ct.drug[3, 2:Cycles]))/N
mean.C[3]<-rd.factor*(sum(mean.3C[1,1:3])) + cDrug.tot[3]

for(d in 1:3){
        mean.Tm[d,1]<-sum(pi[d,2:Cycles, 1]) / N
        mean.Tm[d,2]<-sum(pi[d,2:Cycles, 2]) / N
        mean.totTm[d]<-sum(mean.Tm[d,])

        for(s in 1:S){
```

```
                        mean.3U[d,s]<-sum(ut[d,2:Cycles, s])/N/12
                }
                mean.U[d]<-sum(mean.3U[d,1:3])
        }

        Cost.diff<-mean.C[2] - mean.C[1]
        Util.diff<-mean.U[2] - mean.U[1]
        ICER<-Cost.diff/Util.diff

for (g in 1:21)  {
        Rc[g]<-(g-1)*5000

        IncNetbenefit[g]<-Rc[g]*Util.diff- Cost.diff
        ProbCE[g]<-step(IncNetbenefit[g])

        NB[1,g]<-Rc[g]*mean.U[1]- mean.C[1]
        NB[2,g]<-Rc[g]*mean.U[2]- mean.C[2]
        NB[3,g]<-Rc[g]*mean.U[3]- mean.C[3]

        for(j in 1:3){
          pCE[j,g]<-equals(rank(NB[,g],j),3)
   }
   Prob.C[g]<-pCE[1,g]
   Prob.T[g]<-pCE[2,g]
   Prob.P[g]<-pCE[3,g]
  }

}
```

Appendix F : Program codes for Chapter 5

Box F1: R codes for data simulation

```
#Simulation of patients in Orme study
library(msm)
analysis.seed<-546321

#-----------------------------------------------
#Function to simulate categorical data
x.cat<-function(p.var, indx){
  x<-rmultinom(n, size = 1, prob = p.var)
  v<-diag(indx);   temp<-v %*% x
  x.var<-colSums(temp)

  return(x.var)
}

#-----------------------------------------------
#Simulate data
#Define number of subjects required
n<-2048  ### number of binomial simulations
e<-1   #number of sample

#Gender
p<-24.7 / (24.7+74.5)
x<-rbinom(n, e, p) ; gender<-x

#Relapse during last 3 months
p<-0.289
x<-rbinom(n, e, p) ; relapse<-x

# Test results - binomial case
#(t=(table(x)/n)*100   ### Create a distribution table with proportions

#Education
l.edu<-c("Secondary school", "College/sixth form", "University/Polytechnic degree", "PG degree")
den.p.edu<-0.322+0.265+0.297+0.101
p.edu<-c(0.322/den.p.edu, 0.265/den.p.edu, 0.297/den.p.edu, 0.101/den.p.edu)
i.edu<-1:length(p.edu)
edu<-x.cat(p.edu, i.edu)
(table(edu)/n)*100

#MS type
l.ms<-c("RRMS", "SPMS", "PPMS")
p.ms<-c(0.355, 0.372, 0.273)
i.ms<-1:length(p.ms)
ms<-x.cat(p.ms, i.ms)
(table(ms)/n)*100

#EDSS
n.edss<-c(28, 151, 180, 77, 193, 323, 396, 309, 210, 165, 16)
p.edss<-n.edss/n
i.edss<-c(0:6, 6.5, 7:9)
x<-rmultinom(n, size = 1, prob = p.edss)
edss<-colSums(diag(i.edss) %*% x)
table(edss)
sim.edss<-table(edss)
```

```
testsum<-sum(abs(n.edss-c(sim.edss)));testsum
# Simple Bar Plot for EDSS
#barplot(table(edss), main="", xlab="EDSS")

#EQ-5D
m.eq5d<-0.87 + c(0, -0.071, -0.165, -0.296, -0.26, -0.352, -0.412, -0.408, -0.573, -0.919, -1.065)
w.ci.eq5d<-c(0.176, 0.188, 0.187, 0.203, 0.187, 0.184, 0.186, 0.188, 0.193, 0.197, 0.291)
se.eq5d<-w.ci.eq5d/(2*1.959963985)
mat.sd.eq5d<-diag(sqrt(n.edss)) %*% se.eq5d
sd.eq5d<-as.vector(mat.sd.eq5d)

#Steps to reduce SD using 95%RI to be within range [-0.594, 1]
#uri.sd<-(1.0-m.eq5d)/1.959963985
#lri.sd<-(m.eq5d-(-0.594))/1.959963985
#v<-c(1:11)
#min.sd.eq5d<-sapply(v, function(x) min(lri.sd[x], sd.eq5d[x], uri.sd[x]))
min.sd.eq5d<-sd.eq5d

#-----------------------------------------------
#Truncated Normal distribution
v<-c(1:11)
norm.mat.r.eq5d<-t(sapply(v, function(x) rtnorm(n, mean=m.eq5d[x], sd = min.sd.eq5d[x], lower= -
0.594, upper=1)))
norm.mat.eq5d<-x*norm.mat.r.eq5d
normeq5d<-colSums(norm.mat.eq5d)


#-----------------------------------------------
# Normal distribution
v<-c(1:11)
utnorm.mat.r.eq5d<-t(sapply(v, function(x) rnorm(n, mean=m.eq5d[x], sd = min.sd.eq5d[x])))
utnorm.mat.eq5d<-x*utnorm.mat.r.eq5d
utnormeq5d<-colSums(utnorm.mat.eq5d)
set1utnormeq5d<-utnormeq5d

// Steps to reduce the datset and re-compute the SD
newvec<-cbind(edss, utnormeq5d); newvec[1:15, ]
newset<-subset(newvec, (utnormeq5d>=-0.594 & utnormeq5d <= 1), select = c(edss, utnormeq5d));
newset[1:5, ]

newedss<-newset[,"edss"]
newutn<-newset[,"utnormeq5d"]
newtest<-t(sapply(v, function(x) c(length(newutn[newedss==w[x]]), m.eq5d[x],
mean(newutn[newedss==w[x]]), min.sd.eq5d[x], sd(newutn[newedss==w[x]])))); newtest
(newsd<-sapply(v, function(x) sd(newutn[newedss==w[x]])))

utnorm.mat.r.eq5d<-t(sapply(v, function(x) rnorm(n, mean=m.eq5d[x], sd = newsd[x])))
utnorm.mat.eq5d<-x*utnorm.mat.r.eq5d
utnormeq5d<-colSums(utnorm.mat.eq5d)


#-----------------------------------------------
#Log-normal transformation EQ-5D ---> y ---> z

mu.new<-log((m.eq5d+0.594)/(1-m.eq5d)) + ((min.sd.eq5d)^2)*(-1*(m.eq5d+0.594)^-2 +(1-m.eq5d)^-
2)/2
var.new<-((min.sd.eq5d)^2)* ((1/(m.eq5d+0.594))+(1/(1-m.eq5d)))^2
sd.new<-sqrt(var.new)

#Simulate lognormal distribution
```

```
v<-c(1:11)
z<-t(sapply(v, function(x) rnorm(n, mean=mu.new[x], sd = sd.new[x])))
#z is log normal.... keep z for later analyses
mat.z<-x*z
zeq5d<-colSums(mat.z)

# back transform:
y<-(exp(z)-0.594)/(1+exp(z))
#hist(y, breaks=25)

mat.eq5d<-x*y
eq5d<-colSums(mat.eq5d)


#Using sd from UT normal distribution
mu.new<-log((m.eq5d+0.594)/(1-m.eq5d)) + (newsd^2)*(-1*(m.eq5d+0.594)^-2 +(1-m.eq5d)^-2)/2
var.new<-(newsd^2)* ((1/(m.eq5d+0.594))+(1/(1-m.eq5d)))^2
sd.new<-sqrt(var.new)

#Simulate lognormal distribution
v<-c(1:11)
set.seed(52720)
z<-t(sapply(v, function(x) rnorm(n, mean=mu.new[x], sd = sd.new[x])))
#z is log normal.... keep z for later analyses
mat.z<-x*z
zeq5dutn<-colSums(mat.z)

# back transform:
y<-(exp(z)-0.594)/(1+exp(z))
#hist(y, breaks=25)

mat.eq5d<-x*y
eq5dutn<-colSums(mat.eq5d)


#-----------------------------------------------
#Beta distribution

#Conversion y --> x for beta-distribution
m.eq5d.beta<-(m.eq5d +0.594)/(1.594)
sd.eq5d.beta<-(newsd)/(1.594)                                    #using sd from utnormal
#sd.eq5d.beta<-(min.sd.eq5d)/(1.594)

(alpha<-(m.eq5d.beta^2)*(1-m.eq5d.beta)/(sd.eq5d.beta^2) - m.eq5d.beta)
(beta<-alpha*(1-m.eq5d.beta)/m.eq5d.beta)

v<-c(1:11)
mat.eq5d.beta<-t(sapply(v, function(x) rbeta(n, shape1=alpha[x], shape2=beta[x], ncp = 0)))
hist(mat.eq5d.beta)

mat.utbeq5d<-x*mat.eq5d.beta
utbetaeq5d<-colSums(mat.utbeq5d)
hist(utbetaeq5d)
hist(1-utbetaeq5d)
hist(log(-utbetaeq5d))
hist(exp(utbetaeq5d))

mat.b.eq5d<-1.594*mat.eq5d.beta -0.594
mat.eq5d<-x*mat.b.eq5d
betaeq5d<-colSums(mat.eq5d)
```

```
betaeq5d<-1.594*utbetaeq5d -0.594


#Constructed using 95%RI = [-0.594, 1]
#Steps to reduce SD using 95%RI to be within range [-0.594, 1]
uri.sd<-(1.0-m.eq5d)/1.959963985
lri.sd<-(m.eq5d-(-0.594))/1.959963985
v<-c(1:11)
ri.sd.eq5d<-sapply(v, function(x) min(lri.sd[x], sd.eq5d[x], uri.sd[x]))
sd.eq5d.beta<-(ri.sd.eq5d)/(1.594)
(alpha<-(m.eq5d.beta^2)*(1-m.eq5d.beta)/(sd.eq5d.beta^2) - m.eq5d.beta)
(beta<-alpha*(1-m.eq5d.beta)/m.eq5d.beta)

v<-c(1:11)
set.seed(52720)
mat.eq5d.beta<-t(sapply(v, function(x) rbeta(n, shape1=alpha[x], shape2=beta[x], ncp = 0)))
mat.b.eq5d<-1.594*mat.eq5d.beta -0.594
mat.eq5d<-x*mat.b.eq5d
s1betaeq5d<-colSums(mat.eq5d)
```

Box F2: WinBUGS codes for linear regression (1st degree polynomial and 2nd degree polynomial for illustration)

```
model{
  for (i in 1:num) {
    eq5d[i]~dnorm(mu[i],prec)
    cedss[i]<-(edss[i] - 5)
      #1st degree polynomial
      mu[i]<-alpha+beta1*cedss[i]
      #2nd degree polynomial
      mu[i]<-alpha+beta1*cedss[i]+beta2*cedss[i]*cedss[i]
  }
  prec~dgamma(0.001,0.001)
  sd<-1/sqrt(prec)
  alpha~dnorm(a.m,a.prec)

  #1st degree polynomial
  beta1~dnorm(b1.m,b1.prec)
  #2nd degree polynomial
  beta1~dnorm(b1.m,b1.prec)
  beta2~dnorm(b2.m,b2.prec)

  #estimates:
  for (w in 1:10) {
    #1st degree polynomial
    lnorm.eq5d[w]<-alpha + beta1*(w-6)
    #2nd degree polynomial
    lnorm.eq5d[w]<-alpha + beta1*(w-6) + beta2*(w-6)*(w-6)
    eq5d.res[w]<-(exp(lnorm.eq5d[w]) - 0.594)/(1+exp(lnorm.eq5d[w]))
  }

    #1st degree polynomial
    lnorm.e65.eq5d<-alpha + beta1*(6.5 - 5)
    #2nd degree polynomial
    lnorm.e65.eq5d<-alpha + beta1*1.5 + beta2*pow(1.5,2)
  eq5d.e65.res<-(exp(lnorm.e65.eq5d) - 0.594)/(1+exp(lnorm.e65.eq5d))
}
```

Appendix G : Publications and posters

This appendix contains the publications and posters presentations resulting from the work carried out in this thesis.

Research papers:

Tan SH, Bujkiewicz S, Sutton A, Dequen P, Cooper N. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. *J Health Serv Res Policy* 2013;18(4):224-32.

Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ. Novel presentational approaches were developed for reporting network meta-analysis. *J Clin Epidemiol* 2014;67(6):672-80.

Poster presentations:

Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ. Novel Presentational Approaches for Reporting Network Meta-Analysis. *The International Society of Clinical Biostatistics, Munich 2013.*

Tan SH, Bujkiewicz S, Abrams, KR. Bivariate indirect comparison meta-analysis model in economic evaluation of cancer treatments. *The International Society for Pharmacoeconomics and Outcomes Research Annual European Congress, Dublin, 2013*

Journal of Health Services Research & Policy
18(4) 224–232
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1355819613498379
jhsrp.rsmjournals.com
SAGE

Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons

Sze Huey Tan[1], Sylwia Bujkiewicz[2], Alexander Sutton[3], Pascale Dequen[4] and Nicola Cooper[5]

Abstract

Objectives: To establish current guidance and practice in UK on presentation of indirect comparison and mixed treatment comparison analyses; to provide recommendations to improve indirect comparison/mixed treatment comparison reporting and to identify research priorities for improved presentation.

Methods: Existing institutional guidance for conducting indirect comparison/mixed treatment comparison alongside current practice in health technology assessment was reviewed. Reports published in UK by the Health Technology Assessment programme since 1997, which utilized indirect comparison/mixed treatment comparison methods, were reviewed with respect to the presentation of study data, statistical models and results. Recommendations for presentation were developed.

Results: Guidance exists that provide the details necessary to conduct a successful indirect comparison/mixed treatment comparison analysis but recommendations on presentation are limited. Of 205 health technology assessment reports that contained evidence synthesis for effectiveness, 19 used indirect comparison/mixed treatment comparison methods. These reports utilized numerous presentational formats from which the following key components were identified: network table/diagram for presenting data; model description to allow reproducibility; and tables, forest plots, matrix tables and summary forest plots for presenting a range of results. Recommendations were developed to ensure that reporting is explicit, transparent and reproducible. Approaches most understandable by non-technical decision makers, and areas where future research is required, are outlined.

Conclusions: There is no standard presentational style used in UK for reporting indirect comparison/mixed treatment comparison, and the use of graphical tools is limited. Standardization of reporting and innovation in graphical representation of indirect comparison/mixed treatment comparison results is required.

Keywords

indirect treatment comparisons, mixed treatment comparisons, reporting

Introduction

In health technology assessment (HTA), it is necessary to collate information on effectiveness and cost-effectiveness for influencing health policy. Information should ideally be comprehensive and obtained from all relevant, well-managed and documented data sources including randomised controlled trials (RCTs), observational studies, expert opinions and bench research. Evidence synthesis of clinical effectiveness between treatment interventions has largely been evaluated using standard pairwise meta-analysis[1,2] using studies that compared the same two interventions of interest. Although it may be possible to address the comparative

[1]Research Student, Department of Health Sciences, University of Leicester, Leicester, UK
[2]Lecturer of Population and Public Health Sciences, Department of Health Sciences, University of Leicester, Leicester, UK
[3]Professor of Medical Statistics, Department of Health Sciences, University of Leicester, Leicester, UK
[4]Research Associate, Department of Health Sciences, University of Leicester, Leicester, UK
[5]Professor of Health Care Evaluation Research, Department of Health Sciences, University of Leicester, Leicester, UK

Corresponding author:
Sze Huey Tan, Department of Health Sciences, University of Leicester, 2nd Floor (Room 211), Adrian Building, University Road, Leicester LE1 7RH, UK.
Email: sht10@le.ac.uk

effectiveness between treatments and judge whether the new treatment is better than placebo or standard care, this type of analysis is limited. Often decision makers are interested in assessing the comparative effectiveness of multiple interventions, but trials comparing all technologies of interest may not be available. And even when head-to-head trials do exist, there is increasing interest to incorporate data from other related studies to inform the estimate of effectiveness. Hence, evidence synthesis methods that enable direct and/or indirect comparisons of multiple treatments are required to elucidate how these compare in terms of effectiveness and cost-effectiveness.

Indirect treatment comparison (IC) and mixed treatment comparison (MTC) (also known as network meta-analysis) are recent developments in evidence synthesis that allow for the comparison of multiple treatments simultaneously.[3–8] IC and MTC are now acknowledged methodologies by HTA agencies worldwide including the National Institute for Health and Care Excellence (NICE) in England and Wales, the Canadian Agency for Drugs and Technologies in Health, the French Haute Autorité de la Santé and the Pharmaceutical Benefits Advisory Committee in Australia, as well as emerging national agencies in Austria, Brazil, Colombia, Cuba and Ireland.

Since 2004, NICE guidance[9] has emphasized that, in HTA appraisals, evidence synthesis using meta-analysis of head-to-head RCTs is the preferred method. If no head-to-head RCTs are available, the guidance permits the use of ICs so long as the potential bias in its use is appropriately explored and reported. In 2008, NICE re-emphasized the preference for synthesis results from head-to-head RCTs where available.[10] However, it also stated that MTC analyses results may be included even if head-to-head trials were available, if it is justified that the MTC analysis will add information that is not available from the head-to-head trials. Furthermore, recent guidance (only available online for consultation at the time of writing) further strengthens NICE's position regarding the use of IC/MTC in HTA appraisals.[11]

Internationally recognized guidelines for good reporting of standard pairwise meta-analysis of RCTs, such as the PRISMA statement (formerly known as QUOROM), have been developed.[12,13] While much of the meta-analysis reporting guidance is also highly relevant to the IC/MTC context (including advice on clarity of labels and legends, and use of reasonable sizes for symbols and lines in tables and figures), the inherently greater complexity of the latter approaches presents further challenges for reporting such analyses.

Most specific IC/MTC guidance is embedded in institutional HTA evaluation documentation. In UK, the NICE guidance states that all data used for estimates of effectiveness should be presented in tabular form with the source of the data clearly stated.[9,10] It also states that, for IC/MTC analyses, the evidence may be presented in either tabular or diagrammatic form and should be reported as both relative and absolute effectiveness estimates. This is re-enforced by more detailed advice that complements their guidance on evidence synthesis (though their content is non-mandatory when making submissions).[14–20] Another detailed source of guidance on conducting IC/MTC analyses is the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force good practice documents.[21,22] While these provide many of the details necessary to conduct a successful IC/MTC analysis, specific details and recommendations on presentational formats, particularly of the data and results, are limited.

Our objectives were the first to review guidance on the presentation of IC/MTC analyses for institutional guidelines and to assess what has previously been done in practice in UK by reviewing HTA reports, some of which were commissioned to inform NICE appraisals. Second, to provide recommendations on how to improve future reporting of IC/MTC analyses. And, third, to identify research priorities for improving the presentation of IC/MTC analyses.

Methods

HTA programme reports published by the NIHR in UK between 1997 and 2011 were reviewed by one of the authors (NJC). The reports that used IC and/or MTC methodology for evidence synthesis were identified. Using a standardized data extraction form (see Appendix 1), these reports were examined to establish the approaches taken in the reporting of IC/MTC as regards:

Input data—presentation of the number of interventions, study level data and the relationship structure of the interventions and the studies included in the analysis;
Methods—specification of Bayesian or frequentist statistical models, software used and, where appropriate, presentation of prior distributions used (and assessment of their influence on the results via sensitivity analyses) and assessment of model convergence;
Results—presentation of relative effects, absolute effects, probability of treatment being best and/or ranking of interventions.

Throughout this review process, there was a particular emphasis on the use of tables and graphical presentational approaches.

Results

Systematic review of the HTA reports

Of 608 UK NIHR HTA reports published, 375 contained systematic reviews of effectiveness and/or cost-effectiveness of which 205 contained evidence synthesis and 19 (including 10 of which informed NICE appraisals) utilized IC/MTC methodology (Figure 1). All 19 reports were published after 2004, the year in which NICE guidance[9] recommended the use of IC analysis when no head-to-head RCT data are available for HTA. Eight reports used IC methods and 11 used MTC methods to synthesize the effectiveness data (Table 1). The evaluation of approaches used for presenting IC/MTC results in this work is focused on the presentation of effectiveness although the use of IC and MTC methodologies for the synthesis of adverse events was seen in six reports. Eighteen reports included only RCTs; the exception used both observational cohort studies and RCTs but performed sensitivity analysis excluding observational data. The number of interventions used in the IC/MTC analysis ranged from 3 (minimum required for IC/MTC analysis) to 15. There was no major 'lumping' of treatment interventions by drug class although varying drug doses were grouped together in some reports. Complete study summary data (unless excluded for confidentiality reasons) used for the analysis were provided for 17 reports.

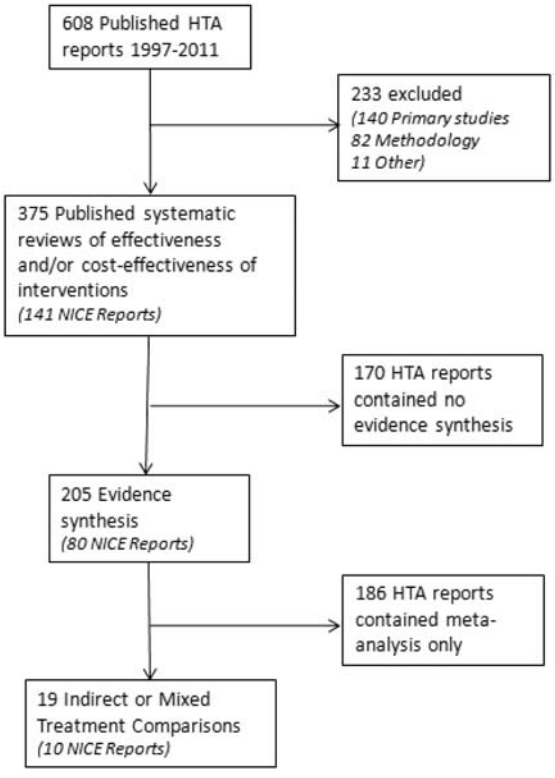

Figure 1. Flowchart of HTA review selection.

Presentation of IC/MTC data

Eleven of the 19 HTAs used a network table (see Appendix 2) to display the treatment comparisons (columns) considered by each trial (rows). Only three included all data used in the synthesis as elements in the network table. The rest used ticks, cross marks, shading or patient numbers to indicate what treatment interventions were investigated in each RCT; although, in the latter cases, data may have been presented for the relevant trials in other sections of the report. Four reports used network diagrams (Appendix 2) to display similar information on the treatment comparisons considered by the included trials. This type of diagram graphically displays all the treatment interventions included in the IC/MTC and links these treatments with lines if the comparison of the treatments exists in at least one of the studies. Only two reports used both network table and network diagram. The remaining six reports did not show the network of trials used.

Presentation of IC/MTC synthesis model and its implementation

All reports discussed the method and the rationale for the use of IC/MTC. Sixteen reports utilized the Bayesian framework for the statistical model. Seven reports presented the statistical model in the main report text while five presented it in an appendix. Four reports did not present the statistical model but referenced published models. The remaining three gave a short description of the statistical model used. Checks for inconsistencies in the IC/MTC network(s) were assessed in two reports using informal methods (recently developed formal methods of using deviance information criteria or node-splitting technique[23] were not applied).

In view of the computationally intensive nature of IC/MTC analysis, the software used for the analysis was reviewed. Out of the 19 reports, 16 used the WinBUGS statistical software version 1.4.[24] The others did not specify explicitly the software used but reported that either SAS, STATA, RevMan or StatsDirect was used for the presented meta-analysis results that were reported together with the IC results. For the 16 reports that used WinBUGS software, 14 defined the Bayesian prior distributions. All of these used vague prior distributions, of which five conducted sensitivity analysis to assess the influence of the choice of prior distribution on the results obtained. Checking

Table 1. IC/MTC presentation in the 19 reports.

HTA report	Appraisal	Method	No. of Interventions	Data presentation			Results presentation					
				Network table	Network diagram	Not presented	Matrix table	Table (RET)	Table (AET)	Summary forest/ caterpillar plot	Text only	Unclear
Vol 15, No 40. (2011) (Squires)	NICE	MTC[a]	6	✓[b]	✓		✓			✓		
Vol 15, No 39. (2011) (Lewis)	Non-NICE	MTC[c]	15		✓		✓			✓		
Vol 15, No 31. (2011) (Greenhalgh)	Non-NICE	MTC[a,e]	4		✓[f]			✓				
Vol 15, No 19. (2011) (Bhattacharya)	Non-NICE	IC[a]	4			✓				✓		
Vol 15, No 10. (2011) (Rodgers)	NICE	IC[a]	4			✓			✓			
Vol 14, No 40. (2010) (Imamura)	Non-NICE	MTC[c]	14	✓				✓		✓		
Vol 14, No 24. (2010) (McKenna)	Non-NICE	IC[d]	4	✓				✓				
Vol 14, No 17. (2010) (McDaid)	Non-NICE	MTC[a,e]	4	✓[f]	✓			✓				
Vol 14, No 2. (2010) (Thompson-Coon)	NICE	IC[a]	3			✓		✓				
Vol 13, No 58. (2009) (Burch)	NICE	IC[a]	3			✓		✓				
Vol 13, No 34. (2009) (Ara)	Non-NICE	MTC[a]	5			✓		✓				
Vol 11, No 39. (2007) (Soares-Weiser)	Non-NICE	MTC[a]	8	✓					✓			
Vol 11, No 2. (2007) (Collins)	NICE	IC[a]	IC: 3	✓							✓	
		MTC[d,e]	MTC: 8									
Vol 10, No 46. (2006) (Woolacott)	NICE	MTC[a]	8	✓				✓				
Vol 10, No 38. (2006) (Brown)	Non-NICE	IC[a,e]	7			✓		✓				
Vol 10, No 23. (2006) (King)	NICE	MTC[d,e]	6	✓					✓			
Vol 10, No 9. (2006) (Main)	NICE	MTC[d,e]	Analysis 1: 3	✓[g]								✓
			Analysis 2: 6									
Vol 10, No 31. (2006) (Riemsma)	NICE	IC[a]	3	✓[b]					✓			
Vol 8, No 19. (2004) (Bridle)	NICE	MTC[d]	6	✓[b]					✓			
Total	10 NICE	8 IC	Range: 3–15	11	4	6	2	9	5	4	1	1
	9 Non-NICE	11 MTC										

RET: relative effect table (e.g. Table 3); AET: absolute effect table (e.g. Table 4).
[a]Reported in Clinical effectiveness chapter.
[b]Reported study sample sizes/outcome data/both in network tables.
[c]Reported in MTC chapter.
[d]Reported in Cost-effectiveness chapter.
[e]Adverse events were analysed and used the same IC/MTC unless otherwise stated.
[f]Reported in Appendix.
[g]Reported on one table for both analyses.

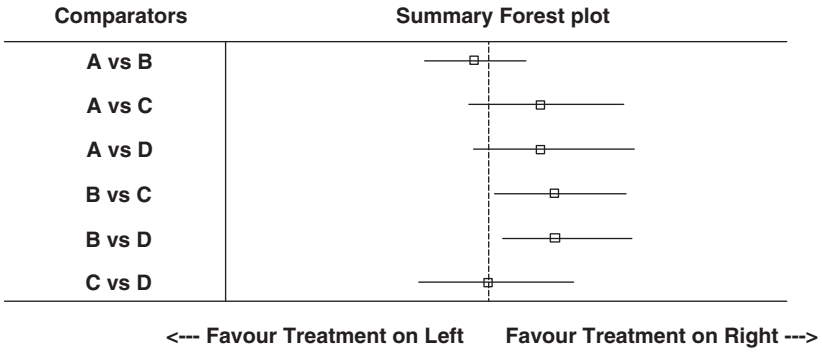

Figure 2. Summary Forest plot.

Table 2. Matrix table of relative effects.

	Mixed treatment comparison			
Standard Meta-Analysis	Intervention A	$OR_{A\text{-}B\_MTC}$ (95% CrI)	$OR_{A\text{-}C\_MTC}$ (95% CrI)	$OR_{A\text{-}D\_MTC}$ (95% CrI)
	$OR_{A\text{-}B\_MA}$ (95% CrI)	Intervention B	$OR_{B\text{-}C\_MTC}$ (95% CrI)	$OR_{B\text{-}D\_MTC}$ (95% CrI)
	$OR_{A\text{-}C\_MA}$ (95% CrI)	$OR_{B\text{-}C\_MA}$ (95% CrI)	Intervention C	$OR_{C\text{-}D\_MTC}$ (95% CrI)
	Not calculable	$OR_{B\text{-}D\_MA}$ (95% CrI)	$OR_{C\text{-}D\_MA}$(95% CrI)	Intervention D

Key: $OR_{A\text{-}B\_MTC}$ = Odd ratio of A vs B using Mixed Treatment Comparison.
$OR_{A\text{-}B\_MA}$ = Odd ratio of A vs B using Meta-analysis.

of the convergence of the MCMC sampler in the Bayesian IC/MTC analysis was reported by 9 out of the 16 reports using WinBUGS and 14 reports included WinBUGS codes in their appendices.

Presentation of IC/MTC results

IC/MTC effectiveness results were presented for 18 of the 19 reports. Two had a section specifically for the IC/MTC effectiveness analysis while 13 presented the IC/MTC analysis in the effectiveness and 4 in the cost-effectiveness chapters of the report (i.e. IC/MTC was only used in the cost-effectiveness modelling).

Most reports (16) presented the results of the IC/MTC in tables (Table 1), with four of these also presenting either summary forest or caterpillar plots (Figure 2). Across the 16 reports, three different table formats were used:

- two reports used a matrix table of relative effects (Table 2), which contains all permutations of treatment comparisons for both IC/MTC and pairwise meta-analysis, separated by the off-diagonal;
- nine used a relative effect table (Table 3), which summarizes pooled ratios/weighted mean differences of selected treatment comparisons relevant to the HTA and
- five used an absolute effect table (Table 4), which presents the posterior probability of rates (response, rejection, etc.) for all treatment interventions in the IC/MTC by the use of a specified underlying baseline rate.

Thirteen HTAs reported comparative effectiveness estimates, of which eight reported all permutations of pair-wise comparison results from the IC/MTC analysis, five concentrated either on active treatments compared with placebo (or no treatment or standard care) or on active treatments of interest compared to one another.

IC/MTC analysis results were reported either as pooled relative effects (odds ratios, hazard ratios, weighted mean differences) or rates (response rates, withdrawal rates) with either 95% confidence interval (for frequentist evidence synthesis) or 95% credible interval (for Bayesian evidence synthesis). Five HTAs reported the probability that each treatment was the most effective of which four of these presented the 'best' statistic in tables and one reported it only in the main text of the report. None of the published reports provided tables or graphs that ranked the technologies in terms of effectiveness as has been presented elsewhere.[25]

Table 3. Relative effects table.

Treatment comparators		Mixed treatment comparison		Standard meta-analysis	
		Mean	95% CrI	Mean	95% CrI
Intervention A	Intervention B	$OR_{A\text{-}B\_MTC}$	(95% CrI)	$OR_{A\text{-}B\_MA}$	(95% CrI)
Intervention A	Intervention C	$OR_{A\text{-}C\_MTC}$	(95% CrI)	$OR_{A\text{-}C\_MA}$	(95% CrI)
Intervention A	Intervention D	$OR_{A\text{-}D\_MTC}$	(95% CrI)	Not calculable	Not calculable
Intervention B	Intervention C	$OR_{B\text{-}C\_MTC}$	(95% CrI)	$OR_{B\text{-}C\_MA}$	(95% CrI)
Intervention B	Intervention D	$OR_{B\text{-}D\_MTC}$	(95% CrI)	$OR_{B\text{-}D\_MA}$	(95% CrI)
Intervention C	Intervention D	$OR_{C\text{-}D\_MTC}$	(95% CrI)	$OR_{C\text{-}D\_MA}$	(95% CrI)

Key: $OR_{A\text{-}B\_MTC}$ = Odd ratio of A versus B using mixed treatment comparison.
$OR_{A\text{-}B\_MA}$ = Odd ratio of A versus B using meta-analysis.

Table 4. Absolute effects table.

Treatments	Mixed treatment comparison		Standard meta-analysis	
	Mean	95% CrI	Mean	95% CrI
Intervention A	Eff_{A_MTC}	(95% CrI)	Eff_{A_MA}	(95% CrI)
Intervention B	Eff_{B_MTC}	(95% CrI)	Eff_{B_MA}	(95% CrI)
Intervention C	Eff_{C_MTC}	(95% CrI)	Eff_{C_MA}	(95% CrI)
Intervention D	Eff_{D_MTC}	(95% CrI)	Eff_{D_MA}	(95% CrI)

Key: Eff_{A_MTC} = Absolute treatment effects estimates of A using mixed treatment comparison.
Eff_{A_MA} = Absolute treatment effects estimates of A using meta-analysis.

All except two reports used the IC/MTC evidence synthesis results to inform their economic decision model. Of those that did, one used a subset of RCTs used for effectiveness and another used more RCTs than those included in the effectiveness evaluation.

After we completed this review, a further systematic review of international literature[26] of 42 MTC analyses was in broad agreement with ours. They found that the presentation of results was much more likely to be tabular (89.5%) rather than graphical (21.1%) formats. Thus, while our sample of IC/MTCs was limited to HTA reports in UK, it is reassuring that a review of reports from a wider area obtained similar findings.

Discussion

Recommendations to improve reporting of IC/MTC analyses

In our recommendations for IC/MTC reporting, we consider data, model and results (Table 5) and also cover (i) the qualities that make good IC/MTC analysis reporting, (ii) the most appropriate and informative presentation methods, and (iii) the target audiences for the information.

In any statistical analysis, the credibility of the results depends on the quality of the data used and the appropriateness of the model adopted. It is, therefore, imperative that a clear description of the data and statistical models is presented to ensure transparency and reproducibility. Therefore, we recommend that studies used in the IC/MTC analysis should be clearly presented with, at a minimum, references properly cited, and the data used in the analysis clearly stated or outcome data (e.g. for a binary outcome, number of events and number of patients in arm) from which the outcome measures used (e.g. log-odds ratios and standard errors) can be calculated. A tabular format is a good way of presenting this information. A network diagram is an excellent way of visualizing the relationships between the studies and interventions under evaluation. Since this could be derived from a tabular description of all the studies, of the form described above, it is not strictly essential, but where space allows, we encourage its inclusion. However, if any included studies have more than two treatment arms, then it is not possible to derive a network table from a network diagram, in addition to not being able to identify citations to specific studies.

For the statistical model, this should be fully described with associated algebra together with analysis

Table 5. Table of recommendations for IC/MTC analysis reporting.

Component	Qualities	Presentation method recommended (domains fulfilled)	Target audience
Data	D1. Studies included in analysis clearly presented and references cited	Network table, with the *structure* of Figure in Appendix 2, but with outcome data presented in the table cells/references to studies included.	Academics
	D2. Treatments compared in each study reported	Network table provides transparency of the trial data used in the analysis (D1–D3)	Decision-makers
	D3. Treatment data / effect sizes (with a measure of uncertainty) for each included study reported		
Model	M1. Statistical model described typically with algebraic representation	Full analysis codes including data with data code sheet. This can be provided in Appendix or as online supplementary materials.	Academics
	M2. Citation given to the model used	This allows reproducibility of the results and contains the model used and data included in the analysis (M1–M3)	Statistical analysts
	M3. Analysis code presenting model (and data)		
Results	R1. Results of interest (given the aim of the study) reported	Tables (see Table 3 for an example) and summary forest plot of results (see Figure 2) of interest (R1)	Academics
	R2. (Relative) Comparison of all treatments	Matrix tables (see Table 2 for an example) with Summary Forest Plot/caterpillar plot (see Figure 2 for an example) (R1, R2)	Decision-makers
	R3. Probability best statistics/ranking of treatments	Tables or 'rankogram' presenting probability best statistics and ranking (R3)	

code (including data with a data code sheet explaining the data structure) either in the main text or as an appendix. If report space is limited, then either citation to the model specification published elsewhere would be required or supplementary material describing the model provided online.

The presentation of results from the analysis relies largely on the question of interest: which treatment is best, what is the treatment effect for a specific comparison or perhaps all comparisons with either usual care or all treatments in the network being of interest. IC/MTC analyses can answer these questions by providing estimates of the treatment effects for any comparisons included in the network, the probability that each treatment is best and the ranking distribution of each treatment. It may not be feasible and/or desirable to report all of them in a report manuscript, and therefore the focus of the analysis should be well defined to guide the choice of the most appropriate statistics to report. This issue strengthens the desirability for making data/analysis code available to enable readers to obtain results for any aspect of the analysis not reported. Hence, prescriptive requirements on reporting results are not possible; however, we encourage the use of graphical and tabular approaches to reporting for ease of interpretation. This is consistent with the draft 2012 NICE guidance.[11]

Finally, the choice on what type of results to report and which tools to use depends on the audience. For example, whereas academics may be interested in all three components of the IC/MTC analysis, the statistical analysts would be expected to focus more on the model specification and decision makers on the transparency of the data and clarity of the results of interest.

The above recommendations should be used in conjunction with the good practice in IC/MTC methods documents recently published by an ISPOR Task Force[21,22] and NICE's Decision Support Unit.[14–20] The latter provides a checklist for reviewers' of MTCs and goes beyond the issue of methods and results reporting including sections on the definition of the decision problem and embedding the synthesis in a probabilistic cost-effectiveness model. Finally, it should be acknowledged that we make no recommendations regarding the reporting of advanced issues such as variable study quality, subgroup analysis or inconsistency analysis, due to the infancy of this methodology.

Research priorities

IC/MTC techniques, by their nature, often include a larger number of studies, require more complicated statistical analysis models and produce larger arrays of results when compared with standard meta-analysis. For example, an MTC including five different treatment regimens generates 10 pairwise comparisons and this increases to 45 pairwise comparisons when 10 different treatment regimens are included. Interpreting such large arrays of results and how individual studies influence them can be challenging. Indeed, consistency and clarity in the presentation of results will aid the interpretability of the results while clear presentation will enable better understanding of the relationship between treatment comparators and facilitate sensitivity analysis, subgroup analysis and network inconsistencies analysis. More work needs to be done, since summary figures highlighting the most important results (e.g. those interventions which have the highest probabilities of being best and ranking statistics) would be an improvement on forest plots. A good model for academic papers would be an approach of the type taken by Trelle et al.[27] where multiple extensive appendices supplement a concise article. There have already been some published MTCs with very large networks for example one that included 12 treatments and 117 RCTs.[28] Optimal presentational strategies for MTC may vary depending on, among other things the size of the dataset, making rigid reporting guidelines challenging. Added to this is the acknowledgement of individuals' personal preferences for numeric or graphical displays, which may be coupled with their perspective of wanting to draw inferences from the analysis or make decisions based on it.

Other innovations in presentation of IC/MTC analyses may be valuable to explore further. For example, the widths of the lines on network diagrams have been set to define the number of trials contributing direct comparative estimates, and the size of the treatment boxes/nodes made proportional to the number of randomized subjects (sample size).[28] Rankograms, that graphically represent the distribution and probability of rankings when compared to all other interventions in the network, have been presented.[25,28] Similarly, cumulative ranking probability plots have been proposed.[25] In an attempt to communicate variable quality of the evidence for each direct and MTC comparison, coloured circles have been added in the cells of the Matrix table of results.[29] Finally, in a similar vein to the work undertaken by Bax et al.[30] for pairwise meta-analysis, an evaluation of how different audiences is able to interpret different presentational approaches would seem a valuable exercise.

Funding

This study was funded by Department of Epidemiology, Janssen Research and Development.

References

1. Borenstein M, Hedges LV, Higgins JPT, et al. *Introduction to meta-analysis*. Chichester: Wiley, 2009.
2. DerSimonian R and Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007; 28: 105–114.
3. Caldwell DM, Ades AE and Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331: 897–900.
4. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med* 2003; 22: 2995–3016.
5. Ades AE, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* 2006; 24: 1–19.
6. Higgins JP and Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996; 15: 2733–2749.
7. Lu G and Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23: 3105–3124.
8. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002; 21: 2313–2324.
9. NICE. *Guide to the methods of technology appraisal*. London: National Institute for Health and Clinical Excellence (NICE), 2004.
10. NICE. *Guide to the methods of technology appraisal*. London: National Institute for Health and Clinical Excellence (NICE), 2008.
11. NICE. *Draft: guide to the methods of technology appraisal*. London: National Institute for Health and Clinical Excellence (NICE), 2012. http://www.nice.org.uk/aboutnice/howwework/devnicetech/TAMethodsGuideReview.jsp (2012, accessed 15 October 2012).
12. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700.
13. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; 339: b2535.
14. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 1: introduction to evidence synthesis for decision making. Available from http://www.nicedsu.org.uk (2011, accessed 31 January 2012).
15. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials, http://www.nicedsu.org.uk (2011, accessed 31 January 2012).
16. Dias S, Sutton AJ, Welton NJ, et al. NICE DSU technical support document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment, http://www.nicedsu.org.uk (2011, accessed 31 January 2012).

17. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 4: inconsistency in networks of evidence based on randomised controlled trials, http://www.nicedsu.org.uk (2011, accessed 31 January 2012).
18. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 5: evidence synthesis in the baseline natural history model, http://www.nicedsu.org.uk (2011, accessed 31 January 2012).
19. Dias S, Sutton AJ, Welton NJ, et al. NICE DSU technical support document 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices, http://www.nicedsu.org.uk (2011, accessed 31 January 2012).
20. Ades AE, Caldwell DM, Reken S, et al. NICE DSU technical support document 7: evidence synthesis of treatment efficacy in decision making: a reviewer's checklist, http://www.nicedsu.org.uk (2012, accessed 1 May 2012).
21. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR task force on indirect treatment comparisons good research practices, part 1. *Value Health* 2011; 14: 417–428.
22. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices, part 2. *Value Health* 2011; 14: 429–437.
23. Dias S, Welton NJ, Caldwell DM, et al. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; 29: 932–944.
24. Spiegelhalter D, Thomas A, Best N, et al. WinBUGS user manual, Version 1.4 ed. Cambridge, England: MRC Biostatistics Unit, 2003.
25. Salanti G, Ades AE and Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; 64: 163–171.
26. Coleman CI, Phung OJ, Cappelleri JC, et al. Use of mixed treatment comparisons in systematic reviews. methods research report. (Prepared by the University of Connecticut/Hartford Hospital Evidence-based Practice Center under Contract No. 290-2007-10067-I.) AHRQ Publication No. 12-EHC119-EF. Rockville, MD: Agency for Healthcare Research and Quality, www.effectivehealthcare.ahrq.gov/reports/final.cfm (2012, accessed 15 October 2012).
27. Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *BMJ* 2011; 342: c7086.
28. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746–758.
29. Dumville JC, Soares MO, O'Meara S, et al. Systematic review and mixed treatment comparison: dressings to heal diabetic foot ulcers. *Diabetologia* 2012; 55: 1902–1910.
30. Bax L, Ikeda N, Fukui N, et al. More than numbers: the power of graphs in meta-analysis. *Am J Epidemiol* 2009; 169: 249–255.

ELSEVIER

Journal of Clinical Epidemiology 67 (2014) 672−680

Journal of Clinical Epidemiology

Novel presentational approaches were developed for reporting network meta-analysis

Sze Huey Tan[a,b,*], Nicola J. Cooper[a], Sylwia Bujkiewicz[a], Nicky J. Welton[c], Deborah M. Caldwell[c], Alexander J. Sutton[a]

[a]*Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK*
[b]*Division of Clinical Trials and Epidemiological Sciences, National Cancer Centre Singapore, 11 Hospital Drive, 169610, Singapore*
[c]*School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK*

Accepted 11 November 2013; Published online 20 February 2014

Abstract

Objectives: To present graphical tools for reporting network meta-analysis (NMA) results aiming to increase the accessibility, transparency, interpretability, and acceptability of NMA analyses.

Study Design and Settings: The key components of NMA results were identified based on recommendations by agencies such as the National Institute for Health and Care Excellence (United Kingdom). Three novel graphs were designed to amalgamate the identified components using familiar graphical tools such as the bar, line, or pie charts and adhering to good graphical design principles.

Results: Three key components for presentation of NMA results were identified, namely relative effects and their uncertainty, probability of an intervention being best, and between-study heterogeneity. Two of the three graphs developed present results (for each pairwise comparison of interventions in the network) obtained from both NMA and standard pairwise meta-analysis for easy comparison. They also include options to display the probability best, ranking statistics, heterogeneity, and prediction intervals. The third graph presents rankings of interventions in terms of their effectiveness to enable clinicians to easily identify ''top-ranking'' interventions.

Conclusions: The graphical tools presented can display results tailored to the research question of interest, and targeted at a whole spectrum of users from the technical analyst to the nontechnical clinician. © 2014 Elsevier Inc. All rights reserved.

Keywords: Network meta-analysis; Graphical displays, presentational tools, summary forest plot, ranking, probability best

1. Introduction

Until recently, systematic reviews and health technology assessments (HTAs) have been limited to pairwise comparisons of interventions where direct evidence exists. However, often there is an array of candidate interventions relevant to the clinical question of interest, thus an analysis comparing all the relevant interventions may be more appropriate and useful to decision makers. Methodology to address this issue, which has increasingly been applied, is network meta-analysis (NMA; also known as mixed [or multiple] treatment comparisons) [1−4]. Despite the increase in the use of NMA, there is no established graphical presentational standard for reporting the results of NMA analogous to the forest plot [5] for standard pairwise meta-analysis (PWMA) [6,7].

Herein, we propose three novel graphical tools that aim to present NMA results in a clear and concise manner that combine both graphs and numerical estimates for optimal interpretation of NMA results and with built-in alternative display options to satisfy the needs of different audiences. General principles of graphical excellence for presenting data [8−10], in a manner that highlight and organize the data effectively, were used. This included reducing non-data ink; enhancing data ink; and grouping, prioritizing, and sequencing the data.

2. What is NMA?

The NMA is a recent development in evidence synthesis that extends the functionality of standard PWMA to allow for a simultaneous and coherent comparison of multiple interventions using an evidence base of trials that individually may

Conflicts of interests: A.J.S., N.J.C., and S.B. have received a research grant from Janssen Research and Development under which the research presented here was conducted; A.J.S. does consultancy giving advice on statistical analyses for Janssen Research and Development before receiving the grant. S.H.T., N.J.W., and D.M.C. have no potential conflicts of interests to declare.

* Corresponding author. Tel.: (+44) 116 229 7255; fax: (+44) 116 229 7250.
E-mail address: sht10@leicester.ac.uk (S.H. Tan).

0895-4356/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jclinepi.2013.11.006

What is new?

- Network meta-analyses generate large amounts of outputs that make reporting of key results challenging, leading to variable reporting styles and often suboptimal reporting of the results.
- Three graphical tools are proposed: two reporting the key results of NMA (alongside pairwise meta-analysis results), whereas the third summarizes the overall ranking of the interventions in terms of effectiveness.
- These graphical tools are designed to be tailored to display results relevant to the research question of interest, and the different formats are aimed to target both analysts and clinicians.
- Standardizing graphical tools for presenting NMA results would increase the acceptability, accessibility, transparency, and interpretability of NMA analyses.
- Software for the implementation of the graphical tools are freely available.

not compare all the treatment options of interest. Advantages of NMA include: (1) preservation of within-trial randomization when combining randomized controlled trials (RCTs) evidence (ie, NMA is performed using the relative effectiveness results of randomized arms of interventions from each trial included in the network—hence there is no breaking of randomization when synthesizing the results), (2) transparency of the framework (ie, no need for ''back of the envelope'' indirect comparisons based on a series of PWMAs), and (3) potential reduction of uncertainty owing to the inclusion of more data.

Owing to the inherent feature of NMA to compare multiple interventions simultaneously, there has been rapid growth in the number of published clinical articles that use NMA for the synthesis of evidence from clinical trials, as well as, tutorial articles that focus on educating clinicians and methodologists alike on the fundamentals of NMA and how to interpret NMA results presented in journal articles. For example, Salanti [11] summarizes what the principles of NMA are, and its benefits and concerns as a next generation evidence synthesis tool. Articles by Dias et al. [12,13] provided technical guidance on the conduct of NMA through the use of tutorial examples, which included useful program codes to facilitate the analysis and enhance the understanding. Other tutorial articles with greater relevance to clinicians on understanding the core concepts of NMA, interpreting results from published NMA, and hence applying it to real-life clinical situation were published recently in medical journals, for example, by Mills et al. [14,15] and Cipriani et al. [16].

Given the many advantages and the increased accessibility by the publication of the tutorials, the popularity and use of NMA have increased. However, the NMAs generate large numbers of results compared with PWMA; for example, an NMA including five different treatment regimens generates 10 pairwise comparisons; and this increases to 45 pairwise comparisons when 10 different treatment regimens are included. Presenting such large numbers of results can be challenging, especially when NMA is used to evaluate a number of different outcome measures within the area of interest. Two recent reviews on the reporting of NMA results highlighted the variability in reporting styles [7,17] in terms of both the content (eg, relative effect estimates, the probability that a treatment is most effective compared with all other treatments included in the network analysis [referred to subsequently as probability best], and so on) and presentational form (eg, table, text, and graph), and called for additional guidance and presentational tools for reporting NMA results to aid ease of interpretability.

3. Which NMA results are important?

A recent review by Tan et al. [7] on the reporting of NMA results in UK National Institute for Health Research HTA reports found that the most often reported NMA results included relative effects of comparative pairs of interventions, absolute effects of interventions, and probability best, all of which are recommended in the published NMA methods guidance documents by agencies such as the National Institute for Health and Care Excellence (NICE) [18] or International Society For Pharmacoeconomics and Outcomes Research [19,20]. Another statistic used in the reporting of NMAs, although not reported in the HTA reports reviewed, is the order of preference of an intervention among a number of interventions (ie, the ranking of an intervention, where the probability that an intervention is rank 1 is the probability best statistic). The ranks may be presented as summary statistics (eg, mean/median rank and surface under the cumulative ranking curve [SUCRA] [21]) or graphical representations of the distribution of ranks (eg, rankograms/barplots) indicating the probability that a given intervention is first, second, third best, and so on when compared with all other interventions in the network. In addition to the above, PWMA results are reported in the HTA reports, sometimes alongside NMA results to allow informal consistency checks to be made. Prediction intervals (the interval indicating the likely location for the underlying effect in a new study), although not routinely reported, have recently been advocated [22] for the reporting of the impact of heterogeneity in evidence synthesis.

4. Data set

As an illustrative example to present the graphical tools developed, we selected a recently published study that used

NMA to investigate the use of tocolytic therapy for preterm child delivery [23]. This published NMA included 95 RCTs and considered eight classes of drugs for the treatment of preterm delivery (See Fig. 2 of Haas et al. [23] for the network of interventions and trials included in the NMA). The primary outcome measurement was 48-h delay in delivery and the analysis was performed on the odds ratio scale using 55 RCTs. Other secondary outcomes were also analyzed in the study; but for the illustration of the graphs proposed in this article, only the results of the primary outcome measure will be displayed; the graphs for the other outcomes will have a similar display format.

In the original article [23], key analysis results of the primary endpoint, such as NMA and PWMA odds ratios, probability best, and rankogram were presented separately using tables and figures (Table 1 and Figs. 3 and 7 in the original article—also see Appendix B at www.jclinepi.com of this article). The graphical tools proposed here are designed to consolidate the key results into a single figure that enable easier referencing of results for the authors and ease of interpretation of the results for clinicians and academics.

5. Graphical tools

In this section, three graphical tools are presented that aim to amalgamate the important NMA results—identified in the section ''Which NMA results are important?''—to aid readability and maximize interpretation in NMA reports. Two of the graphs present relative effects of comparative pairs of interventions, probability best, ranking statistics, and heterogeneity estimates. They also present the results of the PWMA alongside the NMA results to allow informal checks for consistency of results to be made easily. The primary aim of the third plot was to give a simple summary of the order of preference of interventions in terms of effectiveness.

The different graphical displays were developed with different target audiences in mind. With academics and statisticians/analysts in mind, the main objective was to graphically present all key NMA results on a single graph, while ensuring interpretability through clear presentation; this also aimed to help meet restrictions on the number of tables and figures often enforced by research journals (using graphs 1 and 2 below). Although completeness of NMA results presentation may be desired by the academics and analysts, clinicians and decision makers in health care are more likely to be interested in visualizing the overall conclusions of the analysis by presenting the rankings of all interventions in terms of their effectiveness (ie, highlighting ''top-ranking'' interventions; graph 3).

5.1. Graph 1: summary forest plot matrix

5.1.1. Description

The first plot, referred to as the summary forest plot (SFP) matrix, is shown in Fig. 1. The plot design is similar to a scatterplot matrix often used for the investigations of correlations. Along the diagonals, the interventions included in the network are displayed. These interventions may be ordered, for example, by their median rank, as is done here, to highlight the most relevant comparisons by placing them at the top of the graph. Below the diagonal, in the lower triangle of the plot, SFPs for all possible combinations of the intervention pairs analyzed in the NMA—in black color—are presented above the PWMA results—in gray color—to aid visual assessment of consistency between the two analyses (the intervention labeled horizontally to the right of the plot is compared with the intervention labeled vertically above and clear labeling of the axes is given for each ''plot element'' on the bottom of the matrix). The SFPs display the point estimates of effect size (drawn as a square) with 95% confidence/credible intervals (CrIs) and 95% prediction intervals (shown by two-tiered error bars). Any summary plot without a gray-colored estimate indicates a comparison for which no head-to-head trials exist. The corresponding numerical estimates of comparative effectiveness are presented above the diagonal in the upper triangle and are presented as a ''mirror image'' to the SFPs taking the diagonal as the mirror line. To assist in understanding the heterogeneity of the studies in the network, the numerical estimate of between-study variance (ie, heterogeneity) is reported below the matrix. Alternating shadings of each plot element is used to improve readability (a technique often used in rail/bus timetables). Also included in the matrix, along the diagonal, are the median ranks together with rankograms, which provide the full probability distribution of rankings for each intervention.

For the example in Fig. 1, the drug class of prostaglandin inhibitors (row 1) is most likely the best intervention with the highest median rank (1) and probability best statistic as shown on the rankogram (0.80—the height of the density at rank 1 [x-axis]). The effectiveness of the interventions ranked second and third based on median rank (magnesium sulfate and calcium channel blockers) relative to prostaglandin inhibitors are given by the odds ratio of 0.53 (95% CrI: 0.24−1.20) and 0.51 (95% CrI: 0.20−1.50), respectively. The lower triangle of the SFP matrix allows the reader to easily identify pairs of interventions for which there were no head-to-head trials. In the example presented in Fig. 1, the drug class of prostaglandin inhibitors is compared directly, in head-to-head trials, with all interventions except for the drug classes: others, oxytocin receptor blockers, and nitrates (as indicated by the lack of PWMA estimate below the NMA estimate). This graph allows readers/clinicians to interpret key results of an NMA using a single plot, to compare the NMA and PWMA results and to identify which pairs of interventions could not be compared in a PWMA (owing to a lack of head-to-head trials). These functionalities, together with the prediction intervals presented in the graphs, could be used to guide potential areas of future clinical trials/epidemiological research studies.

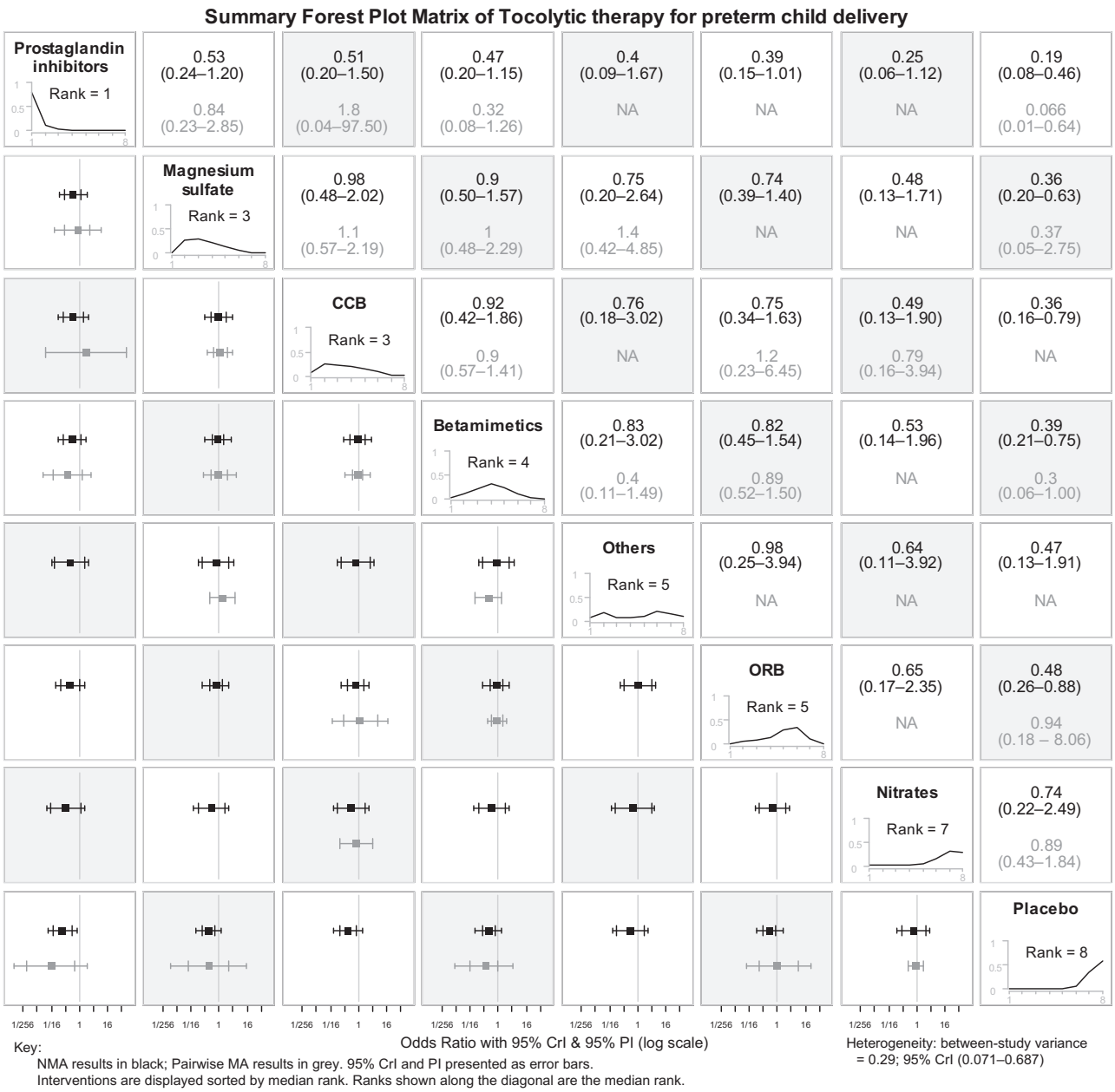

Fig. 1. Summary forest plot matrix. CCB, calcium channel blocker; ORB, oxytocin receptor blocker; NMA, network meta-analysis; CrI, credible interval; PI, prediction interval.

5.1.2. Advantages and limitations

Traditional forest plots display individual study effects together with the summary estimates to enable readers to assess the effects of each study, how different they are from one another and from the summary estimates, and their influence on the summary estimates. As much as it is desirable to display individual studies used in an NMA, it is cumbersome as the number of studies included in an NMA can often be large. Instead the graphical tools developed aimed to use the traditional forest plots, familiar to a great number of audiences in the medical area, to display the summary estimates from both NMA and PWMA and placing them side by side. The deliberate placement of the PWMA alongside the NMA results is to allow clinicians to directly address the question that naturally arises with NMA, that is, how different the results of NMA (that uses a network of trials) are compared with the results of traditional PWMA of head-to-head trials.

Along the diagonal, key NMA summary statistics (such as the median ranks together with rankograms, which provide the full probability distribution of rankings for each intervention) that are commonly reported separately are included in the graph. By including these statistics on the same plot as the relative effects, it enables the reader to instantly identify which intervention is most likely to be the best and read its comparative

effects with all the other remaining interventions (ordering on the median rank statistics further facilitates this by ensuring the ''best'' interventions are placed at the top of the plot).

Owing to the matrix square design of the SFP matrix, we believe that it works best for networks that are of moderate size (<10 interventions). Displaying NMA results of larger network will evidently require the SFP matrix to be separated into pages in the multiples of 2, hence reducing the readability and ease of interpretation of the NMA results. As such, we have included options to sort in terms of key NMA summary statistics and print a user-specified range of interventions, which will be discussed in greater details in the display options section.

5.1.3. Display options

The SFP matrix shown in Fig. 1 is one variant of the many that can be displayed. In its simplest form, the SFP matrix contains only the NMA and PWMA SFPs and estimates (with 95% CrI), with only the intervention names displayed along the diagonal and the heterogeneity estimates presented. Prediction intervals as shown in Fig. 1 can be optionally included in the graph. Further NMA results components such as the ranking and probability best statistics can be optionally included in the graph and displayed along the diagonal as shown in Fig. 1 where the rankograms were displayed. Ranking statistics can be displayed in the form of (1) rankogram with median rank, (2) bar chart with mean rank, or (3) the SUCRA estimates with cumulative ranking probability plots. Probability can best be displayed with a pie chart with the probability estimates.

Apart from the display of key NMA results components, options to sort or reduce the number of interventions displayed in the graphs are available (with caution notes displayed as footnotes in the graphs to remind readers of the actual number of interventions used in the NMA to produce the displayed results). Although we recommend the presentation of all pairwise comparisons in the network, we also acknowledge that it is sometimes necessary to display a reduced set of interventions, especially in the case of large networks. It may be helpful to clinicians and decision makers to restrict presentation of the NMA results to that of the top 5 or 10 ranking interventions when that the network contains say 20 interventions or more. The NMA components that can be used for sorting the results are (1) median rank, (2) mean rank, (3) SUCRA percentages, (4) probability best, and (5) relative treatment effect compared with the treatment coded as 1 (which is commonly placebo or standard of care) in the analysis. Footnotes in the graphs can also be removed wherever it is necessary. Illustrative examples of SFP matrix plots in its simplest format and with SUCRA percentages are shown in Appendix A at www.jclinepi.com.

5.2. Graph 2: SFP table

5.2.1. Description

The second graph, referred to as the SFP table, is shown in Fig. 2. This plot uses the presentational style of the traditional forest plot where the numerical estimates are reported alongside the SFPs. The SFP table presents results for all possible combinations of intervention comparisons with each intervention in the second column compared with the intervention listed in the first column. Similar to the SFP matrix previously described, the interventions have been ordered by their median rank. The third column reports the number of head-to-head trials that compare the two interventions listed in columns 1 and 2 (a feature not incorporated in the SFP matrix). In column 4, the numerical estimates of the relative effects with corresponding 95% CrI are presented for the NMA with the PWMA results directly below in gray. Finally, column 5 presents the SFPs; again the PWMA results are presented below the NMA results to allow visual assessment of consistency between the two analyses. Similar to the SFP matrix, the display of the prediction intervals alongside the CrIs on the SFP is optional. An estimate of heterogeneity across the trials included in the network is also presented. Median ranks of all interventions are also reported in this graph, numerically and graphically using a slider bar format (full rankograms, as presented on the SFP matrix, were problematic for the SFP table and difficult to read).

5.2.2. Advantages and limitations

One advantage of this plot over the matrix format is that the reference line of the SFPs for all pairwise comparisons is drawn on the same vertical line, hence facilitating the assessment of differences in comparative estimates and their precision between treatment pairs. Another advantage of this reporting style is that the NMA results from large networks can be reported more easily with the SFP table extending to multiple pages, where necessary.

Key NMA summary statistics such as the median rank, mean rank, SUCRA percentages, and probability best statistics are presented in the top first box of NMA results. This is a result of the reduction in comparative pairs by one as the primary comparator moves to the next intervention in the NMA. As such for an NMA with eight interventions (like in our example data), there will only be seven boxes on the SFP table because the last intervention (in our example—placebo) would have been compared with all preceding interventions and will not be listed in column 1 (the primary comparator column). Besides not having the last intervention in column 1, the size of the boxes decreases as it moves toward the next intervention, making it challenging to include rankograms on the graph. This is not only a limitation of this graph design but is also an advantage of this graph as the key NMA summary statistics had to be placed in the top box and this, in turn, allows readers to compare the interventions without having to flip through pages of the table when the network is large.

Summary Forest Plot Table of Tocolytic therapy for preterm child delivery

	Comparators	Median Rank	H-H Trials	Odds Ratio (95% CrI)	Summary Forest Plot (log Scale)
Prostaglandin inhibitors	Magnesium sulfate	3	3	0.53 (0.24–1.20) 0.84 (0.23–2.85)	
	CCB	3	1	0.51 (0.20–1.50) 1.85 (0.04–97.50)	
	Betamimetics	4	3	0.47 (0.20–1.15) 0.32 (0.08–1.26)	
	Others	5	0	0.40 (0.09–1.67) NA	
	ORB	5	0	0.39 (0.15–1.01) NA	
	Nitrates	7	0	0.25 (0.06–1.12) NA	
	Placebo	8	2	0.19 (0.08–0.46) 0.07 (0.01–0.64)	
Magnesium sulfate	CCB		5	0.98 (0.48–2.02) 1.13 (0.57–2.19)	
	Betamimetics		7	0.90 (0.50–1.57) 1.00 (0.48–2.29)	
	Others		1	0.75 (0.20–2.64) 1.43 (0.42–4.85)	
	ORB		0	0.74 (0.39–1.40) NA	
	Nitrates		0	0.48 (0.13–1.71) NA	
	Placebo		3	0.36 (0.20–0.63) 0.37 (0.05–2.75)	
CCB	Betamimetics		14	0.92 (0.42–1.86) 0.90 (0.57–1.41)	
	Others		0	0.76 (0.18–3.02) NA	
	ORB		2	0.75 (0.34–1.63) 1.18 (0.23–6.45)	
	Nitrates		1	0.49 (0.13–1.90) 0.79 (0.16–3.94)	
	Placebo		0	0.36 (0.16–0.79) NA	
Betamimetics	Others		1	0.83 (0.21–3.02) 0.40 (0.11–1.49)	
	ORB		6	0.82 (0.45–1.54) 0.89 (0.52–1.50)	
	Nitrates		0	0.53 (0.14–1.96) NA	
	Placebo		4	0.39 (0.21–0.75) 0.30 (0.06–1.00)	
Others	ORB		0	0.98 (0.25–3.94) NA	
	Nitrates		0	0.64 (0.11–3.92) NA	
	Placebo		0	0.47 (0.13–1.91) NA	
ORB	Nitrates		0	0.65 (0.17–2.35) NA	
	Placebo		2	0.48 (0.26–0.88) 0.94 (0.18–8.06)	
Nitrates	Placebo		1	0.74 (0.22–2.49) 0.89 (0.43–1.84)	

Median Rank: 8 (Bottom) … 1 (Top)

1/256 1/64 1/16 1/4 1 4 16 64

Odds Ratio with 95% CrI & 95% PI (log scale)

Heterogeneity: between-study variance = 0.29; 95% CrI (0.071–0.687)

Key: NMA results in black; Pairwise MA results in grey. 95% CrI and PI presented as error bars. Interventions are displayed sorted by median rank.

Fig. 2. Summary forest plot table. H-H trials, head-to-head trials; CCB, calcium channel blocker; ORB, oxytocin receptor blocker; NMA, network meta-analysis; CrI, credible interval; PI, prediction interval.

5.2.3. Display options

The SFP table shown in Fig. 2 is one variant of the many that can be displayed. In its simplest form, the SFP table contains only the NMA and PWMA SFPs and estimates (with 95% CrI), and the number of head-to-head trials for each pair of intervention comparisons. Prediction intervals as shown in Fig. 2 can be optionally included in the graph. Other NMA results components such as the ranking and probability best statistics can be optionally included in the graph and displayed in the first set of intervention comparisons as shown in Fig. 2 where the median ranks were displayed. Choice of display of the ranking statistics are (1) median rank presented using slider bar, (2) mean rank presented using slider bar, and (3) SUCRA percentages. Probability is best displayed with a pie chart alongside the probability estimates.

Similar to SFP matrix, options to sort or reduce the number of interventions displayed in the graphs are available (with caution notes displayed). The NMA results components that can be used for sorting the results are (1) median rank, (2) mean rank, (3) SUCRA percentages, (4) probability best, and (5) relative treatment effect compared with the treatment coded as 1 (which is commonly placebo or standard of care) in the analysis. Illustrative examples of SFP table plots in its simplest format and with SUCRA percentages are shown in Appendix A at www.jclinepi.com.

5.3. Graph 3: median rank chart

5.3.1. Description

The third graph, shown in Fig. 3, presents the median ranks of all interventions included in the NMA with the aim to ''simplify'' the presentation of rankings in NMA. A color intensity scheme is used in this graph to help draw attention to the best treatment(s) (using black ink in the lightest zone at the top of the chart) while simultaneously highlight the worst treatment(s) (in the darkest zone at the bottom of the chart). In our example (Fig. 3), the drug class of prostaglandin inhibitors is most likely to be the best with a median rank of 1, whereas the drug classes of nitrates and placebo are the worst, and the five other interventions have similar rankings between these extremes.

5.3.2. Advantages and limitations

As this graph allows all interventions included in the NMA to be presented in a single graph that can be printed on a single page, we believe that it is a particularly useful graphical tool when the network contains a large number of interventions. This graph provides readers with only the median ranking of the interventions; hence unlike the SFP matrix and SFP table, this does not provide quantitative information on the differences in efficacy estimates between the interventions included in the analysis.

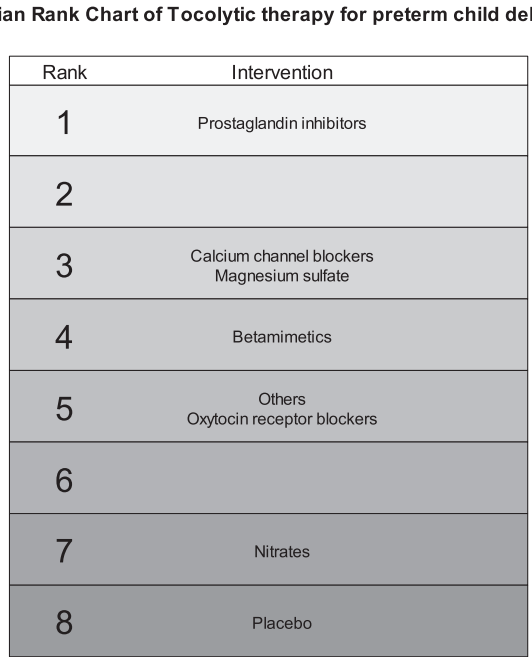

Fig. 3. Median rank chart.

6. Software

Functions for creating the graphs in the form presented in this article have been written in the R software language and is available for download from: https://www2.le.ac.uk/departments/health-sciences/research/biostats/sb-supplementary-materials/nma-graphics.

7. Discussion

In this article, we have presented three graphical tools to aid clear presentation and facilitate interpretation of NMA results. The SFP matrix and SFP table provide a comprehensive presentation of the important NMA and PWMA results displayed on a single plot. These plots not only enable easy comparison of NMA and PWMA results but also assist to reduce the number of tables and/or figures required for all relevant results to be presented in the main text of a journal article where space is often limited. The median rank chart complements the SFP matrix or the SFP table by providing a visual summary of each intervention's median ranking within the network of interest, thus enabling decision makers and clinicians to easily identify the ''top-ranking'' intervention(s) in terms of effectiveness. The graphs have been developed to display relative effectiveness results and are reported here on the odds ratio scale; however, they can also be used to present other outcome measures (such as continuous data, hazard ratios, and so on). An example of its use for the presentation of NMA results on continuous outcome data is shown in Appendix C at www.jclinepi.com.

Visual design principles were applied in the development of the graphs. The NMA results presented in the

SFP matrix and SFP table combine three main groups of results, namely (1) the SFP graphs, (2) the numerical estimates corresponding to the SFPs, and (3) the ranking or probability best statistics. In the SFP matrix, the most important intervention (eg, usual care/placebo or the top-ranking intervention when sorted by ranking) is usually at the top left hand corner and hence the numerical estimates were strategically placed in the upper triangle of the matrix plot. This allows readers to read the relative effectiveness of interventions compared with the most important intervention easily as reading from left to right is generally the order that readers will scan a page. This therefore allows the SFPs to be placed in the lower triangle where the x-axis for the plots can be placed at the bottom of the matrix, which is conventional with the usual placement of the x-axis on graphs. The ranking or probability best statistics are placed along the diagonal with the intervention names in an enclosure so that readers can readily know what intervention the statistics correspond to. The NMA and PWMA results for each intervention comparative pairs are grouped and placed in an enclosure to allow the assessment of consistency of the results.

In the SFP table, the three main groups of data are presented from left to right. First, the intervention names together with the ranking or probability best statistics; second, the numerical estimates of the relative effectiveness; and lastly, the SFPs. In this design, the texts help to complement and enhance the SFPs that follow. Enclosures in the form of boxes present NMA results grouped by the reference intervention, allowing readers to easily recognize that all SFPs and numerical estimates within an enclosure are compared with the same reference intervention.

As both the graphs are developed for the presentation of NMA results, the NMA SFPs and numerical estimates are presented in stronger (black) ink to highlight the main results, whereas the PWMA results are presented in lighter (gray) ink, displayed for comparison. The intensity of the colors of the enclosures and axes, which do not represent the key results, are reduced to a minimum, whereas light intermittent shading of enclosures in the SFP Matrix is used to improve readability.

The median rank chart presents the top-ranking intervention at the top, using the concept that readers will read from top to bottom, so attention is drawn to the top-ranking intervention first. Also, the top-ranking intervention is written in black ink in the lightest background shading compared with the worse intervention in the darkest background shading, using the visual perception concept of contrast to highlight the most important result [9].

There has been an evolution of reporting standards initially for PWMA [24] and more recently for NMA [19,20]. Furthermore, technical support documents [12,25−30] (commissioned by NICE), and a series of evidence synthesis for medical decision-making tutorial articles [13,31−36] have recently been published providing technical details of the implications and implementation of NMA methodology as well as guidance on reporting. These all highlight the need for a clear description of the NMA statistical model, and its assumptions, together with model fit statistics, including checks for inconsistency. Additionally, presentation of the evidence structure, in the form of a network diagram [37], is also recommended. We believe that the graphical tools presented in our article improve existing methods to report the results of NMA and as such complement the aforementioned guidance documents. The graphs proposed focused mainly on the presentation of single outcome but can potentially be adopted to present multiple outcomes in the future. Ultimately, our hope is that such displays will be recommended in updated guidance published in the future.

Acknowledgments

The authors thank David Haas (Indiana University) for providing the data for the example presented in this article. They are also grateful to the statisticians who attended the Methods in Meta-analysis meeting at the Royal Statistical Society in London on December 14, 2012 for their valuable comments on our graphs. Finally, they thank Soledad Cepeda and Jesse Berlin (Janssen Research and Development) for their useful comments on an earlier version of this manuscript and the two anonymous reviewers for their comments and suggestions.

Funding: The development of the graphical tools was funded by Department of Epidemiology, Janssen Research and Development. N.J.W. and D.M.C. were supported by Medical Research Council fellowships.

Appendix

Supplementary material

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.11.006.

References

[1] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897−900. http://dx.doi.org/10.1136/bmj.331.7521.897.

[2] Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006;24(1):1−19.

[3] Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;15:2733−49. doi: 10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.

[4] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105−24. http://dx.doi.org/10.1002/sim.1875.

[5] Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322:1479−80.

[6] Fadda V, Maratea D, Trippoli S, Messori A. Network meta-analysis. Results can be summarised in a simple figure. BMJ 2011;342:d1555. http://dx.doi.org/10.1136/bmj.d1555.

[7] Tan SH, Bujkiewicz S, Sutton AJ, Dequen P, Cooper NJ. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. J Health Serv Res Policy 2013; 18:224−32. http://dx.doi.org/10.1177/1355819613498379.

[8] Cleveland WS. The elements of graphing data. Summit, NJ: Hobart Press; 1994.

[9] Few S. Show me the numbers: designing tables and graphs to enlighten. Oakland, CA: Analytics Press; 2004.

[10] Tufte ER. The visual display of quantitative information. 2nd ed. Cheshire, CT: Graphics Press; 2001.

[11] Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods 2012;3(2):80−97. http://dx.doi.org/10.1002/jrsm.1037.

[12] Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011; Last updated August 2011. Available at http://www.nicedsu.org.uk. Accessed January 31, 2012.

[13] Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making 2013;33:607−17. http://dx.doi.org/10.1177/0272989X12458724.

[14] Mills EJ, Ioannidis JP, Thorlund K, Schunemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 2012;308:1246−53. http://dx.doi.org/10.1001/2012.jama.11228.

[15] Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. BMJ 2013;346:f2914. http://dx.doi.org/10.1136/bmj.f2914.

[16] Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013;159: 130−7. http://dx.doi.org/10.7326/0003-4819-159-2-201307160-00008.

[17] Coleman CI, Phung OJ, Cappelleri JC, Baker WL, Kluger J, White CM, et al. Use of mixed treatment comparisons in systematic reviews. Methods research report. (Prepared by the University of Connecticut/Hartford Hospital Evidence-based Practice Center under Contract No. 290-2007-10067-I.) AHRQ Publication No. 12-EHC119-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012:Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm. Accessed October 15, 2012.

[18] NICE. Guide to the methods of technology appraisal. London, UK: National Institute for Health and Care Excellence (NICE); 2008.

[19] Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14(4):429−37. http://dx.doi.org/10.1016/j.jval.2011.01.011.

[20] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011;14(4):417−28. http://dx.doi.org/10.1016/j.jval.2011.04.002.

[21] Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64:163−71. http://dx.doi.org/10.1016/j.jclinepi.2010.03.016.

[22] Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;342:d549. http://dx.doi.org/10.1136/bmj.d549.

[23] Haas DM, Caldwell DM, Kirkpatrick P, McIntosh JJ, Welton NJ. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. BMJ 2012;345:e6226. http://dx.doi.org/10.1136/bmj.e6226.

[24] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009;151: W65−94. 0000605-200908180-00136 [pii].

[25] Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 1: introduction to evidence synthesis for decision making. 2011. Available at http://www.nicedsu.org.uk. Accessed January 31, 2012.

[26] Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU technical support document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011. Available at http://www.nicedsu.org.uk. Accessed January 31, 2012.

[27] Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU technical support document 4: inconsistency in networks of evidence based on randomised controlled trials. 2011. Available at http://www.nicedsu.org.uk. Accessed January 31, 2012.

[28] Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 5: evidence synthesis in the baseline natural history model. 2011. Available at http://www.nicedsu.org.uk. Accessed January 31, 2012.

[29] Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU technical support document 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. 2011. Available at http://www.nicedsu.org.uk. Accessed January 31, 2012.

[30] Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. NICE DSU technical support document 7: evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. 2012. Available at http://www.nicedsu.org.uk. Accessed May 1, 2012.

[31] Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. Med Decis Making 2013;33:597−606. http://dx.doi.org/10.1177/0272989X13487604.

[32] Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity−subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making 2013;33:618−40. http://dx.doi.org/10.1177/0272989X13485157.

[33] Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making 2013;33:641−56. http://dx.doi.org/10.1177/0272989X12455847.

[34] Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. Med Decis Making 2013;33:657−70. http://dx.doi.org/10.1177/0272989X13485155.

[35] Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. Med Decis Making 2013;33:671−8. http://dx.doi.org/10.1177/0272989X13487257.

[36] Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making 2013;33:679−91. http://dx.doi.org/10.1177/0272989X13485156.

[37] Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. Ann Intern Med 2008;148:544−53.

Bibliography

ACHANA, F. A., et al. 2014. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol,* 14**,** 92.

ADES, A. E. 2003. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med,* 22**,** 2995-3016.

ADES, A. E., et al. 2012. NICE DSU Technical Support Document 7: Evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. Available from http://www.nicedsu.org.uk.

ADES, A. E., et al. 2013. Evidence synthesis for decision making 7: a reviewer's checklist. *Med Decis Making,* 33**,** 679-91.

ADES, A. E., et al. 2010. Network meta-analysis with competing risk outcomes. *Value Health,* 13**,** 976-83.

ADES, A. E., et al. 2006. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics,* 24**,** 1-19.

ALTMAN, D. G. 1991. *Practical Statistics for Medical Research,* United States of America, Chapman & Hall.

ASARIA, M., et al. 2011. Fingolimod for the treatment of relapsing remitting multiple sclerosis: A Single Technology Appraisal. CRD and CHE Technology Assessment Group.

BAFETA, A., et al. 2014. Reporting of results from network meta-analyses: methodological systematic review. *BMJ,* 348**,** g1741.

BERRY, W., et al. 2002. Phase III study of mitoxantrone plus low dose prednisone versus low dose prednisone alone in patients with asymptomatic hormone refractory prostate cancer. *J Urol,* 168**,** 2439-43.

BLACK, W. C. 1990. The CE plane: a graphic representation of cost-effectiveness. *Med Decis Making,* 10**,** 212-4.

BLOOMFIELD, D. J., et al. 1998. Economic evaluation of chemotherapy with mitoxantrone plus prednisone for symptomatic hormone-resistant prostate cancer: based on a Canadian randomized trial with palliative end points. *J Clin Oncol,* 16**,** 2272-9.

BORENSTEIN, M., et al. 2009. *Introduction to Meta-analysis,* Chichester, Wiley.

BORMANN, I. 2013. DigitizeIt 2.0.3 program. Braunschweig, Germany. URL http://www.digitizeit.de.

BOUCHER, R., et al. 2015. Understanding the process of simulating individual patient level data using an Illness-Death Model framework when only summary data are available in order to address treatment switching. *ISPOR European Congress.* Milan.

BRIGGS, A., et al. 2006. *Decision Modelling for Health Economic Evaluation,* United States, Oxford University Press.

BRIGGS, A. & FENN, P. 1998. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ,* 7**,** 723-40.

BRIGGS, A. & SCULPHER, M. 1998. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics,* 13**,** 397-409.

BUCHER, H. C., et al. 1997. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol,* 50**,** 683-91.

BUJKIEWICZ, S., et al. 2015a. Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Stat Med,* (In press).

BUJKIEWICZ, S., et al. 2015b. Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints. *Stat Methods Med Res,* (In press).

BUJKIEWICZ, S., et al. 2013. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Stat Med,* 32**,** 3926-43.

CALDWELL, D. M., et al. 2005. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ,* 331**,** 897-900.

CHAIMANI, A., et al. 2013. Graphical tools for network meta-analysis in STATA. *PLoS One,* 8**,** e76654.

CHEUNG, Y. B., et al. 2009. Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D utility index. *Value Health,* 12**,** 371-6.

CIPRIANI, A., et al. 2013. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med,* 159**,** 130-7.

CLEVELAND, W. S. 1994. *The Elements of Graphing Data,* Summit, NJ, USA, Hobart Press.

COHEN, J. A., et al. 2010. Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. *N Engl J Med,* 362**,** 402-15.

COLEMAN, C. I., et al. 2012. Use of Mixed Treatment Comparisons in Systematic Reviews. Methods Research Report. (Prepared by the University of Connecticut/Hartford Hospital Evidence-based Practice Center under Contract No. 290-2007-10067-I.) AHRQ Publication No. 12-EHC119-EF. Rockville, MD: Agency for Healthcare Research and Quality. Available at www.effectivehealthcare.ahrq.gov/reports/final.cfm. (Date Last Accessed: 15 October 2012).

COLLINS, R., et al. 2007. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. *Health Technol Assess,* 11**,** iii-iv, xv-xviii, 1-179.

CROTT, R. & BRIGGS, A. 2010. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur J Health Econ,* 11**,** 427-34.

DANIELS, M. J. & HUGHES, M. D. 1997. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med,* 16**,** 1965-82.

DAVIS, S., et al. 2012. NICE DSU Report: A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer. Available from http://www.nicedsu.org.uk.

DERSIMONIAN, R. & KACKER, R. 2007. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials,* 28**,** 105-14.

DERSIMONIAN, R. & LAIRD, N. 1986. Meta-analysis in clinical trials. *Control Clin Trials,* 7**,** 177-88.

DIAS, S., et al. 2013a. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making,* 33**,** 607-17.

DIAS, S., et al. 2011a. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Available from http://www.nicedsu.org.uk.

DIAS, S., et al. 2011b. NICE DSU Technical Support Document 6:Embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. Available from http://www.nicedsu.org.uk.

DIAS, S., et al. 2013b. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making,* 33**,** 618-40.

DIAS, S., et al. 2013c. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Med Decis Making,* 33**,** 671-8.

DIAS, S., et al. 2010. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med,* 29**,** 932-44.

DIAS, S., et al. 2011c. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. Available from http://www.nicedsu.org.uk

DIAS, S., et al. 2011d. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. Available from http://www.nicedsu.org.uk.

DIAS, S., et al. 2011e. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model.: Available from http://www.nicedsu.org.uk.

DIAS, S., et al. 2013d. Evidence synthesis for decision making 1: introduction. *Med Decis Making,* 33**,** 597-606.

DIAS, S., et al. 2013e. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making,* 33**,** 657-70.

DIAS, S., et al. 2011f. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. Available from http://www.nicedsu.org.uk.

DIAS, S., et al. 2013f. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making,* 33**,** 641-56.

DICKERSIN, K., et al. 1987. Publication bias and clinical trials. *Control Clin Trials,* 8**,** 343-53.

DIELS, J., et al. 2015. Mapping FACT-P to EQ-5D in a large cross-sectional study of metastatic castration-resistant prostate cancer patients. *Qual Life Res,* 24**,** 591-8.

DOLAN, P. & ROBERTS, J. 2002. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. *Med Care,* 40**,** 442-6.

DUVAL, S. & TWEEDIE, R. 2000. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics,* 56**,** 455-63.

EFRON, B. & TIBSHIRANI, R. 1986. Bootstrap methods for standard errors, conÞdence intervals, and other measures of statistical accuracy. *Statistical Science,* 1**,** 54-75.

EFTHIMIOU, O., et al. 2014. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Stat Med,* 33**,** 2275-87.

EFTHIMIOU, O., et al. 2015. Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostatistics,* 16**,** 84-97.

EGGER, M., et al. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ,* 315**,** 629-34.

ERNST, D. S., et al. 2003. Randomized, double-blind, controlled trial of mitoxantrone/prednisone and clodronate versus mitoxantrone/prednisone and placebo in patients with hormone-refractory prostate cancer and pain. *J Clin Oncol,* 21**,** 3335-42.

FENWICK, E., et al. 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ,* 10**,** 779-87.

FENWICK, E., et al. 2004. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. *Health Econ,* 13**,** 405-15.

FEW, S. 2004. *Show Me the Numbers: Designing Tables and Graphs to Enlighten,* Oakland, CA, USA, Analytics Press.

FISK, J. D., et al. 2005. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *J Neurol Neurosurg Psychiatry,* 76**,** 58-63.

FOGARTY, E., et al. 2013. Relating health-related Quality of Life to disability progression in multiple sclerosis, using the 5-level EQ-5D. *Mult Scler,* 19**,** 1190-6.

GAROPOULOU, V., et al. 2014. The effect of an aquatic training program on walking ability and quality of life of patients with multiple sclerosis. *Journal of Physical Education and Sport,* 14**,** 106-114.

GARTLEHNER, G. & MOORE, C. G. 2008. Direct versus indirect comparisons: a summary of the evidence. *Int J Technol Assess Health Care,* 24**,** 170-7.

GLASS, G. V. 1976. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher,* 5**,** 3-8.

GOLD, R., et al. 2012. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *N Engl J Med,* 367**,** 1098-107.

GUYOT, P., et al. 2012. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol,* 12**,** 9.

HAAS, D. M., et al. 2012. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. *BMJ,* 345**,** e6226.

HALABI, S., et al. 2009. Progression-free survival as a predictor of overall survival in men with castrate-resistant prostate cancer. *J Clin Oncol,* 27**,** 2766-71.

HIGGINS, J. P. & THOMPSON, S. G. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med,* 21**,** 1539-58.

HIGGINS, J. P., et al. 2003. Measuring inconsistency in meta-analyses. *BMJ,* 327**,** 557-60.

HIGGINS, J. P. & WHITEHEAD, A. 1996. Borrowing strength from external trials in a meta-analysis. *Stat Med,* 15**,** 2733-49.

HOAGLIN, D. C., et al. 2011. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health,* 14**,** 429-37.

HONG, H., et al. 2013. *A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons.* Rockville (MD).

HUTTON, B., et al. 2015. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med,* 162**,** 777-84.

HUTTON, B., et al. 2014. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. *PLoS One,* 9**,** e92508.

JANSEN, J. P., et al. 2011. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health,* 14**,** 417-28.

KANTOFF, P. W., et al. 1999. Hydrocortisone with or without mitoxantrone in men with hormone-refractory prostate cancer: results of the cancer and leukemia group B 9182 study. *J Clin Oncol,* 17**,** 2506-13.

KAPPOS, L., et al. 2014. Quality of life outcomes with BG-12 (dimethyl fumarate) in patients with relapsing-remitting multiple sclerosis: the DEFINE study. *Mult Scler,* 20**,** 243-52.

KAPPOS, L., et al. 2010. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med,* 362**,** 387-401.

KOBELT, G., et al. 2001. Costs and Quality of Life in Multiple Sclerosis: An Observational Study in Germany. *The European Journal of Health Economics,* 2**,** 60-68.

KRISTON, L. 2013. Dealing with clinical heterogeneity in meta-analysis. Assumptions, methods, interpretation. *Int J Methods Psychiatr Res,* 22**,** 1-15.

KUSPINAR, A. & MAYO, N. E. 2013. Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health Qual Life Outcomes,* 11**,** 71.

LAMBERT, P. C., et al. 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med,* 24**,** 2401-28.

LEWIS, S. & CLARKE, M. 2001. Forest plots: trying to see the wood and the trees. *BMJ,* 322**,** 1479-80.

LIBERATI, A., et al. 2009a. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med,* 151**,** W65-94.

LIBERATI, A., et al. 2009b. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ,* 339**,** b2700.

LIMONE, B. L., et al. 2013. Estimation of the effect of dalfampridine-ER on health utility by mapping the MSWS-12 to the EQ-5D in multiple sclerosis patients. *Health Qual Life Outcomes,* 11**,** 105.

LU, G. & ADES, A. E. 2004. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med,* 23**,** 3105-24.

LU, G., et al. 2011. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Research Synthesis Methods,* 2**,** 43-60.

LU, L., et al. 2012. Cost-effectiveness analysis of degarelix for advanced hormone-dependent prostate cancer. *BJU Int,* 109**,** 1183-92.

LUMLEY, T. 2002. Network meta-analysis for indirect treatment comparisons. *Stat Med,* 21**,** 2313-24.

LUNN, D. J., et al. 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility *Statistics and Computing,* 10**,** 325-337.

MILLS, E. J., et al. 2012. How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA,* 308**,** 1246-53.

MILLS, E. J., et al. 2013. Demystifying trial networks and network meta-analysis. *BMJ,* 346**,** f2914.

MITOSEK-SZEWCZYK, K., et al. 2014. Quality of life in Polish patients with multiple sclerosis. *Adv Med Sci,* 59**,** 34-8.

MOHER, D., et al. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ,* 339**,** b2535.

MORENO, S. G., et al. 2009. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. *BMJ,* 339**,** b2981.

MOSS-MORRIS, R., et al. 2013. A randomized controlled trial of cognitive behavioral therapy (CBT) for adjusting to multiple sclerosis (the saMS trial): does CBT work and for whom does it work? *J Consult Clin Psychol,* 81**,** 251-62.

NICE 2004. Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence (NICE).

NICE 2008. Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence (NICE).

NICE 2013. Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence (NICE).

NOVARTIS PHARMACEUTICALS UK LTD 2011. Fingolimod for the treatment of relapsing-remitting multiple sclerosis in adults: NICE Single Technology Appraisal.

O'BRIEN, B. J. & BRIGGS, A. H. 2002. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Stat Methods Med Res,* 11**,** 455-68.

ORME, M., et al. 2007. The effect of disease, functional status, and relapses on the utility of people with multiple sclerosis in the UK. *Value Health,* 10**,** 54-60.

OUDARD, S., et al. 2005. Multicenter randomized phase II study of two schedules of docetaxel, estramustine, and prednisone versus mitoxantrone plus prednisone in patients with metastatic hormone-refractory prostate cancer. *J Clin Oncol,* 23**,** 3343-51.

OWEN, R. K., et al. 2015. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value Health,* 18**,** 116-26.

PARMAR, M. K., et al. 1998. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med,* 17**,** 2815-34.

PEARSON, K. 1904. Report on Certain Enteric Fever Inoculation Statistics. *BMJ,* 2**,** 1243-1246.

PENNINGTON, B. & DAVIS, S. 2014. Mapping from the Health Assessment Questionnaire to the EQ-5D: the impact of different algorithms on cost-effectiveness results. *Value Health,* 17**,** 762-71.

PETERS, J. L., et al. 2008. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol,* 61**,** 991-6.

PETERS, J. L., et al. 2010. Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 173**,** 575-591.

PETRYLAK, D. P., et al. 2004. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *N Engl J Med,* 351**,** 1513-20.

PUHAN, M. A., et al. 2014. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ,* 349**,** g5630.

PUTZKI, N., et al. 2009. Quality of life in 1000 patients with early relapsing-remitting multiple sclerosis. *Eur J Neurol,* 16**,** 713-20.

R CORE TEAM 2012. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. URL http://www.R-project.org/.

RABIN, R. & DE CHARRO, F. 2001. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med,* 33**,** 337-43.

REESE, J. P., et al. 2013. Preference-based Health status in a German outpatient cohort with multiple sclerosis. *Health Qual Life Outcomes,* 11**,** 162.

RILEY, R. D., et al. 2007. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med,* 26**,** 78-97.

RILEY, R. D., et al. 2011. Interpretation of random effects meta-analyses. *BMJ,* 342**,** d549.

RILEY, R. D., et al. 2015. Multivariate meta-analysis using individual participant data. *Res Synth Methods,* 6**,** 157-74.

RUCKER, G., et al. 2008. Arcsine test for publication bias in meta-analyses with binary outcomes. *Stat Med,* 27**,** 746-63.

SALANTI, G. 2012. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res. Synth. Method,* 3**,** 80-97.

SALANTI, G., et al. 2011. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol,* 64**,** 163-71.

SALANTI, G., et al. 2014. Evaluating the quality of evidence from a network meta-analysis. *PLoS One,* 9**,** e99682.

SALANTI, G., et al. 2008. Exploring the geometry of treatment networks. *Ann Intern Med,* 148**,** 544-53.

SANDBLOM, G., et al. 2004. A population-based study of pain and quality of life during the year before death in men with prostate cancer. *Br J Cancer,* 90**,** 1163-8.

SANOFI-AVENTIS 2005. Sponsor submission to the National Institute for Health and Clinical Excellence: Taxotere (docetaxel) in metastatic hormone refractory prostate cancer (mHRPC) [industry submission].

SONG, F., et al. 2011. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ,* 343**,** d4909.

SONNENBERG, F. A. & BECK, J. R. 1993. Markov models in medical decision making: a practical guide. *Med Decis Making,* 13**,** 322-38.

SPIEGELHALTER, D., et al. 2003. WinBUGS User Manual. Version 1.4 ed. Cambridge, England: MRC Biostatistics Unit.

STATACORP 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

STERNE, J. A. & EGGER, M. 2001. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol,* 54**,** 1046-55.

STERNE, J. A., et al. 2001. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ,* 323**,** 101-5.

STURTZ, S., et al. 2005. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software,* 12**,** 1-16.

SULLIVAN, S. M., et al. 2014. What guidance are researchers given on how to present network meta-analyses to end-users such as policymakers and clinicians? A systematic review. *PLoS One,* 9**,** e113277.

SUTTON, A. J., et al. 2000. *Methods for Meta-Analysis in Medical Research,* Chichester, Wiley.

TAN, S. H., et al. 2013. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. *J Health Serv Res Policy,* 18**,** 224-232.

TAN, S. H., et al. 2014. Novel presentational approaches were developed for reporting network meta-analysis. *J Clin Epidemiol,* 67**,** 672-80.

TANNOCK, I. F., et al. 2004. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *N Engl J Med,* 351**,** 1502-12.

TANNOCK, I. F., et al. 1996. Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian randomized trial with palliative end points. *J Clin Oncol,* 14**,** 1756-64.

TECKLE, P., et al. 2013. Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health Qual Life Outcomes,* 11**,** 203.

THOMPSON, S. G. & HIGGINS, J. P. T. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine,* 21**,** 1559-1573.

TUFTE, E. R. 2001. *The Visual Display of Quantitative Information,* Cheshire, CT, USA, Graphics Press.

VAN HOUWELINGEN, H. C., et al. 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med,* 21**,** 589-624.

VANEY, C., et al. 2012. Robotic-assisted step training (lokomat) not superior to equal intensity of over-ground rehabilitation in patients with multiple sclerosis. *Neurorehabil Neural Repair,* 26**,** 212-21.

WARREN, F. C., et al. 2014. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. *Stat Med,* 33**,** 2449-66.

WEI, Y. & HIGGINS, J. P. 2013. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med,* 32**,** 1191-205.

ZETTL, U. K., et al. 2013. Burden of disease in multiple sclerosis patients with spasticity in Germany: mobility improvement study (Move I). *Eur J Health Econ*.