Using multicentre RCT-based IPD to populate decision analytic CE models for location-specific decision making: a Bayesian approach

Saramago, Pedro*

* Centre for Health Economics, The University of York, Alcuin A, York, YO10 5DD, UK, Tel: +44 (0)1904 321453; Fax: +44 (0)1904 321402; E-mail: prsg500@york.ac.uk

Abstract

Objectives Applying the Bayesian approach to the analysis of IPD from multicentre/ multinational RCTs with the aim of estimating location-specific parameters to populate decision models for location-specific decision making.

Methods Multilevel or hierarchical modelling is the analytical framework used to handle hierarchical cost-effectiveness data. Hierarchical modelling was developed in a Bayesian framework, that is, the estimation of the parameters was performed by MCMC, which was used to populate the economic decision model. Bayesian probabilistic modelling was used to evaluate the decision problem and Bayesian shrinkage estimation procedures were used to obtain location-specific CE estimates.

Results Using data from a recently conducted economic analysis of the RITA 3 trial, location-specific cost-effectiveness measures were obtained and compared to the trial-wide results. For the analysed centres, the centre-specific CEPs showed higher variability in mean differential cost and mean differential QALY estimates compared to the trial wide results, with the latter having longer left tail estimate distribution. The majority of the location-specific ICER results show higher cost per QALY for the intervention strategy compared to the trial wide results (approx. £41,400/QALY). With respect to centre-specific CEACs, the curves for the selected centres display great variability across centres in cost-effectiveness for given values of the threshold, λ. If the decision maker is willing to pay £50,000 for an additional QALY, the probability that the intervention strategy is cost-effective is, for instance, 0.34 for centre 37, compared to the 0.65 for the trial wide results.

Conclusions This work showed two important results. Firstly, it was demonstrated, through the use of one illustrative example, how a trial-based CEA may be implemented within a Bayesian framework and evaluated using Gibbs sampling MCMC methods. In particular it has provided the 'building blocks' for extending the modelling framework to allow the incorporation of more relevant evidence: (i) data may be added in a prior distribution format; and (ii) data from different study designs (e.g. RCTs, observational studies together with expert judgement). Secondly, it was demonstrated how Bayesian hierarchical modelling could be used to estimate more appropriate cluster-specific parameters for use in DAMs where IPD from a multi-location trial are available. Bayesian hierarchical modelling estimates can be used to explore correctly the variability between centres/countries of the cost-effectiveness results allowing the correct quantification of uncertainty by adjusting the standard errors to reflect the estimates variability both within and between locations.

Keywords Bayesian methods; Multilevel/hierarchical modelling; cost-effectiveness analysis; decision analytic models; individual patient data.

This work is considered a shorter version of the dissertation presented for the degree of Master of Science in Health Economics – Centre for Health Economics, University of York, September 2008.

I would like to acknowledge my Master thesis supervisor Dr. Andrea Manca for all his support and guidance.

---

1 Introduction

The main purpose of health care economic evaluation (EE) is to assess the economic consequences of health interventions, programmes or services with the aim of informing decisions regarding resource provision within health systems operating under a fixed budget [7]. Economic analysis of health interventions concerns choices that are consequence of financial pressures, budget constraints and resource scarcity. The National Institute for Health and Clinical Excellence (NICE) in the UK is an example of an institution that uses EE to support efficient resource allocation.

Cost-effectiveness analysis (CEA) is the most commonly used EE method, where effectiveness is commonly measured in terms of Quality Adjusted Life Years (QALY). The main application of CEA is in supporting reimbursement decisions made by health care providers regarding health technologies. CEA evaluates technologies to find which one minimizes the cost of generating a given level of health, or which one maximizes the level of health within a specified budget [10].

In order to inform NICE decision-making process, an EE is required to address two main questions [1]. Firstly, with the current evidence, is the technology cost-effective? Secondly, would further research correspond to good 'value for money'? To deal with the former, the methodological structure has to follow some specific criteria: (i) the objective function has to be clear and precise; (ii) the comparison of the new technology needs to be judged against all relevant comparators and needs to include all relevant evidence; (iii) there needs to be consistency in costs and benefits perspective (many argue for a societal perspective, however a third party or payer perspective is commonly adopted); (iv) and, finally, it needs to assess the costs and effects of an alternative treatment strategy within an appropriate time horizon. The second question requires that uncertainty regarding an adoption of a decision must be unequivocally characterised [39]. Quantifying the cost of making a wrong decision represents the basis for assessing if whether acquiring further evidence through funding new research is valuable [29].

1.1 Assessing cost-effectiveness

The summary measures of interest to the decision maker are the expected values of both cost and effectiveness outcomes for each treatment strategy. These are commonly aggregated in a distinctive cost-effectiveness (CE) outcome measure as the incremental cost-effectiveness ratio (ICER) $=\Delta C/\Delta E$ (ΔC-mean differential costs; ΔE-mean differential effects), or its reformulation, the net benefit (NB) measure. Decisions on whether to accept/reject the new technology can me made using analytic models, avoiding, this way, a subjective burden. As there is uncertainty around the CE estimates, any decision based on CEA will also be uncertain [30]. A decision model can explicitly represent this uncertainty and quantify it through the use of probabilistic sensitivity analysis (PSA). The objective of probabilistic modelling is to reflect the uncertainty in the input parameters and illustrate its consequences on the outputs of interest [2]. Decision analytic models (DAM) are used to combine information from various sources of information using mathematical relationships [12; 37].

1.2 Multicentre / multinational RCTs

Often, when conducting Randomized Controlled Trials (RCTs), data on resource use and outcomes are gathered in several different sites. The common objective is to generate a generalizable CE estimate which can be applied across locations. This practice implicitly assumes that resource use and effectiveness data are perfectly transferable [36]. The question is how generalizable are the results of a multiple location evaluation to specific sites and their individual health care situations [15; 45].

If a comparison of health services in different locations is performed, it will disclose important differences in a variety of parameters relevant to the decision problem [38]. The emphasis goes to economic variables including resource use and factor prices, technical efficiency and preferences about health states. The between centre variability is expected to affect the level of resource use, unit costs and outcome data observed in the trials. The dataset will therefore have a hierarchical structure with potential correlation in costs and outcomes linked to patients treated within the same location [15; 24].

Studies from various multinational trial-based analyses assume that resource use data are not at all exchangeable between locations, while effectiveness data are. However, despite this methodology being only feasible when a sufficient number of patients were recruited in the location of interest [44], it disregards also that costs and effects are naturally correlated. Consequently, the correct quantification of uncertainty surrounding CE estimates is endangered [26].

Several analytical methodologies have been proposed to analyse multinational trial data and most of these involve regression analysis. Willke *et al* [45] explored the between-country variability by applying a regression model that included country-by-treatment and country-by-outcome interaction terms, which facilitated country-specific estimation of mean differential costs and effects. Manca *et al* [24] extended the NB regression approach [18] to contain the hierarchical structure of economic data in multilocation trials. The use of hierarchical models was thought to provide an ideal pathway to analyze CE individual patient data (IPD) from multiple location trials allowing for between-location variability. Hierarchical models were therefore shown to be able to obtain trial-wide and location-specific estimates of CE measures, while correctly quantifying sampling uncertainty around these mean estimates.

Pinto *et al* [35] and Willan *et al* [44] explored alternative estimation methods to obtain country-specific estimates of CE from summary data derived from a large multinational trial. Hierarchical modelling was used alongside empirical Bayes shrinkage estimation to obtain country-specific mean estimates. Moreover, Manca *et al* [26] recently investigated the use of Bayesian bivariate hierarchical regression modelling to analyze CE IPD collected alongside multinational trials using also empirical Bayes shrinkage estimation methodology.

1.3 Bayesian methods for CEA

Recently, Bayesian DAM techniques, evaluated using Markov Chain Monte Carlo (MCMC) simulation, have started to be applied to EE decision models. Bayesian methods have been primarily supported by Parmigiani *et al* [34] and latter by O'Hagan *et al* [33] in the evaluation of CE utilising clinical trials data whilst allowing for correlations between the costs and effects.

Given that the relationship between the inputs is too complex to return a 'closed form' solution, describing the exact distribution of the estimator for the CE measure, Monte Carlo (MC) techniques are usually employed to propagate uncertainty in the model and to produce an estimate of the joint distribution of the mean differential costs and mean differential benefits. The conventional approach to probabilistic DAM is to use a frequentist 'two-stage' approach. That is, firstly, a series of analyses are performed to obtain point estimates and

uncertainty around parameter estimates – and this constitutes the 'first stage'. The outcomes of these analyses will return all or part of the parameters input estimates of the Markov model, i.e. transition probabilities and outcomes such as costs and utilities. Secondly, distributions characterising the uncertainty over these parameter estimates are explicitly defined, based on the nature of the parameter and the method of estimation previously used [3]. These are then specified in a spreadsheet and randomly sampled from to obtain a set of realizations for the Markov parameters. The Markov model is evaluated for each of these sets and CE is calculated through the MC estimator[1] – the second stage of estimation is then complete. This non-parametric approach to MC simulation is adopted in practice and the empirical distribution is used to represent the distribution of the CE outcome.

An alternative, and the estimation approach advocated in this paper, is to use a 'one-stage' Bayesian modelling approach. This 'one stage' Bayesian approach comprises the acquisition of posterior distributions for the parameters through MCMC estimation procedures. Given this valuable information, there is no need to assume distributions to characterise the uncertainty over the parameter estimates. One can use directly these posterior distributions to obtain a set of realizations of the Markov parameters that will be used to evaluate the Markov model. The Bayesian analyses described herein relaxes therefore the need for some of the distributional assumptions usually attached to probabilistic modelling.

2 Methods

2.1 Decision analytic models design

The aim of CE evaluation of health interventions is to evaluate the distribution of the expected outcomes. DAMs design, for which the evaluations of expected outcomes with an explicit expression is possible, are denoted as cohort or aggregated models. Examples are decision trees and discrete time Markov chains [2]. For both types of DAMs, each event or clinical state is associated with a monetary cost and a measure of benefit such as QALYs.

The majority of applied DAMs in chronic or long-term diseases are aggregated Markov chains, typically discrete time models. When the instant of episode occurrence is pertinent, when events may happen repetitively throughout time or when time risk is integrated in the decision framework, Markov chains are a valuable technique [41]. In EE the use of non-homogeneous Markov chains is common. This framework implies that transition probability (TP) functions are dependent on time [16; 40].

A Markov model encompasses a set of mutually exclusive and collectively exhaustive health states. Each individual in the model must be in one and only one health state at any point in time. At fixed increments of time - Markov cycle length - subjects' transit among the health states according to a set of TPs which can be constant or time-dependent. Health states can be transient (individuals can revisit the state at any time), temporary (individuals can stay in the state for only one cycle), or absorbing (once entered, individuals can never exit the state) [8].

[1] Most probabilistic models assume independence between parameter estimators. Nevertheless, when the covariance between parameters is known it should be incorporated. Techniques such as the Cholesky decomposition of the variance-covariance matrix for an underlying multivariate normal distribution can be engaged to jointly capture uncertainty.

2.2 Use of statistical modelling to analyse IPD-RCT data to estimate model inputs

In cohort models, the simplest inputs available to estimate TPs are proportions (cumulative incidences) or rates (incidence rates) from published sources [28]. However, when IPD is available, regression analysis is conducted to estimate TPs. The regression framework used to estimate TPs is commonly based on parametric distributional assumptions. One can use models such as linear models, generalized linear models, longitudinal models or other modelling methodologies in order to provide adequate inputs to the Markov process. Estimation of TPs is most commonly conducted through maximum likelihood, as illustrated by Craig *et al* [6].

If patient level resource use from a clinical trial are available one can use standard ordinary least squares (OLS) regressions to obtain mean costs for the different alternative technologies or, accounting for the usual skewed behaviour of cost data, one can obtain reliable estimates by using a Log-Normal distribution or generalised linear models with, for instance, an underlying Gamma distribution (distribution constrained on the interval 0 to positive infinity).

In the case of patient level health-related quality-of-life (HRQoL), such as the EQ-5D for instance, one can analyse these data using the Log-Normal or the Gamma distributions on the disutility scale (e.g. 1-utility). Depending on the clinical trial time horizon, one can have utility data at randomization and at other pre-defined points in time. Regression analysis of longitudinal data approach can be employed in order to obtain estimates of HRQoL changes after randomization.

In the estimation procedure of TPs or other model inputs it is possible to consider the hierarchical nature of the data through the usage of multilevel models, applicable, for instance, in the case of multicentre/multinational trials.

2.3 Multilevel/Hierarchical regression analysis

When dealing with hierarchical data structure, such as data gathered alongside a multicentre RCT, multilevel models are the appropriate approach to obtain unbiased estimates of aggregate measures [11]. Implicit in hierarchical data structure is cross-group heterogeneity. This type of heterogeneity might emerge because of unmeasured factors in group *j*, where there are $j = \{1,...,C\}$ clusters in the data. In lieu of non-pooling models, researchers are interested in modelling both level-1 (x_i) and level-2 (z_j) covariate effects and models may include factors measured at both levels.

Multilevel linear and multilevel generalised linear models - Random intercepts and slopes

Considerations of multiple levels of variation lead to models with random-effects. Multilevel models allow the possible relationship between the response variable and covariates [11; 21]. Suppose there is one level-1 factor and one level-2 factor, the unconditional model would be given by:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_i \,, \qquad \text{eq. 1}$$

where β_{1j} is the slope coefficient for variable x_{ij}. The constant term, β_{0j}, randomly varies across units *j*. Accounting for this, the unconditional model is obtained with

$$\beta_{0j} = \beta_0 + u_{0j} \quad \text{and} \quad \beta_{1j} = \beta_1 + u_{1j} \,,$$

and has reduced-form

$$y_{ij} = \left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right) x_{ij} + \varepsilon_i \,, \qquad \text{eq. 2}$$

where β_0 is the intercept estimate; β_1 the slope coefficient for the relationship between x_{ij} and y_{ij}; u_{0j} the level-2 intercept error; u_{1j} the error term for the randomly varying slope coefficient x_{ij}; and ε_i corresponds to the level-1 error term [11].

Multilevel modelling is applied to logistic regression and other generalised linear models in the similar way as with linear regression.

Multilevel survival models

In a multicentre RCT one may be interested in the treatment effects on the survival of patients. Therefore, one has to extend the usual survival analysis to hierarchical survival models with, for instance, a Proportional Hazards Model (PHM) with random-effects to investigate the centre effect on the efficacy of the treatment as well as on the baseline. This may be understood as a natural extension of the usual mixed-effects model to survival analysis [11; 27].

The fixed-effect model inherently assumes that the centres comprise the entire population of interest. A more realistic assumption is that the centres are random samples from a larger population. For the random-effect survival model let's assume that there are *C* distinct centres with n_j patients from the j^{th} centre. Let t_{ij} be the survival time for the i^{th} patient from the j^{th} centre ($j=1,...,C$; $i=1,...,n_j$). A PHM is assumed for the effects of covariates and centres:

$$h_{ij}\left(t \mid x_{ij}\right) = h_0\left(t\right)\psi\left(x_{ij}\right) = h_0\left(t\right)\exp\left(\left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right)x_{ij}\right). \qquad \text{eq. 3}$$

In the above model $h_{ij}\left(t \mid -\right)$ represents the hazard for the ij^{th} patient conditional on the random-effects u_{0j} and u_{1j} of the explanatory matrix x_{ij}, $h_0\left(t\right)$ is an unknown baseline hazard, β_1 is the fixed-effect corresponding to the covariate x_{ij}. The random components u_{0j} and u_{1j} represent the deviation of the j^{th} centre from a baseline hazard (baseline risk).

With a large number of observations for each centre, one could estimate each centre parameter. However, in practice, one has limited data and must borrow strength across centres to make inferences about either u_{0j} and u_{1j}, so it is usually assumed that the random-effects are independent variables drawn from a family of distributions. This assumption implies that one can learn about one centre parameter by understanding the variability in parameters across the population. Thus, the model is completed by the distributional assumption about the random-effects and a variety of specifications for this distribution can be applied [27].

Multilevel linear mixed models in a longitudinal data framework

With longitudinal data a time element is added to the data and there are repeated measurements for each individual observation, for instance, HRQoL data collected at several time points alongside RCTs. Longitudinal modelling allows one to look at dynamic relationships of individuals and also allows one to control for *unobserved cross-section heterogeneity*. Longitudinal data are closely related to multilevel/hierarchical data, being themselves hierarchically structured by individual [19; 20; 46].

One may be interested in modelling datasets where there is a multilevel structure and, therefore, to have several random effect levels [11]. Adopting the longitudinal framework, the individual may be a level-2 in, for instance, a three-level linear model. Nevertheless, the existence of several levels in the data may bring some problems (i) due to non-independence between the levels; and (ii) because one wants to investigate the different clusters in each level. Therefore, one is interested in generalizing the mixed effects models towards nested random-effects.

In the case where one has three random effect levels, nested one within others, the linear mixed model is:

$$y_{tij} = X_{tij}\beta + Z_{j,it}u_j + Z_{ji,t}u_{ji} + Z_{tij}u_{tij} + \varepsilon_{tij}, \quad \text{for } j = 1,...,C,\ i = 1,...,N_j \text{ and } t = 1,...,T_{ij} \qquad \text{eq. 4}$$

where, y_{tij} is the dependent variable array for the t^{th} cluster of the level-3, nested on the i^{th} cluster of the level-2, nested on the j^{th} cluster of the level-1; X_{tij} is the fixed effects covariate matrix $\left(n_{tij} \times k\right)$; β is the fixed effect array; $Z_{j,it}$ is the level-1 random-effects covariate matrix; u_j is the level-1 random-effects array (normally independent distributed (NID) with mean *0* and Σ_1 variance-covariance matrix); $Z_{ji,t}$ is the level-2 random-effects covariate matrix; u_{ji} is the level-2 random-effects array nested in level-1 j^{th} random effect (NID with mean *0* and Σ_2 variance-covariance matrix, for different *j's*, *i's* or *t's*); Z_{tij} is the level-3 random-effects covariate matrix; u_{tij} is the level-3 random-effects array nested in level-1 j^{th} random effect and nested in level-2 i^{th} random effect (NID with mean *0* and Σ_3 variance-covariance matrix, for different *j's*, *i's* or *t's*); ε_{tij} is the random errors array (NID with mean *0* and $\sigma^2 I$ variance-covariance matrix, for different *j's*, *i's* or *t's*); C is the level-1 number of clusters; N_j is the level-2 number of clusters nested in the j^{th} level-1 and T_{ij} is the level-3 number of clusters nested in the j^{th} level-1 cluster and nested in the i^{th} level-2 cluster, with u_j, u_{ij}, u_{tij} and ε_{tij} independent.

2.4 The Bayesian Approach

Bayesian methods can be considered as an alternative to the classical approach to statistical inference [23]. The increased feasibility of implementing Bayesian methods has been made possible by the advances made in computer power, and the development of user-friendly software such as WinBUGS [43]. The Bayesian approach is appealing since it provides a flexible modelling framework, allowing the researcher to venture beyond the confines of analyses provided in standard statistical packages and account fully for all forms of model estimation uncertainty [42].

Bayesian inference refers to statistical procedures that model unknown parameters (and also missing and latent data) as random variables. Bayesian inference starts with a prior distribution on the unknown parameters and updates this with the likelihood of the data, yielding a posterior distribution (that when obtained by simulation is non-subjected to distributional assumptions) which is used for inferences and predictions [11].

Therefore, interest lies in the calculus of this posterior distribution $f(\theta \mid x)$ of the parameter θ given the observed data x. According to the Bayes theorem, the posterior distribution can be written as

$$f(\theta \mid x) = \frac{f(x \mid \theta) f(\theta)}{f(x)} \propto f(x \mid \theta) f(\theta), \qquad \text{eq. 5}$$

where θ is a parameter or a parameter array with prior distribution $f(\theta)$, and x a random variable with probability density function, $f(x \mid \theta)$, belonging to the space-parameter Θ [32].

In the Bayesian framework, all parameters must have prior distributions. Most prior distributions are vague/non-informative or are prior models. Opposed to prior models, non-informative priors are intended to allow Bayesian inference for parameters for which not much is known beyond the data included in the analysis at hand [11]. The simplest form of Bayesian inference may use a Uniform prior distribution, so that the posterior distribution is the same as the likelihood function [4]. It is usually performed sensitivity analysis to check if results are stable across a range of different prior distributions [5].

The usual MCMC method used is Gibbs sampling. Gibbs sampling is an iterative MC method for generating samples indirectly from a difficult joint distribution of the model parameters without calculating the density [23]. The mechanism is based only on elementary properties of Markov chains. The basic idea of Gibbs sampling is to partition the set of unknown parameters and then estimate them one at a time, or one group at a time, with each parameter or group of parameters estimated conditional on all the others [44].

Bayesian shrinkage estimation

The common analysis of multicentre RCT datasets would use pooled estimates common to all centres or would split the dataset, with all the statistical limitations attached to it. An alternative approach uses empirical Bayesian shrinkage estimation. If one assumes that the individual centre data is sampled from an underlying Normal distribution, then a pooled random-effects estimate provides an empirical mean for the prior distribution for the centre-specific differences [35; 44]. The estimated difference for a particular centre is then the mean of the posterior distribution, which is given by a variance-weighted linear sum of the prior difference (pooled random-effects estimate) and the observed difference for that centre. That is, the empirical Bayes shrinkage estimator is a weighted sum of the estimate provided by the pooled random effects estimate and the estimate provided by the centre-specific observed difference [24; 44].

The key advantage of this approach is that it affords a gain in statistical efficiency by 'borrowing' information from all locations in the estimation of the difference for an individual centre. The amount of information 'borrowed' depends on the proportion of the total variance that is due to the variance between centres. As this proportion decrease, more information is 'borrowed', and the estimates of the centre-specific difference are 'shrunken' towards the pooled random-effects estimate [35].

3 Motivating Example: The RITA 3 trial

In this section, the use of Bayesian hierarchical regression models to analyse hierarchical datasets from multicentre trials is illustrated using a specific case study: the Intervention Trial of unstable Angina (RITA 3). Although this is a multicentre trial conducted in one country, the analytical principles to apply in multinational studies are the same. A DAM was developed and information from the RITA 3 trial was previously analysed [17] to inform the DAM, although ignoring the hierarchical nature of the data. The analysis published in the original paper was replicated in the current work and adapted to consider the evident hierarchical structure of the trial data, and to obtain location-specific estimates that will populate the DAMs. All the assumptions surrounding the original regressions were maintained, particularly model specification and choice of covariates. Suggestions for improvement and further extensions to the present analysis are discussed in the following chapter.

3.1 Background

The third Randomized Intervention Trial of unstable Angina study aimed at supporting the existing evidence that suggested that an early interventional strategy (routine angiography followed by revascularization if clinically indicated) in the management of patients with non-ST-elevation acute coronary syndrome (NSTE-ACS) could improve health outcomes, but at increased costs, when compared with a conservative strategy (ischemia or symptom-driven angiography). The CE of the intervention in different risk groups was assessed to determine whether the gain in health outcomes justified the increase in costs. Full clinical and economic results have been published elsewhere [9; 17].

Based on data from RITA 3 trial, the economic analysis investigated the heterogeneity in CE in patients with different risk profiles at randomization and the effectiveness of early intervention. The economic model provided a tool to extrapolate the trial results to a relevant lifetime time horizon.

A series of regression models (referred to as equations) were estimated to determine the rates of cardiovascular death (CVD) or non-fatal myocardial infarction (MI) during the index hospitalisation and the remainder of the trial follow-up period. These estimates of effectiveness were then incorporated into the CE model which is based on a short-term decision tree (instantaneous in time) and a long-term Markov structure. The main purpose of the short-term tree was to distribute the analysed cohort over the starting states in the long-term Markov structure and to estimate the short-term costs associated with each treatment strategy. The short and long-term models represent the index hospitalisation and the post-index hospitalisation, respectively. Costs and QALYs were determined for the index hospitalisation and for each state in the long-term Markov structure. The Markov structure is shown in Figure 1. The box [MI/CVD] in the figure indicates that a composite event has occurred during a cycle and does not represent a formal health state since patients are then assigned to either a fatal or non-fatal state based on a separate calculation.

Analysis of effectiveness

A logistic regression model was used to estimate the risk of the combined endpoint of CVD or MI during the index hospitalisation in the short-term decision tree. The index hospitalisation was defined as the time from randomization to hospital discharge (Equation 1 in Figure 1). To estimate the risk of the combined endpoint of CVD or MI during the remainder of the trial period, a time-to-event Weibull PHM was employed with the starting time set at hospital discharge. In extrapolating beyond the period of trial follow-up (5 years), a conservative assumption of no continued treatment effect from the early interventional strategy was made (Equation 2 in Figure 1).

There were insufficient patients in RITA 3 trial to estimate the risk of a second composite endpoint of MI or CVD following a non-fatal MI. Instead, the risks of a first composite endpoint were used, multiplied by the coefficient for the additional proportionate risk for patients who had a non-fatal MI prior to their entry into the RITA 3 trial. A Weibull PHM of risk of a second composite endpoint of CVD or MI was employed (Equation 3 in Figure 1). The hazard of dying from non-cardiovascular causes was estimated using general UK population age-and-sex specific life-tables, adjusted to exclude cardiovascular mortality (ICD10 codes I00 to I99) [13; 31].

A logistic regression model was employed to estimate the proportion of composite endpoints being non fatal. A dummy variable was used to investigate if this proportion was different between the index hospitalisation and the remainder of follow-up (Equation 4 in Figure 1).

Analysis of costs

Comprehensive resource use data were collected in patients in RITA 3 up to one-year follow-up. Two standard OLS regressions were used to determine mean costs for the alternative strategies during the index hospitalisation and for the remainder of the trial. Mean costs were estimated, differentiating between management strategies, for patients with and without a composite endpoint of CVD or MI. When extrapolating beyond one year, the analysis assumed no difference between the treatment strategies in the cost of patients not experiencing the composite event.

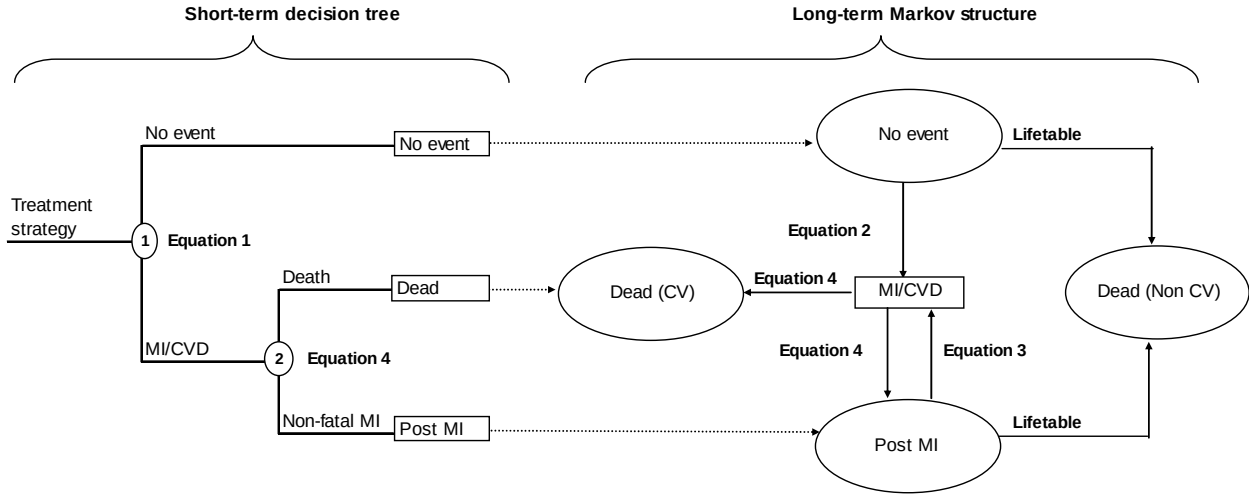

Figure 1. Model structure of the CEA of the RITA 3 trial (MI=myocardial infarction, CV=cardiovascular, CVD=cardiovascular death) [17].

Analysis of HRQoL

HRQoL data were collected in patients in RITA 3 at randomization, 4 months, 1 year, and yearly thereafter, until the 5th year. To estimate QALYs for each treatment strategy, quality adjustment weights (utilities) were required. These were obtained from the trial sample using the EQ-5D instrument, and employing the preferences of the UK general population. A standard OLS regression was employed in order to estimate the mean HRQoL of patients with different risk profiles at randomization. A longitudinal data approach was then employed in order to estimate changes in HRQoL after randomization, differentiating between the two management strategies and whether a composite endpoint of CVD or MI had occurred. For the long-term extrapolation, no difference in HRQoL between the treatment strategies was assumed after the first year in patients not having experienced a composite endpoint.

Covariates

All statistical analyses included previously identified risk factors for cardiac events measured at randomization and randomized treatment. These risk factors were included as covariates in the statistical models and are shown in Table 1.

Covariate	Obs	Mean (std.dev.) or proportion	Min	Max
Age (cathegorical indicator for every 10 years over 60 years of age)	1810	0.887 (0.849)	0	4
Diabetes (indicator of diabetes at study inclusion)	1810	0.135 (0.342)	0	1
Previous MI (indicator of previous MI at study inclusion)	1810	0.277 (0.447)	0	1
Smoker (indicator of smoker at study inclusion)	1810	0.324 (0.468)	0	1
Pulse (discrete indicator for every 5 beats per minute)	1809	7.451 (2.778)	2	20
ST depression (indicator of ST depression at study inclusion)	1810	0.365 (0.481)	0	1
Angina (indicator of angina grade 3 or 4 at study inclusion)	1809	0.359 (0.480)	0	1
Male (indicator of male)	1810	0.623 (0.485)	0	1
Left BBB (indicator of left bundle branch block at study inclusion)	1810	0.035 (0.185)	0	1
Treat (indicator of randomized to early interventional strategy)	1810	0.494 (0.500)	0	1
Risk score (risk of CVD or MI)	1807	0.194 (0.127)	0.034	0.860

Table 1. Baseline covariates included in the statistical models.

3.2 Applying the Bayesian hierarchical approach

As mentioned in section 2.3, multilevel or hierarchical models are considered the appropriate approach to obtain unbiased estimates of aggregate measures. Therefore, to determine the rates of CVD or non-fatal MI during the index hospitalisation and the remainder of the trial follow-up period, a series of Bayesian hierarchical regressions, accounting for within and between centres variability were estimated. These estimates of effectiveness provided sets of location-specific TPs which can be incorporated into location-specific CE models, which will help decision making about allocation of resources at the local level.

In the original study the CE of the intervention was assessed in different risk groups (5 risk groups) to determine whether the gain in health outcomes justified the increase in costs. However, due to the main objectives of this work and also due to practicality issues, the focus here was made on the first baseline risk group (risk group 1) and also only evidence obtained from the RITA 3 trial was used (in the original analysis meta-analytic techniques were also used in order obtain a combined effectiveness measure which informed some effectiveness model input parameters).

Table A1 in appendix**Table** presents summary data of the covariates included in the regression models by centre. In total, 1810 patients were included in the study distributed across 46 centres (hospitals). The distribution of patients across centres is unbalanced, with a minimum of 1 patient observed in centre 18 and a maximum of 153 in centre 11. The average number of patients per centre is approximately 39. For each covariate the mean value and the standard deviation (std. dev.) are presented. A simple inspection of the summary data by location reveals a great deal of variability in covariates, both within and across the centres.

3.2.1 Software

To implement the proposed analysis, the hierarchical regression models were performed in the freely available software package **R** version 2.7.2 (Copyright © 2009 The **R** Foundation for Statistical Computing) and in the also freely available software package **WinBugs/OpenBugs** version 3.0.3 (Copyright © 2007 Medical Research Council (UK), Imperial College (UK) and RNI Helsinki (Finland)) and compared to the **Stata** results from the original study (Stata version 9.0 – Stata statistical software – StataCorp LP).

The Bayesian hierarchical models were implemented using WinBugs/OpenBugs and linked to the software R through two important R packages: **R2WinBugs/BRugs** and **CodaPkg**. Bayesian MCMC methods were employed with one chain and through a simulation process with 5,000 iterations and a 2,000 iteration burn-in period (except for HRQoL where a 20,000 iteration burn-in period, with several thin rate scenarios tested, was performed due to encountered convergence problems).

The decision-analytic model was programmed and analysed in R and compared to the original model performed in **Microsoft® Excel** (Microsoft Corporation 2003, Redmond, Washington, USA).

3.2.2 Results for effectiveness

Equation 1

A logistic regression model was used to estimate the risk of the combined endpoint of CVD or MI during the index hospitalisation in the short-term decision tree. The Bayesian hierarchical logistic regression model was implemented and to obtain the relevant TPs (probability of

composite endpoint during index admission for conservative and treatment group), from equation 1, the inverse logit transformation was used [14].

Logistic regression

CCIndex	R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects											
Treat	0.417	0.288	0.148	0.425	0.294	-0.143	1.008	0.386	0.308	-0.223	0.980
Age	0.549	0.161	0.001	0.554	0.162	0.243	0.874	0.576	0.165	0.260	0.913
Angina	0.636	0.284	0.025	0.635	0.287	0.068	1.195	0.627	0.286	0.064	1.202
Constant	-4.622	0.334	0.000	-4.671	0.338	-5.355	-4.039	-4.841	0.392	-5.680	-4.159
Random Effects											
σ_{Treat}	-	-	-	-	-	-	-	0.198	0.244	0.012	0.866
σ_{Cnst}	-	-	-	-	-	-	-	0.432	0.370	0.011	1.176
ρ_{Treat_Cnst}	-	-	-	-	-	-	-	0.00142			

*Values in log odds ratios

**5,000 iterations and a 2,000 iteration burn-in period

Table 2. Log-odds ratio of composite endpoint of CVD or MI during index hospitalisation (NHM – non-hierarchical model; HM – hierarchical model).

Data on the Bayesian hierarchical model can be found in columns 9 to 12 in Table 2. Information on between-centre variability, σ_{Cnst}, the treatment random-effect component, σ_{Treat}, and the correlation between the random components, ρ_{Treat_Cnst}, are presented in the three bottom rows. The same framework was used in subsequent models.

Compared to the non-hierarchical models, the fixed-effects estimates from the hierarchical model are similar in terms of magnitude, sign and significance of estimates. It can be identified a decrease on the intercept and treatment estimates (fixed effects), a reflection of the decomposition of the effects in both fixed and random components. The empirical correlation estimate between the random components is considered weak.

The results are similar and consistent across software's, showing that increasing age and severe angina (grade 3 or 4) are associated with an increased risk of a composite endpoint during the index hospitalisation. Although not statistically significant, the early interventional strategy is associated with an increased risk of a composite endpoint during the index hospitalisation (frequentist estimates: odds ratio of 1.517 with p-value = 0.148; Bayesian estimates: odds ratio of 1.530, not sig. at usual levels; Bayesian hierarchical estimates: odds ratio of 1.471, not sig. at usual levels).

The centre-specific random-effects components are shown in Table 3. For simplicity, since there are 46 centres in the trial, here is reported only the random-effects for treatment and intercept of 5 specific centres. These are centres with sample sizes of 17, 153, 65, 94 and 110 respectively, and have been selected to explore the impact of sample size on the model results. It can be observed the differences in the random estimates within and across centres, reflecting the variability within and between-centres of the risk of a composite endpoint during the index hospitalisation.

Logistic regression

CCIndex		WinBugs** - HM			
Centre		mean	std. dev.	95% CrI	
Random Effects					
centre 2	$u_{1j\text{ - Treat}}$	-0.019	0.305	-0.682	0.555
	$u_{0j\text{ - Cnst}}$	-0.103	0.529	-1.365	0.901
centre 11	$u_{1j\text{ - Treat}}$	0.092	0.292	-0.289	0.954
	$u_{0j\text{ - Cnst}}$	0.057	0.361	-0.710	0.886
centre 23	$u_{1j\text{ - Treat}}$	-0.077	0.317	-0.928	0.391
	$u_{0j\text{ - Cnst}}$	-0.146	0.436	-1.267	0.643
centre 37	$u_{1j\text{ - Treat}}$	-0.030	0.261	-0.686	0.504
	$u_{0j\text{ - Cnst}}$	0.122	0.390	-0.585	1.103
centre 40	$u_{1j\text{ - Treat}}$	-0.024	0.243	-0.641	0.463
	$u_{0j\text{ - Cnst}}$	0.382	0.490	-0.196	1.517

**5,000 iterations and a 2,000 iteration burn-in period

Table 3. Centre specific random effects for 5 centres in the trial, results of *Bayesian* hierarchical logistic regression of composite endpoint of CVD or MI during index hospitalisation (HM – hierarchical model).

Equation 2

To estimate the risk of the combined endpoint of CVD or MI during the remainder of the trial period, a time-to-event Weibull PHM was employed with the starting time set at hospital discharge. The Bayesian hierarchical Weibull regression model was implemented according to the following framework: assuming that survival times are Weibull distributed, the regression coefficients β were believed a priori to follow independent Normal distributions with zero mean and vague precision 0.0001. The shape parameter γ for the survival distribution was given a *Gamma(1, 0.0001)* prior, which is slowly decreasing on the positive real line.

The TPs needed to populate the long-term Markov structure were derived from the results of the statistical models, for which the yearly TP of a composite endpoint in Markov cycle *t*, *tp(t)*, is given by $tp(t) = 1 - \exp(e^{\beta x_i}(t-1)^{\gamma} - e^{\beta x_i} t^{\gamma})$.

The non-hierarchical models show that all risk factors, except presence of severe angina, were significant at the 5% level (Table 4). However, this risk factor was very close to significance and was kept in the parametric model. The fact that the shape parameter in the Weibull proportional hazards regression model is less than 1 indicates that the rate of the composite endpoint of CVD or MI declines as time elapses from hospital discharge. The early interventional strategy was associated with a statistically significant lower rate of CVD or MI after the index hospitalisation (frequentist estimates: log-hazard ratio -0.477, 95% CI [-0.767;-0.186]; Bayesian estimates: log-hazard ratio -0.477, 95% CrI [-0.770;-0.159]; Bayesian hierarchical estimates: log-hazard ratio -0.527, 95% CrI [-0.899;-0.228]).

The results of the Bayesian hierarchical model were similar to the non-hierarchical ones. Compared to the NHMs, the fixed-effects estimates obtained were equivalent or vaguely smaller. In addition to severe angina risk factor, gender was found to be non-significant but very close to the assumed significance level. The random-effect standard deviation of the treatment effect is large in magnitude, balanced by a lower treatment fixed-effect. A negative but small correlation between the random components was found.

Parametric proportional hazards model - Weibull regression

	R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects											
Age	0.575	0.087	0.000	0.567	0.087	0.397	0.731	0.563	0.088	0.390	0.736
Diabetes	0.645	0.173	0.000	0.634	0.181	0.271	1.004	0.634	0.174	0.284	0.954
Previous MI	0.386	0.154	0.012	0.382	0.159	0.074	0.682	0.387	0.149	0.095	0.679
Smoker	0.501	0.160	0.002	0.496	0.170	0.157	0.834	0.481	0.161	0.140	0.779
Pulse	0.060	0.024	0.014	0.060	0.025	0.005	0.109	0.052	0.022	0.011	0.096
St depression	0.357	0.149	0.016	0.361	0.144	0.086	0.650	0.357	0.146	0.059	0.630
Angina	0.280	0.149	0.060	0.279	0.158	-0.030	0.583	0.268	0.145	-0.017	0.562
Male	0.316	0.158	0.045	0.329	0.148	0.038	0.624	0.292	0.161	-0.004	0.634
Left BBB	0.682	0.268	0.011	0.649	0.272	0.089	1.174	0.678	0.267	0.121	1.176
Treat	-0.477	0.148	0.001	-0.477	0.154	-0.770	-0.159	-0.527	0.175	-0.899	-0.228
Constant	-4.790	0.302	0.000	-4.837	0.308	-5.440	-4.230	-4.699	0.333	-5.334	-4.100
Shape parameter (γ)	0.579	0.070	-	0.597	0.038	0.519	0.666	0.582	0.041	0.510	0.668
Random Effects											
σ_{Treat}	-	-	-	-	-	-	-	0.206	0.187	0.011	0.667
σ_{Cnst}	-	-	-	-	-	-	-	0.057	0.050	0.009	0.194
ρ_{Treat_Cnst}	-	-	-	-	-	-	-		-0.0314		

*Values in log hazard ratios

**5,000 iterations and a 2,000 iteration burn-in period

Table 4. Log-hazard ratio of composite endpoint of CVD or MI from hospital discharge until end of trial (NHM – non-hierarchical model; HM – hierarchical model).

The centre-specific random-effects components estimates for the 5 centres under analysis can be found in the appendix section (appendix – Table A2).

Equation 3

As mentioned before, equation 2 was used to estimate the risk of a second composite endpoint by updating the covariate for prior MI. Equation 3 coefficient estimates were derived from equation 2 after update.

The Bayesian hierarchical Weibull regression model implemented followed the same framework has the previous survival model.

Equation 4

All the events reported in the RITA trial (comprising a total of 244 first events and 17 second events) were included in the logistic regression model estimating the probability of a composite endpoint being non-fatal.

The results are similar and coherent across software's, showing that this probability was higher during the index hospitalisation than during the follow-up period, reflecting the fact that patients are likely to receive treatment without delay if they experience an MI whilst in hospital (Table 5). The Bayesian logistic regression model implemented followed the same framework has the previous hierarchical logit model.

Logistic regression

Non-fatal MI	R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects											
Index dummy	1.162	0.314	0.000	1.195	0.322	0.577	1.820	1.202	0.318	0.599	1.831
Age	-0.347	0.146	0.017	-0.356	0.147	-0.644	-0.069	-0.366	0.151	-0.668	-0.077
Previous MI	-0.595	0.264	0.024	-0.604	0.266	-1.124	-0.088	-0.614	0.269	-1.155	-0.091
Constant	0.235	0.248	0.344	0.240	0.249	-0.252	0.725	0.255	0.255	-0.228	0.769
Random Effects											
σ_{Cnst}	-	-	-	-	-	-	-	0.135	0.125	0.011	0.463

*Values in log odds ratios

**5,000 iterations and a 2,000 iteration burn-in period

Table 5. Log- odds ratio of a composite endpoint of CVD or MI being non-fatal (NHM – non-hierarchical model; HM – hierarchical model).

Compared to the non-hierarchical models, the results for the hierarchical model show similar results in terms of magnitude, sign and significance of estimates. An increase in the intercept fixed effect estimates is identified. The centre-specific random-effects components are shown in the appendix.

3.2.3 Results for costs

Cost regression 1 - Estimated costs during the index hospitalisation

Despite the skewed behaviour of cost data and the non-negative value constraint, the original model for costs during the index hospitalisation was based on a multiple linear regression which did not take account of these characteristics. The dependent variable was regressed against a set of covariates, result from a backward stepwise covariate selection procedure.

The non-hierarchical models demonstrate similar results, showing that during the index hospitalisation, the early interventional strategy was associated with a higher mean cost (frequentist estimates: mean £5,654, 95% CI [£5,151;£6,157]; Bayesian estimates: mean £5,652, 95% CrI [£5,145;£6,159]; Bayesian hierarchical estimates: mean £5,881, 95% CrI [£5,146;£6,632]) compared with a conservative strategy (Table 6). This additional cost was seen as a result of a higher number of angiographies and revascularizations undertaken in the early interventional arm. After controlling for treatment allocation, a non-fatal MI or death was associated with additional costs of approximately £6,200 and £7,900, respectively, which included the costs for the administration of thrombolytic drugs, revascularisations and longer hospital stay in wards and intensive care.

As in the NHMs, HM covariates such as age, sex, and ST depression were also associated with higher costs during the index hospitalisation. The correlation between the random components was found to be negative and relatively high. The centre-specific random-effects components are shown in the appendix section (appendix - Table A4).

Linear model

Costs index	R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.	std. err.	Pr(>|z|)	mean	std. dev.	95% CrI		mean	std. dev.	95% CrI	
Fixed Effects											
MI index	6221.2	972.1	0.000	6228.6	993.8	4266.1	8230.1	6236.7	942.0	4332.4	8050.7
Dead index	7947.4	1229.4	0.000	7921.5	1211.9	5577.1	10312.0	7874.8	1197.7	5539.0	10219.5
Treat	5653.9	256.4	0.000	5652.4	259.8	5144.5	6159.0	5881.3	376.1	5145.6	6631.8
Male	1034.8	264.6	0.000	1039.2	263.7	513.6	1557.4	1131.0	258.7	624.0	1638.4
Age	878.3	152.6	0.000	876.4	152.0	577.3	1175.8	877.3	152.0	583.8	1176.3
ST depression	1224.4	268.1	0.000	1228.6	267.7	706.0	1764.5	1080.3	269.6	543.5	1605.2
Constant	1778.5	295.3	0.000	1773.8	291.2	1216.0	2355.4	1882.6	329.9	1235.0	2542.9
Random Effects											
σ_{ε}	-	-	-	-	-	-	-	5215.7	89.0	5048.3	5395.7
σ_{Treat}	-	-	-	-	-	-	-	1729.1	362.4	1073.5	2478.3
σ_{Cnst}	-	-	-	-	-	-	-	941.6	233.9	511.6	1448.1
ρ_{Treat_Cnst}	-	-	-	-	-	-	-		-0.20164		

**5,000 iterations and a 2,000 iteration burn-in period

Table 6. Estimated costs during the index hospitalisation (NHM – non-hierarchical model; HM – hierarchical model).

Cost regression 2 - Estimated costs during the follow-up period

The non-hierarchical models demonstrated similar results, showing that during the first year after the index hospitalisation, the early interventional strategy was associated with a lower mean cost (frequentist estimates: mean -£1,106, 95% CI [-£1,562;-£650]; Bayesian estimates: mean -£1,112, 95% CrI [-£1,570;-£657]; Bayesian hierarchical estimates: mean -£1,104, 95% CrI [-£1,588;-£614]) compared with the conservative strategy (Table 7). This reflected the fact that more patients in the conservative strategy had further symptoms that necessitated revascularization during this period.

The results also indicated that patients had a substantially higher mean cost, irrespective of treatment allocation, if they suffered a MI within the previous year (frequentist estimates: mean £5,467, 95% CI [£3,880; £7,020]; Bayesian estimates: mean £5,446, 95% CrI [£3,883;£7,019]; Bayesian hierarchical estimates: mean £5,444, 95% CrI [£3,871;£7,001]) or prior to the trial (frequentist estimates: mean £724, 95% CI [£210; £1,240]; Bayesian estimates: mean £717, 95% CrI [£212;£1,238]; Bayesian hierarchical estimates: mean £694, 95% CrI [£170;£1,208]).

Compared to the non-hierarchical models, the results for the *Bayesian* hierarchical model show that fixed-effects estimates are very similar. Centre-specific random-effects components can be found in the appendix section.

Linear model

Costs follow-up exc.MI/stroke	R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.	std. err.	Pr(>|z|)	mean	std. dev.	95% CrI		mean	std. dev.	95% CrI	
Fixed Effects											
MI year 1	5467.1	804.0	0.000	5445.9	796.6	3882.9	7019.1	5444.4	804.8	3871.1	7000.5
Treat	-1106.1	232.6	0.000	-1112.0	231.5	-1570.3	-657.0	-1103.9	246.7	-1587.7	-614.0
Male	586.2	242.2	0.016	580.7	240.7	102.8	1039.9	603.3	242.2	127.9	1074.4
Angina	1033.8	246.9	0.000	1040.1	247.9	545.9	1528.7	951.3	247.6	468.1	1439.8
Previous MI	724.4	262.4	0.006	717.1	263.7	211.5	1238.4	694.1	263.6	169.9	1207.7
Constant	2734.9	247.6	0.000	2741.2	246.8	2256.0	3230.8	2786.4	269.9	2259.4	3313.9
Random Effects											
σ_{ε}	-	-	-	-	-	-	-	4793.7	81.0	4636.9	4958.3
σ_{Treat}	-	-	-	-	-	-	-	474.8	159.0	243.3	856.1
σ_{Cnst}	-	-	-	-	-	-	-	614.7	164.3	329.0	978.0
ρ_{Treat_Cnst}	-	-	-	-	-	-	-		0.07732		

**5,000 iterations and a 2,000 iteration burn-in period

Table 7. Estimated costs during the follow-up period (NHM – non-hierarchical model; HM – hierarchical model).

3.2.4 Results for HRQoL

HRQoL regression 1 - Estimated baseline utilities

A regression model of EQ-5D at baseline was built to give starting QoL estimate for the population under consideration, assuming the trial sample is representative of the target population.

At randomization, mean HRQoL (in terms of 0 to 1 utilities) were higher for males whereas diabetes, previous MI, ST depression and angina were associated with lower HRQoL (Table 8). Similar results were obtained for the different software's.

The Bayesian normal linear regression model with non-informative priors implemented followed the same framework has the previous Bayesian linear models. To account for not only within but also between-centre variability, a Bayesian hierarchical model was built incorporating a random intercept component. Except for diabetes and presence of previous MI, the results for the hierarchical model show now slightly higher mean estimates and also lost statistical significance. The centre-specific random-effects components are shown in the appendix section (appendix - Table A6).

Linear model

HRQoL baseline	R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.	std. err.	Pr(>|z|)	mean	std. dev.	95% CrI		mean	std. dev.	95% CrI	
Fixed Effects											
Diabetes	-0.050	0.021	0.016	-0.051	0.021	-0.091	-0.009	-0.036	0.019	-0.074	0.003
Previous MI	-0.045	0.016	0.006	-0.045	0.016	-0.078	-0.014	-0.021	0.015	-0.050	0.009
ST depression	-0.066	0.015	0.000	-0.067	0.015	-0.096	-0.038	-0.031	0.014	-0.058	-0.003
Angina	-0.074	0.015	0.000	-0.073	0.015	-0.104	-0.043	-0.073	0.014	-0.100	-0.045
Male	0.072	0.015	0.000	0.072	0.015	0.043	0.101	0.079	0.014	0.053	0.106
Constant	0.693	0.015	0.000	0.693	0.015	0.665	0.722	0.662	0.024	0.615	0.710
Random Effects											
σ_{ε}	-	-	-	-	-	-	-	0.271	0.005	0.262	0.280
σ_{Cnst}	-	-	-	-	-	-	-	0.128	0.017	0.099	0.164
ρ_{Treat_Cnst}	-	-	-	-	-	-	-		-0.03515		

**5,000 iterations and a 2,000 iteration burn-in period

Table 8. Estimated baseline utilities (NHM – non-hierarchical model; HM – hierarchical model).

HRQoL regression 2 - Estimated gain in HRQoL

The original model was fitted using generalized least squares random-effects estimators, where binary covariates were included to represent whether the utility measure was taken at month 4 (D4) or subsequently (D12) and an interaction term for treatment group. Changes in utility at one year were maintained until the end of the follow up period, for patients who do not experience a MI. Binary covariates were also included to indicate whether a MI had occurred recently (that is, within 1 year prior to the time of the follow up interview) (current MI) and a covariate indicating whether a MI had occurred at all prior to the time of the follow-up interview, either before or during the trial (prior MI).

The number of patients with EQ-5D data in the follow-up period was 1,734 and the number of observations was 6,203 indicating that each patient on average had their HRQoL measured 3.5 times. Table 9 shows the results for the multilevel model considering the individual as a cluster, named non-centre hierarchical model (NCHM), and also, in columns 9 to 12, the hierarchical model with time clustered in patients and patients nested in health care centres (centre hierarchical model (CHM)). Table 9 supplies the within-patient standard error and the between-patient standard error.

Longitudinal data

Change HRQoL	**R - NCHM**			**WinBugs** - NCHM**				**WinBugs** - CHM**			
Covariate	coef.	std. err.	Pr(>|z|)	mean	std. dev.	95% CrI		mean	std. dev.	95% CrI	
Fixed Effects											
D4 x Treat	0.039	0.017	0.020	0.043	0.016	0.012	0.074	1.307	0.017	1.274	1.341
D12	0.038	0.008	0.000	0.015	0.008	-0.001	0.032	0.015	0.009	-0.001	0.035
D12 x Treat	0.018	0.016	0.238	0.024	0.015	-0.005	0.053	1.287	0.010	1.266	1.306
Prior MI	-0.010	0.016	0.521	-0.018	0.016	-0.049	0.013	-0.020	0.012	-0.044	0.002
Current MI	-0.035	0.022	0.109	-0.029	0.022	-0.074	0.014	-0.031	0.023	-0.081	0.009
Constant	0.044	0.013	0.001	0.040	0.012	0.015	0.063	0.028	0.022	-0.010	0.075
Random Effects											
σ_{ε}	0.033	-	-	0.174	0.002	0.169	0.179	0.174	0.002	0.169	1.78
$\sigma_{Cnst\_patient}$	0.089	-	-	0.003	0.000	0.003	0.004	0.003	0.000	0.003	0.004
σ_{Treat_centre}	-	-	-	-	-	-	-	1.281	0.137	1.042	1.577
σ_{Cnst_centre}	-	-	-	-	-	-	-	0.117	0.020	0.085	0.161
$\rho_{Treat\_Cnst\_centre}$	-	-	-	-	-	-	-		0.03867		

**5,000 iterations and a 2,000 iteration burn-in period

Table 9. Estimated gain in HRQoL (NCHM – non-centre hierarchical model; CHM – centre hierarchical model).

The R model results are similar, however, the estimates differ in magnitude and significance from the NCHM obtained in WinBugs due to chain convergence problems. Despite the efforts to improve convergence, by increasing the burn-in period or by changing the thinning rate, auto-correlation was still evident. Therefore, an improvement in the original model's covariate structure is recommended.

However, the results for the NCHM reveal that in both treatment strategies HRQoL was improved at 4 months although an incremental gain of the early interventional strategy compared with the conservative strategy was observed. Between 4 and 12 months, HRQoL was improved further in both treatment strategies, although the incremental gain of the early interventional strategy is non-significant, at the common levels of significance. A recent MI was associated with a decrement in HRQoL regardless of treatment allocation and a previous MI prior to study inclusion was associated with a smaller HRQoL decrement, but, nevertheless, also both non-significant at the usual significance levels.

The Bayesian hierarchical model considering the centre variability shows more problems of chain convergency. The results differ substantially compared to the other models, especially

for the interaction *D4treat* with treatment covariate and the interaction with treatment group variable for utility measured at and after 12 months, *D12treat*. These changes may be due to the referred omitted variable and the chain convergence problems. Therefore, any interpretation of this model estimates should be performed with caution.

These statements are supported by the estimate for the standard error of the centre-level treatment random-effect, σ_{Treat_centre} (mean of approximately 1.280, 95% CrI 1.042 – 1.577). The considerably high value of the estimate reflects the presence of unexplained variability that is being captured here. The centre-specific intercept random-effects components are shown in the appendix.

3.2.5 Cost-effectiveness

The expected (mean) costs and health outcomes of both strategies were combined into an ICER, which is interpreted as the additional cost of generating an additional unit of health outcome (QALY). Many health care systems compare the ICER with a threshold value (λ) to establish whether the strategy should, in principle, be recommended for implementation. NICE in the UK uses a threshold of around £20,000 per QALY gained. Cost-effectiveness was estimated over patients' lifetimes using a UK health service perspective.

Location-specific cost-effectiveness results

Figure 2 illustrates the joint CE density plotted on the Cost Effectiveness Plane (CEP) for 5 of the centres (hospitals) in the RITA 3 trial, namely centres 2, 11, 23, 37 and 40. In the figure, the presence of the trial wide results is for comparison reasons. It can be observed that in all centre-specific CEP plots the majority of the simulation results are located in the *NE* quadrant, indicating the same conclusion of the trial wide results: more effective intervention strategy than the comparator, but at higher costs.

The centre-specific CEPs show higher variability in mean differential cost estimates compared to the trial wide results. For instance, centre 2 CEP depicts a range in mean differential costs from approximately £0 to £8,000, with an estimated average of approximately £4,950. The inclusion of only 17 patients in this centre may be an explanation for the evident large uncertainty attached to the cost estimates. In centre 37 (94 patients), the mean differential cost estimates are on average higher than the trial wide and also higher than other centre estimates (average of approximately £7,750, 95% CrI £6,040 - £9,465). See Table 10 for details on mean differential costs, mean differential QALYs and ICERs at the trial wide and at the centre level.

The centre-specific CEP plots also show high variability of mean differential QALY estimates, with longer left tail estimate distribution compared to the trial wide results. For all centre-specific CEPs one can observe that the majority of the simulated results are concentrated in the range of 0 and 0.2 values of the incremental QALY estimates.

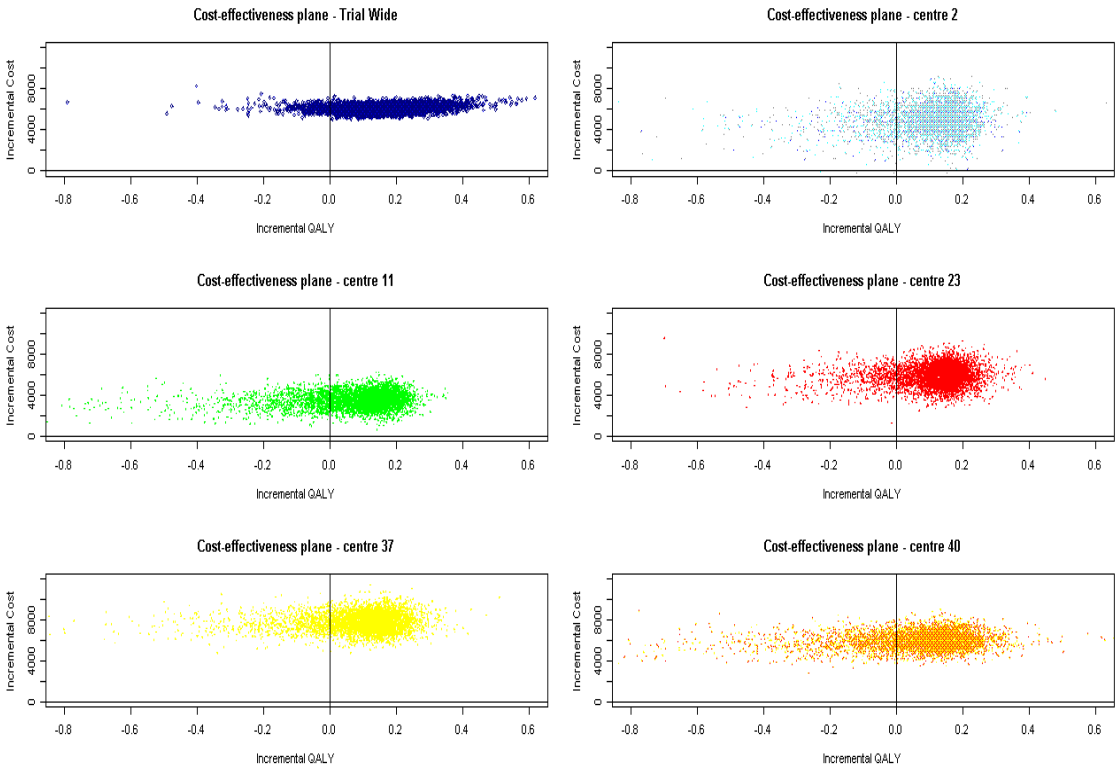

Figure 2. Cost-effectiveness planes of the RITA 3 model with trial wide results and centre-specific results for centres 2, 11, 23 37 and 40, respectively.

	Δ Costs (£) (95% CrI)	**Δ QALYs** (95% CrI)	**ICER (£/QALY)**
Trial wide	6,418 (5,426 ; 6,753)	0.155 (-0.058 ; 0.268)	41,406
centre 2	4,949 (2,286 ; 7,612)	0.120 (-0.129 ; 0.368)	41,239
centre 11	3,551 (2,002 ; 5,100)	0.090 (-0.214 ; 0.394)	39,458
centre 23	5,879 (3,985 ; 7,773)	0.132 (-0.092 ; 0.356)	44,539
centre 37	7,752 (6,039 ; 9465)	0.111 (-0.158 ; 0.381)	69,830
centre 40	5,951 (4,370 ; 7,532)	0.086 (-0.272 ; 0.444)	69,168

Table 10. Trial wide and centre-specific estimated differential costs and QALYs (95% credibility intervals) and ICERs estimates (centres 2, 11, 23, 37 and 40, respectively).

Similar features are revealed in terms of the Cost Effectiveness Acceptability Curves (CEACs) for these 5 centres (Figure 3). Once again, the curves display great variability across centres (hospitals) in CE for given values of the threshold, λ. This variability appears greatest at the values of λ ranging from £20,000 to £60,000, although caution is required here as this observation is based on only those selected centres displayed. For example, the probability of

the intervention strategy being cost-effective, at a ceiling ratio of £50,000, is approximately 0.65 applying the trial wide results with single-level specification. The corresponding probability for centre 37 is 0.34 and for centre 40 *is* 0.43. The observed maximum probability that the intervention is cost-effective for centre 40 is approximately, 0.66 (at λ = £140,000). For centre 23, the maximum is 0.82 (at λ = £140,000). For values of λ greater than £34,000, the intervention strategy would probably be considered cost-effective based on the results of centre 11. However, for values of λ less than £70,000, the intervention strategy would probably not be considered cost-effective based on patient cost and outcomes reported for centre 37.

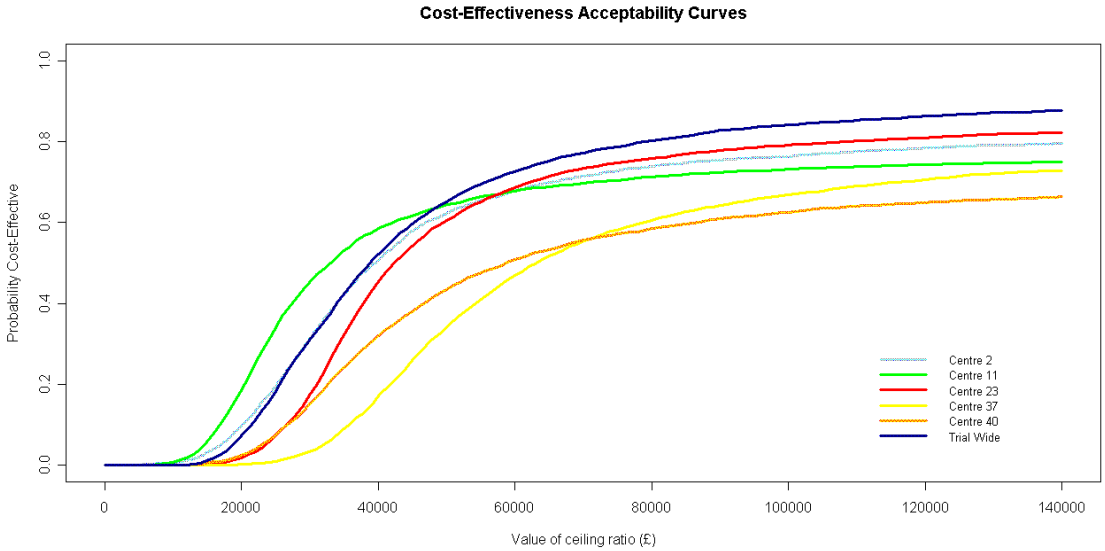

Figure 3. Cost-effectiveness acceptability curve for the trial wide results and centre-specific results for centres 2, 11, 23 37 and 40, respectively.

4 Discussion

This paper has demonstrated the significance of Bayesian hierarchical modelling to estimate cluster-specific parameters for use in DAMs where IPD from a multi-location trial are available. The case-study was based on a multicentre trial in one country, but the methods are equally applicable to the analysis of multinational trials to produce country-specific CE estimates.

The work here presented is still an ongoing project and all results should be considered as preliminary. However, this paper has established, through the use of one illustrative example, how a multicentre trial-based CE analysis may be implemented within a Bayesian framework and evaluated using Gibbs sampling MCMC methods in the software package WinBUGS. The potential advantages of this type of approach are that it (a) removes the need to make parametric distributional assumptions over the obtained parameter distribution (the posterior distributions for each parameter); (b) prior beliefs may be easily incorporated; and (c) may allow a single coherent model to be developed which facilitates the incorporation of all available sources of evidence (whether from RCTs, observational studies and/or expert opinion) to evaluate the CE of alternative interventions.

The 'two-stage' approach may be considered easier to implement than a 'one-stage' decision model, as it splits the analysis into data synthesis and model evaluation, and specialist spreadsheet packages (such as Microsoft Excel) providing a more user-friendly environment

for implementing the latter. However, the main advantage of the suggested modelling approach compared to a 'two-stage' approach is that it allows a single coherent model to be developed to evaluate the CE of alternative interventions.

The extent to which the use of Bayesian hierarchical modelling is decisive in a particular study depends on the proportion of overall variability in CE that takes place between locations. The limitation of regression results obtained from fixed effect models is that they are only valid within the sample of locations that participated in the study. In contrast, random effect models have the property that allows them to be generalisable to the centres outside the study sample that share similar characteristics with the level-2 units participating in the trial.

The analyses presented here can be extended in six important ways. The first would be to rethink the variable selection procedure to be used in the regression models, particularly the backward stepwise selection framework, performed in most of the original models. As mentioned in Judd *et al* [22], given that the data analyst knows more about the data than a computer algorithm, better models can be produced by a better understanding of the data.

The second extension proposed to the framework presented here is to consider the data characteristics in terms of range and skewness. Just as the choice of distribution for probability data was based upon the range of data, cost data are constrained to be non negative and are usually highly skewed. Therefore one should employ the Log-Normal or the Gamma distributions to reflect the skewness often found in cost data, and apply generalised linear mixed models for the analysis of multicentre / multinational cost data.

The third proposed extension to the work presented here is the fact that one should account for the imbalances in baseline utility in the estimation of mean differential HRQoL. The non-inclusion of the baseline utility covariate in the models can result in misleading CE estimates because baseline utility is likely to be strongly related to utility at follow-up, and consequently should be controlled for in estimating differential HRQoL. HRQoL estimates are, therefore, sensitive to small imbalances in mean baseline utilities between the arms of the trials. In addition, given that baseline utilities usually enter directly into the HRQoL calculation, they should represent a strong predictor of HRQoLs [25].

Forthly, alter the short-term part of the decision model to a unique multinomial logit framework. Fifthly, perform sensitivity analysis over obtained parameter estimates by using a variety of non-informative and perceived informative priors. And finally, integrate the extra sources of evidence present in the original model (i.e. evidence from meta-analysis) in the present Bayesian decision model using the promoted framework.

5 Bibliography

[1] Bojke, L., Claxton, K., Palmer, S., Sculpher M. (March 2006). Defining and characterising structural uncertainty in decision analytic models. *CHE Research Paper 9*;

[2] Briggs, A., Sculpher, M. and Claxton, K. (2006). Decision modelling for health economic evaluation. *Handbooks in Health Economic Evaluation Series. Oxford University Press*, First edition;

[3] Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., Brazier, J., O'Hagan, T. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* (14), 339-347;

[4] Cooper, N. J., Abrams, K., Sutton, A. J., Turner, D., Lambert, P. (2003). A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advance breast cancer. *Journal of the Royal Statistical Society* (166), 3, 389-405;

[5] Cooper, N. J., Sutton, A. J., Abrams, K. R., Turner, D., Wailoo, A. (2004). Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Economics* (13), 203-226;

[6] Craig, B. and Sendi, P. (2002). Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics*, 11, (1), 33–42;

[7] Culyer, A. J. and Newhouse, J. P. (2000). Handbook of Health Economics. *North-Holland* Elsevier;

[8] Drummond, M. and McGuire, A. (2001). Economic evaluation in health care: merging theory with practice. *Oxford University Press*;

[9] Fox, K., Poole-Wilson, P., Henderson, R., *et al*, for the Randomized Intervention Trial of unstable Angina (RITA) investigators. (2002). Interventional versus conservative treatment for patients with unstable angina or non-ST-elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. Randomized intervention trial of unstable angina. *Lancet*, 360, 743-751;

[10] Garber, A. and Phelps, C. (1997). Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* (16), 1–31;

[11] Gelman, A., Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Cambridge University Press*;

[12] Gold, M., Siegel, J., Russell, L. and Weinstein, M. (1996). Cost-effectiveness in health and medicine. *Oxford University Press*, First edition;

[13] Government actuary department. Interim life tables. (2006). Expectation of life for males in the United Kingdom, based on data for the years 2002-2004. (online at http://www.gad.gov.uk);

[14] Greene, W. (2003). Econometric Analysis. *Prentice Hall*, Fifth edition, New York;

[15] Grieve, R., Nixon, R., Thompson, S., Normand, C. (2005) Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Economics*, (14), 185–196;

[16] Hawkins, N., Sculpher, M. and Epstein, D. (2005). Cost-effectiveness analysis of treatments for chronic disease: using R to incorporate time dependency of treatment response. *Medical Decision Making*, 25, (5), 511–519;

[17] Henriksson, M., Epstein, D., Palmer, S., Sculpher, M., Clayton, T., Pocock, S., Henderson, R., Buxton, M. and Fox, K. (2008). The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. *BMJ Heart*, (94), 717-723;

[18] Hoch, J., Briggs, A., Willan, A. (2002). Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics*, (11), 415–430;

[19] Jones, A. (2000). Health Econometrics. in Culyer, A. and Newhouse, J., Handbook of Health Economics. *North Holland*. Elsevier;

[20] Jones, A. (2005). Applied econometrics for health economists – a practical guide. *Office of Health Economics*;

[21] Jones, B. (2008). Multilevel modelling. in Box-Steffensmeier, J., Brady, H. and Collier, D. The Oxford handbook of political methodology. *Oxford University Press*;

[22] Judd, C. and McClelland, G. (1989). Data analysis: a model comparison approach. *New York: Harcourt Brace Jovanovich*;

[23] Lee, P. M. (1997). Bayesian Statistics. *Arnold*: London;

[24] Manca, A., Rice, N., Sculpher, M., Briggs, A. (2005). Assessing generalizability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics*. (14), 471–485;

[25] Manca, A., Hawkins, N. and Sculpher, M. (2005). Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Economics, (14), 487–496.

[26] Manca, A., Lambert, P., Sculpher, M. and Rice, N. (2007). Cost-effectiveness analysis using data from multinational trials: the use of bivariate hierarchical modelling. Medical Decision Making, (27), 471-490;

[27] Matsuyama, Y., Sakamoto, J. and Ohashi, Y. (1998). A Bayesian hierarchical survival model for the institutional effects in a multi-centre cancer clinical trial. *Statistics in Medicine*, (17): 1893-1908;

[28] Miller, D. and Homan, S. (1994). Determining transition probabilities: confusion and suggestions. *Medical Decision Making*, 14, 52-58;

[29] National Institute for Health and Clinical Excellence (2004). A guide for manufacturers and sponsors: contributing to a technology appraisal. *London: National Institute for Health and Clinical Excellence*;

[30] National Institute for Health and Clinical Excellence (2007). Briefing paper for methods review workshop on exploring uncertainty. *Institute's Decision Support Unit*, London: National Institute for Health and Clinical Excellence (online at: http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/technology_appraisal_process_guides.jsp);

[31] National Statistics. (2003). Review of the registrar general on deaths by cause, sex and age, in England and Wales. London: *National Statistics*;

[32] Ntzoufras, I. (2009). Bayesian modelling using WinBUGGS. Wiley series in computational statistics. *Wiley*;

[33] O'Hagan, A., Stevens, J. W., Montmartin, J. (2001) Bayesian cost-effectiveness analysis from clinical trial data. *Statististics in Medicine* (20), 733–753;

[34] Parmigiani, G., Samsa, G., Ancukiewicz, M., Lipscomb, J., Hasselblad, V. and Matchar, D. B. (1997). Assessing uncertainty in cost-effectiveness analyses. *Medical Decision Making* (17), 390–401;

[35] Pinto, E., Willan, A., O'Brien, B. (2005). Cost-effectiveness analysis for multinational clinical trials. *Statistics in Medicine*, (24), 1965–1982;

[36] Sculpher, M., Pang, F., Manca, A., Drummond, M., Golder, S., Urdahl, H., Davies, L. and Eastwood, A. (2004). Generalisability in economic evaluation studies in healthcare: a review and case studies. *Technical Report 49, Health Technology Assessment*;

[37] Sculpher, M., Claxton, K., Drummond, M., McCabe, C. (2006). Whither trial-based economic evaluation for health care decision making?. *Health Economics* (15), 677–687;

[38] Sculpher, M., Drummond, M. (2006). Analysis Sans Frontiéres: can we ever make economic evaluations generalizable across jurisdictions?. *Pharmacoeconomic*, 24, (11), 1087-1099;

[39] Smith, P., Ginnelly, L., Sculpher, M. (2005). Health Policy and Economics – Opportunities and Challenges. *Open University Press – State of Health Series*;

[40] Soares, M. (2007). Stochastic processes in cost-effectiveness in health care – MSc thesis;

[41] Sonnenberg, F. and Beck, J. (1993). Markov models in medical decision making: a practical guide. *Medical Decision Making*, 13, (4), 322–338;

[42] Spiegelhalter, D. (1998). Bayesian graphical modelling: a case study in monitoring health outcomes. *Journal of the Royal Statistics Society Series*, (47), 115–133;

[43] Spiegelhalter, D. J., Thomas, A., Best, N. G., Lunn, D. (2003). WinBUGS Version 1.4 User Manual. *Medical Research Council*, Biostatistics Unit: Cambridge;

[44] Willan, A., Pinto, E., O'Brien, B., *et al*. (2005) Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. *Health Economics*, (14), 327–338;

[45] Willke, R., Glick, H., Polsky, D., Schulman, K. (1998) Estimating country-specific cost-effectiveness from multinational clinical trials. Health Economics, (7), 481–493;

[46] Wooldridge, J. (2002). Econometric analysis of cross-section and panel data. *MIT Press*;

6 Appendix

Centre	Obs (n_j)	Age mean (std.dev.)	Diabetes mean (std.dev.)	Previous MI mean (std.dev.)	Smoker mean (std.dev.)	Pulse mean (std.dev.)	ST depression mean (std.dev.)	Angina mean (std.dev.)	Male mean (std.dev.)	Left BBB mean (std.dev.)	Treat mean (std.dev.)	Risk score mean (std.dev.)
1	39	0.74 (0.81)	0.12 (0.33)	0.15 (0.36)	0.43 (0.50)	7.51 (3.21)	0.41 (0.49)	0.48 (0.50)	0.64 (0.48)	0.00 (-)	0.51 (0.50)	0.18 (0.10)
2	17	0.94 (0.89)	0.11 (0.33)	0.17 (0.39)	0.35 (0.49)	7.00 (1.83)	0.41 (0.50)	0.35 (0.49)	0.82 (0.39)	0.00 (-)	0.47 (0.51)	0.19 (0.11)
3	101	0.80 (0.82)	0.16 (0.37)	0.31 (0.46)	0.32 (0.47)	8.26 (2.66)	0.61 (0.48)	0.45 (0.50)	0.60 (0.49)	0.06 (0.25)	0.50 (0.50)	0.21 (0.12)
4	21	0.76 (0.88)	0.19 (0.40)	0.33 (0.48)	0.47 (0.51)	8.61 (2.57)	0.33 (0.48)	0.42 (0.50)	0.52 (0.51)	0.04 (0.21)	0.47 (0.51)	0.18 (0.11)
5	32	1.34 (0.86)	0.12 (0.33)	0.25 (0.44)	0.18 (0.39)	7.21 (2.53)	0.53 (0.50)	0.31 (0.47)	0.71 (0.45)	0.06 (0.24)	0.50 (0.50)	0.26 (0.16)
6	33	0.63 (0.78)	0.00 (0.00)	0.21 (0.41)	0.48 (0.50)	6.48 (2.69)	0.12 (0.33)	0.54 (0.50)	0.84 (0.36)	0.00 (-)	0.45 (0.50)	0.15 (0.08)
7	84	1.29 (0.84)	0.09 (0.29)	0.16 (0.37)	0.16 (0.37)	8.39 (3.45)	0.44 (0.49)	0.31 (0.46)	0.58 (0.49)	0.04 (0.21)	0.50 (0.50)	0.23 (0.14)
8	42	0.88 (0.80)	0.19 (0.39)	0.33 (0.47)	0.21 (0.41)	6.47 (3.03)	0.28 (0.45)	0.42 (0.50)	0.59 (0.49)	0.04 (0.21)	0.45 (0.50)	0.18 (0.12)
9	21	0.90 (0.88)	0.04 (0.21)	0.28 (0.46)	0.47 (0.51)	9.47 (3.23)	0.28 (0.46)	0.42 (0.50)	0.47 (0.51)	0.09 (0.30)	0.52 (0.51)	0.24 (0.16)
10	6	0.33 (0.51)	0.16 (0.40)	0.66 (0.51)	0.66 (0.51)	6.83 (3.18)	0.50 (0.54)	0.33 (0.51)	0.33 (0.51)	0.00 (-)	0.50 (0.54)	0.17 (0.09)
11	153	0.84 (0.83)	0.11 (0.32)	0.23 (0.42)	0.44 (0.49)	7.32 (2.46)	0.29 (0.45)	0.26 (0.44)	0.63 (0.48)	0.01 (0.11)	0.50 (0.50)	0.17 (0.10)
12	38	0.60 (0.67)	0.26 (0.44)	0.42 (0.50)	0.44 (0.50)	7.44 (2.36)	0.13 (0.34)	0.50 (0.50)	0.65 (0.48)	0.02 (0.16)	0.47 (0.50)	0.19 (0.10)
13	14	0.71 (0.72)	0.21 (0.42)	0.28 (0.46)	0.50 (0.51)	6.42 (2.10)	0.42 (0.51)	0.71 (0.46)	0.42 (0.51)	0.00 (-)	0.42 (0.51)	0.18 (0.09)
14	65	0.96 (0.86)	0.12 (0.33)	0.26 (0.44)	0.26 (0.44)	8.26 (3.26)	0.44 (0.50)	0.16 (0.37)	0.60 (0.49)	0.03 (0.17)	0.49 (0.50)	0.20 (0.13)
15	33	0.72 (0.76)	0.15 (0.36)	0.18 (0.39)	0.45 (0.50)	6.75 (2.65)	0.06 (0.24)	0.39 (0.49)	0.57 (0.50)	0.00 (-)	0.51 (0.50)	0.16 (0.11)
16	53	0.75 (0.83)	0.11 (0.32)	0.15 (0.36)	0.47 (0.50)	7.01 (2.52)	0.28 (0.45)	0.30 (0.46)	0.69 (0.46)	0.00 (-)	0.50 (0.50)	0.15 (0.09)
17	72	0.83 (0.75)	0.18 (0.38)	0.40 (0.49)	0.30 (0.46)	7.15 (2.60)	0.77 (0.41)	0.58 (0.49)	0.66 (0.47)	0.05 (0.23)	0.50 (0.50)	0.23 (0.15)
18	1	0.00 (-)	0.00 (-)	0.00 (-)	0.00 (-)	10.0 (-)	0.00 (-)	0.00 (-)	0.00 (-)	0.00 (-)	0.00 (-)	0.08 (-)
19	45	0.66 (0.73)	0.08 (0.28)	0.31 (0.46)	0.35 (0.48)	6.57 (2.12)	0.06 (0.25)	0.42 (0.49)	0.68 (0.46)	0.02 (0.14)	0.48 (0.50)	0.15 (0.11)
20	78	0.78 (0.84)	0.07 (0.26)	0.19 (0.39)	0.30 (0.46)	7.78 (3.32)	0.51 (0.50)	0.21 (0.41)	0.60 (0.49)	0.01 (0.11)	0.50 (0.50)	0.17 (0.13)
21	77	0.80 (0.76)	0.23 (0.42)	0.35 (0.48)	0.41 (0.49)	7.53 (2.70)	0.37 (0.48)	0.19 (0.39)	0.66 (0.47)	0.02 (0.16)	0.49 (0.50)	0.19 (0.12)
22	23	1.00 (0.73)	0.21 (0.42)	0.30 (0.47)	0.17 (0.38)	8.65 (2.79)	0.34 (0.48)	0.21 (0.42)	0.52 (0.51)	0.08 (0.28)	0.52 (0.51)	0.20 (0.17)
23	65	1.32 (1.01)	0.13 (0.34)	0.30 (0.46)	0.23 (0.42)	6.80 (2.12)	0.23 (0.42)	0.26 (0.44)	0.66 (0.47)	0.01 (0.12)	0.49 (0.50)	0.21 (0.15)
24	21	0.95 (0.92)	0.04 (0.21)	0.38 (0.49)	0.19 (0.40)	6.61 (2.50)	0.33 (0.48)	0.38 (0.49)	0.52 (0.51)	0.09 (0.30)	0.47 (0.51)	0.19 (0.12)
25	10	0.30 (0.48)	0.00 (0.00)	0.20 (0.42)	0.40 (0.51)	8.30 (2.75)	0.20 (0.42)	0.60 (0.51)	0.40 (0.51)	0.10 (0.31)	0.70 (0.48)	0.11 (0.04)
26	31	0.83 (0.86)	0.25 (0.44)	0.29 (0.46)	0.29 (0.46)	8.09 (3.20)	0.32 (0.47)	0.38 (0.49)	0.64 (0.48)	0.06 (0.25)	0.45 (0.50)	0.19 (0.15)
27	10	1.10 (0.87)	0.20 (0.42)	0.30 (0.48)	0.30 (0.48)	7.80 (1.54)	0.50 (0.52)	0.10 (0.31)	0.60 (0.51)	0.00 (-)	0.50 (0.52)	0.20 (0.13)
28	27	0.85 (0.86)	0.07 (0.26)	0.33 (0.48)	0.18 (0.39)	7.07 (2.26)	0.37 (0.49)	0.59 (0.50)	0.74 (0.44)	0.00 (-)	0.51 (0.50)	0.19 (0.13)
29	17	1.00 (0.93)	0.11 (0.33)	0.23 (0.43)	0.29 (0.47)	7.58 (3.37)	0.52 (0.51)	0.11 (0.33)	0.52 (0.51)	0.00 (-)	0.47 (0.51)	0.19 (0.13)
30	64	0.68 (0.61)	0.17 (0.38)	0.35 (0.48)	0.42 (0.49)	6.96 (2.46)	0.54 (0.50)	0.34 (0.48)	0.57 (0.49)	0.04 (0.21)	0.50 (0.50)	0.18 (0.10)
31	12	0.91 (0.90)	0.08 (0.28)	0.41 (0.51)	0.25 (0.45)	6.08 (1.78)	0.33 (0.49)	0.50 (0.52)	0.50 (0.52)	0.00 (-)	0.50 (0.52)	0.19 (0.19)
32	55	1.38 (1.11)	0.07 (0.26)	0.18 (0.38)	0.16 (0.37)	6.92 (2.23)	0.23 (0.42)	0.49 (0.50)	0.50 (0.50)	0.03 (0.18)	0.49 (0.50)	0.21 (0.14)
33	29	0.93 (0.88)	0.06 (0.25)	0.31 (0.47)	0.31 (0.47)	7.62 (2.93)	0.55 (0.50)	0.24 (0.43)	0.62 (0.49)	0.00 (-)	0.48 (0.50)	0.19 (0.10)
34	10	0.40 (0.69)	0.10 (0.31)	0.50 (0.52)	0.30 (0.48)	9.30 (3.49)	0.20 (0.42)	0.40 (0.51)	0.60 (0.51)	0.00 (-)	0.50 (0.52)	0.15 (0.05)
35	13	1.30 (1.03)	0.07 (0.27)	0.15 (0.37)	0.38 (0.50)	5.76 (1.53)	0.23 (0.43)	0.53 (0.51)	0.76 (0.43)	0.00 (-)	0.46 (0.51)	0.21 (0.12)
36	19	1.00 (0.74)	0.21 (0.41)	0.26 (0.45)	0.36 (0.49)	7.42 (3.35)	0.68 (0.47)	0.42 (0.50)	0.42 (0.50)	0.21 (0.41)	0.52 (0.51)	0.23 (0.16)
37	94	0.96 (0.79)	0.19 (0.39)	0.28 (0.45)	0.25 (0.43)	7.55 (2.76)	0.37 (0.48)	0.23 (0.42)	0.56 (0.49)	0.02 (0.14)	0.50 (0.50)	0.18 (0.11)
38	31	0.93 (0.81)	0.09 (0.30)	0.25 (0.44)	0.22 (0.42)	7.67 (2.91)	0.29 (0.46)	0.16 (0.37)	0.64 (0.48)	0.03 (0.18)	0.45 (0.50)	0.17 (0.09)
39	5	1.20 (0.83)	0.40 (0.54)	0.40 (0.54)	0.20 (0.44)	7.60 (2.60)	0.20 (0.44)	0.60 (0.54)	0.80 (0.44)	0.20 (0.44)	0.40 (0.54)	0.27 (0.20)
40	110	0.79 (0.75)	0.12 (0.33)	0.24 (0.43)	0.28 (0.45)	7.50 (2.81)	0.19 (0.39)	0.39 (0.49)	0.59 (0.49)	0.04 (0.20)	0.50 (0.50)	0.16 (0.10)
41	24	0.58 (0.83)	0.16 (0.38)	0.37 (0.49)	0.45 (0.50)	7.62 (2.14)	0.29 (0.46)	0.37 (0.49)	0.66 (0.48)	0.12 (0.33)	0.50 (0.51)	0.20 (0.15)
42	39	0.79 (0.83)	0.10 (0.30)	0.30 (0.46)	0.15 (0.36)	7.05 (2.62)	0.25 (0.44)	0.69 (0.46)	0.66 (0.47)	0.05 (0.22)	0.46 (0.50)	0.18 (0.09)
43	37	0.73 (0.76)	0.10 (0.31)	0.32 (0.47)	0.32 (0.47)	7.67 (3.07)	0.45 (0.50)	0.18 (0.39)	0.86 (0.34)	0.00 (-)	0.51 (0.50)	0.19 (0.13)
44	8	0.50 (1.06)	0.00 (0.00)	0.00 (0.00)	0.50 (0.53)	6.50 (2.13)	0.12 (0.35)	0.50 (0.53)	0.62 (0.51)	0.12 (0.35)	0.62 (0.51)	0.12 (0.08)
45	21	1.00 (1.04)	0.00 (0.00)	0.23 (0.43)	0.38 (0.49)	6.90 (2.36)	0.23 (0.43)	0.52 (0.51)	0.71 (0.46)	0.04 (0.21)	0.47 (0.51)	0.20 (0.16)
46	10	1.40 (1.07)	0.10 (0.31)	0.50 (0.52)	0.20 (0.42)	6.90 (3.03)	0.10 (0.31)	0.40 (0.51)	0.50 (0.52)	0.00 (-)	0.40 (0.51)	0.23 (0.20)

Table A1. Covariates included in the statistical models by centre.

Weibull regression

		WinBugs** - HM			
Centre		mean	std. dev.	95% CrI	
Random Effects					
centre 2	$u_{1j\ -\ Treat}$	0.021	0.258	-0.506	0.649
	$u_{0j\ -\ Cnst}$	-0.007	0.076	-0.178	0.146
centre 11	$u_{1j\ -\ Treat}$	0.082	0.300	-0.275	0.608
	$u_{0j\ -\ Cnst}$	0.003	0.066	-0.145	0.157
centre 23	$u_{1j\ -\ Treat}$	-0.043	0.239	-0.639	0.426
	$u_{0j\ -\ Cnst}$	0.010	0.070	-0.124	0.179
centre 37	$u_{1j\ -\ Treat}$	-0.131	0.269	-0.861	0.235
	$u_{0j\ -\ Cnst}$	-0.007	0.067	-0.171	0.129
centre 40	$u_{1j\ -\ Treat}$	0.021	0.213	-0.402	0.527
	$u_{0j\ -\ Cnst}$	0.005	0.071	-0.143	0.182

**5,000 iterations and a 2,000 iteration burn-in period

Table A2. Random effects components of 5 centres, results of *Bayesian* hierarchical Weibull PHM of composite endpoint of CVD or MI from hospital discharge until end of trial (HM – hierarchical model).

Logistic regression

Non-fatal MI		WinBugs** - HM			
Centre		mean	std. dev.	95% CrI	
Random Effects					
centre 2	$u_{0j\ -\ Cnst}$	-0.011	0.178	-0.436	0.346
centre 11	$u_{0j\ -\ Cnst}$	-0.014	0.154	-0.389	0.302
centre 23	$u_{0j\ -\ Cnst}$	0.065	0.187	-0.203	0.582
centre 37	$u_{0j\ -\ Cnst}$	-0.012	0.163	-0.409	0.323
centre 40	$u_{0j\ -\ Cnst}$	-0.058	0.177	-0.549	0.200

**5,000 iterations and a 2,000 iteration burn-in period

Table A3. Random effects components of 5 centres, results of *Bayesian* hierarchical logistic regression of composite endpoint of CVD or MI being non-fatal (HM – hierarchical model).

Linear model

Costs index		WinBugs** - HM			
Centre		mean	std. dev.	95% CrI	
Random Effects					
centre 2	$u_{1j\ -\ Treat}$	-855.9	1333.2	-3585.8	1645.5
	$u_{0j\ -\ Cnst}$	-586.9	801.8	-2220.6	934.5
centre 11	$u_{1j\ -\ Treat}$	-2261.3	821.9	-3891.6	-670.6
	$u_{0j\ -\ Cnst}$	-1219.1	555.7	-2335.3	-176.3
centre 23	$u_{1j\ -\ Treat}$	39.4	975.2	-1869.4	1943.4
	$u_{0j\ -\ Cnst}$	-0.361	616.7	-1231.9	1220.4
centre 37	$u_{1j\ -\ Treat}$	1905.5	876.7	193.4	3610.6
	$u_{0j\ -\ Cnst}$	178.2	567.7	-932.9	1304.1
centre 40	$u_{1j\ -\ Treat}$	128.1	817.5	-1450.9	1745.5
	$u_{0j\ -\ Cnst}$	-175.2	539.6	-1254.1	838.2

**5,000 iterations and a 2,000 iteration burn-in period

Table A4. Random effects components of 5 centres, results of *Bayesian* hierarchical linear regression of costs during the index hospitalisation (HM – hierarchical model).

Linear model					
Costs follow-up exc.MI/stroke		**WinBugs** - HM**			
Centre		mean	std. dev.	95% CrI	
	Random Effects				
centre 2	$u_{1j\ -\ Treat}$	103.7	483.3	-833.4	1154.7
	$u_{0j\ -\ Cnst}$	95.8	546.1	-952.9	1175.6
centre 11	$u_{1j\ -\ Treat}$	-7.320	397.5	-799.8	775.7
	$u_{0j\ -\ Cnst}$	-142.2	372.5	-892.4	587.1
centre 23	$u_{1j\ -\ Treat}$	76.8	440.3	-750.6	966.7
	$u_{0j\ -\ Cnst}$	-133.5	442.1	-1013.0	733.6
centre 37	$u_{1j\ -\ Treat}$	-209.2	422.6	-1121.6	580.0
	$u_{0j\ -\ Cnst}$	-144.8	417.8	-972.8	669.8
centre 40	$u_{1j\ -\ Treat}$	30.39	412.0	-774.2	874.0
	$u_{0j\ -\ Cnst}$	-358.9	405.0	-1184.6	392.3

**5,000 iterations and a 2,000 iteration burn-in period

Table A5. Random effects components of 5 centres, results of *Bayesian* hierarchical linear regression of costs during the follow-up period (HM – hierarchical model).

Linear model					
HRQoL baseline		**WinBugs** - HM**			
Centre		mean	std. dev.	95% CrI	
	Random Effects				
centre 2	$u_{0j\ -\ Cnst}$	-0.102	0.062	-0.224	0.018
centre 11	$u_{0j\ -\ Cnst}$	0.131	0.030	0.072	0.189
centre 23	$u_{0j\ -\ Cnst}$	-0.071	0.038	-0.145	0.003
centre 37	$u_{0j\ -\ Cnst}$	0.046	0.034	-0.019	0.112
centre 40	$u_{0j\ -\ Cnst}$	0.141	0.033	0.077	0.203

**5,000 iterations and a 2,000 iteration burn-in period

Table A6. Random effects components of 5 centres, results of *Bayesian* hierarchical linear regression of baseline utilities (HM – hierarchical model).

Longitudinal data					
Change HRQoL		**WinBugs** - CHM**			
Centre		mean	std. dev.	95% CrI	
	Random Effects				
centre 2	$u_{1j\ -\ Treat\_centre}$	-1.103	0.130	-1.360	-0.858
	$u_{0j\ -\ Cnst\_centre}$	0.111	0.075	-0.032	0.262
centre 11	$u_{1j\ -\ Treat\_centre}$	-1.320	0.047	-1.412	-1.232
	$u_{0j\ -\ Cnst\_centre}$	-0.090	0.037	-0.162	-0.0155
centre 23	$u_{1j\ -\ Treat\_centre}$	-1.191	0.070	-1.337	-1.059
	$u_{0j\ -\ Cnst\_centre}$	0.056	0.048	-0.037	0.147
centre 37	$u_{1j\ -\ Treat\_centre}$	-1.333	0.056	-1.448	-1.225
	$u_{0j\ -\ Cnst\_centre}$	-0.019	0.043	-0.099	0.072
centre 40	$u_{1j\ -\ Treat\_centre}$	-1.333	0.056	-1.448	-1.225
	$u_{0j\ -\ Cnst\_centre}$	-0.065	0.041	-0.149	0.016

**5,000 iterations and a 2,000 iteration burn-in period

Table A7. Random effects components of 5 centres, results of *Bayesian* hierarchical panel data regression of the gain in HRQoL (HM – hierarchical model).