USING MULTICENTRE RCT-BASED INDIVIDUAL PATIENT LEVEL DATA TO POPULATE DECISION ANALYTIC COST-EFFECTIVENESS MODELS FOR LOCATION-SPECIFIC DECISION MAKING

A Dissertation Presented

for the Degree of

MASTER OF SCIENCE

IN HEALTH ECONOMICS

by

Pedro Saramago Gonçalves

THE UNIVERSITY of York

CHE
Centre For Health Economics

September 2008

Acknowledgements

I sincerely thank Dr. Andrea Manca for the priceless help and expert supervision of the placement project. Andrea Manca's immense support and constant enthusiasm over this project was responsible for it to be carried out at an extremely high motivation level, which I am sure it will continue throughout the upcoming PhD stage.

I thank Martin Henriksson for providing all the documentation on the economic analysis performed over the RITA 3 trial dataset.

Special thanks are also due to Marta Soares and Dr. Claire McKenna for their constructive comments and suggestions.

I am extremely grateful to Professor Mark Sculpher and Professor Karl Claxton for giving me the opportunity to work within the Team for Economic Evaluation and Health Technology Assessment (TEEHTA) of the Centre for Health Economics (CHE). I also thank all the members of the CHE team, especially the TEEHTA team for welcoming me into their group.

I thank Professor Andrew Jones and Professor Karl Claxton for their excellent coordination and organisation of the graduate programme in Health Economics at the University of York.

Financial support throughout the year from Merck Sharp & Dohme Portuguese Foundation is also gratefully acknowledged.

An immense thank you is due to my beautiful and gifted partner Marta Soares for her colossal support and for making my existence so joyful.

My last word of thanks goes to my father, mother, sister and grandmother for their constant love and support.

.

Contents

Tables

Figures

Abbreviations and definitions

AFTM - Accelerated failure life-time model

BLUE - Best linear unbiased estimator

CE - Cost-effectiveness

CEA - Cost-effectiveness analysis

CEAC - Cost-effectiveness acceptability curve

CEP - Cost-effectiveness plane

CHM - Centre hierarchical model

CrI - Credibility interval

CV - Cardiovascular

CVD - Cardiovascular death

DAM - Decision analytic model

EE - Economic evaluation

GLM - Generalized linear model

GLS - Generalized least squares

HM - Hierarchical model

HRQoL - Health related quality of life

HR - Hazard ratio

HTA - Health Technology Assessment

ICC - Intraclass correlation coefficient

ICER - Incremental cost-effectiveness ratio

IID - Independent and identically distributed

IPD - Individual patient data

MCMC - Markov chain Monte Carlo

MI - Myocardial infarction

MLE - Maximum likelihood estimation

MC - Monte Carlo

NB - Net benefit

NCHM - Non-centre hierarchical model

NHB - Net health benefit

NHM - Non-hierarchical model

NMB - Net monetary benefit

NID - Normally and independently distributed

NICE - National Institute for Health and Clinical Excellence

NSTE-ACS - Non-ST-elevation acute coronary syndrome

OLS - Ordinary least squares

PDF - Probability density function

PHM - Proportional hazards model

PSA - Probabilistic sensitivity analysis

QALY - Quality adjusted life years

QML - Quasi maximum likelihood

RCT - Randomized controlled trial

RITA 3 - Third randomized intervention trial of unstable angina

STD DEV – Standard deviation

SE - Standard error

TP - Transition probability

Using multicentre RCT-based individual patient level data to populate decision analytic cost-effectiveness models for location-specific decision making

Abstract

Objectives: To develop methodology for the analysis of individual patient level data from multicentre/multinational randomized controlled trials 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 and *Bayesian* shrinkage estimation procedures were used to obtain location-specific cost-effectiveness 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 cost-effectiveness planes 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 incremental cost-effectiveness ratio 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 cost-effectiveness acceptability curves, 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 thesis shows how *Bayesian* hierarchical modelling can be used to estimate more appropriate cluster-specific parameters for use in decision analytic models where individual patient level data 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.

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 [1]. 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 (QALYs). 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 [2].

In order to inform NICE decision-making process, an EE is required to address two main questions [3]. 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 [4]. Quantifying the cost of making a wrong decision represents the basis for assessing if whether acquiring further evidence through funding new research is valuable [5].

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 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. When a trade-off situation is raised, decision rules should be applied [10]. If the ICER is used, it is interesting to assess the probability of its estimates being smaller than fixed values of willingness to pay (predefined threshold), $ICER < \lambda$ (with $\Delta E>0$). In these circumstances, the intervention is cost-effective in relation to the comparator. With respect to the NB framework, it is attractive to evaluate the probability of its estimates being positive. The new technology would be accepted if: $NMB = \lambda \cdot \Delta E - \Delta C > 0$ (net monetary benefit (NMB)) or equivalently $NHB = \Delta E - \Delta C/\lambda > 0$ (net health benefit (NHB)) [13].

Subjective decisions can be avoided by using analytic models. As there is uncertainty around the cost-effectiveness (CE) estimates, any decision based on CEA will also be uncertain [12]. 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 [11]. Decision analytic models (DAM) are used to combine information from various sources of information using mathematical relationships [8, 9]. Often 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. In that case, *Monte Carlo* (MC) methods can be used to propagate uncertainty in the model over the expected outcome measure. This methodology entails randomly sampling from the distribution of the expected input parameters and recording the realisations of the results of the CE measure iteratively. The CE pairs accumulated can be used to produce an estimate of the joint distribution of the mean difference costs and mean difference benefits. 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.

Figure 1 illustrates possible examples of the joint CE density plotted on the incremental cost-effectiveness plane (CEP). The CEP is a Cartesian coordinate system used to display cost and effectiveness differences in relation to a reference therapy. On the *x-axis* are effect differences (ΔE) and on the *y-axis* cost differences (ΔC).

Considering a reference treatment $\{R\}$, interventions located in the *NW* quadrant (*A*) are a reflection of lower expected incremental benefits and higher incremental costs; therefore the adoption of this intervention is not recommended. On the opposite end, alternatives on the *SE* quadrant lead to gains in effectiveness and reduced costs; interventions in this outline, such as *D*, should be supported due to dominance over *R*. Interventions located on the *SW* quadrant (*C*) are often subject of ethical debate. Finally, often new technologies, such as *B*, have incremental results located in the *NE* quadrant of the CEP, which are more effective but with greater associated costs, and as mentioned above the trade-off situation arises.

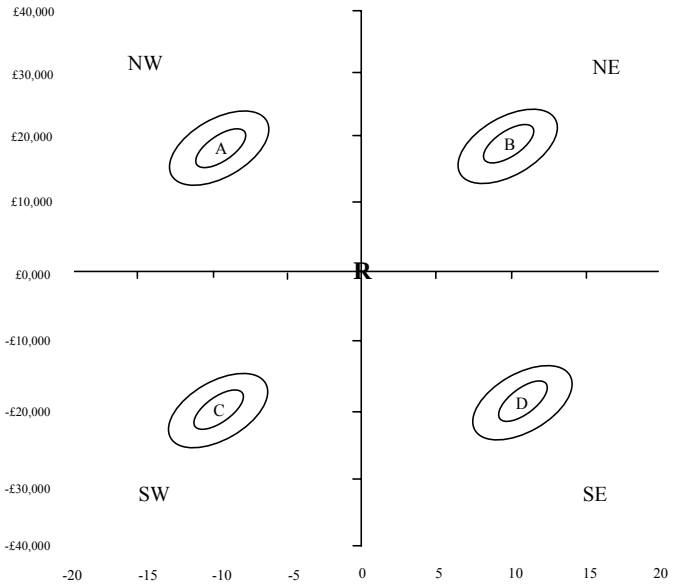

Figure 1. Cost-effectiveness scenarios illustrated on the incremental cost-effectiveness plane

.

The difficulties to visualize alternative thresholds and to assess the probability of *CE* are two problems faced when considering CEP representation [5]. Considerations on the uncertainty surrounding a decision to accept/reject the new technology can be based on a graphical illustration named cost-effectiveness acceptability curve (CEAC). The CEAC is more informative than confidence intervals, and has a natural *Bayesian* interpretation. In this framework, the CEAC represents the probability of the new therapy being cost-effective [20] and is estimated through the proportion of MC samples which lie below a specific threshold.

1.2 Trials *vs.* models: the false dichotomy

The development of DAMs is currently seen as vital to the process of Health Technology Assessment (HTA) in general, occupying a nuclear position in the technology appraisal process at NICE [5]. Because of the specific framework around trial based analysis and the requirements for EE for decision making, DAMs are increasingly being used to inform policy decisions regarding the optimum allocation of health care expenditures. Models have occasionally been characterised as substitute to trials, however such a statement is a reflection of misunderstanding of their particular function. Randomized controlled trials (RCT*)* provide estimates of particular parameters in a specific group of patients in a particular health care environment over a particular timeframe. Decision models supply a configuration within which data from a variety of sources can be focussed to inform a precise decision problem for a defined population and perspective. The dissimilarities involving measurement achieved by trials and decision making supported by analytical structures highlights that models and trials are complements, not substitutes [8].

The development of statistical methods in this area has strengthened the advantages of analysing patient-level data in CEA. Recent examples of applied work employing regression analysis to individual patient data (IPD) from RCTs to populate stochastic DAMs are: Briggs *et al* [22] using the EUROPA trial data, Henriksson *et al* [23] using the RITA 3 trial data and Briggs *et al* [24] using the GOAL trial data. However, data used to populate DAMs often come from multinational/multicentre RCTs. This offers the opportunity to develop location-specific CE models and consequently location-specific estimates of CE measures.

1.3 Multicentre / multinational *RCTs*

Often, when conducting 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 [7]. The question is how generalizable are the results of a multiple location evaluation to specific sites and their individual health care situations [15, 19].

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 [21]. 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 [6, 19].

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 [17], it disregards also that costs and effects are naturally correlated. Consequently, the correct quantification of uncertainty surrounding CE estimates is endangered [16]. The use of hierarchical models provides an ideal pathway to analyze CE IPD from multiple location trials allowing for between-location variability.

Several analytical methodologies have been proposed to analyse multinational trial data and most of these involve regression analysis. Willke *et al* [15] 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* [6] extended the net benefit regression approach [14] to contain the hierarchical structure of economic data in multilocation trials. Hierarchical models were 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* [18] and Willan *et al* [17] 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. Manca *et al* [16] recently investigated the use of *Bayesian* bivariate hierarchical regression modelling to analyze individual patient level CE data collected alongside multinational trials using also empirical *Bayes* shrinkage estimation methodology.

1.4 Motivation and outline of the thesis

The particular focus of this project is to develop methodology for the analysis of multicentre/multinational RCTs with the aim of estimating location-specific parameters to populate the decision model. Multilevel or hierarchical modelling is the technique employed in this work to analyse the multicentre data and produce input parameters for the model.

There are various reasons why one might consider a multilevel modelling framework, whether for purposes of causal inference, study of variation or prediction of outcomes. Multilevel models (*i*) account for individual- and centre-level variation in estimating centre-level regression coefficients; (*ii*) models variation among individual-level regression coefficients; (*iii*) estimates regression coefficients for particular centres.

The RITA 3 trial data was used as a case study to illustrate the methodology proposed herein. The DAM used in the CE analysis of this trial involved a short-term and a long-term model; with the latter having tunnel states used to incorporate time dependency linked to particular health states on disease progression. Using IPD from this trial, multilevel regression estimates are obtained using a *Bayesian* approach. *Bayesian* hierarchical estimates are used to obtain transition probabilities, which are employed to define a location-specific *Markov* model. Location-specific estimates of CE were obtained, weighting the information available within each centre through the global estimates.

This thesis begins by summarizing the features of *Markov* modelling and subsequently the different procedures for obtaining transition probability estimates. It gives a description of single-level and multilevel regression models, focussing thereafter on the estimation through *Bayesian* methods. A case study is presented, and the results obtained by *Bayesian* hierarchical modelling are discussed. The final section summarizes the contribution of this written work and discusses possible extensions of the present work that can be dealt with in future research.

2 Methods

2.1 Model Design: *Markov* model

The aim of CE evaluation of health interventions is to evaluate the distribution of the expected outcomes. DAMs designs, for which the evaluation 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 [11]. 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 [25]. In EE the use of non-homogeneous *Markov* chains is common. This framework implies that transition probability (TP) functions are dependent on time [27, 29].

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) [28].

2.1.1 Markov Chains: a statistical description

Markov chains are understood as one of the simplest possible cases of stochastic processes. Given a probability space $\{\Omega, F, M\}$, a stochastic process (or random process) is a collection of *Y*-valued random variables indexed by a set time *T*. That is, a stochastic process *F* is a collection of $\{Y_t\}_{t\in T}$. In a discrete time Markov chain, a one step transition function can be defined as $M_{i,j}^{t,t+1} = P\left[Y_{t+1} = j \middle| Y_t = i\right]$, representing the probability of Y_{t+1} being in state $j \in S$, given that at Y_t the process was in state $i \in S$, where S is a countable state space $S_{\{i,j\}}$.

Given the present state, the future and past states are independent. This assumption is regularly termed as the *Markov property* or *Markovian assumption*, meaning that the process is "memoryless" for previous cycles [26]. Another assumption in a *Markov* model is that all individuals residing in a health state are identical and any degree of heterogeneity within a health state will cause some degree of bias. In practice, when issues of heterogeneity are believed to be important, health states should be defined according to the underlying heterogeneity factor [28]. Considering estimates of the TP matrix given by the information sources, *Markov* models are used in evaluating the non-conditional probabilities of being in the different states defined.

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 [30]. However, when IPD is available, regression analysis is conducted to estimate transition probabilities. The regression framework used to estimate TPs is commonly based on parametric distributional assumptions. One can use models such as *linear models* (section 2.2.1.1), *generalized linear models* (section 2.2.1.2), *longitudinal models* (section 2.2.2.4) 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* [31].

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.

The main estimation procedures used in this thesis are the *maximum likelihood estimation* and the *Bayesian* framework. In addition to the *ordinary least squares estimation*, a detailed description of these estimation procedures is given in the technical appendix - section A2.

2.2.1 Single level regression analysis

Most econometric efforts in health economics focus on finding the model that appropriately fits the available data. Succinctly, regression analysis involves the estimation and the evaluation of the relationship between a variable of interest (dependent variable) and one or more other variables (independent variables). In this context, several estimation frameworks will be described next.

2.2.1.1 Classical linear regression

Considering a random sample of *n* independently and identically distributed (i.i.d.) observations $y_1,...,y_n$ each of which is normally distributed with mean μ and variance σ^2, but with variant means with $E[y_i] = \beta_0 + \sum_k^K \beta_k x_{ki}$, with *i=1,...,n* and *k=1,...K*. The classical linear regression analysis assumes that the relationship between an outcome, or dependent variable, *y*, and the explanatory variables or independent variables, x_i*'s*, can be summarised by a regression function [33]. The regression function is typically assumed to be a linear function of the *x* variables and of a random error term, *ε*. This relationship can be written using the following shorthand notation,

$$y_i = \beta_0 + \sum_k \beta_k x_{ki} + \varepsilon_i , \qquad \text{eq. 1}$$

where the terms β_k are *regression coefficients*, and the intercept β_0 is the mean of *y* when *x* equals zero and β_k is the change in the mean of *y* when x_k increases by one unit. The *random error* or *error term* $\varepsilon_i = y_i - \left(\beta_0 + \sum_k \beta_k x_{ki} \right)$ captures all of the variation in *y* that is not explained by the x_k variables. The study of statistical inference of the classical regression model requires the errors to be [32]: *independent*, *homoskedastic, uncorrelated*, and *normally distributed*.

2.2.1.2 Generalized Linear Modelling

Generalized linear modelling (GLM) is a framework for statistical analysis that includes linear and logistic regression as special cases. While in linear regression, as it was described in the previous section, it is proposed a model with $\mu = X\beta$ where the dependent variable is normally distributed with mean $X\beta$ and covariance matrix format of $\sigma^2 I_{n\times k}$ (identity link), in GLM framework it is proposed a model with $\eta = g(\mu) = X\beta$ where the dependent variable has distribution belonging to the *Exponential family*. It is assumed that additionally to the identity function *I* there are other

possible functions able to perform the link between μ and $X\beta$. To carry out statistical inference, the *Normal* distribution is not the only one considered given that assumptions are broadened to the *Exponential* family.

A GLM usually involves the following components [32]:

(i) a data vector $y=(y_1,...,y_n)$, with distribution belonging to the *Exponential* family;

(ii) predictors X and coefficients β : $X=\begin{bmatrix} 1 & x_{11} & ... & ... & x_{k1} \\ ... & ... & ... & ... & ... \\ 1 & x_{kn} & ... & ... & x_{kn} \end{bmatrix}$ and $\beta' \equiv (\beta_0,\beta_1,...,\beta_k)$ (considering β_0 as the intercept term) , obtaining a linear predictor $\eta = X\beta$;

(iii) a monotonic function *g* – the link function $\eta = g(\mu)$, yielding a vector of transformed data $\hat{y} = g^{-1}(X\beta)$ that are used to model the data.

It is common to define the link function *g* that allows one to have $g(\mu) = X\beta$. The more common links are the *canonical link* or *logit*, the *probit*, and the *complementary log-log link function.*

It is often the case in health data that the outcome of interest is measured as a binary variable, usually taking values of either one or zero. Often this binary variable will indicate whether an individual/patient is a *participant* or a *non-participant*. The examples include: health care utilisation, presence or absence of particular disease under study [34].

The next section will focus mainly on the logistic regression which is one of the standard ways to model the mentioned binary outcomes.

2.2.1.2.1 Logistic regression

If a binary outcome *y*, depends on a set of explanatory variables *x*, then the conditional expectation of *y* given *x*, in other words the value of *y* that individuals with characteristics *x* are likely to report on average, is

$$E(y \mid x) = 0 \cdot P(y=0 \mid x) + 1 \cdot P(y=1 \mid x) = P(y=1 \mid x) = F(x), \qquad \text{eq. 2}$$

A simple way to model binary data is to use a linear function: for which one can use the shorthand notation of section 2.2.1.2, $F(x) = x\beta$. However, the possibility of predicted probabilities outside

the range *[0,1]* creates a problem of logical inconsistency. It is common to use a non-linear function for *F(.)*. The popular choices are the "*S*" curves, that are bounded between *[0,1]* despite the values of the independent variables. The most common choices of these "*S*" curves are the *logit* and *probit* models. The probability functions for the *probit* and *logit* models both are similar in appearance, although the *logit* model gives more weight to the tails of the distribution [34; 36].

Logit model (as well as the *probit* model) is often encouraged in terms of a latent variable specification. This assumes that there is some continuous latent variable *y** that determines participation [34]. Let,

$$\begin{aligned} y_i &= 1 \quad \text{iff } y_i^* > 0 \\ &= 0 \quad \text{otherwise} \end{aligned} \qquad \text{where, } \; y_i^* = x_i\beta + \varepsilon_i \,, \qquad \text{eq. 3}$$

and, for a symmetrically distributed error term ε with distribution function *F(.)*,

$$P\left(y_i = 1 \mid x_i\right) = P\left(y_i^* > 0 \mid x_i\right) = P\left(\varepsilon > -x_i\beta\right) = F\left(x_i\beta\right). \qquad \text{eq. 4}$$

In this specified framework, assuming that ε_i follows a standard logistic distribution, gives the *logit* model. The *log-likelihood* for a sample of independent observations $y_i \sim Bernoulli\left(p_i\right)$ is,

$$\ell = \sum_i LogL\left(y_i \mid x_i\beta\right) = \sum_i \left\{\left(1-y_i\right)\log\left(1-F\left(x_i\beta\right)\right) + y_i \log\left(F\left(x_i\beta\right)\right)\right\}, \qquad \text{eq. 5}$$

or equivalently, using the equalities $P\left(y_i = 1 \mid x_i\right) = p_i$ and $\text{logit}\left(p_i\right) = \beta_k X_i$:

$$\begin{aligned} \ell &= \sum_i LogL\left(y_i \mid \beta_0,...,\beta_k\right) = \sum_i \left\{ y_i \log\left(\frac{p_i}{1-p_i}\right) + \log\left(1-p_i\right)\right\} = \\ &= \sum_i \left\{ y_i\left(\beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki}\right) - \log\left(1+\exp\left(\beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki}\right)\right)\right\} \end{aligned} \qquad \text{eq. 6}$$

If there are n_i responses associated to the same observation leading to $y_i \sim Binomial\left(n_i, p_i\right)$, the derivations are similar ($n_i p_i$ substitutes p_i).

These models are usually estimated by *maximum likelihood estimation* (MLE). One usually wants to find the values of $\beta \equiv \left\{\beta_0,...,\beta_k\right\}$ that maximize the log-likelihood function. The MLE procedure is presented in the technical appendix – section A2.2.

The coefficients in a logistic regression can be challenging to interpret because of the non-linearity. In the *logit* model the β coefficients can be interpreted in terms of *log-odds ratios*, a concept that is commonly used in biostatistics and epidemiology. Because of the particular functional form of the standard *logistic* distribution the *odds ratio* simplifies to $\frac{P(y_l = 1 \mid x)}{P(y_i = 0 \mid x)} = \exp(x\beta_k)$ and therefore the coefficients can be interpreted in terms of changes in the *log odds ratio* $\log\left(\frac{P(y_l = 1 \mid x)}{P(y_i = 0 \mid x)}\right) = x\beta_k$ [32, 35].

2.2.1.3 Survival Analysis

Some response variables in health economics come in the form of a duration, which is the time elapsed from a well-defined *time origin* until the occurrence of a particular *event* occurs or *end-point*. In applied research, the time origin will often correspond to the recruitment of an individual into an experiment study, such as a clinical trial to compare two or more health technologies or treatments. If, for instance, the end-point is the death of a patient, the resulting data are literally survival times (this is the main reason for the name 'survival analysis'). However, data of a similar form can be obtained when the end-point is not fatal. Examples are the recurrence of disease symptoms or the relief of pain. In these cases, observations are usually referred as *time to event* data [37].

Two main reasons for survival data not being opened to standard statistical procedures are the fact that generally the data is not symmetrically distributed (usually are *positively skewed*) and also because survival times are frequently *censored*. The survival time of a subject is said to be censored when the end-point of interest has not been observed for that individual. The most common censoring type is *right censoring*. This may happen because the data from a study are to be analysed at a point in time when some individuals are still alive or because they are *lost to follow-up*. Two other forms of censoring are *left-censoring* and *interval censoring*. The former is encountered when the actual survival time of an individual is less than that observed. In the latter, individuals are known to have experienced an event within an interval of time [37].

2.2.1.3.1 Survival and hazard functions

In summarizing survival data, there are two functions of vital importance, namely the *survivor function* and the *hazard function*. The actual survival time of an individual can be regarded as the value of a random variable, *T* or survival time, which can take any non-negative value. *T* has a probability distribution with underlying p.d.f. *f(t)*. The survivor function, *S(t)*, is defined to be the probability that the survival time is greater that or equal to *t*:

$$S(t) = P(T \geq t) = 1 - F(t). \qquad \text{eq. 7}$$

The survivor function can be used to represent the probability that an individual survives from the origin to some time beyond *t*.

The hazard function is widely used to express the risk or hazard of death at some time *t*, and it is obtained from the probability that an individual dies at time *t*, conditional on he or she having survived to that time. Considering the conditional probability that the random variable associated with an individual's survival time *T*, lies between *t* and $t+\delta t$, the hazard function is defined as:

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t} \right\} \equiv f(t) \cdot \frac{1}{S(t)}. \qquad \text{eq. 8}$$

Methods for estimating the survivor function and hazard function can be *non-parametric/ distribution-free*, *semi-parametric* or *parametric*. While the first two methods do not require specific assumptions to be made about the underlying distribution of the survival times, the last method requires so. The most known non-parametric method to estimate the abovementioned functions for a single sample of survival data are the *Kaplan-Meier* estimates. In most studies that give rise to survival data, supplementary information will also be recorded on each individual. A typical example would be a clinical trial to compare the survival times of patients who receive one or other of two treatments. In order to explore the relationship between the survival experience and the (explanatory) variables additionally gathered in the study, an approach based on statistical modelling can be used [37].

Most of the regression models in survival analysis belong to one of the following two classes: (*i*) Proportional hazard models (PHM), or (*ii*) Accelerated failure life-time models (AFTM). The AFTM is used in circumstances where the *proportional hazards assumption* (hazard of death at any given time for an individual in one group is proportional to the hazard at that time for a similar individual in the other group) is not reasonable. Due to its versatility, the *Cox-regression* model is the most widely used regression model in survival analysis. In this type of models, the underlying hazard function is not specified (therefore it is considered a semi-parametric method) but is a PHM [37].

Nevertheless, in EE one is interested in specifying parametrically the distribution of *T*. Therefore, the parametric proportional hazards model is described in the next section with emphasis on the *Weibull* distribution setup.

2.2.1.3.2 Parametric proportional hazards model – the *Weibull* distribution

The PHM applicability is widespread in the analysis of survival data, despite having relatively few probability distributions for the survival times that can be used (*Weibull* and *Gompertz* distributions are the most commonly used).

Given that the *Weibull* distribution plays a central role in the analysis of survival data, the PHM based on the *Weibull* distribution is considered in more detail below.

In applied research, the assumption of an unvarying hazard functions or equivalently, exponentially distributed survival times, is rarely plausible. A more general form of the hazard function is such that

$$h(t) = \lambda\gamma t^{\gamma-1}, \quad \text{for} \quad 0 \le t < \infty, \qquad \text{eq. 9}$$

where γ is known as the shape parameter and λ as the scale parameter. In the particular case where $\gamma=1$, the hazard function takes the constant value λ and the survival times have an exponential distribution. For values $0<\gamma<1$ it decreases (decreasing duration dependence) and for values $\gamma>1$ it increases monotonically (increasing duration dependence).

The survivor function is given by $S(t) = \exp\left(-\lambda t^{\gamma}\right)$ and the corresponding p.d.f. is then $f(t) = \lambda\gamma t^{\gamma-1}\exp\left(-\lambda t^{\gamma}\right)$, for $0 \le t < \infty$. In the *Weibull* regression model and under the PHM, the hazard of death at time t for the i^{th} individual is given by $h_i(t;x) = \exp(\beta x_i)h_0(t)$, where x_i is the value of X for the i^{th} individual. The hazard function for a *Weibull* regression model is $h(t;x) = \lambda\gamma t^{\gamma-1}\exp(\beta x_i)$, with scale $\lambda\exp(\beta x_i)$ and shape parameter γ. The survival function is therefore $S(t;x) = \exp\left(-\lambda t^{\gamma}\exp(\beta x_i)\right)$.

Mathematical details on the survival and hazard functions and on the parametric proportional hazards models can be found in the technical appendix - section A1.

2.2.2 Multilevel/Hierarchical regression analysis

2.2.2.1 Multilevel linear models

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 [32]. 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. The models discussed in the previous section do not explicitly account for the heterogeneity associated with cluster j, but can be extended to this setup [32].

When a hierarchical structure of data collection is found one can analyse data through a (i) complete-pooling model (single classical regression ignoring the cluster information); (ii) no-pooling model or 'fixed-effect' model (single classical regression that includes group indicators but not group-level predictors and where the number of *degrees of freedom* consumed may be high); (iii) separate models (separate classical regressions for each cluster, also constrained by the sample size n_i); two-step analysis (starts with no-pooling or separate models, then fits a classical group-level regression), or (iv) a multilevel model [32].

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 level and estimate. However, with this setup the researcher is limited in interpreting the conditional effects of level-2 factors have on level-1 factors, which is often the central interest. The problem with this model is that it fails to account for the fact that the overall variance is not only a function of variance among the level-1 units, but potentially also variation among the level-2 units. The model assumption of deterministic or non-stochastic variation is considered a failure with respect to explicitly accounting for this cross-level variability [38].

Random intercept models

Considerations of multiple levels of variation lead to models with *random-effects*. Let's consider the following simple random intercept model:

$$y_{ij} = \beta_{0j} + \varepsilon_i \,. \qquad \text{eq. 10}$$

In contrast to the standard regression analysis, the intercept is subscripted by cluster j, implying β_{0j} is variable across the j groups. The difference between this model and the previous 'fixed-effects'

model described in section 2.2.1 is the way β_{0j} is treated [38, 32]. In this approach β_{0j} is considered a random coefficient, following a pre-defined probability distribution, typically, the normal distribution:

$$\beta_{0j} \sim N\left(\mu_\beta, \sigma_\beta^2\right), \quad \text{for } j = 1,....,C\,. \qquad \text{eq. 11}$$

The mean effect is estimated, although it is assumed to have some random variability around it, attributable to unmeasured level-2 factors.

Treating β_{0j} as a function of a systematic and a random component, one has:

$$\beta_{0j} = \beta_0 + u_{0j}\,, \qquad \text{eq. 12}$$

where β_0 is the mean effect across sample and u_{0j} the residual of group *j* (i.e. of level-2 unit) from the mean. One can control for random variation in *y* due to level-2 factors. In the multilevel modelling framework, eq. 10 gives the level-1 model and eq. 12 gives the level-2 model. The multilevel model yields [38, 32]:

$$y_{ij} = \beta_0 + u_{0j} + \varepsilon_i\,, \qquad \text{eq. 13}$$

which has a straightforward interpretation: β_0 gives the mean; ε_i gives the level-1 errors; and u_{0j} the level-2 error. This additional error term separates this model from the standard regression model. Having now two sources of residual variation, a ratio of these two variances can be constructed. This ratio is commonly referred to as the *intraclass correlation coefficient* (ICC) [6]. ICC is a statistic that summarises the degree of dependency in nested observations. For the random intercept case:

$$ICC = \rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_\varepsilon^2}\,. \qquad \text{eq. 14}$$

The ICC can be interpreted as the proportion of the total variance that can be attributed to between-cluster variation (the level-2 units). Hence, the total variance in the model is

$$Var\left(y_{ij}\right) = \sigma_{u_0}^2 + \sigma_\varepsilon^2\,, \qquad \text{eq. 15}$$

and considering the typical distributional assumptions, one has

$$\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right) \qquad \text{and} \qquad u_{0j} \sim N\left(0, \sigma_{u_0}^2\right)$$

Random intercepts and random slopes

The next step in complexity of multilevel modelling is to allow the possible relationship between the response variable and covariates [32, 38]. Suppose there is one level-1 factor and one level-2 factor, the unconditional model would be given by:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_i \,, \qquad \text{eq. 16}$$

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

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

and has reduced-form $y_{ij} = \left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right) x_{ij} + \varepsilon_i \,,$ eq. 17

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 [32].

2.2.2.2 Multilevel generalized linear models: the multilevel logistic regression framework

Multilevel modelling is applied to logistic regression and other GLMs in the same way as with linear regression. Error terms are added to the model corresponding to different sources of variation in the data.

Random intercept models

Recalling that in the logistic regression model one has for a sample of independent observations y_i , for $\beta \equiv \{\beta_0,...,\beta_k\}$ and for the vector of explanatory variables X_i:

$$\log\left(\frac{p_i}{1-p_i}\right) = \text{logit}(p_i) = \beta X_i \,. \qquad \text{eq. 18}$$

If a multilevel model is fitted, it allows the prediction of y_i within each cluster, while also allowing for systematic differences between groups [32]:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \text{logit}(p_{ij}) = \beta_{0j} + \beta X_i, \quad \text{for } i = 1,...,n \text{ and } j = 1,...,C \qquad \text{eq. 19}$$

where *ij* indexes the cluster correspondent to observation *i*, X_i is a vector of explanatory variables.

Random intercepts and random slopes

In a multilevel context, a varying-intercept, varying slope *logit* model is given by:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \text{logit}(p_{ij}) = \beta_{0j} + \beta_{1j} X_{ij}, \quad \text{for } i = 1,...,n \text{ and } j = 1,...,C \qquad \text{eq. 20}$$

where, as in the previous expression, X_{ij} represents the independent variables array. The cluster-level intercepts and slopes that are themselves modelled given the average cluster u_j:

$$\beta_{0j} = \beta_0 + u_{0j} + \varepsilon_{\beta_0}, \quad \text{for } j = 1,...,C \quad \text{and}$$

$$\beta_{1j} = \beta_1 + u_{1j} + \varepsilon_{\beta_1}, \quad \text{for } j = 1,...,C \,,$$

with errors ε_{β_0} , ε_{β_1} having mean *0*, variances $\sigma^2_{\beta_0}$, $\sigma^2_{\beta_1}$, and correlation ρ, all estimated from data [32].

2.2.2.3 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 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 [39, 32].

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:

$$h_{ij}\left(t \mid x_{ij}\right) = h_0\left(t\right)\psi\left(x_{ij}\right) = h_0\left(t\right)\exp\left(\left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right)x_{ij}\right). \qquad \text{eq. 21}$$

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). Through this PHM framework the PHM based on the *Weibull* distribution is easily derived using the expressions obtained in section 2.2.1.3.2 [39].

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 [39].

Any continuous distribution with positive support, a unit mean, and finite variance can be used, like for instance, the *Gamma* distribution, the *Inverse Normal* or the *Log-Normal* distribution. The choice of the *Gamma* distribution is the standard for survival data due to mathematical tractability. This group of *mixture models* are called the *Gamma frailty models*. The frailty approach therefore produces a mixture model in that the conditional distribution can be described, for instance, by the *Weibull* distribution, while the mixture distribution is described by the *Gamma* distribution [38].

Frailty models can be considered as '*shared'* or '*unshared'* models. The distinction between the two types relies on the assumption of how the frailty is 'distributed' in the data. Shared or grouped frailty models assume that similar observations share the same frailty, even as that frailty may vary from centre-to-centre [38].

2.2.2.4 Multilevel linear mixed models in a longitudinal data framework

All the models described in section 2.2.1 are applied to cross-section data, where each individual is observed only once. 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 [33, 34, 35].

Lets consider the standard linear regression model presented in eq. 1 and assume one has repeated measurements ($t=1,....,T$) for a sample of n individuals ($i=1,....,n$),

$$y_{it} = \sum_k \beta_k x_{kit} + u_i + \varepsilon_{it}, \text{ for } k = 0,...,K . \qquad \text{eq. 22}$$

Here the dependent variable y is observed for individual, i, in each of the points in time, t. Similarly, the explanatory variables x are observed at each point in time. Some of these variables will be *time varying* (for example, an individual's utility at different points of time). Others may be *fixed* or *time invariant* (such as an individual's gender or treatment group in a RCT). The error term of the regression equation has been split into two components. The first, u_i is an individual-specific unobservable effect - the unobserved characteristics of the individual i that remains constant over time. The second term, ε_{it}, is a random error term representing idiosyncratic shocks that change across t as well as across i. It is assumed that u_i and ε_{it} are uncorrelated with each other. The presence of a common individual effect means that the values of the dependent variable for each individual will tend to cluster together.

One may be interested in modelling datasets where there is a multilevel structure and, therefore, to have several random effect levels [32]. Adopting the abovementioned longitudinal framework, including another level in the hierarchical structure, the individual becomes now level-2 in, for instance, a three-level linear model, whereas the individual was previously considered level one. 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, following the parametric format previously used, is:

$$y_{tij} = X_{tij}\beta + Z_{j,it}u_j + Z_{ji,t}u_{ji} + Z_{tij}u_{tij} + \varepsilon_{tij}, \quad \text{for } j = 1,...,C,\ i = 1,...,N_j \text{ and } t = 1,...,T_{ij} \qquad \text{eq. 23}$$

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.

3 Motivating Example: The RITA 3 trial

In this section, the use of multilevel 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 [23] 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 [41, 23].

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 or non-fatal 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 2. 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 cardiovascular death (CVD) or myocardial infarction (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 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. 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 2).

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 2). 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) [42, 43]. These probabilities are shown in appendix (appendix - Table A1).

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 2).

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.

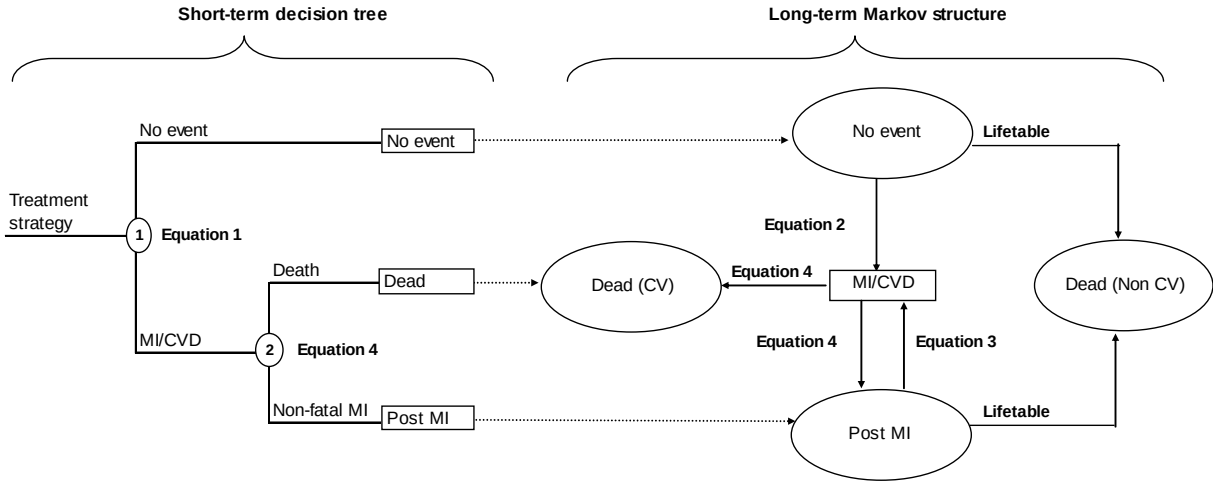

Figure 2. Model structure of the cost-effectiveness analysis of the RITA 3 trial (MI=myocardial infarction, CV=cardiovascular, CVD=cardiovascular death) [23].

Health-related quality of life

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 Multilevel analysis

As mentioned in section 2.2.2, multilevel or hierarchical models are 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 hierarchical regressions, accounting for within and between centres variability were estimated. These estimates of effectiveness provided sets of location-specific transitions probabilities 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.

Table 2 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.

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 2. Covariates included in the statistical models by centre.

3.2.1 Software

To implement the proposed analysis, the non-hierarchical regression models were performed in the freely available software package **R** version 2.7.1 (Copyright © 2008 The **R** Foundation for Statistical Computing) and in the also freely available software package **WinBugs/OpenBugs** version 3.0.3 (Copyright © 2008 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** and **CodaPkg**.

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

3.2.2 Results of effectiveness

Equation 1 *- Logistic regression model of risk of cardiovascular death or myocardial infarction during the index hospitalisation*

Equation 1.1 - logit model, probability of a composite event

In the original model the composite endpoint was regressed against the treatment binary covariate, the age categorical variable and the severe angina (grade 3 or 4) indicator covariate. The selection of variables was based on a backward stepwise selection procedure, forcing treatment into the model. The variable selection process considered is taken as controversial; however the debate on its adequacy is, for the moment, beyond the objectives of this work (see discussion section).

The logistic regression model applied was as followed:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{treat} treat + \beta_{age} age + \beta_{angina} angina + \varepsilon_i, \quad \text{for } i = 1,...,1809. \qquad \text{eq. 24}$$

The results of the non-hierarchical models (implemented in Stata, R and WinBugs) are described in columns 2 to 11 of Table 3. For the first two software's, the coefficient estimates, the standard errors estimates and the p-values are shown. For WinBugs, the mean estimates, standard deviation and 95% credibility intervals are shown. The hierarchical model was implemented in WinBugs using *Bayesian Markov chain Monte Carlo* (MCMC) methods with one chain and through a simulation process with 5,000 iterations and a 2,000 iteration burn-in period. A summary of the WinBugs outputs can be found in the appendix (appendix – Figures A1 and A2). The same framework was used in subsequent models.

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 (odds ratio of 1.520, *p-value* = 0.148).

To account for both within and between-centre variability, a hierarchical *logit* model was built incorporating a random intercept and a random slope for the treatment. This setup allows for both the intercept and the slope of the regression on treatment to vary randomly across centres:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0ij} + \beta_{ij} treat + \beta_{age} age + \beta_{angina} angina + \varepsilon_{ij}, \qquad \text{eq. 25}$$

$$\text{for } i = 1,...,1809 \text{ and } j = 1,...,46$$

with the centre-level components, intercepts and slopes, decomposed as follows:

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

Logistic regression

CCIndex	**Stata - NHM**			**R - NHM**			**WinBugs** - NHM**				**WinBugs** - HM**			
Covariate	coef.*	std. err.	Pr(>|z|)	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects														
Treat	0.417	0.288	0.148	0.417	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	0.549	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	0.636	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	-4.622	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 3. 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 12 to 15 in Table 3. 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 centre-specific random-effects components are shown in Table **4**. 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 4. 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 1.2 - logit model, probability of a composite event by continuous risk defined risk scores from RITA 3

In the original model the composite endpoint was regressed against the treatment binary covariate, the risk score and the interaction between risk at randomization and treatment effect term.

The logistic regression model applied was as followed:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{treat} treat + \beta_{risk} risk + \beta_{risk\times treat} risk \times treat + \varepsilon_i, \qquad \text{eq. 26}$$

$$\text{for } i = 1,...,1807.$$

Again, the results of the non-hierarchical statistical modelling (Table 5) are similar across software's. The inclusion of an interaction between baseline risk and treatment effect shows that a

higher risk is associated with a decreasing odds ratio of a composite endpoint during the index hospitalisation.

As before, a hierarchical model was built incorporating a random intercept and a random slope for the treatment effect:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right) treat + \beta_{risk} risk + \beta_{risk \times treat} risk \times treat + \varepsilon_{ij}, \quad \text{eq. 27}$$

$$\text{for } i = 1,...,1807 \text{ and } j = 1,...,46.$$

Logistic regression

CCIndex	**Stata - NHM**			**R - NHM**			**WinBugs** - NHM**				**WinBugs** - HM**			
Covariate	coef.*	std. err.	Pr(>|z|)	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects														
Treat	0.567	0.539	0.293	0.567	0.539	0.293	0.581	0.545	-0.497	1.653	0.553	0.546	-0.516	1.657
Risk score	3.638	1.245	0.003	3.638	1.245	0.003	3.565	1.248	0.917	5.895	3.674	1.258	1.235	6.193
Treat **x** Risk score	-0.424	1.740	0.807	-0.424	1.740	0.807	-0.410	1.752	-3.791	3.190	-0.400	1.751	-3.844	2.979
Constant	-4.593	0.408	0.000	-4.593	0.408	0.000	-4.624	0.410	-5.484	-3.847	-4.785	0.454	-5.778	-3.969
Random Effects														
σ_{Treat}	-	-	-	-	-	-	-	-	-	-	0.166	0.238	0.009	0.890
σ_{Cnst}	-	-	-	-	-	-	-	-	-	-	0.427	0.347	0.016	1.184
ρ_{Treat_Cnst}	-	-	-	-	-	-	-	-	-	-		0.09986		

*Values in log odds ratios

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

Table 5. Log-odds ratio of composite endpoint of MI or CVD during the index hospitalisation including an interaction between risk at randomization and treatment effect (NHM – non-hierarchical model; HM – hierarchical model).

As before, the *Bayesian* hierarchical model identifies a decrease on the intercept and treatment estimates (fixed effects). The correlation estimated between the random intercept and random slope components is higher compared to the one from Equation 1.1. The centre-specific random-effects components estimates for the 5 centres under analysis can be found in the appendix section (appendix – Table A2).

Equation 2 *- Weibull proportional hazards model of risk of cardiovascular death or myocardial infarction during the remainder of trial*

Equation 2.1 - Weibull model, composite endpoint index admission to end of follow-up

In the original model the *Weibull* PHM model applied was as follows:

$$\log\left(HR_i\right) = \beta_0 + \beta_{age}age + \beta_{diab}diab + \beta_{prevMI}prevMI + \beta_{smok}smok + \beta_{pulse}pulse + \beta_{STdep}STdep + \beta_{angi}angi + \beta_{male}male + \beta_{leftB}leftB + \beta_{treat}treat + \varepsilon_i, \text{ for } i = 1,...,1756. \quad \text{eq. 28}$$

The model was obtained through a backward stepwise covariate selection procedure. The fact that the shape parameter, γ, in the *Weibull* statistical model is less than *1* (approximately *0.58*) indicates that the rate of the composite endpoint of CVD or MI declines as time elapses from hospital discharge. The results of the non-hierarchical and hierarchical *Weibull* models are shown in Table 6.

A proportional hazards *Weibull* hierarchical model was built incorporating a random intercept and a random slope for the treatment:

$$\log\left(HR_{ij}\right) = \left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right)treat + \beta_{age}age + \beta_{diab}diab + \beta_{prevMI}prevMI + \beta_{smoke}smoke + \beta_{pulse}pulse + \beta_{STdep}STdep + \beta_{angi}angi + \beta_{male}male + \beta_{leftB}leftB + \varepsilon_{ij}, \text{ for } i = 1,...,1756 \text{ and } j = 1,...,46. \quad \text{eq. 29}$$

The non-hierarchical models show that all risk factors, except presence of severe angina, were significant at the 5% level. However, this risk factor was very close to significance and was kept in the parametric model. The early interventional strategy was associated with a statistically significant lower rate of CVD or MI after the index hospitalisation (hazard ratio of approximately 0.620, *p-value* ≈ 0.001).

The results of the *Bayesian* hierarchical model were similar to the non-hierarchical one. 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

	Stata - NHM			R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.*	std. err.	Pr(>|z|)	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects														
Age	0.575	0.087	0.000	0.575	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	0.645	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	0.386	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	0.501	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	0.024	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	0.357	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	0.280	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	0.316	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	0.682	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	0.148	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	-4.790	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	0.040	-	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 6. Log-hazard ratio of composite endpoint of CVD or MI from hospital discharge until end of trial (NHM – non-hierarchical model; HM – hierarchical model).

Equation 2.2- Weibull model evaluating a composite event

The non-hierarchical *Weibull* PHM model applied was as follows:

$$\log\left(HR_i\right) = \beta_0 + \beta_{treat} treat + \beta_{risk} risk + \beta_{risk \times treat} risk \times treat + \varepsilon_i, \quad \text{for } i = 1,...,1755. \qquad \text{eq. 30}$$

The results of the statistical models including an interaction between baseline risk and treatment effect are shown in Table 7. Although not statistically significant, the interaction model shows that the positive treatment effect is more pronounced in patients with higher baseline risk. The hazard ratio of a first composite endpoint in the remainder of the trial is close to 1 when the risk score is tending towards 0 and approximately 0.21 when the risk score tends towards 1.

A *Bayesian* proportional hazards *Weibull* hierarchical model was built incorporating a random intercept and a random slope for the treatment:

$$\log\left(HR_{ij}\right) = \left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right) treat + \beta_{risk} risk + \beta_{risk \times treat} risk \times treat + \varepsilon_{ij}, \quad \text{eq. 31}$$
$$\text{for } i = 1,...,1755 \text{ and } j = 1,...,46.$$

Parametric proportional hazards model - Weibull regression

	Stata - NHM			R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.*	std. err.	Pr(>|z|)	coef.*	std. err.	Pr(>|z|)	mean*	std. dev.	95% CrI		mean*	std. dev.	95% CrI	
Fixed Effects														
Treat	-0.035	0.279	0.900	-0.035	0.279	0.900	-0.011	0.265	-0.519	0.472	0.025	0.282	-0.497	0.571
Risk score	4.925	0.475	0.000	4.925	0.475	0.000	4.944	0.451	4.107	5.905	5.024	0.482	4.068	5.987
Treat **x** Risk score	-1.518	0.878	0.084	-1.518	0.878	0.084	-1.596	0.841	-3.257	0.113	-1.749	0.876	-3.511	-0.182
Constant	-3.986	0.183	0.000	-3.986	0.183	0.000	-4.018	0.178	-4.421	-3.717	-4.051	0.178	-4.412	-3.688
Shape parameter	0.580	0.040	-	0.580	0.070	-	0.590	0.039	0.511	0.671	0.592	0.034	0.519	0.659
Random Effects														
σ_{Treat}	-	-	-	-	-	-	-	-	-	-	0.153	0.163	0.010	0.562
σ_{Cnst}	-	-	-	-	-	-	-	-	-	-	0.056	0.051	0.008	0.206
ρ_{Treat_Cnst}	-	-	-	-	-	-	-	-	-	-		0.20103		

*Values in log hazard ratios

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

Table 7. Log-hazard ratio of composite endpoint of CVD or MI from hospital discharge to end of trial including an interaction between risk at randomization and treatment effect (NHM – non-hierarchical model; HM – hierarchical model).

The *Bayesian* hierarchical model shows that, despite non-significant, the treatment effect estimate now has a positive impact on composite endpoint from hospital discharge until end of trial. Although also not statistically significant in the non-hierarchical model, the interaction term estimate in the HM is considered statistically significant now at 5% significance level, showing a more prominent positive treatment effect in patients with higher baseline risk. A high correlation is found between the random components introduced.

Equation 3 *- Weibull proportional hazards model of risk of a second composite endpoint of cardiovascular death or myocardial infarction*

Equation 2.1 was used to estimate the risk of a second composite endpoint by updating the covariate for prior myocardial infarction (Table 6). The risk of a second composite endpoint of CVD or MI was estimated to be about 50% higher than the risk of a first composite endpoint. Using the results from the *Weibull* models estimated in equation 2.1, imposed a logical time dependency, as patients were getting further away from their MI in the model. As mentioned in the background section, this was achieved by employing tunnel states for the first *5* years after a non-fatal MI. The assumption that the hazard in year 5 and after was constant was employed, adjusted for age as patients get older in the model.

Equation 4 *- Logistic regression model of the proportion of composite endpoints being non-fatal*

In the original model the composite endpoint being non-fatal was regressed against the index hospitalization binary covariate, the age categorical variable and the previous MI indicator covariate. The selection of variables was based on a backward stepwise selection procedure.

Therefore, the logistic regression model applied was as follows:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{index} index + \beta_{age} age + \beta_{prevMI}\, prevMI + \varepsilon_i, \quad \text{for } i = 1,...,275. \qquad \text{eq. 32}$$

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 8).

To account for not only within but also between-centre variability, a hierarchical model in WinBugs was built incorporating a random intercept and a random slope for the treatment:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \left(\beta_0 + u_{0j}\right) + \beta_{index} index + \beta_{age} age + \beta_{prevMI}\, prevMI + \varepsilon_{ij}, \quad \text{for } i = 1,...,275 \text{ and } j = 1,...,46. \qquad \text{eq. 33}$$

Logistic regression

Non-fatal MI	**Stata - NHM**			**R - NHM**			**WinBugs** - NHM**				**WinBugs** - HM**			
Covariate	coef.*	std. err.	Pr(>|z|)	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	1.162	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	-0.347	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	-0.595	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	0.235	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 8. 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 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 linear regression model applied was as follows:

$$\begin{aligned} \cos t\_index_i = \beta_0 + \beta_{treat} treat + \beta_{MIindex} MIindex + \beta_{dead\_index} dead\_index + \\ \beta_{male} male + \beta_{age} age + \beta_{STdep} STdep + \varepsilon_i, \quad \text{for } i = 1,...,1808. \end{aligned} \qquad \text{eq. 34}$$

For the non-hierarchical and hierarchical models built in WinBugs, a transformation of the dependent variable had to be made as the estimation procedure doesn't accept large values.

The non-hierarchical models demonstrate similar results, showing that during the index hospitalisation, the early interventional strategy was associated with a higher mean cost (mean of approximately £5,650, 95% CrI £5,145 - £6,159) compared with a conservative strategy (Table 9). 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 £6,225 and £7,940, respectively, which included the costs for the administration of thrombolytic drugs, revascularisations and longer hospital stay in wards and intensive care.

Linear model

Costs index	Stata - NHM			R - NHM			WinBugs** - NHM				WinBugs** - HM			
Covariate	coef.	std. err.	Pr(>|z|)	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	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	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	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	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	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	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	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 9. Estimated costs during the index hospitalisation (NHM – non-hierarchical model; HM – hierarchical model).

A linear hierarchical model was built incorporating a random intercept and a random slope for the treatment:

$$\begin{aligned} cost\_index_{ij} = &\left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right) treat + \beta_{MIindex} MIindex + \\ &\beta_{dead\_index} dead\_index + \beta_{male} male + \beta_{age} age + \\ &\beta_{STdep} STdep + \varepsilon_{ij}, \quad \text{for } i = 1,...,1808 \text{ and } j = 1,...,46. \end{aligned} \qquad \text{eq. 35}$$

As in the NHMs, 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 A7).

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

In the original model the cost dependent variable was regressed against a set of covariates, also as a result of a backward stepwise covariate selection procedure.

The linear regression model applied was as follows:

$$cost\_follow_i = \beta_0 + \beta_{treat} treat + \beta_{MIyear1} MIyear1 + \beta_{male} male + \beta_{angina} angina + \beta_{prevMI} prevMI + \varepsilon_i, \text{ for } i = 1,...,1808. \quad \text{eq. 36}$$

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 (mean of approximately -£1,110, 95% CrI -£1,570 to -£660) compared with the conservative strategy (Table 10). 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 (mean of approximately £5,450, 95% CrI £3,880 - £7,020) or prior to the trial (mean of approximately £720, 95% CrI £210 - £1,240).

A linear hierarchical model was built incorporating a random intercept and a random slope for the treatment:

$$cost\_follow_{ij} = \left(\beta_0 + u_{0j}\right) + \left(\beta_1 + u_{1j}\right) treat + \beta_{MIyear1} MIyear1 + \beta_{male} male + \beta_{angina} angina + \beta_{prevMI} prevMI + \varepsilon_{ij}, \text{ for } i = 1,...,1808 \text{ and } j = 1,...,46. \quad \text{eq. 37}$$

Compared to the non-hierarchical models, the results for the *Bayesian* hierarchical model show that fixed-effects estimates are very similar (Table 10).

Linear model

Costs follow-up exc.MI/stroke	**Stata - NHM**			**R - NHM**			**WinBugs** - NHM**				**WinBugs** - HM**			
Covariate	coef.	std. err.	Pr(>|z|)	coef.	std. err.	Pr(>|z|)	mean	std. dev.	95% CrI		mean	std. dev.	95% CrI	
Fixed Effects														
MI year 1	5466.8	804.0	0.000	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	-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	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	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	724.4	262.4	0.006	717.1	263.7	211.5	1238.4	694.1	263.6	169.9	1207.7
Constant	2734.8	247.6	0.000	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 10. Estimated costs during the follow-up period (NHM – non-hierarchical model; HM – hierarchical model).

3.2.4 Health related quality of life

For the following regressions on HRQoL a note should be made about the non-inclusion of the baseline utility covariate in the original models. One should account for the imbalances in baseline utility in the estimation of mean differential HRQoL. Failure to control for this imbalance can result in misleading CE estimates. Therefore, by not including this covariate in the regression model, the researcher will be faced with an omitted variable problem. However, as stated above, the idea was to replicate the original set of regressions hierarchically without performing judgements about its adequacy [48].

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.

Baseline utility was regressed against a set of covariates:

$$baseline\_u_i = \beta_0 + \beta_{diab} diab + \beta_{prevMI} prevMI + \beta_{STdep} STdep + \beta_{angina} angina + \beta_{male} male + \varepsilon_i, \text{ for } i = 1,...,1800. \quad \text{eq. 38}$$

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 11). Similar results were obtained for the different software's.

Linear model

HRQoL baseline	**Stata - NHM**			**R - NHM**			**WinBugs** - NHM**				**WinBugs** - HM**			
Covariate	coef.	std. err.	Pr(>|z|)	coef.	std. err.	Pr(>|z|)	mean	std. dev.	95% CrI		mean	std. dev.	95% CrI	
Fixed Effects														
Diabetes	-0.050	0.021	0.016	-0.050	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	0.006	-0.045	0.016	-0.078	-0.014	-0.021	0.015	-0.050	0.009
ST depression	-0.066	0.015	0.000	-0.066	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	-0.074	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	0.000	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	0.000	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 11. Estimated baseline utilities (NHM – non-hierarchical model; HM – hierarchical model).

To account for not only within but also between-centre variability, a *Bayesian* hierarchical model was built incorporating a random intercept component:

$$\begin{aligned} baseline\_u_{ij} = &\left(\beta_0 + u_{0j}\right) + \beta_{diab} diab + \beta_{prevMI} prevMI + \\ &\beta_{STdep} STdep + \beta_{angina} angina + \beta_{male} male + \varepsilon_{ij}, \\ &\text{for } i = 1,...,1800 \text{ and } j = 1,...,46. \end{aligned} \qquad \text{eq. 39}$$

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 (Table 11).

HRQoL regression 2 - Estimated gain in health-related quality of life

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 model implemented was:

$$\begin{aligned} \Delta HRQoL_{it} = &\beta_{0it} + \beta_{D4\times treat} D4 \times treat + \beta_{D12} D12 + \beta_{D12\times treat} D12 \times treat + \\ &\beta_{priorMI} priorMI + \beta_{currentMI} currentMI + \varepsilon_{it}, \\ &\text{for } i = 1,...,1734 \text{ and } t \in \{4,12,24,48,60\}. \end{aligned} \qquad \text{eq. 40}$$

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 12 shows the results for the multilevel model considering the individual as a cluster, named non-centre hierarchical model (NCHM), and also, in columns 12 to 15, the hierarchical model with time clustered in patients and patients nested in health care centres (centre hierarchical model (CHM)). Table 12 supplies the within-patient standard error and the between-patient standard error.

Longitudinal data

Change HRQoL	Stata - NCHM			R - NCHM			WinBugs** - NCHM				WinBugs** - CHM			
Covariate	coef.	std. err.	Pr(>|z|)	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	0.039	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	0.038	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	0.015	0.235	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	0.510	-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	0.110	-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	0.000	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	-	-	0.033	-	-	0.174	0.002	0.169	0.179	0.174	0.002	0.169	1.78
$\sigma_{Cnst\_patient}$	0.087	-	-	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 12. Estimated gain in HRQoL (NCHM – non-centre hierarchical model; CHM – centre hierarchical model).

The Stata and 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 built was as follows:

$$\begin{aligned}\Delta HRQoL_{ijt} = &\left(\beta_0 + \gamma_{0i} + u_{0j}\right) + u_{1j} treat + \beta_{D4\times treat} D4 \times treat + \\ &\beta_{D12} D12 + \beta_{D12\times treat} D12 \times treat + \beta_{priorMI}\, priorMI + \\ &\beta_{currentMI} currentMI + \varepsilon_{ijt}, \\ &\text{for } i = 1,...,1734,\ j = 1,...,46 \text{ and } t \in \{4,12,24,48,60\}.\end{aligned} \qquad \text{eq. 41}$$

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 convergency 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 incremental cost-effectiveness ratio, 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.

Trial-wide cost-effectiveness results

Considering only the baseline characteristics of risk group 1, the mean incremental cost per QALY was approximately £41,000 (base-case analysis results). Figure 3 illustrates the obtained joint CE density plotted on the cost-effectiveness plane. It can be observed that the majority of the simulation results are located in the *NE* quadrant of the CEP, denoting that the intervention strategy is more effective than the comparator, but also more costly.

Although the CEP does not allow an easy quantitative interpretation, if a threshold of £20,000 to £30,000 (offered by the NICE guidelines [5]) were to be considered, one would conclude that the new technology would not be regarded as cost-effective.

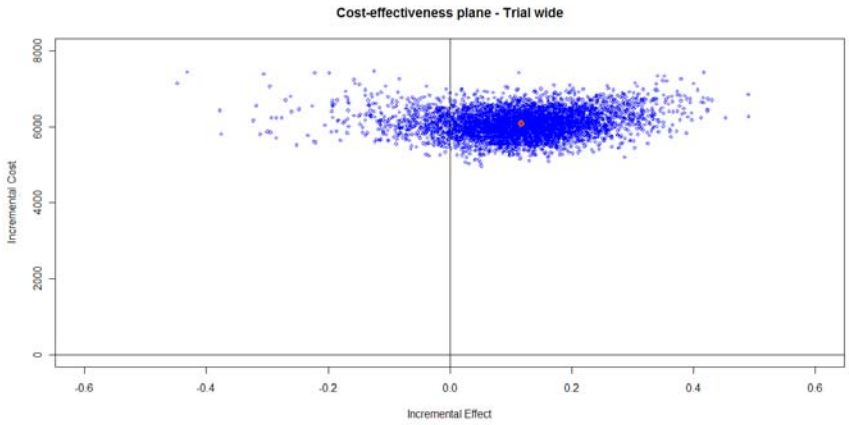

Figure 3. Cost-effectiveness plane of the RITA 3 model risk group 1 with trial wide results.

Considerations of the uncertainty surrounding a decision to reject the new technology can be based on the CEAC. Figure 4 depicts the CEAC for the current case study. It shows the typical "ogive" shape, characteristic of that observed when the joint density of mean differential costs and mean differential effects is contained mainly in the *NE* quadrant [20]. Interpretation is straightforward since the probability that the intervention strategy is cost-effective (*p*) and the associated error probability (*1-p*) can be read off the *y-axis* for any particular threshold. For common threshold values within the range of £20,000 to £30,000, the probability of the intervention being cost-effective is 0.09 to 0.31.

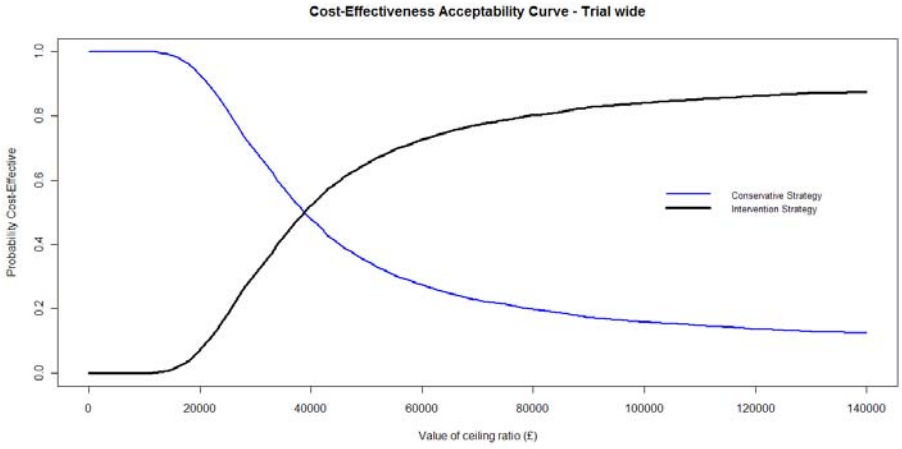

Figure 4. Cost-effectiveness acceptability curve of the RITA 3 model risk group 1 with trial wide results.

Location-specific cost-effectiveness results

Figure 5 illustrates the joint CE density plotted on the 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 13 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.

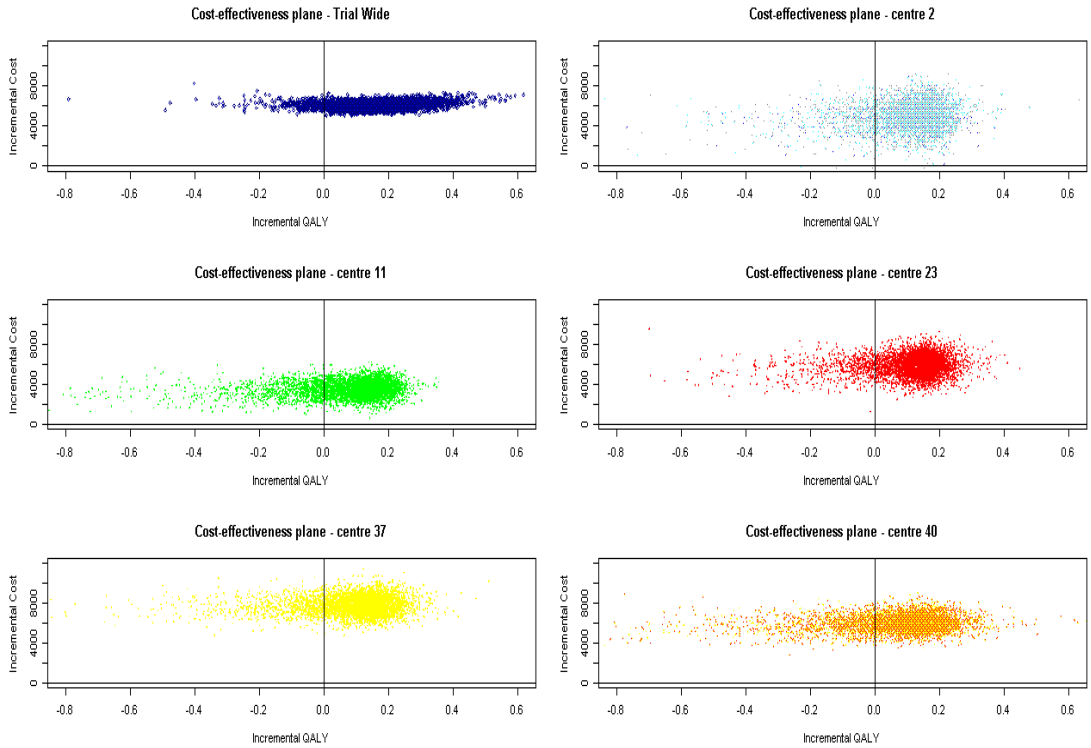

Figure 5. Cost-effectiveness planes of the RITA 3 model with trial wide results and centre-specific results for centres 2, 11, 23 37 and 40, respectively.

	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 13. 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 for these 5 centres (Figure 6). Once again, the curves display great variability across centres (hospitals) in cost-effectiveness 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.

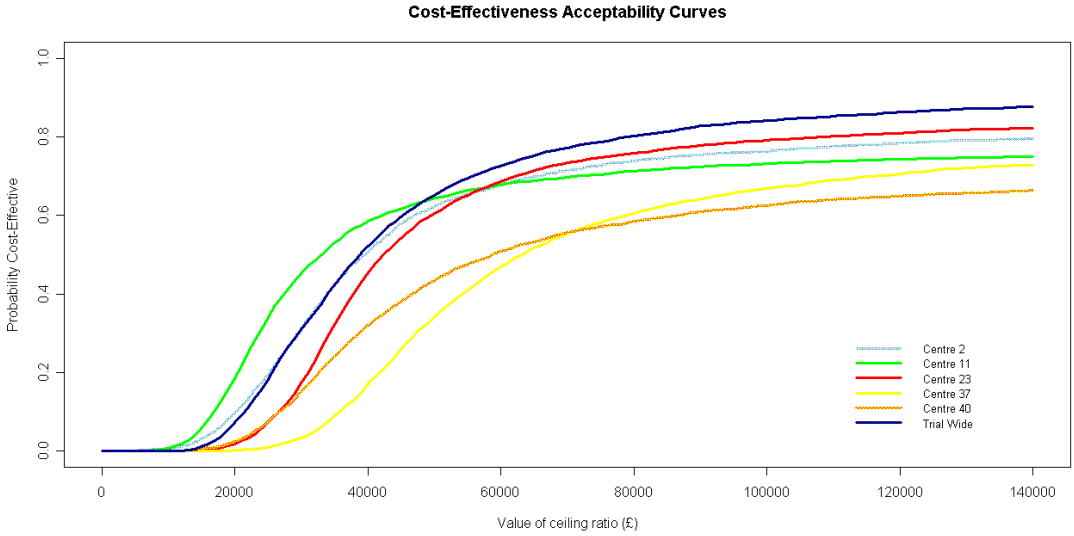

Figure 6. Cost-effectiveness acceptability curve for the trial wide results and centre-specific results for centres 2, 11, 23 37 and 40, respectively.

4 Discussion

This thesis has demonstrated the use 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 cost-effectiveness estimates. 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 three 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. The stepwise variable selection method may not be the most appropriate for the following reasons [44, 45, 46]: (i) the method yields confidence intervals for effects and predicted values that are falsely narrow; (ii) it gives biased regression coefficients that need shrinkage (the coefficients for remaining variables are too large; (iii) it has severe problems in the presence of collinearity; (iv) the number of candidate predictor variables affects the number of noise variables that gain entry to the model; and (v) the size of the sample has little practical importance in determining the number of authentic variables contained in the final model. As mentioned in Judd *et al* [47], 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 ("*S*" shaped logistic function), cost data are constrained to be non negative and are usually highly skewed. Therefore one should employ the *Log-Normal* or the *Gamma* distribution 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 [48].

5 Bibliography

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

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

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

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

[5] 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*;

[6] 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;

[7] 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*;

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

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

[10] Drummond, M., O'Brien, B., Stoddard, G., Torrance G. (2001). Methods for the economic evaluation of health care programs. *Oxford University Press*, First edition;

[11] 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;

[12] 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);

[13] Claxton, K. (1999). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, (18), 342-364;

[14] 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;

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

[16] 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;

[17] 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;

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

[19] 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;

[20] Fenwick, E., O'Brien, B., Briggs, A. (2004). Cost-effectiveness acceptability curves – facts, fallacies and frequently asked questions. *Health Economics*, (13), 405-415;

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

[22] Briggs, A., Mihaylova, B., Sculpher, M., Hall, A., Wolstenholme, J., Simoons, M., Deckers, J., Ferrari, R., Remme, W., Bertrand, M., Fox, K. on behalf of the EUROPA Trial Investigators (2007). Cost effectiveness of perindopril in reducing cardiovascular events in patients with

stable coronary artery disease using data from the EUROPA study. *BMJ Heart*, (93), 1081-1086;

[23] 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;

[24] Briggs, A., Bousquet, J., Wallace, M., Busse, W., Clark, T., Pedersen, S., Bateman, E. on behalf of the GOAL Investigators Group. (2006). Cost-effectiveness of asthma control: an economic appraisal of the GOAL study. *Allergy*, (61), 531-536;

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

[26] Kijima, M. (1997). Markov processes for stochastic modelling. Chapman & Hall;

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

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

[29] 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;

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

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

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

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

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

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

[36] Cameron, A., Trivedi, P. (2005). Microeconometrics – methods and applications. *Cambridge Universtity Press*;

[37] Collet, D. (2003). Modelling survival data in medical research. *Chapman & Hall / CRC*;

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

[39] 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;

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

[41] 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;

[42] 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);

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

[44] Altman, D. and Andersen, P. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, (8), 771–783;

[45] Derksen, S. and Keselman, H. (1992). Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*. (45), 265–282;

[46] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, B Series, (58), 267–288;

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

[48] 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.

Appendix A – Technical appendix

A1 Survival Analysis

A1.1 Survival and hazard functions

In summarizing survival data, there are two functions of vital importance, namely the *survivor function* and the *hazard function*. The actual survival time of an individual can be regarded as the value of a random variable, *T* or survival time, which can take any non-negative value. *T* has a probability distribution with underlying p.d.f. *f(t)*. The distribution function of *T* is given by,

$$F(t) = P(T < t) = \int_0^t f(u)\,du, \qquad \text{eq.A-1}$$

and represents the probability that the survival times is less than some value *t*.

The survivor function, *S(t)*, is defined to be the probability that the survival time is greater than or equal to *t*:

$$S(t) = P(T \geq t) = 1 - F(t). \qquad \text{eq.A-2}$$

The survivor function can be used to represent the probability that an individual survives from the origin to some time beyond *t*.

The hazard function is widely used to express the risk or hazard of death at some time *t*, and it is obtained from the probability that an individual dies at time *t*, conditional on he or she having survived to that time. Considering the conditional probability that the random variable associated with an individual's survival time *T*, lies between *t* and $t+\delta t$, the hazard function is defined as:

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t} \right\}. \qquad \text{eq.A-3}$$

From the previous equation, $h(t)\cdot\delta t$ is the approximate probability that an individual dies in the interval *(t, t+δt)*, conditional on that person having survived to time *t*. From this definition, one can derive some useful relationships between the survivor function and the hazard function:

$$h(t)=f(t)\cdot\frac{1}{S(t)} \qquad \text{given the equality } \frac{P(t\le T<t+\delta t)}{P(T\ge t)}\equiv\frac{F(t+\delta t)-F(t)}{S(t)}$$

and also the derivative of $F(t)$, $f(t)=\lim_{\delta t\to 0}\left\{\frac{F(t+\delta t)-F(t)}{\delta t}\right\}$.

Consequently,

$$h(t)=-\frac{d}{dt}\{\log S(t)\} \quad \text{and therefore} \quad S(t)=\exp\{-H(t)\},$$

where $H(t)=\int_0^t h(u)\,du$ is the *cumulative hazard function*.

A1.2 Parametric proportional hazards model

The PHM applicability is widespread in the analysis of survival data, despite having relatively few probability distributions for the survival times that can be used (*Weibull* and *Gompertz* distributions are used the most).

Let's define a vector $x\equiv(x_1,x_2,...,x_j)'$ of explanatory variables. If $h_0(t)$ is the hazard function for an individual for whom the values of all the explanatory variables that make up the vector *X* are zero, the function $h_0(t)$ is called the *baseline hazard function*. The hazard function for the i^{th} individual can be written as

$$h_i(t)=\psi(x_i)h_0(t), \qquad \text{eq.A-4}$$

where $\psi(x_i)$ is a function of the values of the vector of explanatory variables for the i^{th} individual. There are several possible choices for $\psi(x_i)$, but the choice $\psi(x_i)=\exp\left(\sum_k \beta_k x_{ki}\right)$ is the most commonly used in survival data. Consequently, the general proportional hazard becomes

$$h_i(t)=\exp\left(\sum_k \beta_k x_{ki}\right)h_0(t), \text{ for } k=0,...,K\,, \qquad \text{eq.A-5}$$

or, if regarded as a linear model for the logarithm of the hazard ratio $\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \sum_k \beta_k x_{ki}$.

Notice the absence of the constant term in the linear component of the PHM. In this framework the survival function becomes

$$S(t) = \left[S_0(t)\right]^{\psi(x_i)} \qquad \text{where} \quad S_0(t) = \exp\left(-\int_0^t h_0(u)\,du\right). \qquad \text{eq.A-6}$$

A2 Estimation procedures

A2.1 Ordinary least squares

Classical regression methods usually employ the *ordinary least squares* (OLS) estimation procedure. This common estimation methodology is briefly described below.

If the number of data observations *n* exceeds the number of predictors, *k*, it is not commonly possible to find an estimated vector β_k that gives a perfect fit. The usual estimation goal is to choose the estimate $\hat{\beta}_k$ that minimizes the *sum of squares of the residuals*. The vector $\hat{\beta}$ that minimizes it is called the *least squares estimate* and is usually represented by (in matrix notation) $\hat{\beta} = (X'X)^{-1} X'Y$.

The β_0 and β_k sampling properties of the OLS estimators are referred as *Best Linear Unbiased Estimators* (BLUE), even when errors ε_i are not normally distributed.

The errors come from a distribution with mean 0 and variance σ^2, which can be estimated from the residuals.

A2.2 Maximum likelihood

GLMs usually employ *maximum likelihood estimation* (*MLE*) procedure. This common estimation methodology is briefly described below.

MLE is extensively used in health economics, predominantly in nonlinear models involving qualitative or limited dependent variables. MLE advantageous properties, such as consistency and asymptotic normality, rely on the model being completely and accurately specified. MLE family includes the quasi-maximum likelihood (QML) methods, which share the properties of MLE without having to uphold the statement that the model is correctly specified.

Let's cover a simple example of MLE and afterwards a more complex one with the normality assumption.

Consider an i.i.d. sample of *Bernoulli* trials, each with outcomes 0 or 1 with probabilities *1-β* and *β*, respectively [34]. The sample log-likelihood function, for n_0 zeros and n_1 ones, is

$$\ell(\beta \mid x) = \log\left(L(\beta \mid x)\right) = n_0 \log(1-\beta) + n_1 \log(\beta). \qquad \text{eq.A-7}$$

The MLE of β, $\hat{\beta}$, is the value that maximizes $\ell(\beta \mid x)$, that is $\frac{\partial \ell(\beta \mid x)}{\partial \beta} = 0$, which when solved is the sample proportion: $\hat{\beta} = \frac{n_1}{n_0 + n_1}$.

If the error terms are normally distributed, so that $y_i \sim N\left(X_i\beta, \sigma^2\right)$ for each *i*. The least squares estimate vector $\hat{\beta}$ is the *maximum likelihood estimate*. The *likelihood* of a regression model is defined as the probability of the data given the parameters and inputs. Therefore,

$$L(y \mid \beta, \sigma, X) = p(y \mid \beta, \sigma, X) = \prod_i^n N\left(y_i \mid X_i\beta, \sigma^2\right), \qquad \text{eq.A-8}$$

where $N(.|.,.)$ represents the normal *probability density function* (p.d.f.).

The *log-likelihood* is derived as,

$$\ell = \log L(y \mid \beta, \sigma, X) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \sum_i^n \frac{\left[y_i - X_i\beta\right]^2}{2\sigma^2}. \qquad \text{eq.A-9}$$

The MLE of α and β_j must minimize $\sum_i^n \left[y_i - X_i\beta\right]^2$ and so equal OLS. The MLE of σ^2 is $\hat{\sigma}^2 = n^{-1}\sum_i \hat{\varepsilon}_i^2 = \textit{method of moments}$ estimate.

A2.3 The *Bayesian* approach

For some models, such as the linear mixed-effects model, the integral involved with the likelihood function has a closed form. The ordinary iterative algorithms for maximizing the likelihood are used to obtain MLE or restricted MLE for unknown model parameters. However, for most non-linear models such as the logistic and PHM, the likelihood function does not have

an analytically tractable form. Hence, likelihood inference requires either an analytical approximation or a numerical evaluation [32].

Multilevel inferences can be formulated in a *non-Bayesian framework*, however all multilevel models are *Bayesian* in the sense of assigning probability distributions to the varying regression coefficients. *Bayesian* methods can be considered as an alternative to the classical approach to statistical inference [32]. The *Bayesian* approach is appealing since it is a flexible modelling framework, allowing the researcher to venture beyond frequentist models, and analyses provided in standard statistical packages, and additionally account fully for all forms of model estimation uncertainty [40]. A key difference between the two approaches is that *Bayesian* methods allow external information to be incorporated beyond that included directly in the model.

The challenge in fitting a multilevel model is in estimating a data-level regression (including the coefficients for all the cluster indicators) along with the cluster-level model. *Bayesian inference* is understood as the most direct way of obtaining this [18, 17]. The distinction between *Bayesian* and *non-Bayesian* multilevel models arises only for estimating the non-varying coefficients and the variance parameters.

Bayesian inference refers to statistical procedures that model unknown parameters (and also missing and latent data) as random variables. *Bayesian* inference is understood as a generalization of the OLS and MLE. *Bayesian* inference starts with a prior distribution on the unknown parameters and updates this with the likelihood of the data, yielding a posterior distribution which is used for inferences and predictions [32]. In a continuous framework one has:

$$p(\theta \mid x) \propto L(\theta \mid x) \cdot p(\theta), \qquad \text{eq.A-10}$$

where θ is a parameter or a parameter array with prior distribution $p(\theta)$, and x a random variable with *p.d.f.* $f(x|\theta)$, belonging to the space-parameter Θ.

The usual Markov chain Monte Carlo (MCMC) method used is *Gibbs sampling*. *Gibbs sampling* is an iterative *Monte Carlo* method for generating samples indirectly from a difficult joint distribution of the model parameters without calculating the density. The mechanism is based only on elementary properties of *Markov chains* (described in section 2.1) [18]. 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 [17].

In a *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 [32]. The simplest form of *Bayesian* inference uses a *Uniform* prior distribution, so that the posterior distribution is the same as the likelihood function.

Bayesian shrinkage estimation

As mentioned earlier, 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 [18, 17]. 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 [6, 17].

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 [18].

Let $\hat{\theta}_j$ be the observed between-treatment difference for centre j ($j=1,\ldots,C$). Assuming that $\hat{\theta}_j \sim N\left(\theta_j, \sigma_j^2\right)$ and that $\theta_j \sim N\left(\theta, \sigma^2\right)$, the empirical *Bayes* shrinkage estimator is a weighted average of $\hat{\theta}_j$ and $\hat{\theta}$. The weights used are the proportion of total variance due to between-centre and within-centre, respectively.

$$\tilde{\theta}_j = \frac{\sigma^2}{\sigma^2 + \sigma_j^2} \cdot \hat{\theta}_c + \frac{\sigma_j^2}{\sigma^2 + \sigma_j^2} \cdot \hat{\theta}, \qquad \text{eq.A-11}$$

where σ_j^2 is the within-centre variance of $\hat{\theta}_j$ and σ^2 is the between-centre variance of the mean differences [17].

Appendix B

Age	Men	Women	Age	Men	Women
45	0.0017	0.0013	**73**	0.0246	0.0167
46	0.0019	0.0016	**74**	0.0277	0.0187
47	0.0022	0.0017	**75**	0.0296	0.0194
48	0.0023	0.0018	**76**	0.0326	0.0216
49	0.0027	0.0020	**77**	0.0360	0.0239
50	0.0027	0.0022	**78**	0.0396	0.0263
51	0.0029	0.0024	**79**	0.0436	0.0290
52	0.0032	0.0026	**80**	0.0462	0.0303
53	0.0034	0.0028	**81**	0.0500	0.0334
54	0.0037	0.0032	**82**	0.0545	0.0375
55	0.0041	0.0033	**83**	0.0607	0.0418
56	0.0047	0.0036	**84**	0.0684	0.0479
57	0.0052	0.0041	**85**	0.0764	0.0523
58	0.0057	0.0043	**86**	0.0830	0.0576
59	0.0064	0.0048	**87**	0.0895	0.0641
60	0.0071	0.0052	**88**	0.0993	0.0717
61	0.0077	0.0057	**89**	0.1083	0.0798
62	0.0085	0.0061	**90**	0.1187	0.0910
63	0.0093	0.0067	**91**	0.1263	0.1010
64	0.0100	0.0075	**92**	0.1406	0.1114
65	0.0120	0.0075	**93**	0.1522	0.1232
66	0.0121	0.0084	**94**	0.1641	0.1323
67	0.0135	0.0092	**95**	0.1948	0.1601
68	0.0148	0.0103	**96**	0.2068	0.1727
69	0.0167	0.0114	**97**	0.2278	0.1840
70	0.0177	0.0118	**98**	0.2386	0.1996
71	0.0199	0.0133	**99**	0.2488	0.2128
72	0.0222	0.0149	**100**	0.2727	0.2311

Table A1. UK population age-and-sex specific life-tables, adjusted to exclude cardiovascular mortality.

Equation 1 *- Logistic regression model of risk of cardiovascular death or myocardial infarction during the index hospitalisation*

Equation 1.1 - logit model, probability of a composite event

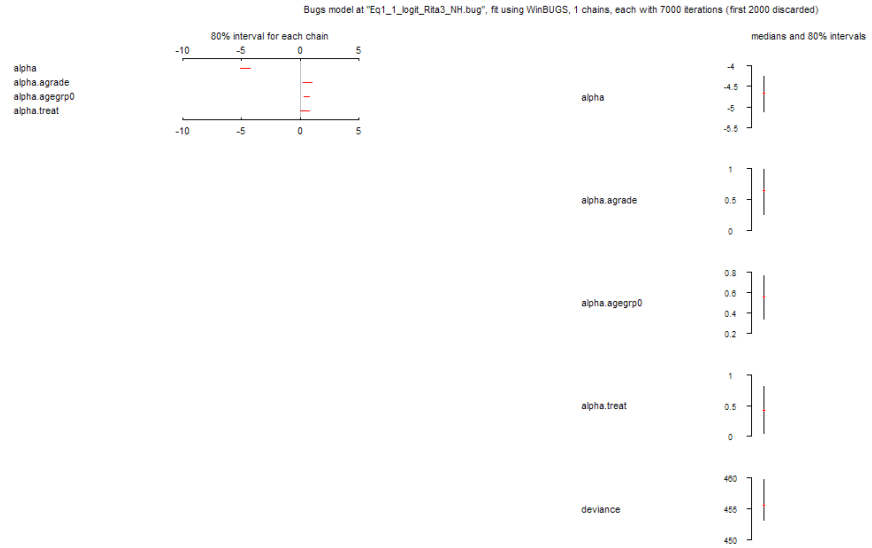

Figure A1. Equation 1.1, non-hierarchical model, WinBugs output.

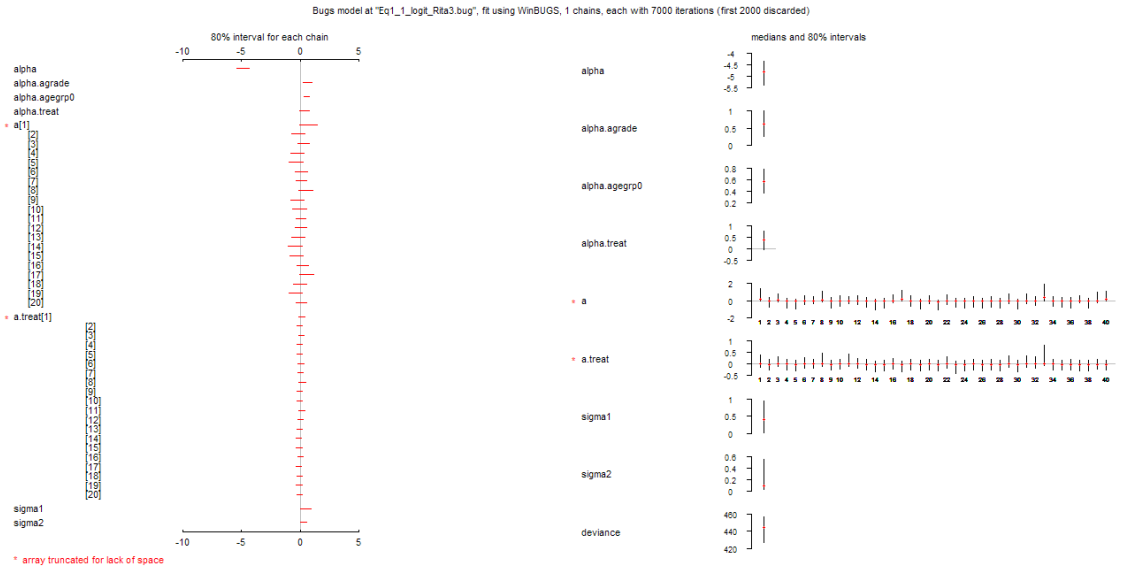

Figure A2. Equation 1.1, hierarchical model, WinBugs output.

Equation 1.2 - logit model, probability of a composite event by continuous risk defined risk scores from RITA 3

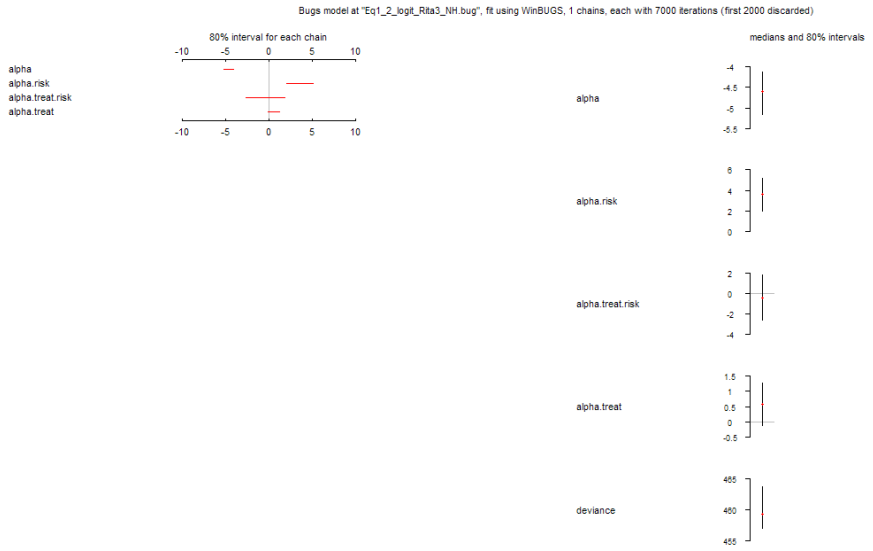

Figure A3. Equation 1.2, non-hierarchical model, WinBugs output.

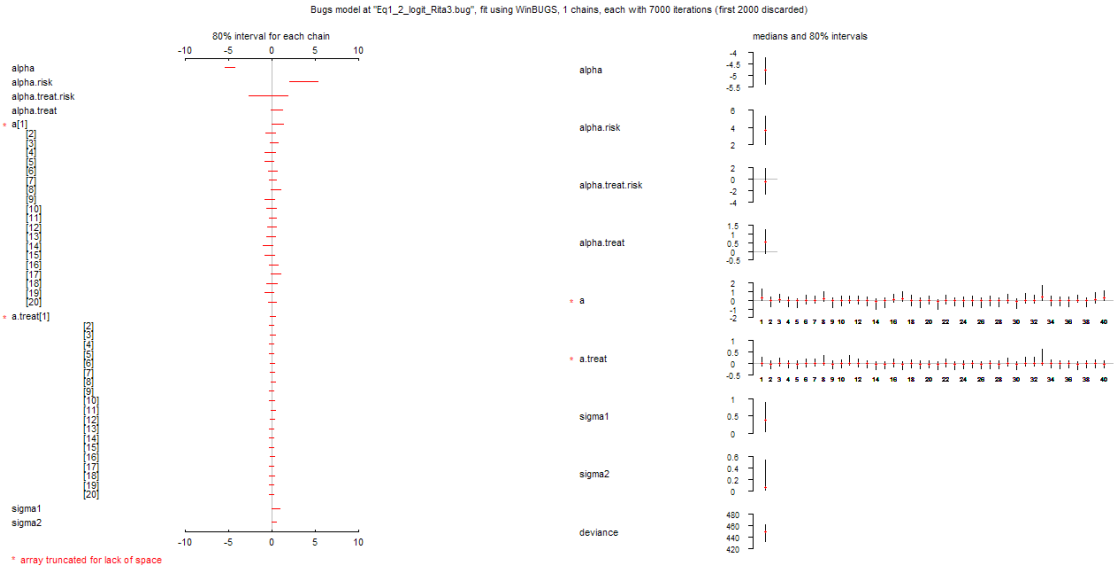

Figure A4. Equation 1.2, hierarchical model, WinBugs output.

Logistic regression

CCIndex		WinBugs** - HM			
Centre		mean	std. dev.	95% CrI	
Random Effects					
centre 2	$u_{1j\text{ - Treat}}$	-0.021	0.275	-0.736	0.524
	$u_{0j\text{ - Cnst}}$	-0.097	0.513	-1.338	0.858
centre 11	$u_{1j\text{ - Treat}}$	0.081	0.276	-0.239	0.926
	$u_{0j\text{ - Cnst}}$	0.046	0.351	-0.648	0.865
centre 23	$u_{1j\text{ - Treat}}$	-0.049	0.282	-0.836	0.367
	$u_{0j\text{ - Cnst}}$	-0.114	0.423	-1.142	0.672
centre 37	$u_{1j\text{ - Treat}}$	-0.031	0.243	-0.664	0.422
	$u_{0j\text{ - Cnst}}$	0.119	0.384	-0.567	1.033
centre 40	$u_{1j\text{ - Treat}}$	-0.013	0.233	-0.591	0.503
	$u_{0j\text{ - Cnst}}$	0.398	0.487	-0.181	1.557

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

Table A2. Random effects components of 5 centres, results of *Bayesian* hierarchical logistic regression of composite endpoint of CVD or MI during index hospitalisation including an interaction between risk at randomization and treatment effect (HM – hierarchical model).

Equation 2 *- Weibull proportional hazards model of risk of cardiovascular death or myocardial infarction during the remainder of trial*

Equation 2.1 - Weibull model, composite endpoint index admission to end of follow-up

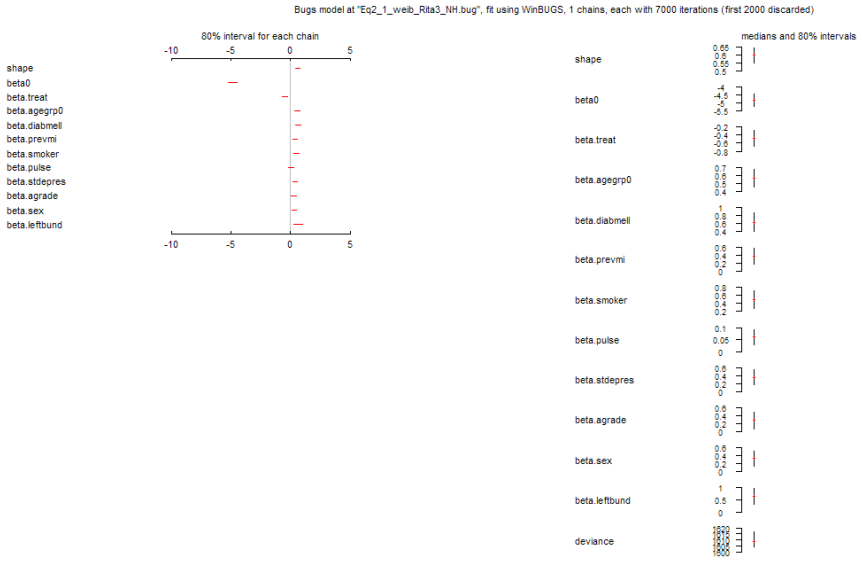

Figure A5. Equation 2.1, non-hierarchical model, WinBugs output.

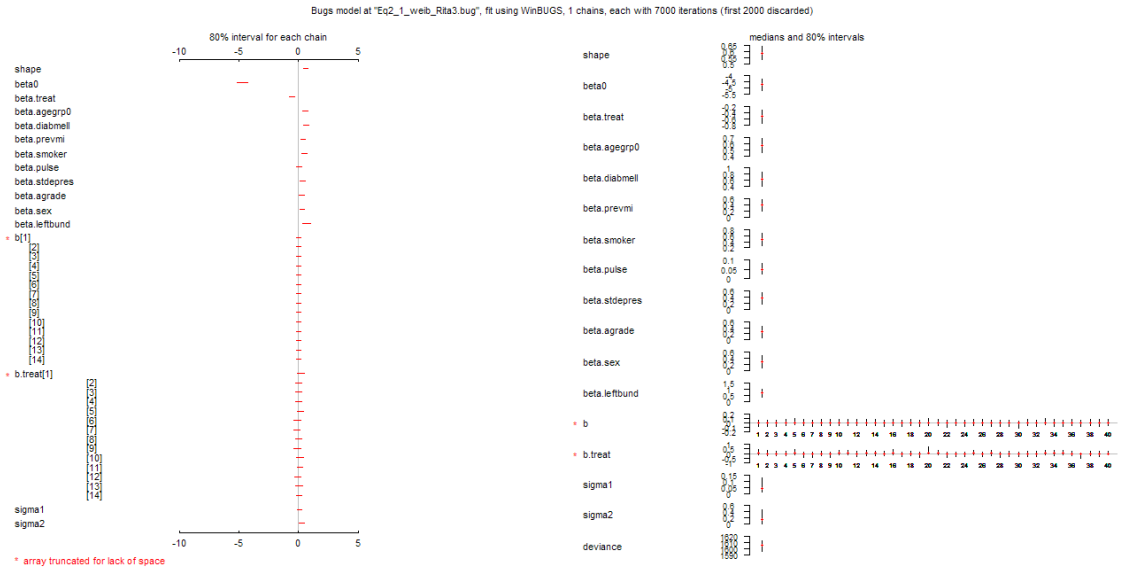

Figure A6. Equation 2.1, hierarchical model, WinBugs output.

Weibull regression

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

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

Table A3. 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).

Equation 2.2- Weibull model evaluating a composite event

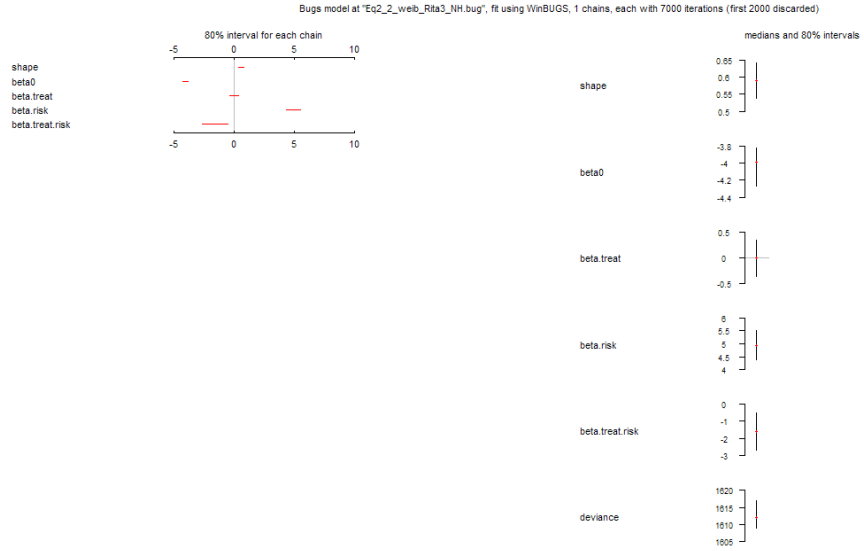

Figure A7. Equation 2.2, non-hierarchical model, WinBugs output.

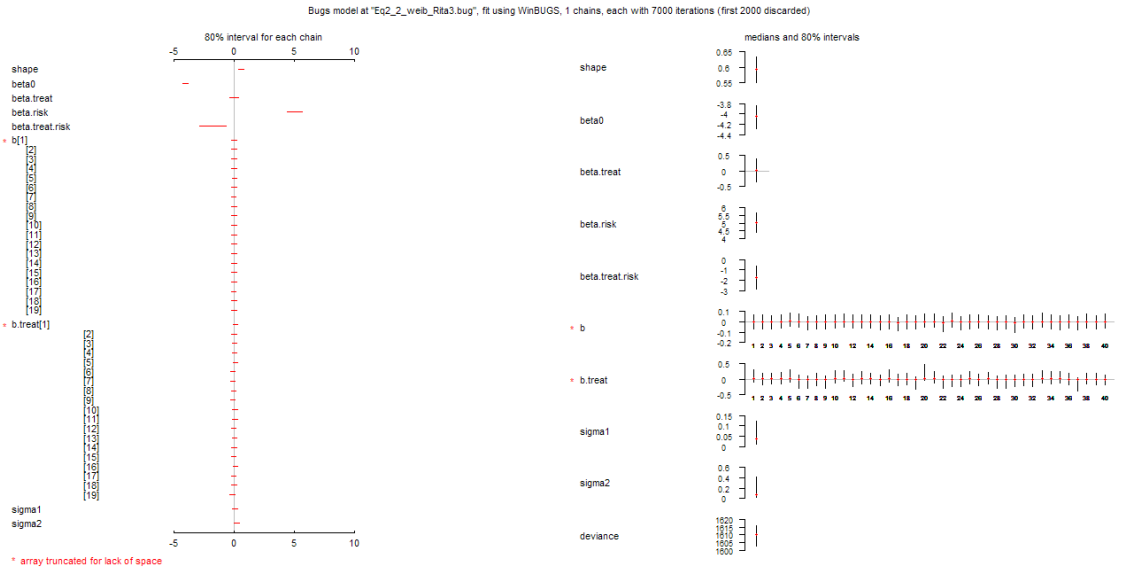

Figure A8. Equation 2.2, hierarchical model, WinBugs output.

Weibull regression

		WinBugs** - HM			
Centre		mean	std. dev.	95% CrI	
	Random Effects				
centre 2	$u_{1j\text{ - Treat}}$	0.008	0.212	-0.471	0.524
	$u_{0j\text{ - Cnst}}$	-0.003	0.070	-0.167	0.141
centre 11	$u_{1j\text{ - Treat}}$	0.050	0.168	-0.222	0.520
	$u_{0j\text{ - Cnst}}$	0.008	0.064	-0.122	0.171
centre 23	$u_{1j\text{ - Treat}}$	-0.028	0.181	-0.495	0.335
	$u_{0j\text{ - Cnst}}$	0.013	0.073	-0.115	0.204
centre 37	$u_{1j\text{ - Treat}}$	-0.097	0.221	-0.747	0.160
	$u_{0j\text{ - Cnst}}$	-0.008	0.074	-0.176	0.134
centre 40	$u_{1j\text{ - Treat}}$	-0.016	0.176	-0.462	0.353
	$u_{0j\text{ - Cnst}}$	0.005	0.069	-0.132	0.172

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

Table A4. Random effects components of 5 centres, results of *Bayesian* hierarchical Weibull PHM of composite endpoint of CVD or MI from hospital discharge to end of trial including an interaction between risk at randomization and treatment effect (HM – hierarchical model).

Equation 4 *- Logistic regression model of the proportion of composite endpoints being non-fatal*

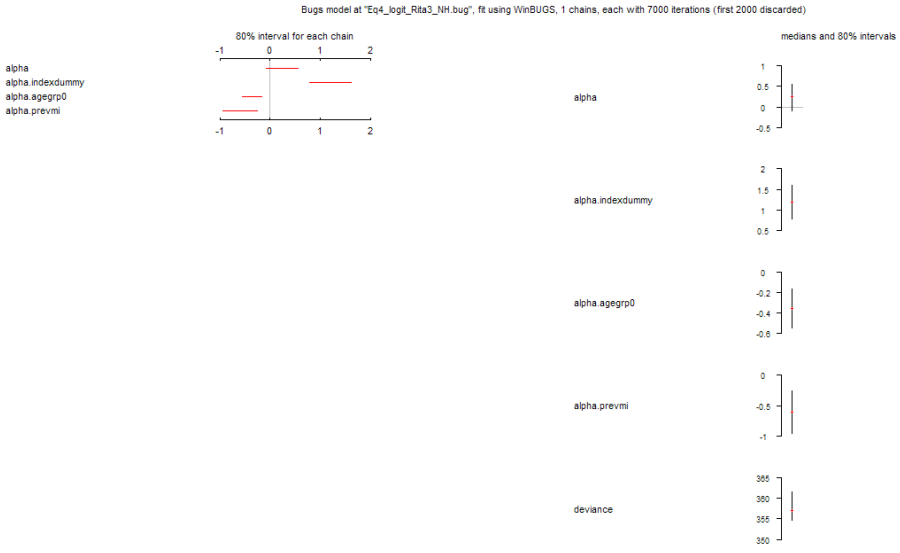

Figure A9. Equation 4, non-hierarchical model, WinBugs output.

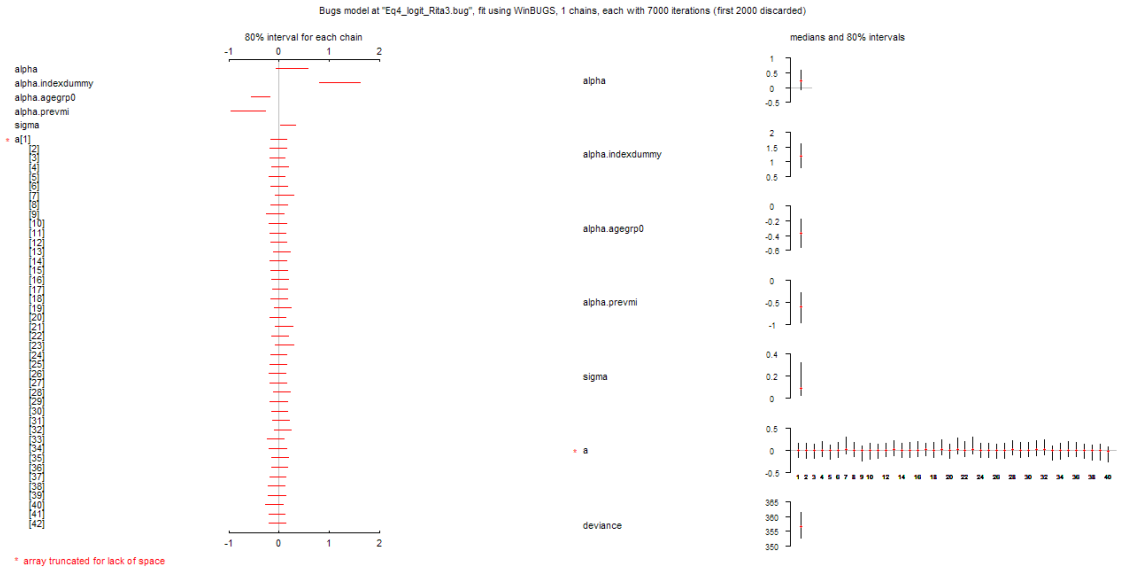

Figure A10. Equation 4, hierarchical model, WinBugs output.

Logistic regression

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

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

Table A5. 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).

Costs

Cost regression 1 - Estimated costs during the index hospitalisation

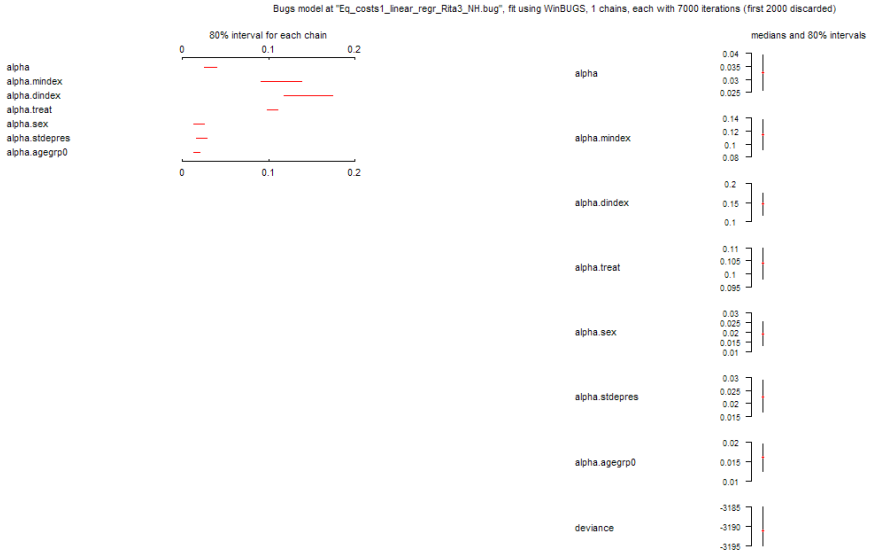

Figure A11. Costs regression 1, non-hierarchical model, WinBugs output.

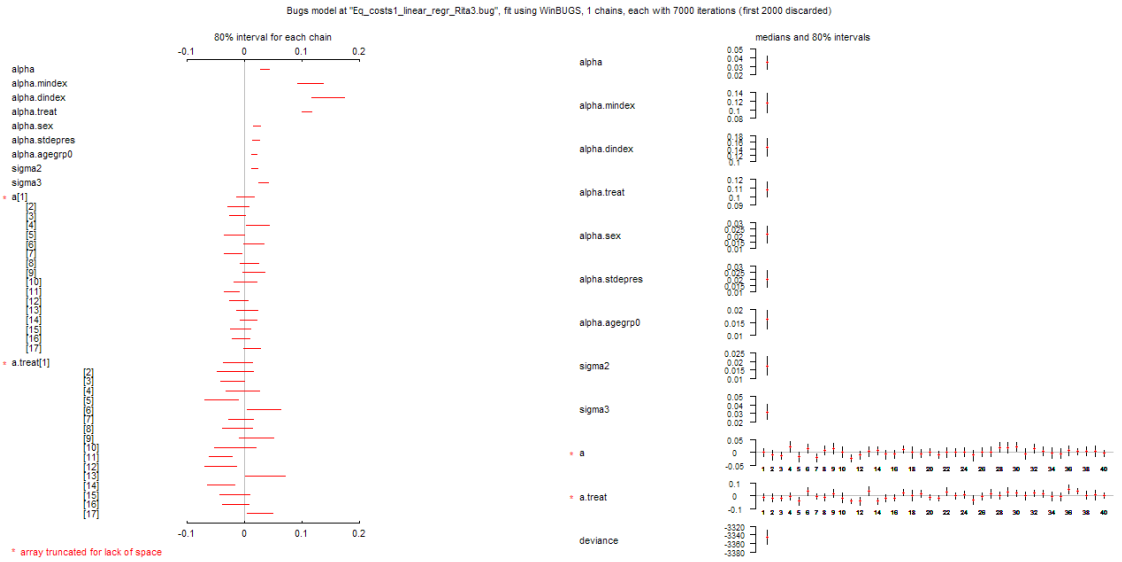

Figure A12. Costs regression 1, hierarchical model, WinBugs output.

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 A6. Random effects components of 5 centres, results of *Bayesian* hierarchical linear regression of costs during the index hospitalisation (HM – hierarchical model).

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

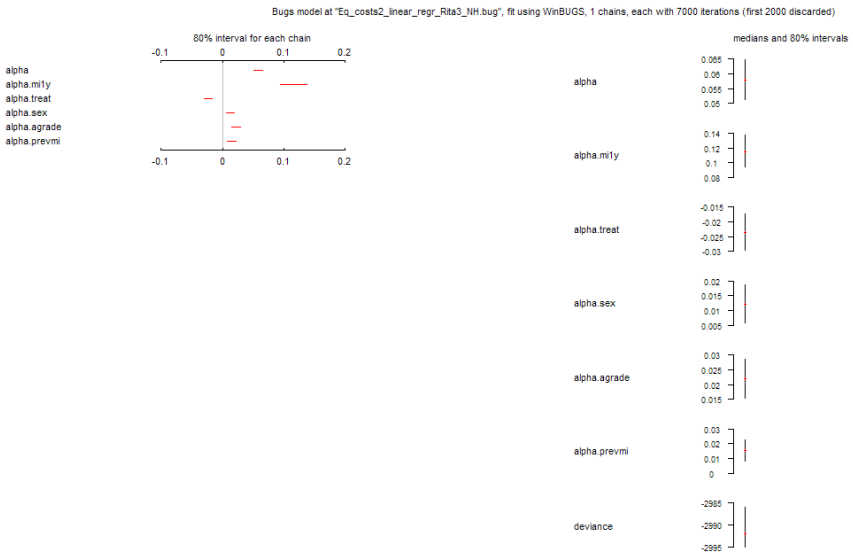

Figure A13. Costs regression 2, non-hierarchical model, WinBugs output.

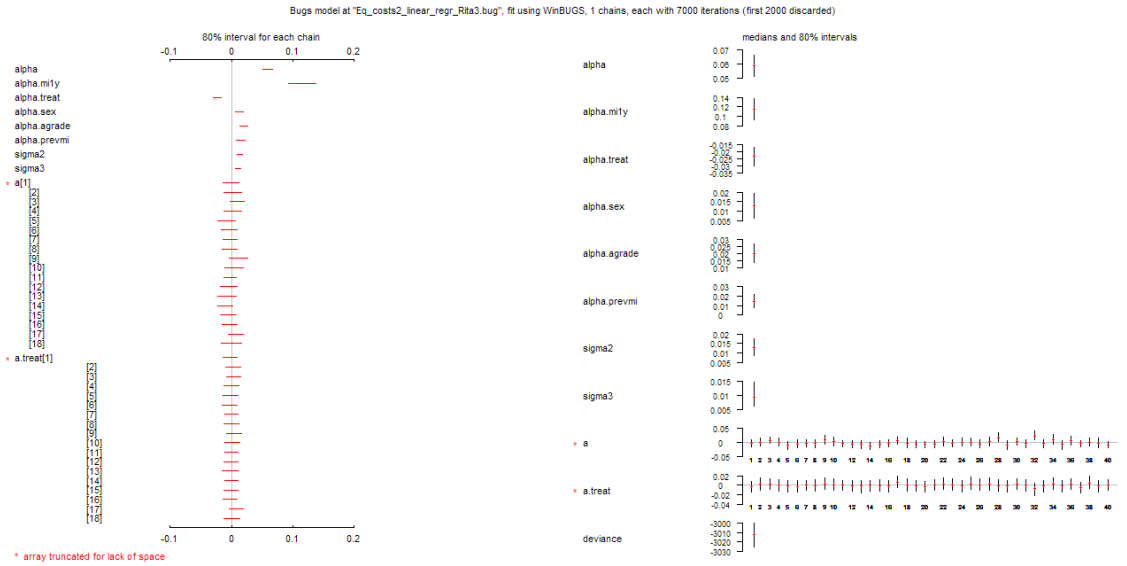

Figure A14. Costs regression 2, hierarchical model, WinBugs output.

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

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

Table A7. Random effects components of 5 centres, results of *Bayesian* hierarchical linear regression of costs during the follow-up period (HM – hierarchical model).

HRQoL regression 1 - Estimated baseline utilities

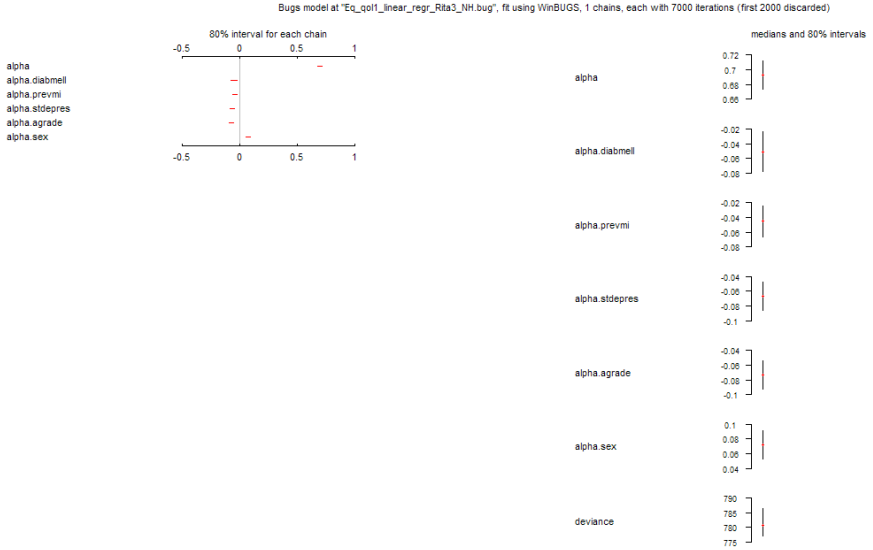

Figure A15. HRQoL regression 1, non-hierarchical model, WinBugs output.

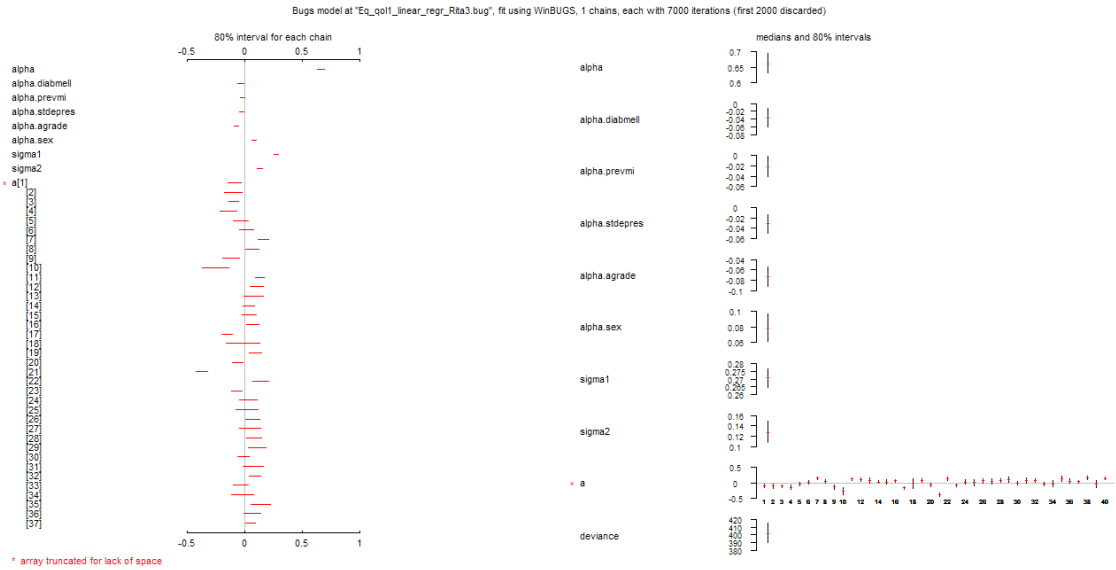

Figure A16. HRQoL regression 1, hierarchical model, WinBugs output.

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 A8. Random effects components of 5 centres, results of *Bayesian* hierarchical linear regression of baseline utilities (HM – hierarchical model).

HRQoL regression 2 - Estimated gain in health-related quality of life

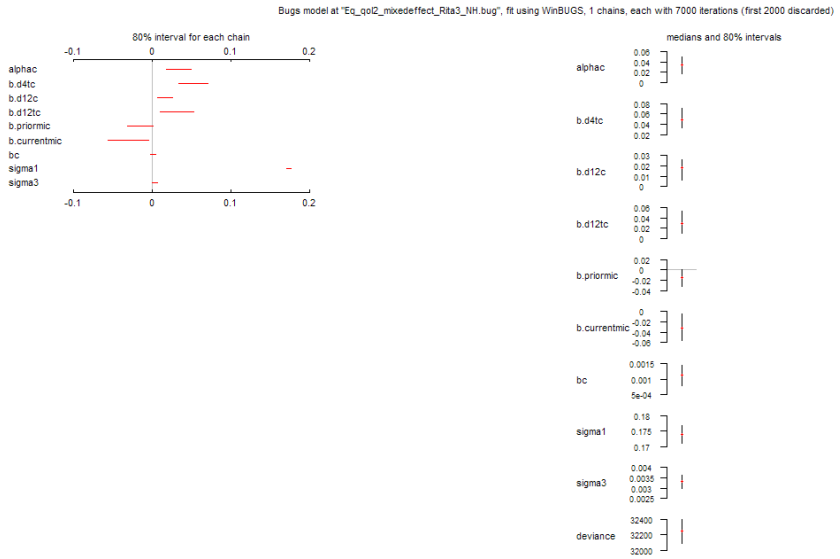

Figure A17. HRQoL regression 2, non-hierarchical model, WinBugs output.

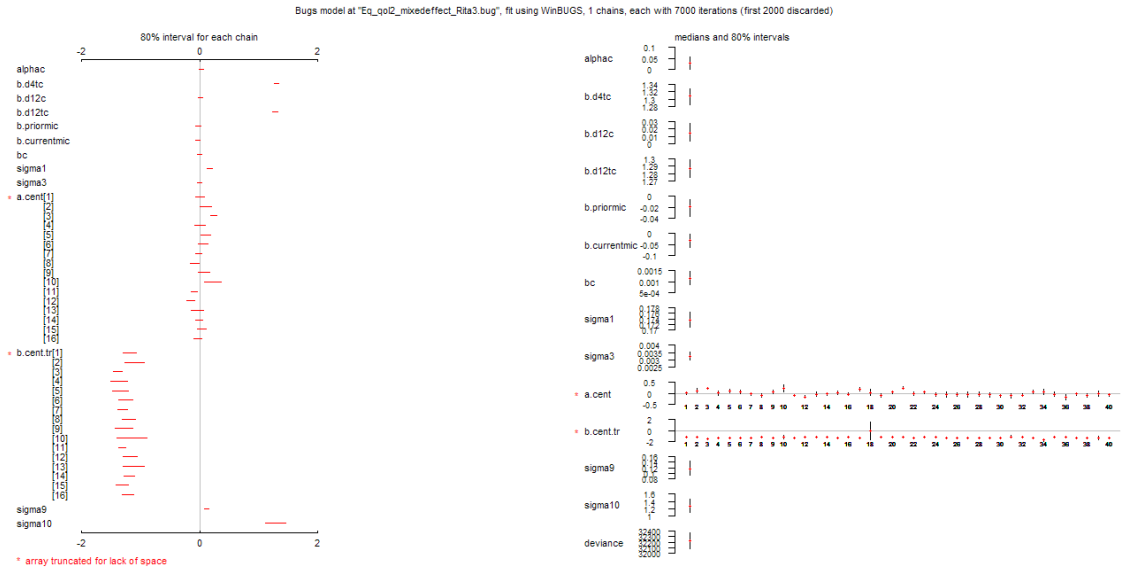

Figure A18. HRQoL regression 2, hierarchical model, WinBugs output.

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 A9. Random effects components of 5 centres, results of *Bayesian* hierarchical panel data regression of the gain in HRQoL (HM – hierarchical model).

Location-specific cost-effectiveness results

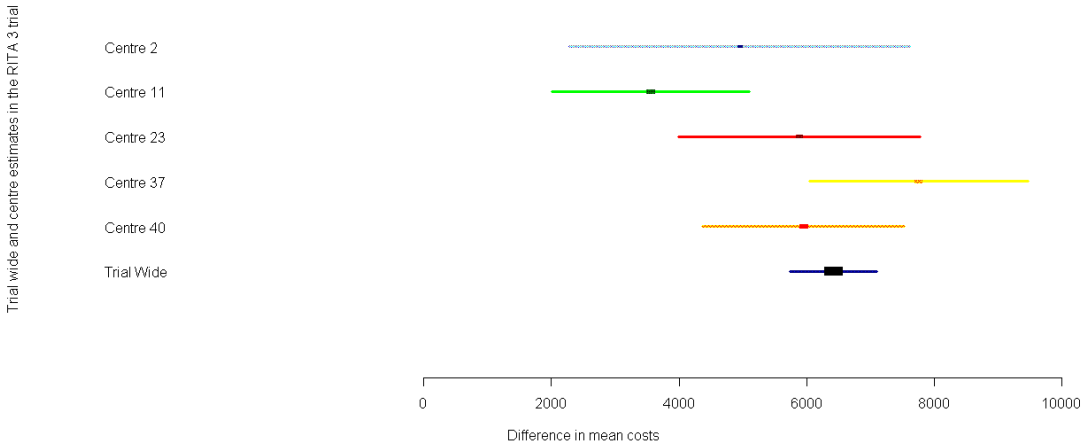

Figure A19. Trial wide and centre-specific estimated cost differences (centres 2, 11, 23, 37 and 40, respectively). Markers indicate trial wide and centre-specific mean differential cost estimates, and horizontal bars across the markers represent 95% credibility intervals.

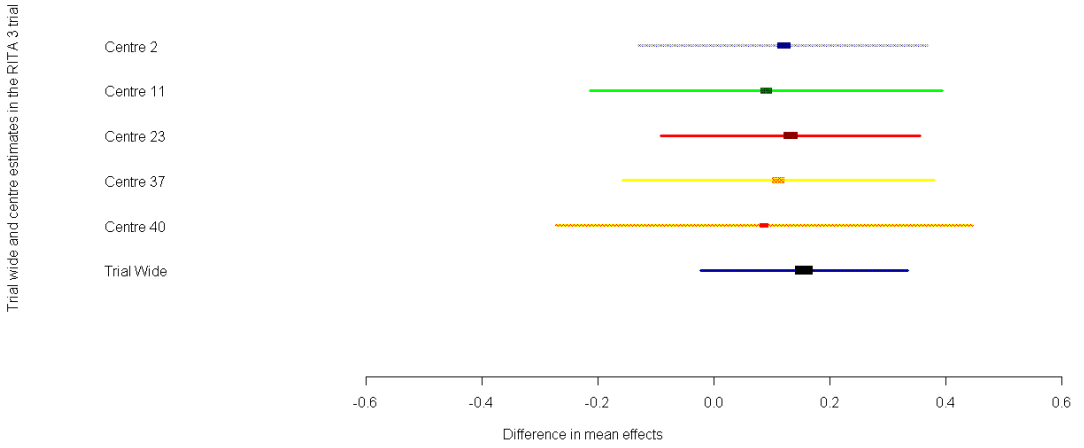

Figure A20. Trial wide and centre-specific estimated effect differences (centres 2, 11, 23, 37 and 40, respectively). Markers indicate trial wide and centre-specific mean differential effect estimates, and horizontal bars across the markers represent 95% credibility intervals.