

TEMPORAL UNCERTAINTY IN COST-EFFECTIVENESS DECISION MODELS:

METHODS TO ADDRESS THE UNCERTAINTIES THAT ARISE WHEN THE
APPROPRIATE ANALYSIS TIME HORIZON EXCEEDS THE EVIDENCE TIME
HORIZON IN COST-EFFECTIVENESS DECISION MODELS AS APPLIED TO
HEALTHCARE INTERVENTIONS

Ronan Mahon

Ph.D.

University of York
Economics and Related Studies

September 2014

ABSTRACT

The problem of predicting outcomes over time and expressing uncertainty about the future is one common to many scientific disciplines. For cost-effectiveness analysis used to aid resource allocation decisions in healthcare, this problem presents itself in the form of a disparity between the evidence time horizon (which is typically short-term) and the appropriate analysis time horizon (which is often long-term). To date, this problem has been primarily characterised as one of a need to extrapolate, i.e. an imperative to interpret the available short-term evidence and project this into the long-term in order to plug the evidence gap. Furthermore, the issue has been strongly associated with estimations of survival, but less so with other measures of disease progression, with estimates of cost, or with estimates of health-related quality of life.

This thesis strives to take a broad and thoughtful approach to examining the general problem of a dearth of evidence pertaining to the long-term. It is argued that this problem is most accurately and most usefully thought of as one of uncertainty. As such, in this thesis, the term 'temporal uncertainty' is employed. Consideration is given to the nature of temporal uncertainty and when it is of significance in the context of decision making with evidence development. Where a full expression of temporal uncertainty is necessary in order to make an informed decision, a number of approaches are described and appraised. Caution is advised in relation to extrapolating evidence over time due to the implicit assumption that outcomes in the short-term are good predictors of outcomes in the long-term. It is recommended that temporal uncertainty be characterised by a single uncertain 'temporal' parameter and incorporated into a probabilistic analysis in order to provide a true estimate of expected cost-effectiveness and to estimate the value of obtaining information that would lessen temporal uncertainty. In the context of these principles, a review of the health technology assessment (HTA) literature reveals that approaches to addressing temporal uncertainty to date have been inconsistent and largely inadequate. The review also makes apparent the full range of model parameters that are regularly exposed to temporal uncertainty and the specific analytical challenges that must be overcome. A motivating example (the RITA-3 decision model) is employed in order to develop and apply methods that appropriately quantify temporal uncertainty for a range of model parameters given the available evidence. The motivating example also facilitates an examination of the effects of expressing temporal uncertainty throughout a decision model. It is found that the replacement of 'conservative' temporal assumptions with expressions of temporal uncertainty alters the adoption recommendation for several of the risk groups under examination, that overall uncertainty around costs and health benefits is greatly inflated, that there is likely to be value in obtaining further information specifically in relation to the long-term temporal nature of certain model parameters and that there may also be value in 'waiting' for further evidence to be revealed if there is the potential for significant irrecoverable costs to be incurred.

In summary, this thesis represents a contribution to the development of methods to aid decision making in healthcare. In particular, the significant issue of temporal uncertainty is expounded and methods to appropriately address temporal uncertainty are developed and demonstrated.

Table of Contents

ABSTRACT.....	2
Table of Contents.....	3
List of Figures.....	5
List of Tables.....	8
Acknowledgements.....	10
Author’s Declaration.....	11
1. CHAPTER 1: INTRODUCTION.....	12
1.1 Introduction.....	12
1.2 Context: Decision-making for Resource Allocation in Healthcare.....	12
1.3 Objectives and Structure of this Thesis.....	17
2. CHAPTER 2: TEMPORAL UNCERTAINTY IN COST-EFFECTIVENESS DECISION MODELS.....	20
2.1 How Temporal Uncertainty Arises in CEA.....	20
2.2 Why/When It Is Important to Characterise Temporal Uncertainty in CEA.....	28
2.3 Expressing Temporal Uncertainty in a Cost-effectiveness Decision Model.....	44
2.4 The Value of Reducing Temporal Uncertainty.....	58
2.5 Conclusions.....	61
3. CHAPTER 3: KEY ANALYTICAL ISSUES IN TEMPOAL UNCERTAINTY AND REVIEW OF METHODS EMPLOYED TO DATE.....	63
3.1 Introduction.....	63
3.2 Current Guidance and Other Relevant Reviews.....	63
3.3 Key Issues When Modelling Temporal Uncertainty.....	69
3.4 Addressing Temporal Uncertainty by Parameter-type.....	77
3.5 Characterisation of Uncertainty Over Time.....	91
3.6 Conclusions & Discussion.....	94
4. THESIS CHAPTER 4: METHODS TO ADDRESS TEMPORAL UNCERTAINTY.....	97
4.1 Chapter 4.1: Overview.....	97
4.2 Chapter 4.2: Baseline Risk.....	117
4.3 Chapter 4.3: Treatment Effect.....	149
4.4 Chapter 4.4: Costs.....	178
4.5 Chapter 4.5: Health-related Quality of Life.....	195
4.6 Chapter 4.6: Overall Results and Discussion.....	209
5. THESIS CHAPTER 5: CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS.....	224

5.1	Conclusions	224
5.2	Limitations.....	228
5.3	Recommendations	229
5.4	Related Issues and Potential Further Research	229
APPENDICES		232
Appendix 1: Details of HTAs Reviewed		232
Appendix 2: RITA-3 Cohort by Risk Profile.....		241
Appendix 3: Application of Expert Elicitation Methodology to RITA-3 example (stylised elicitation)		242
REFERENCES.....		245

List of Figures

Figure 1: Decision modelling and evidence synthesis bridge the gap between the requirements for CEA and the limitation of the available evidence	17
Figure 2: Evidence gap resulting from time horizon mismatch.	23
Figure 3: A Problem of Parameter uncertainty? Where we must estimate expected values and related distributions for successive temporal periods over the unobserved period	25
Figure 4: A problem of structural uncertainty? Where we must consider which assumption regarding how the parameter changes over time is most appropriate.	27
Figure 5: Cumulative Incremental Net Health Benefit Over Time for Example 1.....	34
Figure 6: Cumulative Incremental Net Health Benefit over Time for Example 2	35
Figure 7: Incremental Net Health Benefit over Time for Example 3.....	38
Figure 8: Cumulative Incremental Net Health Benefit over Time showing the various time points and time horizons that relate to decision-making and the relevance of temporal uncertainty	43
Figure 9: Framework indicating the likely required complexity of temporal uncertainty modelling under different circumstances (for simplicity, the framework assumes a two comparator decision problem). Note that ‘non-negligible’ here can be taken to mean: a value of collecting further information that is so much smaller than the cost of collecting further information, the pursuit of further information would not be considered.....	44
Figure 10: The stylised example that will be used throughout Section 3 to explore how temporal uncertainty might be expressed in CEDMs	46
Figure 11: Scenario analysis. Three alternative scenarios (in particular the scenarios that NICE suggest should be explored regarding long-term treatment effect) are implemented in the CEDM..	47
Figure 12: The ‘best fitting’ parametric function is extrapolated over the unobserved period.....	49
Figure 13: Two alternative ‘fits’ to the short-term data are extrapolated over the unobserved period	50
Figure 14: A uniform distribution is applied in an attempt to convey total uncertainty over the unobserved period.....	52
Figure 15: Three possible scenarios incorporated in the probabilistic sensitivity analysis using model averaging.....	56
Figure 16: Temporal uncertainty is fully parameterised so that the long-term curve may take a range of plausible trajectories	57
Figure 17: Illustration of a parametric function being fit to survival data in order to extrapolate beyond the data.....	80
Figure 18: Structure of RITA-3 CEDM.....	100
Figure 19: Available survival data (Kaplan-Meier curves) against full model time horizon	103
Figure 20: Cumulative Incremental Net Health Benefit over time for risk group 3 showing mean CINB over time and 5th and 95th percentiles	105
Figure 21: The ‘proportional hazards’ approach to modelling this component of the observed period	113
Figure 22: The ‘independent arms’ approach to modelling this component of the decision model .	114
Figure 23: Approach of modelling baseline risk and treatment effect separately	115
Figure 24: Baseline risk refers to the transition between the ‘No Event’ health state and the ‘MI/CVD’ health state for patients receiving the conservative treatment strategy.....	117

Figure 25: The baseline survival data from RITA-3	120
Figure 26: “-Log-Log” plots of the baseline survival data	121
Figure 27: Empirical survival data (Kaplan-Meier curves) against the survivor functions of the parametric distributions	124
Figure 28: The resultant yearly transition probabilities over the observed period (5 years) for each distribution.....	125
Figure 29: The long-term survival curves extrapolated directly from the parametric functions described in section 7.1.2	130
Figure 30: Simple Lexis diagram pertaining to the RITA-3 example	134
Figure 31: Lexis diagram conveying rates used under assumption that age can be modelled by applying the within trial age effect.....	135
Figure 32: Transition probabilities resulting from Assumption 1 (for risk group 3)	136
Figure 33: Lexis diagram illustrating Assumption 2 where the hazard rates used over the long-term are the same as those used at the end of trial follow-up, i.e. the effects of aging and time without experiencing an event are assumed to cancel each other out.....	138
Figure 34: Transition probabilities resulting from Assumptions 1 and 2 (for risk group 3).....	139
Figure 35: Transition probabilities resulting from Assumptions 1, 2 and 3 (for risk group 3).....	140
Figure 36: Baseline transition probabilities and percentiles from Weibull extrapolation of the RITA-3 data	143
Figure 37: Desired distribution for β_{Age}	145
Figure 38: Yearly transition probabilities when temporal uncertainty is expressed regarding baseline risk.....	146
Figure 39: Treatment effect is applied to baseline risk in order to compute the risk of transitioning from the ‘No Event’ health state to the ‘MI/CVD’ health state for patients receiving the early interventional treatment strategy	150
Figure 40: Schoenfeld Residuals for alternative treatment strategies in RITA-3.....	153
Figure 41: -Log-Log curves for alternative treatment strategies in RITA-3.....	154
Figure 42: Empirical hazard curves showing treatment effect over the observed period	155
Figure 43: Forest Plot of meta-analysis of treatment effect carried out by Henriksson et al.	157
Figure 44: Illustration of the alternative temporal trajectories of the hazard ratio according to the three scenarios posited.....	160
Figure 45: The elicitation process – based on a diagram from Garthwaite et al. (Garthwaite PH, 2005)	168
Figure 46: Costs are associated with each health state in the long-term Markov portion of the CEDM	179
Figure 47: The trajectory over time of costs for risk group 3 according to 5 alternative assumptions	183
Figure 48: CPI inflation projection: percentage increase in prices on a year earlier (Bank of England, 2013)	187
Figure 49: Simulated long-term inflation rate	188
Figure 50: Temporal uncertainty relating to uncertain future cost reduction	190
Figure 51: Unit costs and utilities are associated with each health state in the long-term Markov portion of the CEDM	196
Figure 52: The trajectory over time of HRQoL per health state for risk group 3 according to 3 alternative assumptions.....	199

Figure 53: Trajectory of HRQoL with age from Ara and Brazier	203
Figure 54: Expression of temporal uncertainty in HRQoL for patient (in risk group 3) who experiences an MI aged 60	206
Figure 55: Evolution of Mean ICER for each characterisation of temporal uncertainty for each risk group	212
Figure 56: Cost-effectiveness planes for original and updated CEDMs.....	212
Figure 57: Change in EVPI as each source of temporal uncertainty was addressed	215
Figure 58: CINB for original and updated CEDMs	217
Figure 59: Frequency chart to be filled in by experts	243
Figure 60: Resultant 'super-distribution' from elicitation	244

List of Tables

Table 1: Summary of guidance relating to temporal uncertainty in 12 countries whose HTA processes are well developed.....	64
Table 2: Nature of the Disease in HTAs using some form of extrapolation.....	71
Table 3: Clinical characteristics of HTAs using some form of extrapolation.....	72
Table 4: Modelling vehicles employed in HTA.....	74
Table 5: Details of long-term estimation of time-to-event parameters.....	83
Table 6: Details of long-term estimation of longitudinal parameters.....	85
Table 7: Details of long-term estimation of cost/resource use parameters.....	88
Table 8: Details of long-term estimation of HRQoL parameters.....	90
Table 9: Details of How Long-term Uncertainty was Characterised.....	91
Table 10: Cost-effectiveness Results from Original CEDM.....	100
Table 11: Summary of Effect of Uncertainty on for each Risk Group.....	101
Table 12: Issues of Temporal Uncertainty in RITA-3 CEDM.....	107
Table 13: How the alternative approaches fare in terms of the three criteria.....	116
Table 14: AIC/BIC scores for each parametric distribution.....	123
Table 15: Death rates from coronary heart disease for males in 2008.....	140
Table 16: Cost-effectiveness results (mean ICER) for alternative assumptions regarding baseline risk and alternative parametric survival functions.....	141
Table 17: Mean ICER for each risk group for each assumption regarding long-term baseline risk....	142
Table 18: Cost-effectiveness Results after temporal uncertainty has been addressed for baseline risk. Compare with Table 10 to observe the change in the results.....	147
Table 19: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for baseline risk. Compare with Table 11 to observe the change in the results.....	148
Table 20: Results of Piecewise Cox Regression.....	155
Table 21: Results (for risk group 3) of sensitivity analysis regarding behaviour of treatment effect (hazard ratio) over unobserved period.....	160
Table 22: Results (for risk group 3) in terms of mean ICERs (£) for two-way sensitivity analysis regarding behaviour of treatment effect and baseline risk over unobserved period.....	161
Table 23: EVPPI for temporal uncertainty relating to treatment effect, given current model settings.....	165
Table 24: Cost-effectiveness Results after temporal uncertainty has been addressed for treatment effect. Compare with Table 18 to observe the change in the results.....	176
Table 25: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for treatment effect. Compare with Table 19 to observe the change in the results.....	177
Table 26: Estimated mean cost per relevant co-variate.....	180
Table 27: Cost-effectiveness results for assumption of constant costs.....	181
Table 28: Cost-effectiveness results for alternative assumptions regarding costs associated with a composite event.....	181
Table 29: Cost-effectiveness results for alternative assumptions regarding the change over time of costs per health state for risk group 3.....	182
Table 30: Cost-effectiveness results for risk group 3 when no age-related increase in costs is assumed versus when an uncertain 10 yearly increase in costs is assumed.....	184

Table 31: Cost-effectiveness results for alternative assumptions regarding discount rates.....	186
Table 32: Cost-effectiveness results when this 'inflation adjusted' discount rates along uncertain inflation approach is taken, versus the standard 'real' discount rates approach	188
Table 33: Cost-effectiveness results with and without UFE	191
Table 34: Cost-effectiveness Results after temporal uncertainty has been addressed for costs. Compare with Table 24 to observe the change in the results	193
Table 35: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for costs. Compare with Table 25 to observe the change in the results.....	194
Table 36: Estimated gain in health-related quality of life (Henriksson et al., 2008)	197
Table 37: Cost-effectiveness results for alternative assumptions regarding the change over time of HRQoL associated with a composite event for risk group 3	198
Table 38: Cost-effectiveness results for alternative assumptions regarding the change over time of HRQoL per health state for risk group 3	199
Table 39: Cost-effectiveness results for risk group 3 for alternative assumptions regarding differential long-term HRQoL in the 'No event' health state	200
Table 40: Cost-effectiveness results for model-averaged characterisation of long-term relative HRQoL in 'No event' health state	202
Table 41: Co-variates in calculation of HRQoL including age.....	202
Table 42: Cost-effectiveness results for an assumption of constant HRQoL vs. the inclusion of an end-of-life decrement	204
Table 43: Mean ICERs for risk group 3 for alternative assumptions regarding change in HRQoL in the 'Post MI' health state	205
Table 44: Cost-effectiveness results for an assumption of constant HRQoL vs. the inclusion of an end-of-life decrement and post MI recovery.....	206
Table 45: Cost-effectiveness Results after temporal uncertainty has been addressed for HRQoL. Compare with Table 34 to observe the change in the results	207
Table 46: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for HRQoL. Compare with Table 35 to observe the change in the results.....	208
Table 47: Summary of model updates	210
Table 48: Cost-effectiveness results for original CEDM vs. updated CEDM	211
Table 49: Probability that the early interventional strategy is cost-effective for the original and updated CEDM for each risk group.....	214
Table 50: EVPI per patient and per population for each risk group	215
Table 51: EVPPI/patient (£) for all uncertain temporal parameters and for each individual uncertain temporal parameter for each risk group	216

Acknowledgements

This thesis was researched and written from 2010 to 2014 at the Centre for Health Economics at the University of York. The research contribution in this PhD thesis was made under the terms of a Career Development research training fellowship issued by the National Institute for Health Research (NIHR) (grant CDF-2009-02-21) to Professor Andrea Manca. Some research was also carried out in the frame of a research project funded by the Medical Research Council (MRC). As such, I would firstly like to thank both the NIHR and the MRC for their financial and administrative support.

I would like to thank the Centre for Health Economics (especially the Team for Economic Evaluation and Health Technology Assessment) and the Department of Economics and Related Studies for their stimulating intellectual environments. Equally I would to thank the many colleagues and friends within both institutions who have helped me greatly throughout the PhD process with their feedback, advice and good company. I would especially like to thank my fellow PhD students in the Centre for Health Economics, Pedro Saramago, Manuel Espinoza and Thomas Patton.

I would like to sincerely thank my supervisor Professor Andrea Manca and the members of my thesis advisory panel, Professor Stephen Palmer, Professor Mark Sculpher and Professor Christopher Jackson. Their time and advice has been invaluable.

To my mother Virginia, my father Christy, my sisters Sinéad, Emer and Orla and my friends in York, in Ireland and elsewhere, thank you for your personal support, encouragement and good humour.

For my wife Áine. I really like her.

Author's Declaration

I declare that this doctoral thesis is the result of my original work. I also affirm that this thesis has not previously been presented to any other university or educational institution for examination. In addition, any views expressed in this document are exclusive responsibility of the author.

I hereby give my authorization for my thesis, if accepted, to be used for photocopying and inter-library loan. Likewise, I give my consent for the title and the abstract to be made available in all academic dissemination sources, making reference to authorship and copyright.

The material developed for this thesis has been communicated at two meetings of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and at a conference of the International Health Economics Association (iHEA).

ISPOR Berlin, 2012: **Mahon, R.**, et al. "PCV59 Where Did the Time Go?: Temporal Uncertainty in Cost-Effectiveness Decision Models." Value in Health **15**(7): A373.

ISPOR New Orleans, 2013: Workshop presentation, available from:

http://www.ispor.org/meetings/neworleans0513/releasedpresentations/W9_Mahon.pdf

iHEA Dublin, 2014: Organised session, abstract available from:

<https://ihea2014.abstractsubmit.org/presentations/12472/>

1. CHAPTER 1: INTRODUCTION

1.1 Introduction

“I think it’s much more interesting to live with uncertainty than to have answers which might be wrong”

Richard Feynman

This introductory chapter lays the foundations to the thesis by describing the decision-making context in which the issue of temporal uncertainty arises, namely, the role of cost-effectiveness analysis (CEA) in healthcare and the use of decision analytic modelling to aid CEA. There follows an outline of the objectives and the structure of this thesis.

1.2 Context: Decision-making for Resource Allocation in Healthcare

1.2.1 Resource Allocation in Healthcare

In healthcare, as in every other sector of the economy, we find the basic economic problem: resources are scarce but human wants and needs are vast, so how to optimally allocate these resources? It is argued that health as an economic good exhibits particular characteristics that make it quite distinct from other economic goods (Culyer, 1971); issues such as equity, consumer rationality and externalities result in a disparity of opinion regarding how the healthcare sector ought to be structured and administrated. Consequently, a variety of healthcare regimes (in terms of financing and provision) exist around the world. In Germany, France and the Netherlands, an approach comprising a private insurance market with a state subsidy is taken. In Canada, Norway and Spain, there exist national-level health insurance systems. In many countries, such as Denmark, Italy, New Zealand and the United Kingdom, national health services provide universal healthcare on behalf of the state (Folland et al., 2004). In England and Wales, to aid resource allocation decisions of the National Health Service (NHS), the National Institute for Health and Care Excellence (NICE) was established in 1999. NICE is tasked with (among other things) ensuring the best use of resources so that patients receive the greatest benefit (Great Britain. Dept. of Health, 1998). Although the research and analysis in this thesis is relevant for many healthcare regimes and advisory bodies,

there is particular focus on the decision-making setting of England and Wales and the advisory role of NICE.

1.2.2 Efficient Allocation

One matter that each kind of regime must consider is that of efficiency (or value for money)¹. For many publicly funded health systems, including the NHS, it is the 'Extra-Welfarist', societal decision-making approach that lays the economic foundation to the resource allocation process, and thus informs the definition of efficiency. In short, this approach takes an exogenously defined societal objective and views the health system's objective as maximising population health given an exogenous budget constraint for healthcare (Sculpher, 2005). In order to allocate healthcare resources 'efficiently' in this framework, it must be that the population health benefits of any health intervention are greater than their opportunity cost, where we think of the opportunity cost as the health benefits attributed to those interventions that are displaced when new interventions that impose costs on the system are imposed (Walker et al., 2007). 'Efficiency' therefore, is a product of a health intervention's benefits and costs, those of its relevant comparators, and the budget allocated to healthcare in the state. Demonstrating the efficiency of health interventions has become a central tenet in the management of modern public health systems.

1.2.3 Cost-effectiveness Analysis to Demonstrate Efficiency

This process of evaluating alternative health interventions in terms of their health benefits and costs is often referred to as 'economic evaluation'. It is important that economic evaluations are consistent, transparent and evidence-based (Drummond et al., 2005) (NICE, 2013). There is debate as to what form of analysis constitutes an appropriate economic evaluation. Such debate relates to the alternative theoretical approaches that may underpin an economic evaluation. The 'extra-welfarist' theoretical approach outlined above is reflected through the use of cost-effectiveness analysis.

Cost-effectiveness analysis (CEA) is the form of economic evaluation specifically recommended by NICE (NICE, 2013). In CEA, the benefits and costs of a health intervention are considered

¹ In economic terms, it is allocative efficiency in particular that is being referred to here, as opposed to technical efficiency (Palmer and Torgerson, 1999)

simultaneously against relative comparators. The cost-effectiveness of any one intervention can then be recorded as x amount of money per y amount of health benefit.

Further (though related) to the question of which theoretical approach should underpin economic evaluation is the question of exactly which costs and which benefits should be included in economic evaluation. It is intuitive that costs falling directly on the healthcare sector ought to be included (e.g. cost of a healthcare programme, cost of equipment, physician visit, etc.). It could be argued that other costs such as those falling on other economic sectors and certain less tangible costs should be included also (e.g. productivity losses and burden on patient and family). These two types of costs could broadly be categorised into 'direct' and 'indirect' costs. Similarly, regarding benefits/disbenefits, the changing health state of the patient(s) would intuitively be accounted for. However, other benefits (that mirror the other costs outlined above) could be accounted for also, e.g. productivity gains, benefits to family, benefits to other sectors. As well as categorising using the direct/indirect dichotomy, a distinction between perspectives is often drawn. For example, a 'healthcare system's perspective' would include only direct costs and benefits, whereas a 'societal perspective' would include costs and benefits that fall outside of the healthcare sector. NICE advocates a 'healthcare system perspective' where only direct health effects are accounted for and only costs relevant for the NHS & PSS (personal social services).

1.2.4 Measures of Health Benefit and Cost

To make consistent resource allocation decisions across clinical areas and kinds of intervention in CEA, there is a need for a generic measure of health outcome. There exist several such measures, e.g. the disability-adjusted-life-year (DALY) (Tan-Torres Edejer and World Health Organization, 2003), the healthy-years-equivalent (HYE) (Mehrez and Gafni, 1989) and the saved-young-life-equivalent (Nord, 1995). However, it is the Quality-Adjusted-Life-Year (QALY), that is the most typically employed measure in CEA (Briggs et al., 2006) and the measure explicitly recommended by NICE (NICE, 2013). The QALY accounts for both survival (life-years) and the health-related quality of life (quality adjusted) and is calculated as the product of these two components (Drummond et al., 2005). The costs should relate to resources under the control of the relevant health bodies (for the England and Wales: the NHS and Personal Social Services (PSS)) and be valued in monetary terms (Walker et al., 2007).

1.2.5 A Decision Criterion for Adoption

With consistent measures of health outcome (the QALY) and costs, competing health interventions can be directly compared in terms of cost-effectiveness. But when can an intervention be said to be cost-effective? Given two alternative treatments A and B, A is said to strictly dominate B (and therefore be deemed cost-effective) if A is less costly and more effective than B, and vice versa². If however, A is more effective than B but also more costly (a more common scenario), what must be considered is the whether the additional (incremental) cost is worth paying for the additional benefits (Walker et al., 2007). In CEA, an incremental cost-effectiveness ratio is calculated according to the following formula:

$$ICER_{AB} = \frac{\text{Costs of A} - \text{Costs of B}}{\text{QALYs of A} - \text{QALYs of B}}$$

The use of treatment A in the health system will be deemed a cost-effective use of resources if the ICER is found to be below a particular threshold (the ICER threshold).

The ICER threshold, in effect, expresses the monetary value of health outcome. More specifically, the threshold represents the maximum acceptable additional cost that can be incurred by the healthcare system in order to fund a healthcare intervention that will result in a health gain of 1 QALY, given that disinvestment from other interventions and services is implicit. The threshold should then, in principle, be determined by the health benefits estimated to be foregone elsewhere in the health system when a new intervention is funded. NICE has employed an ICER threshold range of between £20,000 and £30,000 per QALY (NICE, 2013). NICE has not, to date, provided any empirical evidence for this threshold range. How exactly this threshold should or could be calculated in reality is the subject of much discussion (Culyer et al., 2007) (McCabe et al., 2008). Recent research has endeavoured to estimate the ‘true’ NICE ICER threshold by estimating the relationship between changes in overall NHS spending and changes in mortality and quality of life (Claxton et al., 2013). This research provided a central estimate for the threshold of £12,936 per QALY.

Analogous to the calculation of the ICER in comparison with the threshold is the calculation of ‘net benefit’, the positive or negative calculation of which indicates a positive or negative decision recommendation respectively. The expression for net health or net monetary benefit can be easily

² A treatment B can also be ‘extendedly dominated’ if there is a combination of treatments A and C that would be less costly and more efficacious. When there are more than two competing interventions, an algorithm that removes the dominated treatments and calculates ICERs between the remaining treatments ought to be undertaken. For the remainder of this thesis, for simplicity, only a two-treatment decision problem will be discussed.

derived from the expression for the ICER along with the threshold. For instance, the net health benefit of investing in treatment A instead of treatment B can be expressed as:

$$\text{Net Health Benefit}_{AB} = (\text{QALYs of A} - \text{QALYs of B}) - (\text{Costs of A} - \text{Costs of B})/\text{Threshold}$$

1.2.6 CEA is Informing Two Decisions

When considering the adoption of new health technologies, the choices available to decision makers are not limited to: adopt or do not adopt. There will inevitably be uncertainty surrounding the cost-effectiveness results, and therefore a certain probability that the subsequent adoption decision will not be the correct one, leading to an overall loss of health benefits. As part of its recommendation, advisory bodies may want to request that further evidence be collected³. The CEA then should inform two distinct but related decisions (Claxton et al., 2002):

- (i) Is a health technology cost-effective given the currently available evidence?
- (ii) Should further evidence be sought?⁴

1.2.7 The Development of a Decision Analytic Framework

The key source of evidence employed to inform a CEA often comes in the form of one or more randomised controlled trials (RCTs). While there are many advantages to using RCTs (e.g. randomisation which minimises selection bias), there are numerous limitations to using RCT and other typically available evidence (see Figure 1) as the basis for a CEA (Claxton et al., 2002).

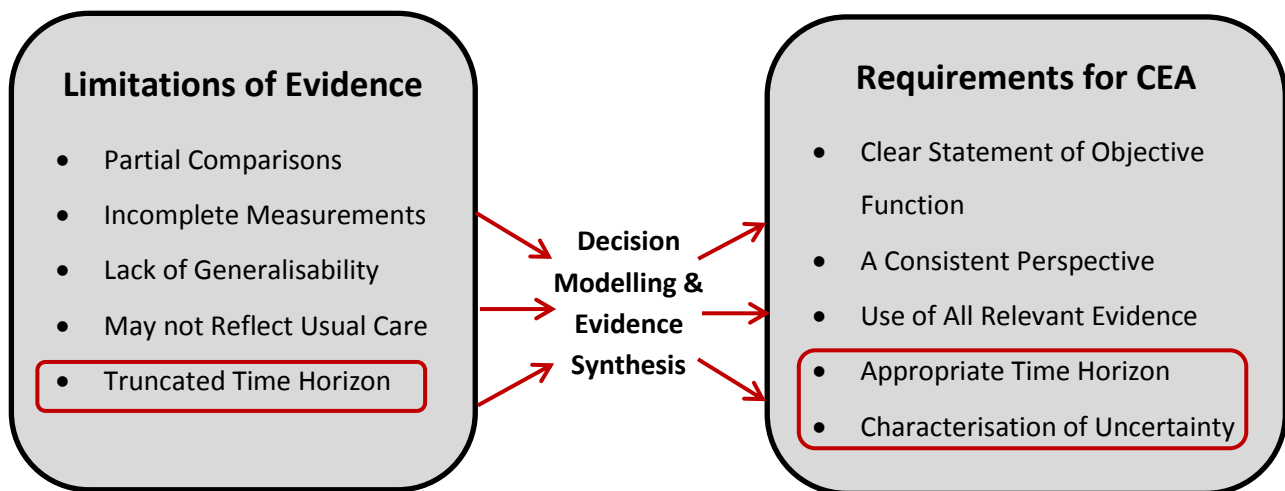
Given the two key questions posed above, a number of requirements arise regarding the appropriate execution of CEA which are at odds with the characteristics typical of RCTs. These requirements have been well articulated elsewhere and include: a consistent perspective, use of all relevant evidence, use of an appropriate time horizon and characterisation of uncertainty (Sculpher et al., 2006, Philips et al., 2006). As a result of these requirements, purely trial-based analyses (that is, analyses that only employ data from a clinical trial and are predicated within the time horizon of that trial) often do not suffice. Additional methods are required in order to provide a complete

³ The full range of decision options available, in principle, under conditions of uncertainty is discussed in Chapter 2.

⁴ At present, addressing the question of the pursuit of further evidence is not a formal requirement for submissions made to NICE.

picture of cost-effectiveness. A decision analytic framework has been developed in recent years that makes use of evidence synthesis and decision modelling in order to bridge the gap between the nature of the available evidence and the requirements for an appropriate CEA, thus facilitating appropriate decision making (Briggs et al., 2006). Analyses that employ decision modelling and/or evidence synthesis can be referred to as cost-effectiveness decision models (CEDMs). Analytic methods that improve CEDMs are continually being developed. Figure 1 illustrates the role of decision modelling and evidence synthesis in CEA. Highlighted are the particular evidence limitation and the particular CEA requirements that give rise to the focus of this research.

Figure 1: Decision modelling and evidence synthesis bridge the gap between the requirements for CEA and the limitation of the available evidence



Note: The items listed are examples, i.e. the lists are not exhaustive. Highlighted are the limitations and requirements that relate to the focus of this research.

1.3 Objectives and Structure of this Thesis

The ultimate goal of this Ph.D. research is to improve the decision making process relating to the allocation of resources in healthcare. It is envisioned that the dissemination of this thesis will lead to the further development of guidance regarding how health interventions are economically assessed and how the results are communicated to decision makers.

This thesis focuses on one important function of decision modelling: addressing the disparity between the time horizon of the primary evidence available and the time horizon deemed appropriate for the analysis. This time horizon mismatch results in an evidence gap for the analysis,

where there is no direct means of estimating long-term values for a range of CEDM input parameters, thus preventing the estimation of the full range of costs and health outcomes pertinent to the decision problem. The thesis will offer a considered examination of the uncertainties that arise when the analysis time horizon exceeds the evidence time horizon, and will endeavour to develop methods that address these uncertainties, thereby facilitating appropriate and efficient decision making. The uncertainties in question will be referred to collectively as ‘temporal uncertainty’.

The three core objectives of this thesis are:

Objective 1: Understanding Temporal Uncertainty

The first objective is to establish the nature and significance of temporal uncertainty in CEA and to determine, at a conceptual level, what would constitute an appropriate approaches for addressing temporal uncertainty in a CEDM. This objective is addressed in Chapter 2 where firstly, it is discussed how temporal uncertainty arises due to a time horizon mismatch and how this may affect a number of model input parameters. Secondly, the importance of quantifying temporal uncertainty is outlined, highlighting circumstances where temporal uncertainty is and is not likely to influence the decisions that CEA is employed to inform. Thirdly, the implications of alternative assumptions regarding knowledge of the unobserved period and the relevance of short-term evidence are examined. Lastly, conceptual methods of accounting for temporal uncertainty in CEDMs are described and appraised.

Objective 2: Examining the Analytic Issues and the Current Methodologies

Moving from the conceptual to the practical, the second objective is to understand the key analytical issues relevant to tackling temporal uncertainty and to identify the areas where specific methodological development is warranted given the methods currently employed in Health Technology Assessment (HTA). This objective is addressed in Chapter 3. Firstly, an overview of current guidance across jurisdictions is given, as well as a summary of recent relevant reviews and studies. Secondly, an outline of key analytical issues is given, with particular focus on the challenge of expressing temporal uncertainty for different types of model parameter. This discussion is accompanied by a thorough review of HTAs conducted in the UK over a six year period in order to surmise and appraise the methods employed to-date. Finally, some thoughts and criticisms are offered regarding the adequacy of current methodology and the requirement for improved methodology and updated guidance.

Objective 3: Developing and Applying Appropriate Methodology and Observing the Consequences

The third objective is to develop and apply methods that appropriately address temporal uncertainty, for a range of model parameters, using the best available evidence. To this end, Chapter 4 comprises a systematic empirical exercise, carried out using an existing cost-effectiveness decision model (RITA-3). This substantial chapter breaks down into several sub-chapters. The first sub-chapter introduces the RITA-3 model and outlines the issues of temporal uncertainty. The second, third, fourth and fifth sub-chapters each focus on different key model parameters. In these sub-chapters, the use and limitations of the available evidence is analysed and means of expressing temporal uncertainty for the particular parameter are developed and applied. In the final sub-chapter, the results of the updated decision model for all risk groups are presented and analysed in comparison to those of the original model. The consequences of addressing temporal uncertainty are then observed and discussed.

The final chapter of this thesis summarises the contribution of this research to the methodological literature in this area, offers some recommendations for future analyses and methods guidance and outlines areas where further research would be fruitful.

Part of this PhD research (in particular, some of the review work outlined in Chapter 3) was carried out in tandem with research conducted as part of an MRC funded cross-institutional project on the subject of extrapolation in cost-effectiveness analysis. The remit of the MRC funded project was to examine and make recommendations on the use of techniques to extrapolate short-term evidence over time to estimate the values of various cost-effectiveness model parameters. The research is ongoing. In contrast, this PhD research takes a broader view and considers extrapolation as one approach to address an evidence gap that may or may not be problematic for the analysis, while considering the problem of a temporal evidence gap as one primarily of uncertainty.

2. CHAPTER 2: TEMPORAL UNCERTAINTY IN COST-EFFECTIVENESS DECISION MODELS

This chapter seeks to introduce, describe and explore the issue of temporal uncertainty in cost-effectiveness analysis (CEA). Firstly, it is discussed how temporal uncertainty arises due to a mismatch between the evidence time horizon and the appropriate analysis time horizon. Secondly, the importance of quantifying temporal uncertainty is outlined, highlighting circumstances where temporal uncertainty is and is not likely to influence the decisions CEA is employed to inform. Thirdly, using a simple stylised example, the implications of alternative assumptions regarding our knowledge of the unobserved period and the relevance of short-term evidence are discussed. Lastly, methods of appropriately expressing temporal uncertainty are explored.

2.1 How Temporal Uncertainty Arises in CEA

In essence, this research concerns itself with one particular limitation of typically available evidence and two particular requirements for CEA. The limitation in question is that of a truncated time horizon given the available evidence, and the requirements in question are those of an appropriate analysis time horizon and the characterisation of uncertainty. This section seeks to describe the nature of the problem for decision making that arises due to the tension between this evidence limitation and these CEA requirements. We can then begin to consider what the role of decision analysis ought to be in addressing this problem and what methods would be best employed and developed.

2.1.1 Appropriate Time Horizon vs. Evidence Time Horizon

The time horizon in a CEA is the time period over which costs and benefits are calculated. To obtain an appropriate estimate of cost-effectiveness, the total differences in costs and health benefits between interventions should be accounted for. Therefore, the appropriate time horizon to use in CEA is the time period over which costs and benefits are expected to differ between the competing interventions. When there are mortality impacts associated with the illness and/or interventions, the appropriate time horizon is likely to be a lifetime time horizon. Generally, the appropriate time horizon is relatively long-term, the notable exceptions being some palliative treatments or certain

acute conditions (Sculpher et al., 2006). It is important to note that it is not the duration of the disease/intervention but the duration of the effects owing to the disease/intervention that must be accounted for. For example the duration of a disease such as acute meningitis is short-term, but because there are risks of mortality and long-term disabilities, a long-term time horizon would be required for any CEA related to this disease. It may be argued that considering very distal costs and effects in healthcare decision-making is futile, as the clinical and decision making contexts are bound to alter significantly over time. However, any shortening of the time horizon could bias the cost-effectiveness decision that needs to be made today. For example, an analysis that does not take into account the distal QALYs gained in young patients who are given life-saving therapy today will underestimate the cost-effectiveness of that therapy. Even if there is major uncertainty regarding conditions in the long-term, it is still necessary to express today's expectation regarding the costs and QALYs that will accrue up to the full time horizon.

In contrast to the typical appropriate analysis time horizon, the time horizon of the key source(s) of evidence is usually short-term. Phase III randomised controlled trials (RCTs), which are often the basis for 'trial plus model' analysis, are typically commissioned primarily to demonstrate efficacy and/or safety and so do not include substantial follow-up data. Although some RCTs are designed to capture long-term effectiveness, the collection of such data is naturally a slow process. The opportunity costs associated with delaying an adoption recommendation mean that analysts often must work with a truncated evidence time horizon.

2.1.2 Consequences for CEA

When the appropriate time horizon exceeds the evidence time horizon, the full range of costs and health effects pertinent to the decision problem cannot be (directly) estimated. As described in Chapter 1, decision modelling can be employed to overcome this, and other, limitations of the available evidence. There are several components of a cost-effectiveness decision model (CEDM) where values over an 'unobserved period' may have to be estimated. Estimates of long-term survival typically receive most attention; however, a range of model input parameters as well as other model components are likely to be affected. Namely:

- (i) Measures of disease progression
 - Both time-to-event and longitudinal
 - Both baseline and relative effect
- (ii) Health-related quality of life

- (iii) Costs and resource use
- (iv) Structural assumptions

The nature of these parameter-types is discussed in detail in Chapter 3. For now, it is important to simply note that all parameter-types and other affected model components are borne in mind as methods to address the temporal evidence gap are developed.

2.1.3 Extrapolation?

Since short-term trial data are often the best (if not the only) relevant evidence available (Charlton, 1991), a temporal evidence gap is commonly overcome by using decision modelling to 'extrapolate' the short-term trial data over time. Extrapolation is clearly a useful procedure, as it exploits the best available evidence and facilitates an estimation of cost-effectiveness over the entirety of the appropriate analysis period. In fact, it is often claimed that extrapolation of evidence is a requirement in CEA when a time horizon mismatch arises (Drummond et al., 2005) (Latimer, 2011) (Sculpher et al., 2006). While extrapolation may play a crucial role in tackling this problem, talking only of a 'need to extrapolate' oversimplifies and may even misrepresent the problem at hand.

Firstly, extrapolating short-term evidence over time may be decidedly inappropriate as there may be little or no relationship between the outcomes over the short-term and outcomes over the long-term. As such, it may be unhelpful to advise analysts that extrapolating evidence over time is the necessary course of action. Extrapolation often involves fitting a parametric function to short-term data and stretching that function over a longer-term period. As a result, the focus of extrapolation modelling is often the optimisation of the functional fit to the data. This element, while not unimportant, distracts from the overriding assumption being imposed, the assumption that extrapolating evidence into another temporal period is in any way appropriate.

Secondly, plugging the evidence gap between the evidence time horizon and the required time horizon need not involve extrapolation over time per se. While extrapolation in CEA can refer to the modelling of surrogate to final outcomes, transferability between sub-groups and other forms of generalisation of evidence, the term extrapolation in CEA is generally associated with the extrapolation of evidence over time in order to infer long-term outcomes from short-term evidence. However, the estimation of long-term outcomes given a dearth of long-term evidence need not literally involve extrapolation of evidence over time. For example, imposing a simple assumption

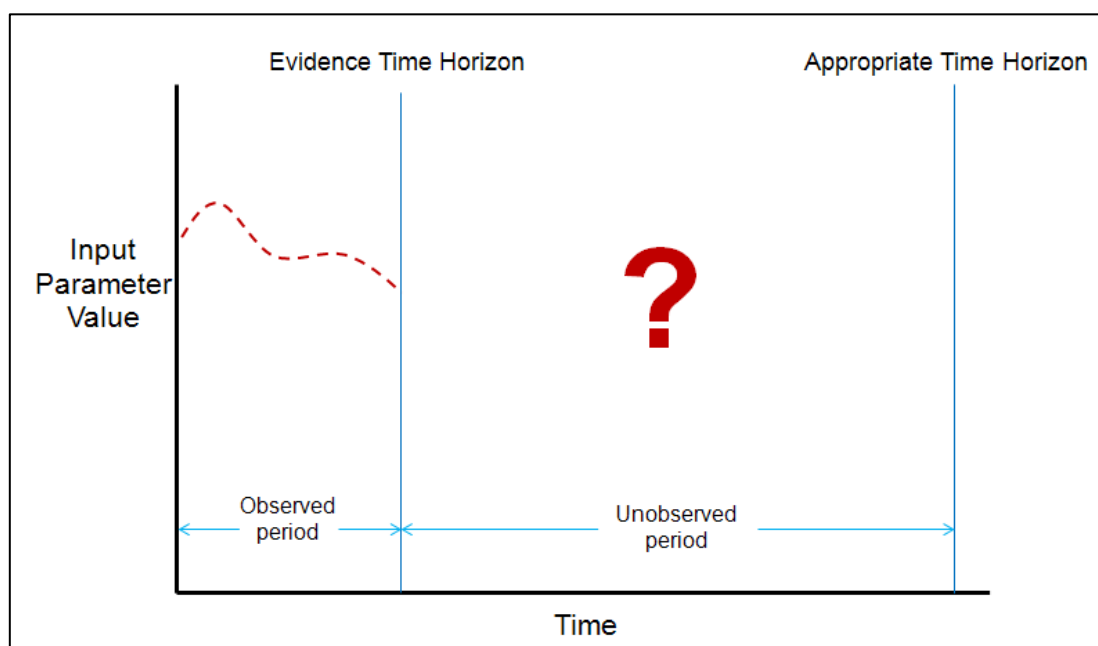
regarding long-term costs or effects, or using expert elicitation may facilitate the completion of a parameter's 'temporal curve' but it would be inaccurate to think of this as extrapolating evidence over time, as the short-term evidence itself may not have been exploited. The term extrapolation therefore is an unhelpful one, as it may cause analysts to automatically rely on short-term evidence to estimate long-term values instead of considering the appropriateness of such an action.

Thirdly, it must be considered whether simply plugging the evidence gap, either by literally extrapolating evidence over time or by some other means, is all that is required. If there is a dearth of direct long-term evidence, there will be significant uncertainty associated with any long-term estimates and as a result, that uncertainty ought to be expressed in the model. It is contended here that a broader characterisation of this problem is required, with a view to developing methods that convey the lack of pertinent evidence available whilst still constructing a useful decision model.

2.1.4 What is the Nature of the Problem?

When the appropriate analysis time horizon exceeds the evidence time horizon, the values for a number of inputs into the analysis (e.g. measure of disease progression, quality of life, costs) must be estimated over an 'unobserved period', i.e. a time period beyond that where evidence exists to directly inform input parameter values (the 'observed period'). This is clearly problematic for the analysis, but the nature of the problem could be thought of in a number of different ways.

Figure 2: Evidence gap resulting from time horizon mismatch.



Note: Values for input parameters can be estimated up to the evidence time horizon (there may even be ‘trend’ as some parameters evolve over time) but there is no, or very little, evidence pertaining to the ‘unobserved period’.

The problem could perhaps most accurately be described as a missing data problem as there is simply an absence of evidence pertaining to a particular time period (between the evidence time horizon and the required time horizon). The literature on missing data in CEA however, generally refers to instances of data missing amongst datasets that cover the period of interest (Briggs et al., 2003, Burton and Altman, 2004), i.e. random censoring, which is not truly the concern here. The issue here is more akin to type I censoring, where after a fixed time point, no data are recorded.

It is asserted here that this problem is most aptly (and most usefully) thought of as one of uncertainty. That is, it may be possible to estimate the values of various model input parameters (as well as the related parameter uncertainty) up to the evidence time horizon, but there is uncertainty regarding their values beyond the evidence time horizon (or regarding their ‘temporal behaviour’ over this period). It is undoubtedly crucial to somehow plug this evidence gap if some estimate of cost-effectiveness is to be generated. However to generate an accurate estimate of cost-effectiveness given current evidence, it is equally crucial to convey in the CEDM the lack of evidence pertaining to the unobserved period. In essence, our current state of knowledge and uncertainty must be reflected as accurately as possible.

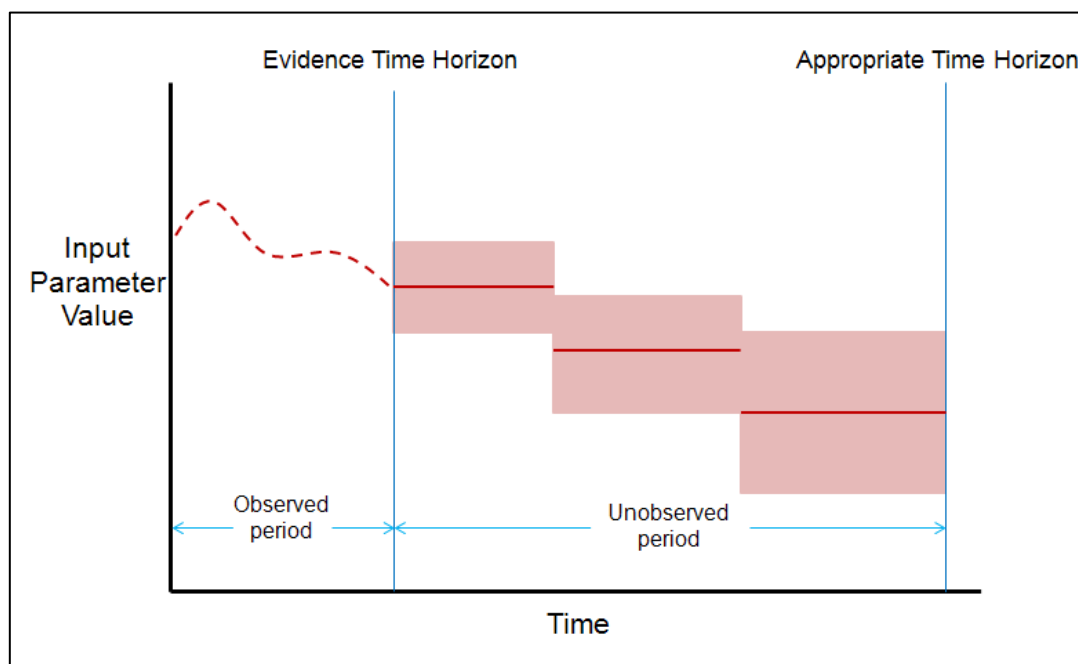
Uncertainty, as found in CEA, has been sub-categorised in a number of ways in the economic evaluation literature (Briggs, 2000, Claxton, 2008, Bilcke et al., 2011). Taking Bilcke et al.’s outline of types of uncertainty, we can begin to consider what type of uncertainty arises when there exists a dearth of long-term evidence in CEA.

Methodological uncertainty refers to the normative views about what approach constitutes optimum decision making (Bilcke et al., 2011). For example, the uncertainty regarding which perspective is appropriate, which costs should be included, which discount rates should be used. In CEA carried out for NICE this form of uncertainty is expected to be small given the existence of clear guidance documents indicating the preferred methods that should be used (NICE, 2013). Although considering a long-term time horizon poses questions regarding long-term discount rates and the inclusion of future costs, questions which could be considered as types of methodological uncertainty, we shall consider these uncertainties to

be negligible as there is clear guidance (in the case of CEAs carried out for NICE) regarding the preferred methodology.

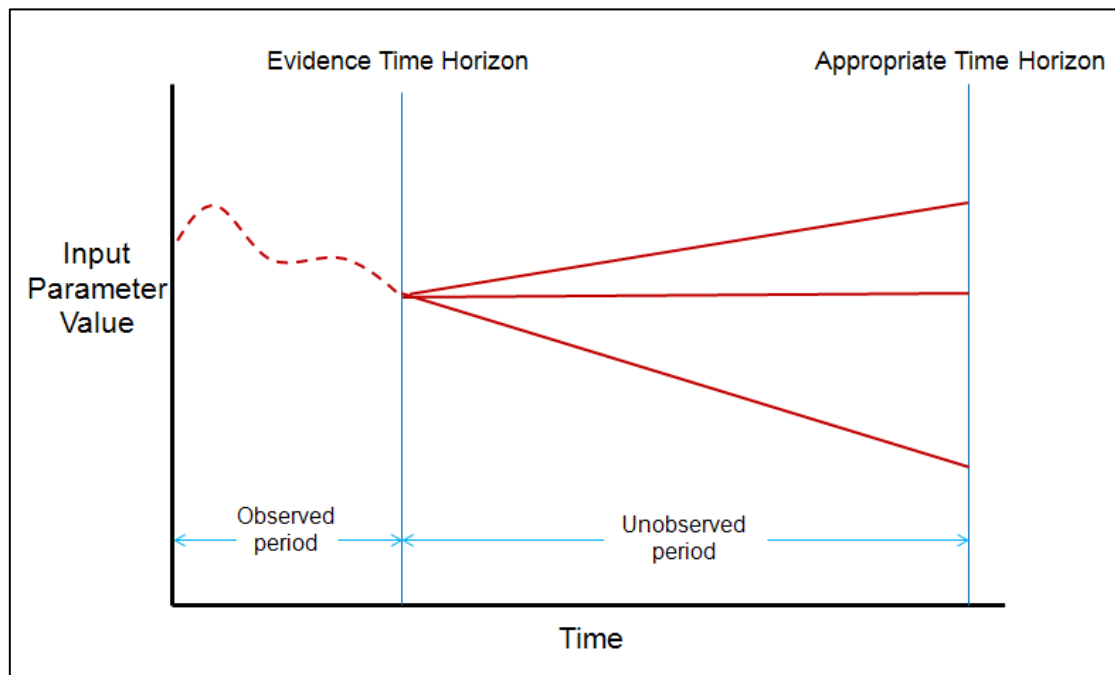
Parameter uncertainty is the uncertainty surrounding the true value of an input parameter (e.g. transition probability, cost, utility value). Given a short-term evidence time horizon, it could be said that there is parameter uncertainty; as for the unobserved period, there is uncertainty in relation to parameter values. However, unlike how parameter uncertainty is usually addressed, this problem does not simply require distributions to be assigned to known point estimates, as even the point estimates are not known for the unobserved period. Moreover, it may not be a single value required for a parameter, but a range of values as time moves forward, i.e. what is effectively required is a separate parameter value per temporal period as illustrated in Figure 3. In other words the uncertainty specifically pertains to the parameter's relationship with time. In this sense, perhaps it is more useful to consider the uncertainty regarding the temporal trajectory of a parameter, rather than the expected values and parameter uncertainties in successive temporal periods. It is important to note that parameter uncertainty extrapolated over time does not equate to uncertainty regarding the behaviour of a parameter over an unobserved period.

Figure 3: A Problem of Parameter uncertainty? Where we must estimate expected values and related distributions for successive temporal periods over the unobserved period



Structural uncertainty refers to the appropriateness of what is imposed by the model framework (Bojke et al., 2009). Recent research has offered methods to address structural uncertainties in CEDMs (Jackson et al., 2011, Russell, 2005). For example Jackson et al. outline a framework whereby single structure analysis, scenario analysis and model averaging are employed depending on the circumstance. The uncertainty arising from a short-term evidence time horizon could be characterised as one of structural uncertainty, as an assumption is required in the model concerning the temporal behaviour of one or more parameters (e.g. the parameter remains fixed at a certain value over the observed period, or the parameter value increases at a certain rate over the unobserved period, etc.). As such, techniques to quantify structural uncertainty may be useful in quantifying the uncertainty under investigation here. To a large extent, the structural uncertainty approach is what is currently used in CEDMs to address this issue. It is common in a CEDM to impose one assumption regarding the long-term behaviour of a parameter in the base-case analysis and then to explore the related uncertainty by applying alternative assumptions in a deterministic sensitivity analysis. However the uncertainty in question may not be well characterised by discrete competing assumptions, rather a continuous measure of the uncertainty may be required. Moreover, the uncertainty arising from a short-term evidence time horizon relates specifically to time, which suggests that a specific methodology related to time may be warranted.

Figure 4: A problem of structural uncertainty? Where we must consider which assumption regarding how the parameter changes over time is most appropriate.



Although all are relevant, the uncertainty arising from a mismatch between the evidence time horizon and the appropriate analysis time horizon does not fit neatly into any of the above categories.

The unique aspect to this source of uncertainty is the role of time. There have been recent calls for a greater focus on the modelling of time and on parameters' relationships with time in CEA (van de Wetering et al., 2012). Those calls are echoed in this thesis. When the appropriate analysis time horizon exceeds the evidence time horizon, the challenge that arises is to convey current expectations and uncertainties regarding the unobserved period. Of course, characterising expectations and uncertainties regarding prospective time periods is inherently difficult. The economist J.M Keynes invoked the concept of 'irreducible uncertainty' to describe the innate difficulty in characterising future outcomes (Keynes and Feinstein, 1921). Yet estimating future outcomes is a requirement for many disciplines. For example, in finance and meteorology, techniques exist that endeavour to characterise expectations regarding future time periods based on historical evidence (Makridakis et al., 1982). While some of these techniques may be applicable to healthcare, it is important to note a key contextual difference between CEA for healthcare and other disciplines which appear to encounter the same problem of uncertainty over time. Elsewhere (in finance and meteorology), time is considered to be cyclical and past trends can be used to estimate

future outcomes (e.g. a certain weather pattern at t has been observed to lead to a certain weather outcome at $t+1$ with probability p , p can be constantly updated as new observations emerge). For the economic evaluation of a healthcare intervention however, time is linear, i.e. it has a beginning (treatment) and an end (end of differential effect of treatment). As such, there is no precedent for measuring the extent to which outcomes (especially clinical outcomes) in one period are related to, or can be used to estimate, outcomes in another period. For example a drug may deliver a reduction in mortality for 3 years but may be ineffectual or even detrimental after 3 years, a characteristic that would not be captured by short-term evidence.

2.1.5 Temporal Uncertainty

Given that this issue relates to a number of well-established areas of cost-effectiveness decision modelling, but is not well captured by any one, it seems desirable to employ a term that represents this unique set of problems so that we may begin to develop methods that address them in a coherent and appropriate way.

Let us define this issue generally as *Temporal Uncertainty*. That is, let temporal uncertainty relate, in general, to the uncertainties that arise when there is a disparity between the required time horizon and time horizon of the primary source(s) of evidence in CEA. This will primarily, though not exclusively, refer to the uncertainties regarding the behaviour/trajectory over time of model input parameters. Addressing temporal uncertainty, therefore, pertains to both estimating expected parameter values over the long-term, as well as quantifying the related uncertainty.

2.2 Why/When It Is Important to Characterise Temporal Uncertainty in CEA

Having established the existence of temporal uncertainty in CEA, what is considered next is the importance of addressing it and whether it may sometimes be less problematic than other times.

2.2.1 Why Quantify Uncertainty (in General)?

The reasons for addressing uncertainty in CEA in general are well established. Addressing and characterising all sources of uncertainty⁵ inherent in a CEA is now a central aspect of CEA. The reasons for this can be summarised as follows.

- (i) To calculate the true expected values of overall costs and effects in a CEA, it is necessary to consider the distributions of the input parameters rather than their point estimates (Claxton, 2008). The reason for this that decision models are typically non-linear (i.e. the output of the model is often a multiplicative function of the input(s)). This is problematic because of the statistical rule that says the expectation of a non-linear transformation does not equal the non-linear transformation of an expectation (Rice, 1995).
- (ii) For certain sources of uncertainty (and this will often be true for temporal uncertainty), the point estimate itself can only be generated by considering the associated uncertainty (e.g. by incorporating and weighting alternative plausible assumptions, we convey what our true expectation is regarding the nature/value of an input parameter). When there is a number of alternative scenarios (e.g. with a structural uncertainty), it will not matter for the adoption decision what the implied expectation is, as long as we have weighted these scenarios appropriately in the analysis in order to produce an estimate of expected cost-effectiveness.
- (iii) Considering uncertainty in a CEA reveals that a technology is expected to be cost-effective only with a particular probability (and therefore with a related error probability)⁶. In other words, there is a probability that after a decision is made, on receipt of further evidence, it transpires that the decision option chosen was in fact not cost-effective. There is thus an expected value associated with collecting further evidence now that would serve to lessen the uncertainty around costs and effects. The value associated with collecting further evidence becomes manifest at times when the acquisition of additional evidence compels the decision maker to change his/her mind, thereby avoiding the incurrence of opportunity costs. The value of collecting further

⁵ By uncertainty here, we do not mean variability (natural variation between patients which is irreducible, also known as first-order uncertainty), nor do we mean heterogeneity (differences between patients with similar characteristics which are explainable), we mean the fact that we cannot know for certain what the costs and outcomes would be if a treatment were to be provided in reality for a particular population of patients (Claxton, 2008).

⁶ This fact, it is important to note, should not imply that the decision maker should concern his/herself with the “statistical significance” of the expected cost-effectiveness but rather with the value of obtaining further evidence (Claxton, 1999)

evidence (i.e. the expected additional net benefits) can be calculated and compared to the cost of obtaining further evidence so as to indicate the worthiness of such an undertaking. This analysis is termed Value of Information (VoI) Analysis.

- (iv) On the completion of a decision model and VoI analysis, *some* decision must be taken there and then. There are potential opportunity costs associated with whatever decision is made, e.g. if a technology is adopted but is then seen to be cost-ineffective at a later date, there may be irrecoverable costs associated with reversing the initial decision. Therefore it is prudent to address the uncertainty around whether the irrecoverable costs expected to be forgone are greater than the additional net benefits of immediate use of the technology, since it may be cost-effective to wait until more evidence is available before endorsing the technology (Claxton, 2008, Eckermann and Willan, 2007).
- (v) A formal requirement to quantify uncertainty can also incentivise manufacturers to lower prices and/or to provide further information. If the characterisation of uncertainty reveals that it is more prudent to obtain further evidence before making an adoption recommendation, a decision maker could then demand that the manufacturer either generates the required further evidence or simply reduces the price of their health technology to the point that the pursuit of further evidence is no longer valuable (Griffin et al., 2011).

It can be concluded from these points that it is always desirable to characterise uncertainty in CEA. However, as temporal uncertainty pertains only to the long-term (unobserved period) we may consider the circumstances under which temporal uncertainty will truly 'matter' in the analysis; or more accurately, what level of modelling might be needed to characterise temporal uncertainty in order to satisfy the remit of the analysis?

2.2.2 Addressing Temporal Uncertainty: When Does It Truly Matter?

Although ideally, all uncertainties in CEA are fully characterised, the consequences of not addressing particular sources of uncertainty can vary by circumstance. It is argued here that the characterisation of temporal uncertainty is crucial under some circumstances, but in other circumstances the impact of temporal uncertainty is negligible and as a result, less complex modelling is required.

In many instances, the costs and effects pertaining to the unobserved period contribute little to total costs and effects, in which case temporal uncertainty may be of little cause for concern. There are several potential reasons for this:

- (i) If the length of the unobserved period relative to the observed period is small, then intuitively fewer unobserved costs and effects are expected. The uncertainty regarding the values of these costs and effects therefore has less of an impact on the CEA results. Relatedly, the 'maturity' of the evidence can dictate the extent to which temporal uncertainty will have an impact. For example, if 90% of patients have died within the observed period, then the health outcomes over the unobserved period (even if lengthy) will have relatively little effect.
- (ii) Another consequence of a relatively short unobserved period is that there is likely to be less uncertainty regarding the values of costs and effects as we approach the analysis time horizon. Generally, there is greater uncertainty the more distal an outcome is. Therefore, if the evidence time horizon is relatively close to the required time horizon, there is likely to be relatively little uncertainty regarding the values of outcomes over the unobserved period.
- (iii) Commonly, a CEDM will chart a progressive disease where the cohort, on average, move to progressively 'worse' health states as time advances. As a result, there are fewer health effects (i.e. QALYs) at stake over longer-term periods.
- (iv) To compound the above point, the existence of discounting (of both health effects and costs) renders outcomes over distal periods less valuable⁷. Often, very distal outcomes (after circa 50 years) have negligible impact on the ICER. To put it another way, one could say that it is common in CEA to observe *diminishing temporal returns*. (This concept is illustrated in Figures 5, 6 & 7).

Research has shown that the assumptions imposed regarding the behaviour of parameters over the unobserved period (let us call these temporal assumptions) can have a marked impact on the key CEA output, i.e. the mean ICER. For instance, Kim and Thompson considered three plausible models

⁷ Both future costs and future effects are discounted in CEA. The use of discounting is based on the assumptions that (i) health is tradable with monetary assets, (ii) health is tradable over time and (iii) there exists a positive time preference. There is debate surrounding whether both costs and health benefits should be discounted at the same rate or whether differential discounting should take place. (Claxton et al., 2006, Brouwer et al., 2005)

for measuring the cost-effectiveness of screening for abdominal aortic aneurysm. These models when extrapolated to a lifetime horizon produced cost-effectiveness estimates ranging from £1600 to £4200 per life-year gained (Kim LG and Thompson S, 2010). Another analysis by Connock et al. demonstrated that the choice of function to represent and extrapolate the short-term data had a significant effect on overall survival gain, and consequently expected cost-effectiveness, in two cost-effectiveness models for cancer drugs (Connock et al., 2011).

Regardless of the impact that alternative temporal assumptions can have on the mean ICER, what ultimately is of consequence is the impact that these assumptions can have on the decisions that the CEA is designed to inform. Recall that the goal of CEA is to address two questions: Should a health technology be adopted given the current evidence? And, should further evidence be sought? It is sometimes the case however, that only an adoption recommendation is required. For example, NICE does not currently require value-of-information analysis as part of its health technology assessment programme.

2.2.2.1 The Adoption Decision

If our sole concern is producing an adoption recommendation, the uncertainty regarding the value of parameters after the observed period is of consequence only if there is reason to believe that the addition of costs and effects over the unobserved period could possibly alter the adoption recommendation. For instance, if the analysis up to the evidence time horizon returns an adoption recommendation, we must consider whether an analysis spanning the full required time horizon could plausibly return a different recommendation. It would arguably be adequate to provide a within-trial estimate of cost-effectiveness if it could be demonstrated that longer follow-up would certainly only confirm the adoption recommendation (Sculpher et al., 2006). It may be that relatively simple analysis can be employed in order to demonstrate that no plausible set of assumptions regarding costs and effects over the unobserved period would be expected to change the adoption recommendation.

It is common in HTA for cost-effectiveness results to be presented for alternative time horizons as a scenario analysis (as we will see in Chapter 3). This may be helpful in conveying the sensitivity of the adoption recommendation to the time horizon imposed. However, such a scenario analysis paints an incomplete picture of how cost-effectiveness evolves over time and of the importance of costs and health benefits accrued over the unobserved period.

To analyse, more thoroughly, the impact of long-term costs and benefits and to assess what rigour of modelling is required to address temporal uncertainty, cumulative incremental net health benefit (CINHB) over time can be calculated. CINHB at time t is calculated as:

$$CINHB_t = \sum_{k=1}^t (Q_{Bk} - Q_{Ak}) - (C_{Bk} - C_{Ak})/\lambda$$

Or alternatively, cumulative incremental net monetary benefit (CINMB) which can be calculated as:

$$CINMB_t = \sum_{k=1}^t (Q_{Bk} - Q_{Ak}) * \lambda - (C_{Bk} - C_{Ak})$$

where A and B are two competing health interventions, C and Q are costs and QALYs respectively and λ is the willingness to pay threshold for an additional unit of health benefit.

This measure, taking into account costs and health benefits and the cost per effect threshold, portrays how the adoption recommendation evolves over time (as we account for more distal costs and health benefits). When $CINHB_t > 0$, intervention B is in the cost-effective 'zone'. Applying the standard adoption decision rule, we get:

If $CINHB_T > 0$

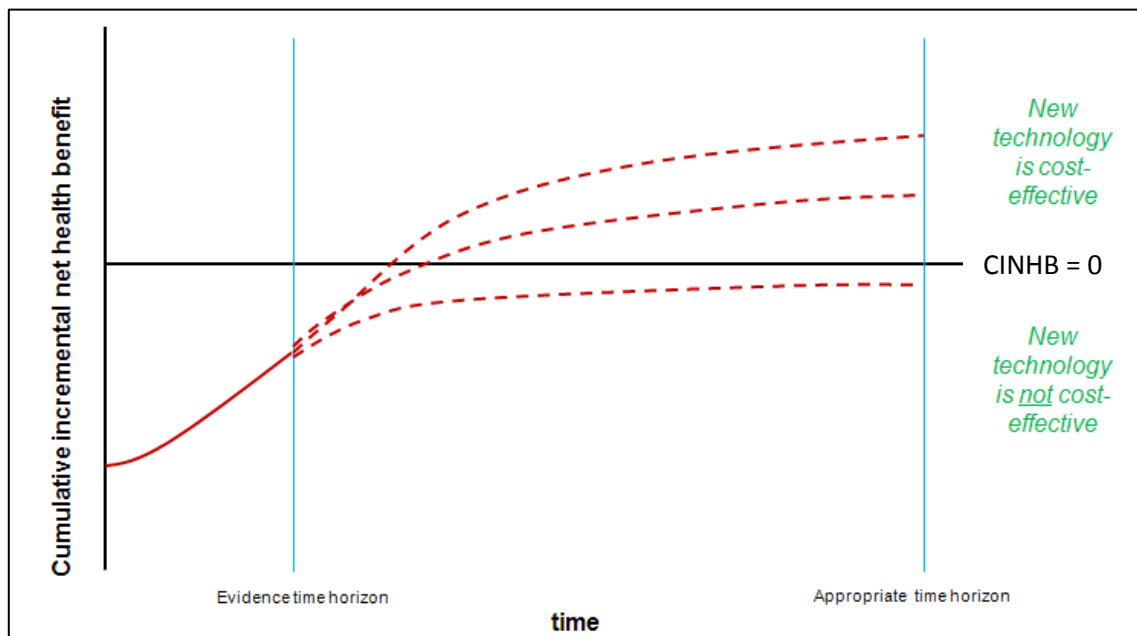
Then *Adopt B*

Where T is the appropriate analysis time horizon.

Let us explore the characteristics and usefulness of CINHB by examining three stylised examples. For simplicity, let us say that we are comparing two technologies; the 'new technology' is both more expensive and more effective than the 'old technology'. A CINHB above zero represents a recommendation of adopting the new technology.

Example 1: In this example, all costs are assumed to be captured over the trial period (for example the technology may be a one-off operation or a short course of drug treatment). It is also assumed that there will be no rebound effect or any other future event that may impact future health benefits. Finally it is assumed that the ‘new technology’ is observed to be cost-ineffective at the evidence time horizon. In this example we expect CINHB to be upward sloping where the more effective technology accrues more net health benefits over time relative to its comparator. Even if the treatment effect is assumed to cease after the trial period, the slope of the CINHB curve ought to be ≥ 0 as the cohort treated with the ‘new technology’ will have experienced less mortality and less morbidity and therefore will accrue more health benefits for the remainder of their lives. The CINHB that represents Example 1 is depicted in Figure 5.

Figure 5: Cumulative Incremental Net Health Benefit Over Time for Example 1



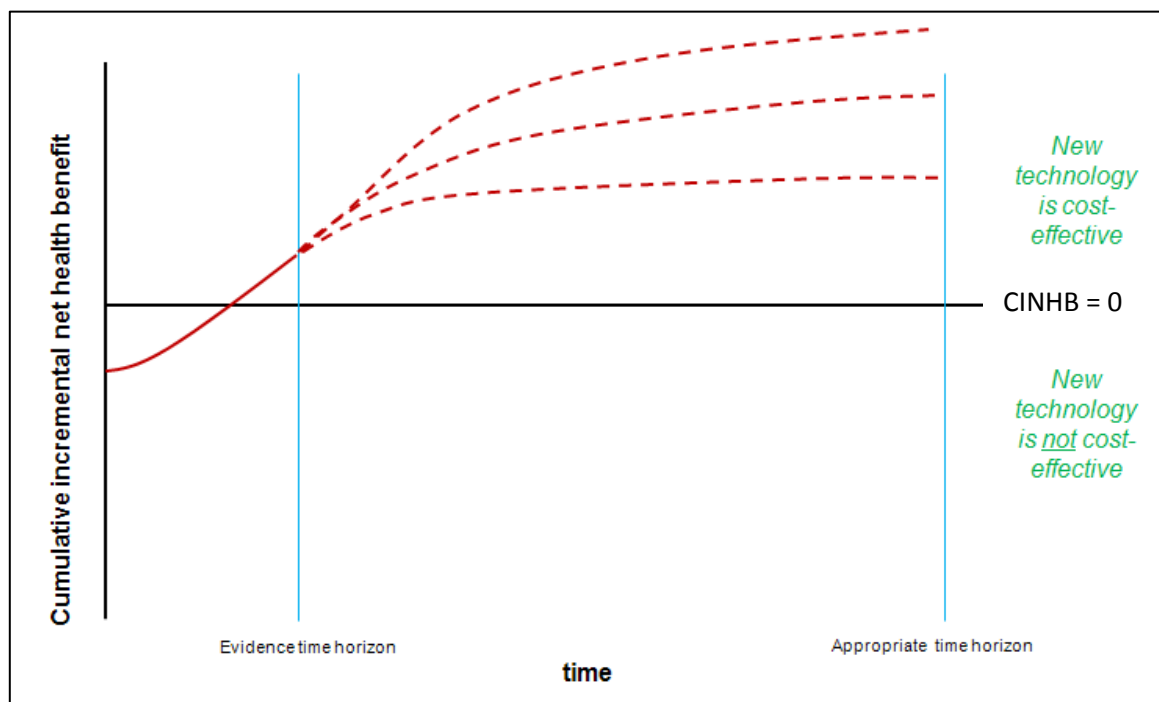
Note: the new technology depends on the net health benefits accrued over the unobserved period in order to be deemed cost-effective and as a result the CEA is highly sensitive to the temporal assumptions imposed. Also the curve of the CINHB illustrates the concept of diminishing temporal returns.

The characteristics described in example 1 are common in health technology assessment. The incremental net health benefits accrue over the observed period but do not reach the point where the new technology is deemed to be cost-effective (i.e. where the CINHB > 0). Although we expect incremental net health benefits to continue to accrue over the unobserved period, it is not certain whether these additional net health benefits will be enough for the new technology to be deemed

cost-effective (as can be seen from three hypothetical realisations depicted in the graph). In other words, since the new technology relies on the net health benefits accrued over the unobserved period in order to be deemed cost-effective, it is possible that the adoption recommendation resulting from the model will depend on the ‘temporal assumptions’ imposed, i.e. the assumptions that dictate the behaviour of the model parameters as we move into long-term periods. As such, in this scenario, it would be vital to address and fully quantify the temporal uncertainty present.

Example 2: Example 2 is identical to Example 1 except that the new technology is observed to be cost-effective at the evidence time horizon. Example 2 is depicted in Figure 6.

Figure 6: Cumulative Incremental Net Health Benefit over Time for Example 2



Note: the new technology is shown to be cost-effective at the evidence time horizon and the CINHB is expected to continue to increase over time

In this example, the new technology is cost-effective at the evidence time horizon. Given the characteristics of the competing technologies, we can be confident that the new technology will remain cost-effective over the unobserved period as it is expected to continue to accrue more net health benefits relative to its comparator over the unobserved period, i.e. we have no reason to believe that CINHB could begin to monotonically decrease at any point. In general therefore, when a

more costly and more effective technology is observed to be cost-effective at the evidence time horizon, it may be possible, by considering whether it is feasible for the CINHB curve to move below zero, to provide an adoption recommendation without estimating parameter values over the unobserved period.

i.e. where E = Evidence time horizon, B/A = new/old technology, K is a point in time

If $CINHB_E > 0, C_{BK} \cong 0, Q_{BK} > Q_{AK}$ for all $K > E$

Then $CINHB_T > 0$, Adopt B

Making these assumptions could only be done under particular circumstances however.

Consideration must be given to the following possibilities:

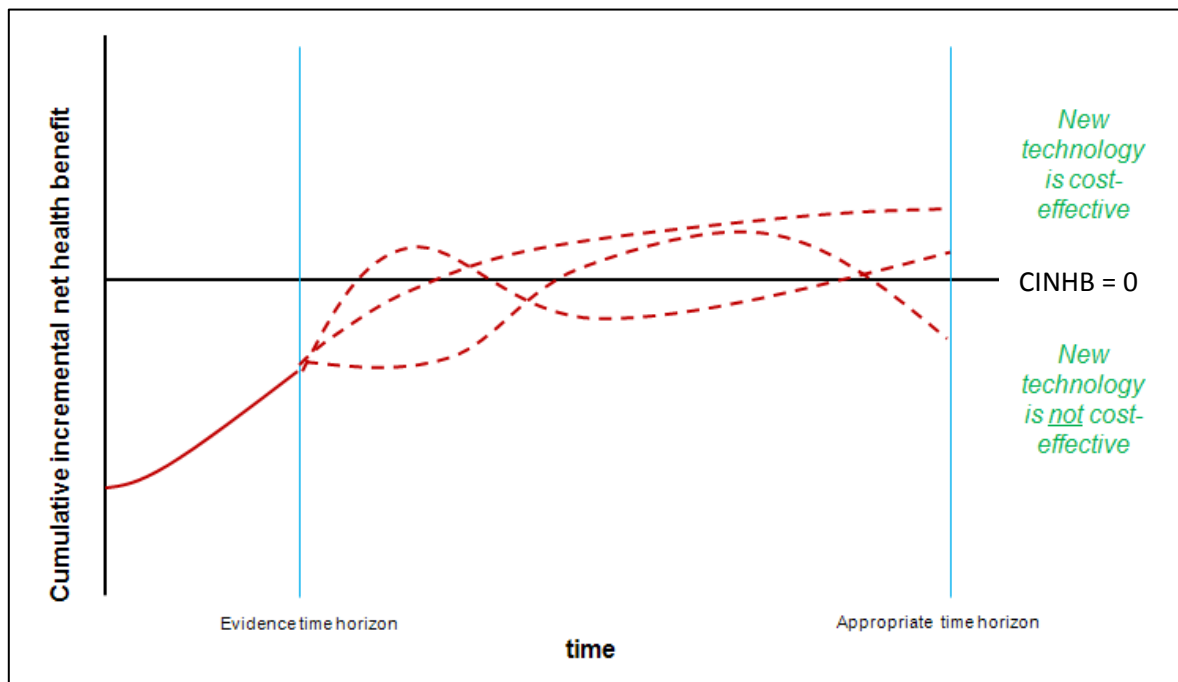
- (i) A rebound effect. When a health technology shows clear clinical superiority to its comparator during say an RCT, it is typically expected to continue to perform at least as well as its comparator over the unobserved period. Occasionally however, there may be reason to believe that there will be a reversal in relative clinical effectiveness sometime in the future. This could, for example, be due to an initially effective treatment simply delaying death or disease progression and when treatment ceases (or the effects wear off), patients die or progress at a faster rate than those receiving competing treatments (Drummond et al., 2005). In the event of a rebound effect, it is possible for the relative negative health effects in the long-term to negate the overall cost-effectiveness of a treatment. The concept of a rebound effect is discussed in depth in Section 4.3.1.2.
- (ii) Significant future costs. Often there will be minimal long-term costs directly associated with a health technology (e.g. the intervention may be a 'one-off' with high upfront costs, or a drug may be administered for a fixed amount of time recorded within the trial). If however, the future costs associated with a technology are expected to be significant over the long-term, these may negate the relative health gains accrued over this time (especially if costs remain fixed but the relative health gains decrease).

(iii) Other future event. There are a number of other possible future contextual changes (for example a price shock, or the entry of a new comparator) that may affect the adoption recommendation when a long-term analysis time horizon is considered. The impact that such an event has on expected cost-effectiveness depends on a number of factors such as: the extent to which costs paid before the change are irrecoverable, whether sequencing between treatments is possible and whether there are mortality impacts. For example if there is a significant reduction in the price of a treatment but patients may sequence from one treatment to another without loss of health benefit, it ought to be possible to simply change the adoption recommendation without incurring irrecoverable costs. Such a possibility would still affect the optimality of the initial decision, but perhaps not drastically.

Notwithstanding the above possibilities, a steadily increasing CINHB that has crossed the threshold (the point of being deemed cost-effective) during the observed period often implies cost-effectiveness at the long-term analysis time horizon. In Example 2, the new technology is the more expensive technology and so cost-effectiveness at the evidence time horizon implies significant health gains over the observed period. If we do not expect any rebound effect, nor any significant future costs, nor any other significant future event, the new technology would be expected to continue to accrue at least as many net health benefits compared to the old technology over the unobserved period, therefore remaining cost-effective. Although it may be sometimes reasonable to make adoption recommendations based on truncated time horizons, in practice, it may be unwise to recommend that analysts take this approach as it is still desirable to quantify temporal uncertainty under all circumstances.

Example 3: This final example simply describes the expected behaviour of CINHB when one or more of the possibilities discussed above (future costs, rebound effect, other future events) affects CINHB over the unobserved period. The existence of one or more of these characteristics renders the assumption that the CINHB curve will remain monotonic invalid. Example 3 is illustrated in Figure 7.

Figure 7: Incremental Net Health Benefit over Time for Example 3



Note: certain model characteristics give rise to the possibility of a non-monotonic CINHB and thus a requirement to characterise temporal uncertainty for the purposes of making an adoption recommendation

Essentially if it is possible for the CINHB curve to be non-monotonic with respect to time, it will be necessary to characterise (to at least some extent) the temporal uncertainty existent in the model.

Given either (a) uncertainty regarding the monotonicity of CINHB over unobserved period, or (b) a monotonically increasing CINHB where the more expensive and more effective technology is not cost-effective at evidence time horizon, some characterisation of temporal uncertainty will be required. However, for the purposes of appropriately producing an adoption recommendation, complex quantification of temporal uncertainty may not be necessary. To test whether full quantification of temporal uncertainty is truly necessary to make an informed adoption recommendation, a range of extreme temporal assumptions could be imposed in order to observe the effect on the mean ICER and the resulting adoption recommendation. Such assumptions ought to include scenarios that represent the bounds of plausibility regarding the temporal behaviour of model parameters. Of course, what is extreme but plausible is a subjective concept and some validation would be necessary for this task.⁸ If it transpired that no plausible assumption resulted in a different adoption recommendation, then one could argue that there is no need to fully characterise the temporal uncertainty. All that would be necessary would be to demonstrate with

⁸ For example, a clinical expert might be consulted to inform what can be deemed plausible.

these extreme assumptions (or perhaps with a single ‘conservative’ assumption) that there is negligible chance of the recommendation being incorrect. The imposition of one or more conservative assumptions is a means of conveying confidence that the resultant adoption recommendation can be given with confidence. For example, a new technology could conservatively be assumed to have minimal clinical effectiveness over the long-term and if still observed to be cost-effective, a positive recommendation can be made with confidence. However, there is a danger that conservative assumptions, when used in a base-case analysis, can lead to an inappropriate calculation of expected relative cost-effectiveness (e.g. the mean ICER). It is important to note that when conservative assumptions are employed (i.e. assumptions that knowingly underestimate or overestimate the value of input parameters), the resultant mean ICER will not be a true reflection of expected relative cost-effectiveness. Furthermore, it is not appropriate to carry out value of information analysis when conservative assumptions are used.

If alternative plausible assumptions do represent different adoption recommendations, it is necessary to investigate further and characterise more thoroughly the existent temporal uncertainty in order to compute expected cost-effectiveness.

2.2.2.2 The Decision to Obtain Further Evidence

Broadly, there are two aspects to how temporal uncertainty can impact decisions related to evidence acquisition. Firstly, temporal uncertainty represents one of several sources of uncertainty that will directly influence the value of obtaining further information given today’s available evidence when considering adoption recommendations for present and future incident populations. It may also be a source of uncertainty for which it is desirable to specifically calculate the value of reducing or eliminating. Secondly, the issue of temporal uncertainty is highly pertinent to the broader relationship between decision-making and time. In this sense, we can examine the impact that temporal uncertainty has on the attempt to strike a balance between the value of obtaining evidence and the value of access to a new technology.

2.2.2.2.1 Temporal Uncertainty and the Value of Further Information

The quantification of uncertainty in general allows an analysis to express the usefulness of current evidence with regard to decision making and to calculate the value of obtaining supplementary

evidence. Even if it is possible to appropriately calculate expected cost-effectiveness (and therefore produce an adoption recommendation) with current evidence, it may be the case that the opportunity costs associated with a possible 'wrong' adoption decision are so substantial that it is more prudent (and indeed cost-effective) to obtain further evidence before a final recommendation is made. The full quantification of temporal uncertainty therefore may cause the analysis to suggest that obtaining some supplementary evidence on the temporal nature of one or more parameters before making an adoption recommendation is the optimal course of action.

A requirement to quantify the value of obtaining further evidence should thus imply a requirement to fully quantify every source of uncertainty including temporal uncertainty. It is possible that the expected ICER with current evidence is so high or so low (i.e. that we are so certain what is the correct adoption recommendation) that there would be negligible value in collecting further information, in which case one could argue that it is again reasonable to simply employ conservative assumptions to aid the adoption decision. In general however, it should be assumed that a requirement to examine the value of information leads to a requirement to fully quantify temporal uncertainty.

2.2.2.2 Irrecoverable Costs, Further Decision Options and the Value of Waiting

Temporal uncertainty represents more than just a further source of uncertainty to be quantified in order to carry out value of information analysis. The issue of uncertainty regarding the value of costs and health benefits over time is inherently linked to the broader issue of decision uncertainty over time. In effect, separate adoption decisions are required for present and future incident populations. What links these separate decisions and makes each relevant for decision-making in the present is the potential for irrecoverable costs, i.e. a coverage decision can always be reversed if it transpires after further evidence collection that the decision was 'wrong' in terms of cost-effectiveness, but it is possible that in implementing such a 'wrong' decision, investments which cannot be recovered (e.g. equipment, facilities, staff training) are incurred⁹. Irrecoverable costs can be avoided by simply waiting until sufficient further evidence is available. However, this approach implies a decision of 'reject' for present and near future incident populations. In order to strike a balance between the

⁹ In making a decision regarding access to a new technology therefore, there are, in fact, three pertinent notions of value: the value of the technology given current evidence; the value of reducing uncertainty about the technology's cost-effectiveness; and the value of any investment (or reversal of that investment) were the new technology to be adopted.

value of evidence regarding the cost-effectiveness of a technology and the value of access to that technology, further decision options (beyond adopt and reject) are, at least in principle, available to decision-makers.

- (i) Approve only in research (OIR), where a technology is approved, but only in the context of further research, e.g. clinical trials (Claxton et al., 2011a).
- (ii) Approve with research (AWR): where a technology is given broad approval, but it is stipulated that further information is collected as patients receive treatment. Long-term approval may be contingent on this information being positive for the technology (Claxton et al., 2011a).

A framework has recently been developed which guides decision-making in the context of evidence development (Walker et al., 2012). The framework outlines circumstances where the use of OIR and AWR type decision are likely to be appropriate, in particular: when more evidence is worthwhile, the required evidence can be generated following an approval decision and there would be a cost associated with reversing the decision at a later date.

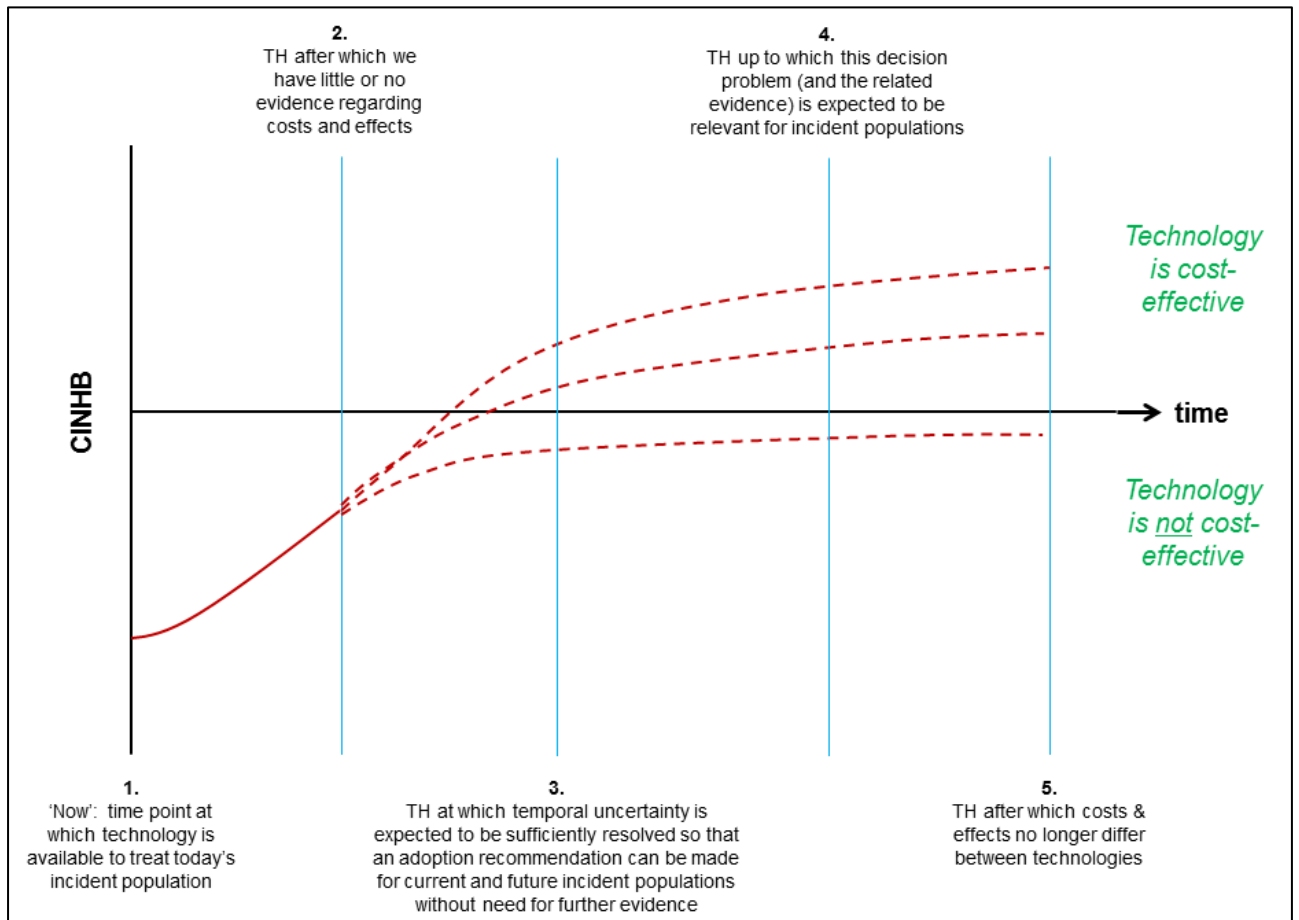
It may be possible to address temporal uncertainty through evidence generated from further research making an OIR/AWR decision potentially appropriate. However, it may also be the case that the evidence required cannot be resolved by research but only over time (e.g. waiting for further trial follow-up to estimate long-term treatment effect). In the latter circumstance, an OIR/AWR decision would not be appropriate as the evidence generated would not be able to address the pertinent uncertainties (unless further trial follow-up is considered as OIR). Rather, a judgment will be required regarding whether in the period of time between the point at which the technology is first available and the point at which the uncertainty regarding the cost-effectiveness of the technology will sufficiently resolve itself, the expected benefits of allowing immediate access to a technology outweighs the potential irrecoverable costs incurred, i.e. there may be value in 'waiting' until further evidence naturally becomes available before making an adoption recommendation. In this sense again, it is apparent that temporal uncertainty is a special case of uncertainty. If we can identify the point at which the uncertainty regarding long-term costs and health benefits ought to sufficiently resolve itself in order to make an adoption recommendation without the need for further evidence, then a calculation can be made comparing the expected benefit in allowing access to the

technology to the expected irrecoverable costs incurred were the positive recommendation to be reversed.

In describing the role temporal uncertainty plays in decision-making, it is helpful to make distinctions between time horizons. Temporal uncertainty on the one hand pertains to the uncertainty regarding costs and health benefits associated with the patients that might be treated today and so one relevant time horizon is that over which those costs and effects are expected to differ (5. in Figure 8 below). On the other hand, as described above, temporal uncertainty can also play a part in determining whether the pursuit of further evidence is worthwhile. In this context, the relevant time horizon is that over which the decision problem is expected to remain relevant or, in a sense, problematic (4. in Figure 8 below). The value of this latter time horizon is itself subject to uncertainty. What must be taken into account is the extent to which future incident populations can benefit from research commissioned in the present day. In reality, evidence generated over the short-term is likely to gradually lose its relevance over the long-term. Thus, a finite time horizon typically acts as a proxy for future changes in technologies, prices and information. Such future changes can be explicitly modelled though this approach presents numerous technical and methodological challenges (Philips Z et al., 2008).

These distinctions between time horizons and the relevance of temporal uncertainty to the value of access/irrecoverable costs trade-off can be illustrated by extending how CINHB is employed.

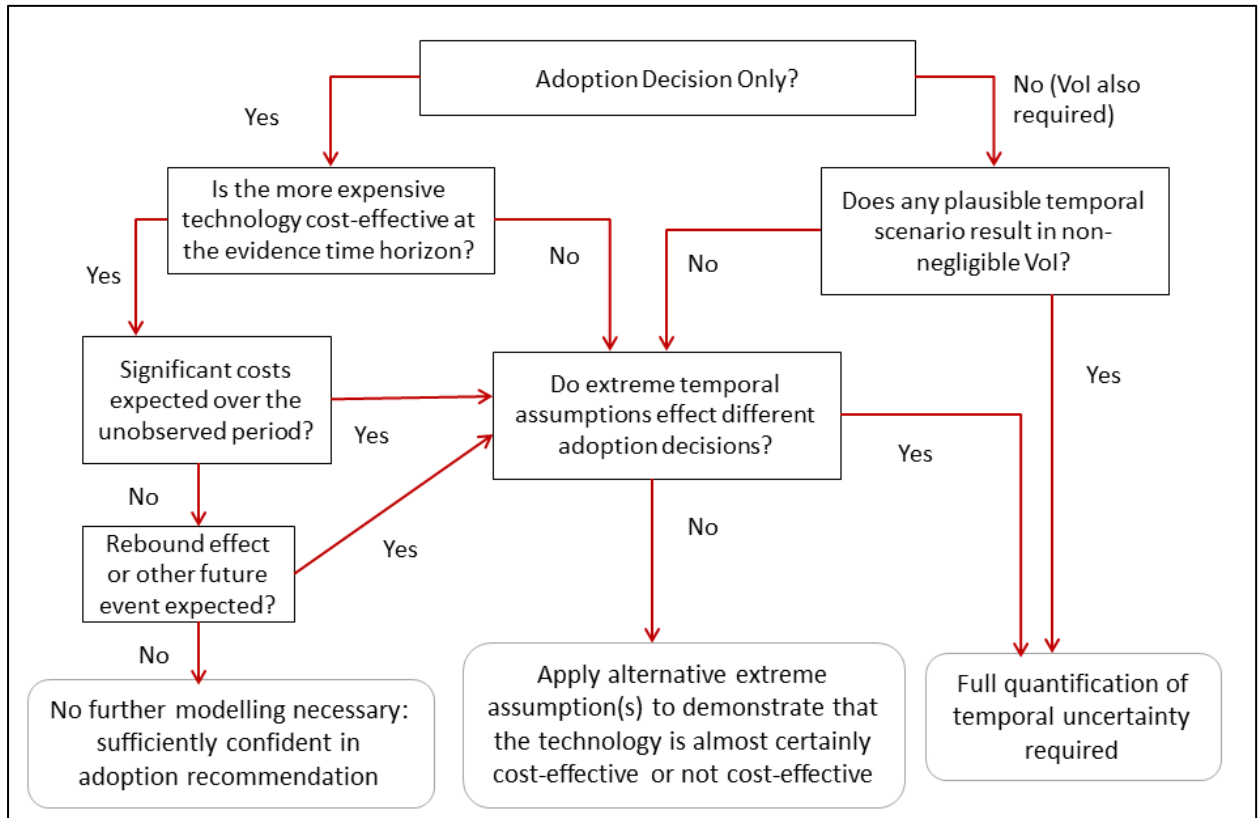
Figure 8: Cumulative Incremental Net Health Benefit over Time showing the various time points and time horizons that relate to decision-making and the relevance of temporal uncertainty



2.2.2.3 Brief Summary

The flow chart in Figure 9 summarises much of the discussion in this section and represents a possible framework regarding the complexity of modelling required to characterise temporal uncertainty depending on the circumstance. To reiterate, it is always desirable to quantify as fully as possible all sources of uncertainty (just the incentive for manufacturers to lower prices and/or provide more evidence is motivation enough to quantify all uncertainty to the greatest extent possible). However, in relation to the decision making a CEA is designed to aid, relatively simple characterisation of temporal uncertainty may, on occasion, be sufficient. Note that the flow chart is designed to illustrate the potential impact of temporal uncertainty in CEA and is not meant as a definite guide to modelling – the scenarios described in the chart will themselves be subject to uncertainty and so the appropriate complexity of temporal uncertainty modelling may be more than indicated.

Figure 9: Framework indicating the likely required complexity of temporal uncertainty modelling under different circumstances (for simplicity, the framework assumes a two comparator decision problem). Note that ‘non-negligible’ here can be taken to mean: a value of collecting further information that is so much smaller than the cost of collecting further information, the pursuit of further information would not be considered.



2.3 Expressing Temporal Uncertainty in a Cost-effectiveness Decision Model

This section explores how the temporal uncertainty of a model parameter in a CEDM might be expressed, assuming that there is a requirement to fully characterise parameter behaviour over the unobserved period¹⁰. The implications of various approaches are discussed with a view to determining what constitutes an appropriate method of expressing temporal uncertainty.

2.3.1 Quantifying Uncertainty in General

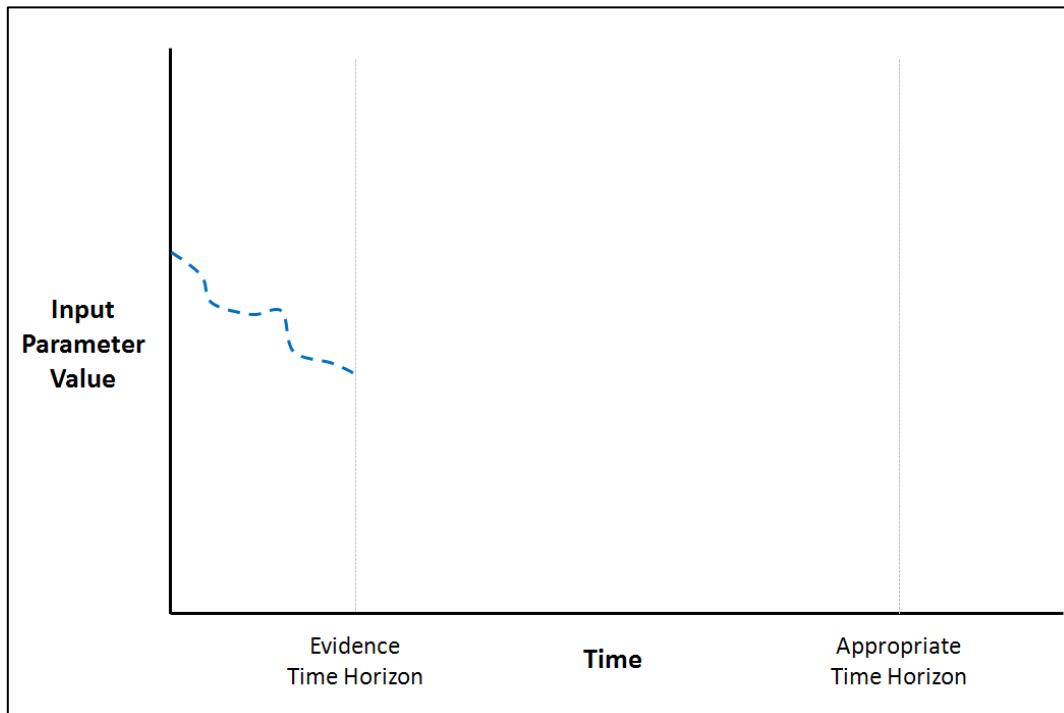
¹⁰ For now, we simply refer some generic model parameter. The particular challenges posed by different parameter ‘types’ are addressed in Chapter 3.

Given the reasons outlined in Section 2.2.1 for quantifying uncertainty in CEA, an appropriate method to do so must be employed. Although there exist the options of scenario analyses and parametric approaches to characterising posterior (decision) uncertainty (e.g. assuming net benefit is normally distributed), probabilistic sensitivity analysis has been shown to appropriately convey the combined impact on decision uncertainty of the uncertainty surrounding model input parameters in a way that allows the estimation of the *expected* cost-effectiveness and the value of obtaining further information (Claxton et al., 2005).

Probabilistic sensitivity analysis (PSA) is a non-parametric means of appraising the impact on decision making of uncertainty pertaining to a range of input parameters. PSA involves characterising model input parameters as probability distributions rather than point estimates in order to express the uncertainty around their 'true' value. Typically, monte carlo simulation is then employed to propagate the uncertainty around the parameter inputs through a decision model to express uncertainty through the distribution of cost-effectiveness results. This can be illustrated using a cost-effectiveness plane and/or a cost-effectiveness acceptability curve (CEAC) (Briggs, 2000). A PSA can also neatly facilitate the estimation of the value of further information as the components needed to estimate, for instance, the expected value of perfect information (i.e. the expected payoff with current information and the expected payoff with perfect information) can naturally be calculated using the output of a PSA. The use of PSA is recommended by NICE and by ISPOR guidance (Briggs et al., 2012, NICE, 2013)

To facilitate the discussion around how temporal uncertainty in particular might be addressed, a simple hypothetical example is constructed. Let us assume that through perfect short-term evidence, we know the value of a model parameter p between $t=0$ and $t=e$ (the evidence time horizon). Let us assume, initially, that there is no further evidence and therefore the value of p between $t=e$ and $t=a$ (the appropriate analysis time horizon) is unknown.

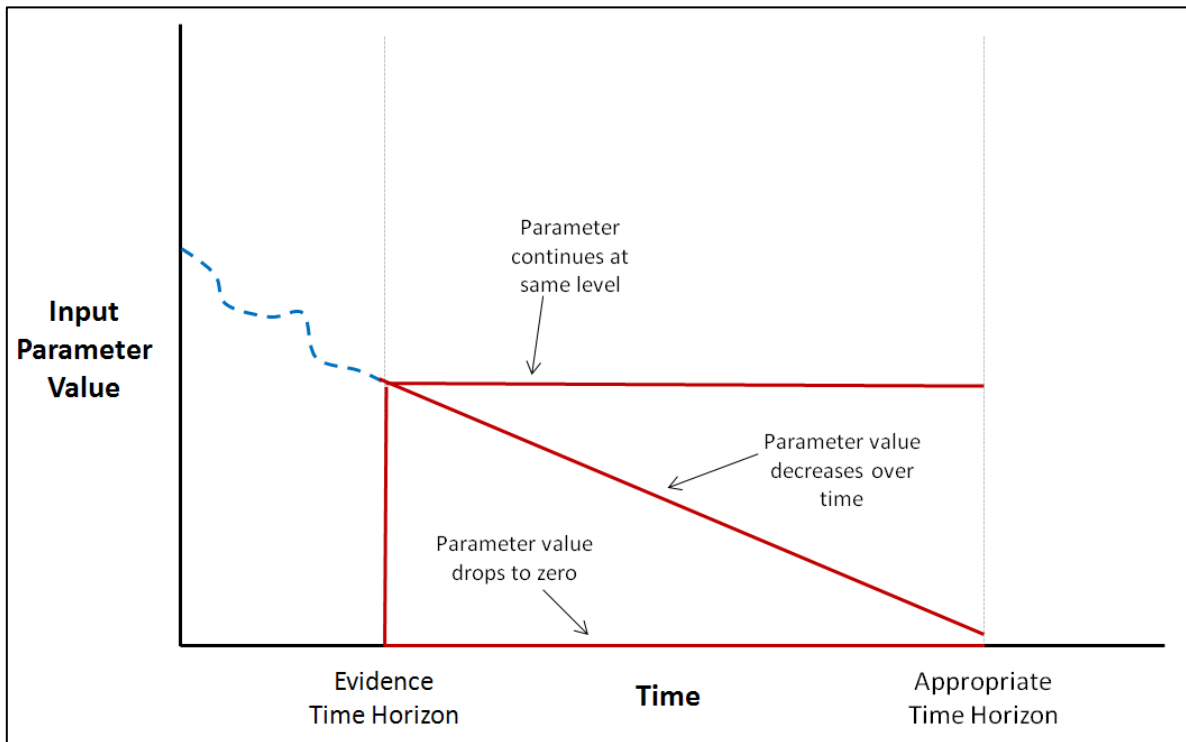
Figure 10: The stylised example that will be used throughout Section 3 to explore how temporal uncertainty might be expressed in CEDMs



2.3.2 Scenario Analysis

It is possible to explore alternative temporal assumptions in a deterministic sensitivity (i.e. simple scenario) analysis. This is akin to considering temporal uncertainty strictly as a problem of structural uncertainty. It has been recommended that structural uncertainties can be suitably characterised by presenting a number of alternative scenarios, with their related cost-effectiveness results, to the decision maker (Weinstein et al., 2003). In relation to long-term treatment effect for example, the current NICE guidelines suggest the consideration and presentation of scenarios where (i) the treatment effect over the unobserved period is nil; (ii) the treatment effect continues at the same level as during the observed period; and (iii) the treatment effect diminishes over time (NICE, 2013). Such a scenario analysis is illustrated in Figure 11.

Figure 11: Scenario analysis. Three alternative scenarios (in particular the scenarios that NICE suggest should be explored regarding long-term treatment effect) are implemented in the CEDM



Although a scenario analysis could convey to decision makers the temporal uncertainty existent in a model, there are some notable problems with this approach.

- (i) There will be uncertainty regarding whether the scenarios chosen represent the truly plausible outcomes and thus convey the temporal uncertainty
- (ii) Its use in CEA implies that the decision maker must carry out some informal weighting of the alternative scenarios in order to make a decision based on expected cost-effectiveness
- (iii) Scenario analysis precludes any estimation of the value of obtaining further information

The crucial limitation, in essence, is that scenario analysis does not truly quantify temporal uncertainty. It has been well documented that uncertainties ought to be quantified in a probabilistic sensitivity analysis (PSA) where possible (Claxton et al., 2005). Scenario analysis is useful for determining the impact of temporal uncertainty on decision-making (as outlined in the previous section), but if the adoption recommendation is found to be sensitive to temporal assumptions, it is

imperative to quantify temporal uncertainty within the a PSA in order to (i) estimate our true expectation regarding cost-effectiveness and (ii) estimate the value of obtaining further evidence before making an adoption recommendation. Given the existence of alternative temporal scenarios, this can be done using model averaging if there exists a means to weight alternative scenarios (discussed further in Section 2.3.4). Where there is no evidence with which to weight scenarios (like in this stylised example) it is not clear how to incorporate the scenarios into the PSA. Jackson et al. suggest (in relation to structural uncertainty) that it is reasonable to employ simple scenario analysis where there are no data available to weight alternative scenarios (Jackson et al., 2011).

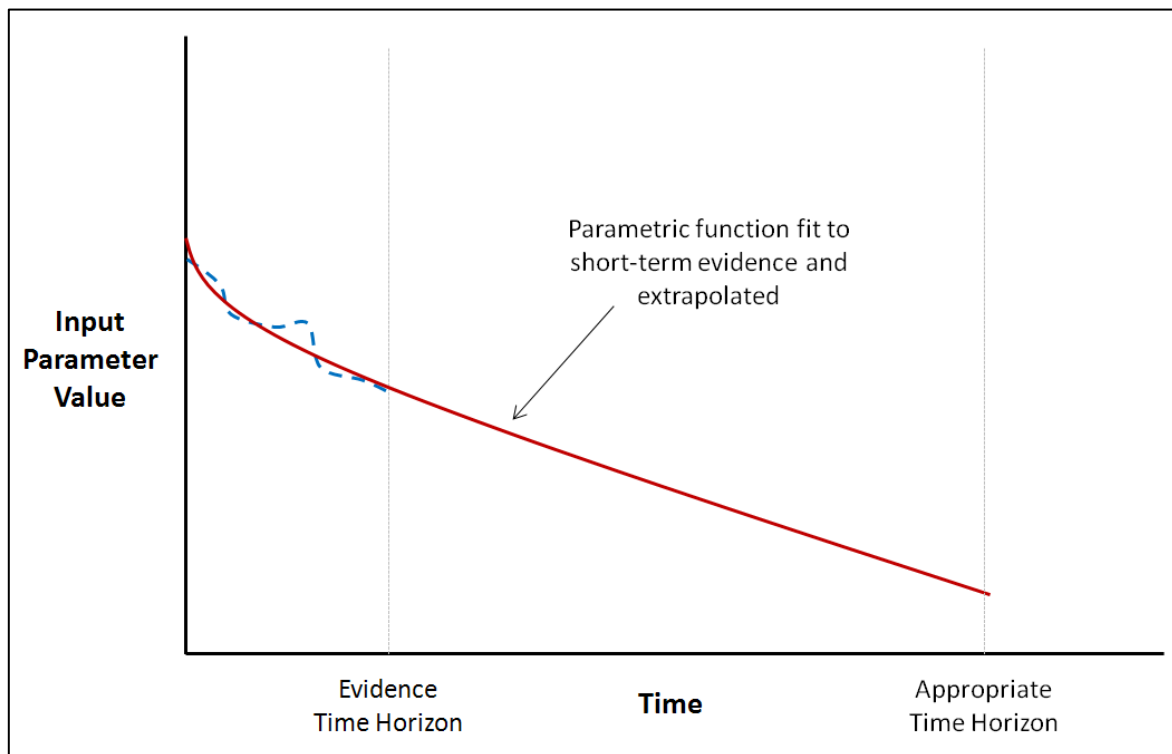
Instead of positing alternative temporal scenarios, an alternative approach would be to consider alternative interpretations of the short-term evidence in order to extrapolate and obtain the inferred parameter behaviour over the long-term.

2.3.3 Extrapolation from Short-term Evidence

Often in CEDMs, parametric functions are fit to short-term evidence. There may be several reasons for this: to facilitate sub-group analysis, to estimate a ‘true’ set of values for the parameter (assuming that the sample short-term evidence is a partial reflection of the true parameter), and to extrapolate the parameter over time (i.e. assume the ‘trend’ observed in the short-term continues into the long-term). Here, we have assumed that our short-term evidence is perfect. Let us also assume that there is no need for sub-group analysis (in fact let us assume that the cohort is homogenous). We can fit a parametric function to the short-term evidence purely for the purposes of characterising the unobserved period by extrapolating. We can choose the parametric function based on the best ‘statistical fit’ to the short-term evidence (e.g. AIC, BIC¹¹). Extrapolation using a parametric function fitted to the short-term evidence is illustrated in Figure 12.

¹¹ The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are measures of the relative quality of a statistical model for a given dataset.

Figure 12: The ‘best fitting’ parametric function is extrapolated over the unobserved period



The values for p over the unobserved period are generated as follows:

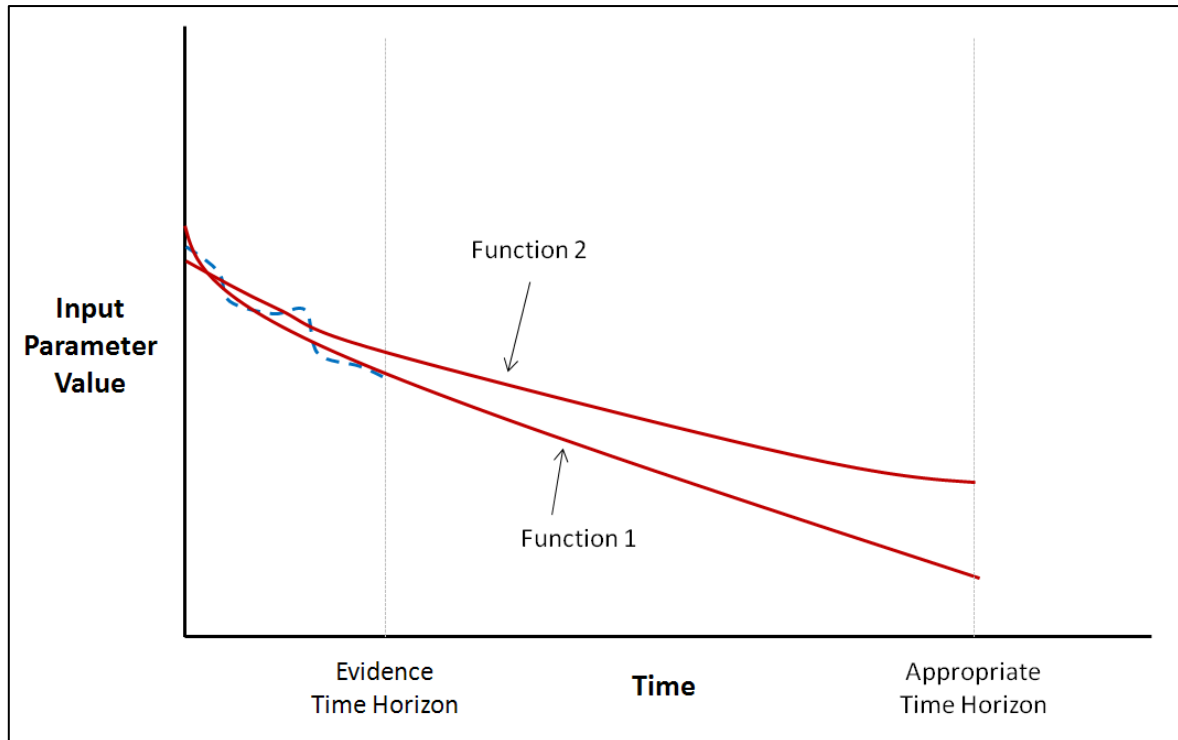
$$\text{For } e < t \leq a, \quad p_t = f(t)$$

i.e. the value of p (a parameter) over the unobserved period is some function of t (=time), that function having been generated from analysis of the short-term evidence. ‘ a ’ is the appropriate time horizon and ‘ e ’ is the evidence time horizon.

It may seem reasonable to consider the extrapolated curve as the expected temporal trajectory of the parameter as this function best represents the short-term evidence. Just using this extrapolation however, would imply in the CEDM that we know with certainty that the parameter will take these values over the unobserved period. Uncertainty regarding the extrapolation could be expressed in this situation by considering multiple functional fits to the short-term evidence (again, based on the ‘best’ statistical fits to the data). These functions would represent alternative interpretations of the short-term evidence and its implied trajectory over the long-term. The validity of extrapolated curves could be considered; for example, those curves that imply illogical or implausible values over the long-term might be dismissed, leaving the ‘valid’ curves. It is possible to incorporate multiple

‘valid’ extrapolations into the main analysis by using model averaging, where the weights associated with each alternative extrapolation would be derived from measures of statistical fit to the short-term evidence (i.e. AIC, BIC) (Jackson C et al., 2009). The approach of using multiple extrapolations and model averaging is illustrated in Figure 13.

Figure 13: Two alternative ‘fits’ to the short-term data are extrapolated over the unobserved period



The values for p over the unobserved period would (through model averaging), be generated as follows (as part of a Monte Carlo Simulation):

$$\text{For } e < t \leq a, \quad p_t = \begin{cases} f(t), & \text{with probability } q \\ g(t), & \text{with probability } (1 - q) \end{cases}$$

This approach produces an expected temporal trajectory and would seem to characterise the related uncertainty. But are the assumptions being imposed appropriate? And is the uncertainty quantified truly a representation of temporal uncertainty?

Crucially, the assumption implicitly imposed when extrapolating evidence over time is that the short-term evidence tells us everything about the long-term. Effectively, the uncertainty quantified by

incorporating multiple parametric functions represents the uncertainty regarding how the short-term evidence should be interpreted, rather than the uncertainty regarding how the parameter evolves after the observed period. There are further questions associated with this approach, such as: What about the space in between or around the extrapolated curves? How many functions should be included? What does this imply about expected values?

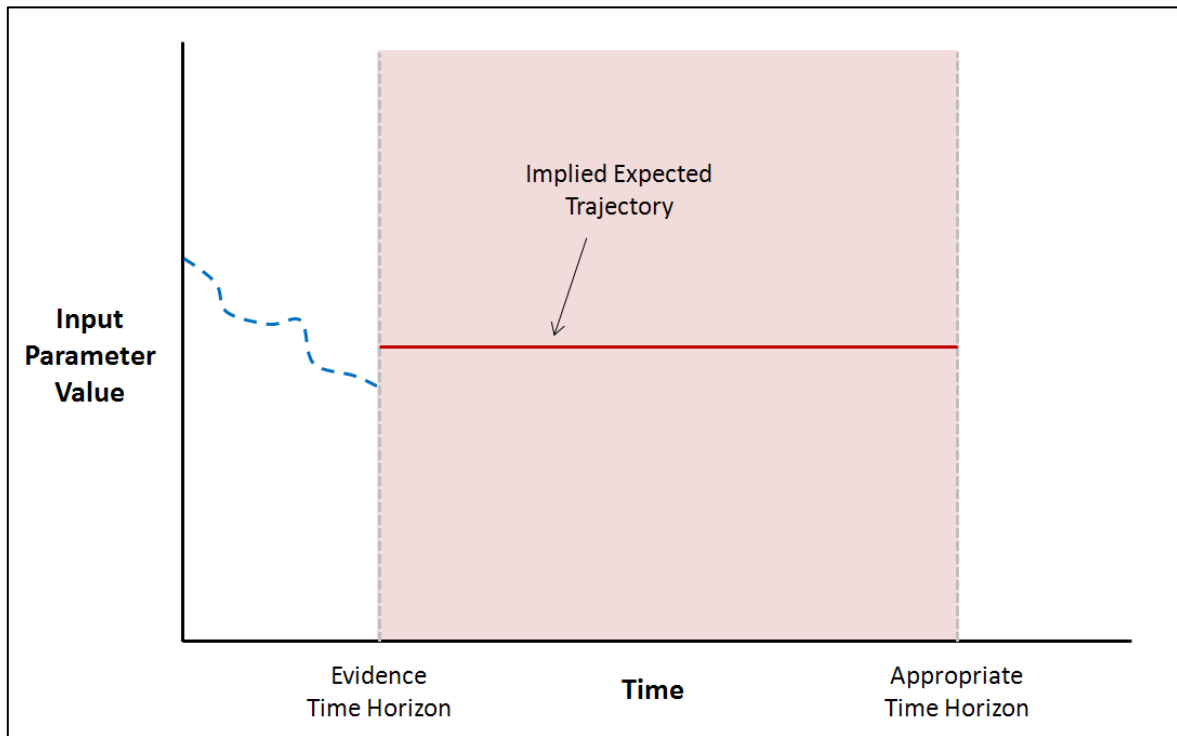
The key concern with this approach is that it does not acknowledge the dearth of evidence pertaining to the long-term. What ought to be expressed is the possibility that something ‘different’ (not captured by the short-term evidence) might happen over the unobserved period. If we do not possess evidence to suggest otherwise, is it most appropriate (or indeed possible) to convey total uncertainty regarding values over the unobserved period?

2.3.4 Assuming No Knowledge of The Unobserved Period

Let us take our assumptions to the other extreme (to those implied by extrapolation); we do not possess any evidence pertaining to the unobserved period, therefore we ought not to make any assumptions regarding parameter values over that period. Is it possible and/or desirable to express this in a CEDM?

A uniform distribution could be applied, where the parameter can take any logical value (assuming the parameter in question cannot by definition fall below a certain value and above another). This approach is illustrated in Figure 14, where the shaded (pink) area represents where on the graph p may take a value over the unobserved period.

Figure 14: A uniform distribution is applied in an attempt to convey total uncertainty over the unobserved period



The values for p over the unobserved period would be generated as follows:

$$\text{For } e < t \leq a, \quad p_t \sim U(0, b)$$

where 0 = the logical lowerbound, b = the logical upperbound and $U()$ = Uniform Distribution

Although this approach manages to express a great amount of uncertainty (that arguably reflects our true ignorance regarding the unobserved period), it has some problems.

- (i) A Uniform Distribution does not truly convey total uncertainty; instead of expressing no expected value, this approach implies that we expect p to be equal to exactly $b/2$.
- (ii) It is not clear whether we ought to impose linear trajectories or completely random trajectories (i.e. for every t , we could redraw from the Uniform Distribution).
- (iii) A Uniform Distribution clearly becomes more problematic when the logical range is infinite or semi-infinite.

In short, it does not appear feasible to produce an appropriate analysis given absolutely no knowledge of the long-term.

Sections 2.3.2 and 2.3.3 represent the alternative extreme assumptions we could take regarding how to treat the unobserved period. Problems have been highlighted with both approaches. It is suggested, at this point, that two realistic assumptions can be made that allow for a more pragmatic 'third way'.

(i) The long-term trend is *partially* explained by the short-term trend

Extrapolation and assuming no knowledge of unobserved period represent the extreme assumptions that could be made regarding the relationship between the temporal trend of the short-term and that of the long-term. In reality, neither assumption is likely to be appropriate (i.e. the short-term evidence neither tells us everything, nor tells us nothing). In most cases, the short-term trend of a model parameter could be assumed to *partially* relate to the long-term trend. Therefore, given the absence of any direct evidence pertaining to the long-term, it is reasonable to assume that the short-term evidence can to *some* extent predict long-term values. However, it will still be necessary to convey temporal uncertainty in order to account for what the short-term evidence does not predict. In some cases, it may be reasonable to allow the *expected* long-term temporal trajectory of a parameter to follow the extrapolated short-term trajectory. Although again, uncertainty ought to be expressed.

(ii) There is some knowledge relating to the long-term

It has been so far assumed that we have no information pertaining to the long-term. In practice, even without trial data there will often be some knowledge of what is plausible regarding parameter values or parameter behaviour over the long-term¹². At this stage, therefore, we will relax the assumption of having absolutely no knowledge of the unobserved period.

The following two sections (2.3.4 & 2.3.5) explore possible methods of expressing temporal uncertainty assuming there exists some information regarding what is plausible over the long-term.

¹² It is important to note the difference between 'logical' and 'plausible' in this context. Logical refers to the values that a parameter could possibly (by definition) take, whereas plausible refers to the values that a parameter could take according to some empirical information.

Essentially, these methods aim to acknowledge the ‘plausible space’ over the unobserved period, whilst conveying a reasonable expected temporal trajectory (possibly based on the extrapolated short-term evidence). The appropriate method will depend on the nature of any supplementary information.

2.3.5 Discrete Scenarios & Model Averaging

It is possible that the available information regarding plausible parameter behaviour over the long-term is such that there is a discrete number of temporal trajectories that the parameter could be expected to take. In this circumstance, it is suitable to characterise temporal uncertainty by incorporating each temporal scenario into the probabilistic sensitivity analysis (PSA).

It was stated in Section 2.3.1 that were temporal uncertainty to exist in this form (discrete alternative scenarios), then model averaging could be used to incorporate the scenarios into the PSA. Bayesian model averaging, in short, involves deciding upon a set of valid alternative models, ascribing relative weights to each model, and averaging across each model to produce a posterior predictive distribution that represents the expected model outcome (plus the related uncertainty if implemented probabilistically) (Leamer, 1978). To quantify the uncertainty, model averaging must be undertaken for each Monte Carlo Simulation carried out as part of the PSA. The key challenge in this process in terms of its application to temporal uncertainty is the generation of the model (or scenario) weights. The use of Bayesian model averaging in cost-effectiveness decision models as a means of characterising structural uncertainties has been outlined by Jackson et al. (Jackson C et al., 2009). They describe how model weights can be derived from measures of ‘goodness of fit’ to the available data (e.g. AIC or BIC). Although temporal uncertainty could be thought of as a type of structural uncertainty (and is often treated as such), the problem for the use of model averaging for addressing temporal uncertainty is clearly the dearth of long-term data with which to ascribe weights to temporal scenarios. Possible methods of ascribing weights to alternative temporal scenarios include:

- (i) Employing external data. If relevant external data are available, these can be employed to validate and judge the likelihood of alternative temporal scenarios.

- (ii) Using measures of statistical fit to short-term data. This would be akin to taking the approach described in Section 2.3.2 where it is assumed the short-term data perfectly inform the long-term trend.

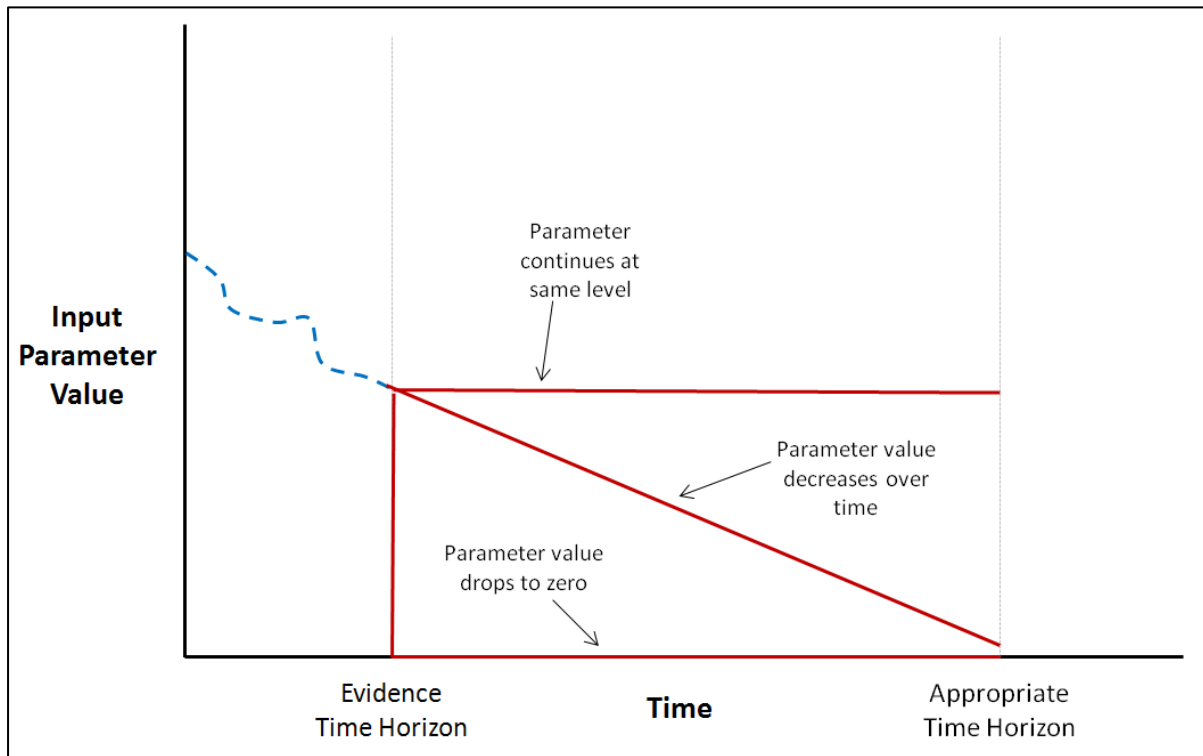
- (iii) Using expert elicitation to directly ascribe weights to the scenarios. This would be an attractive option in the absence of any other means to generate weights. Expert elicitation is discussed further in Sections 3.3.6.3 and 4.3.3.4 .

- (iv) Considering the alternative scenarios as equally likely. This approach is befitting only if there is no other obvious way of weighting scenarios.

The weights generated would then need to be transformed into probabilities such that the sum of probabilities equals 1.

Let us assume that in our stylised example, it is known that after the observed period the parameter could: (i) remain at the same value, (ii) immediately drop to zero, or (iii) continue to decline over time as per the extrapolated curve. This is illustrated in Figure 15.

Figure 15: Three possible scenarios incorporated in the probabilistic sensitivity analysis using model averaging



The values for p over the unobserved period (in the example depicted in Figure 15) using this model averaging approach would be generated as follows:

$$\text{For } e < t \leq r, \quad p_t = \begin{cases} p_e, & \text{with probability } q \\ 0, & \text{with probability } r \\ f(t), & \text{with probability } (1 - q - r) \end{cases}$$

Where p_e = the value of the parameter at the evidence time horizon and $f(t)$ = the extrapolated function that best characterised the short-term evidence.

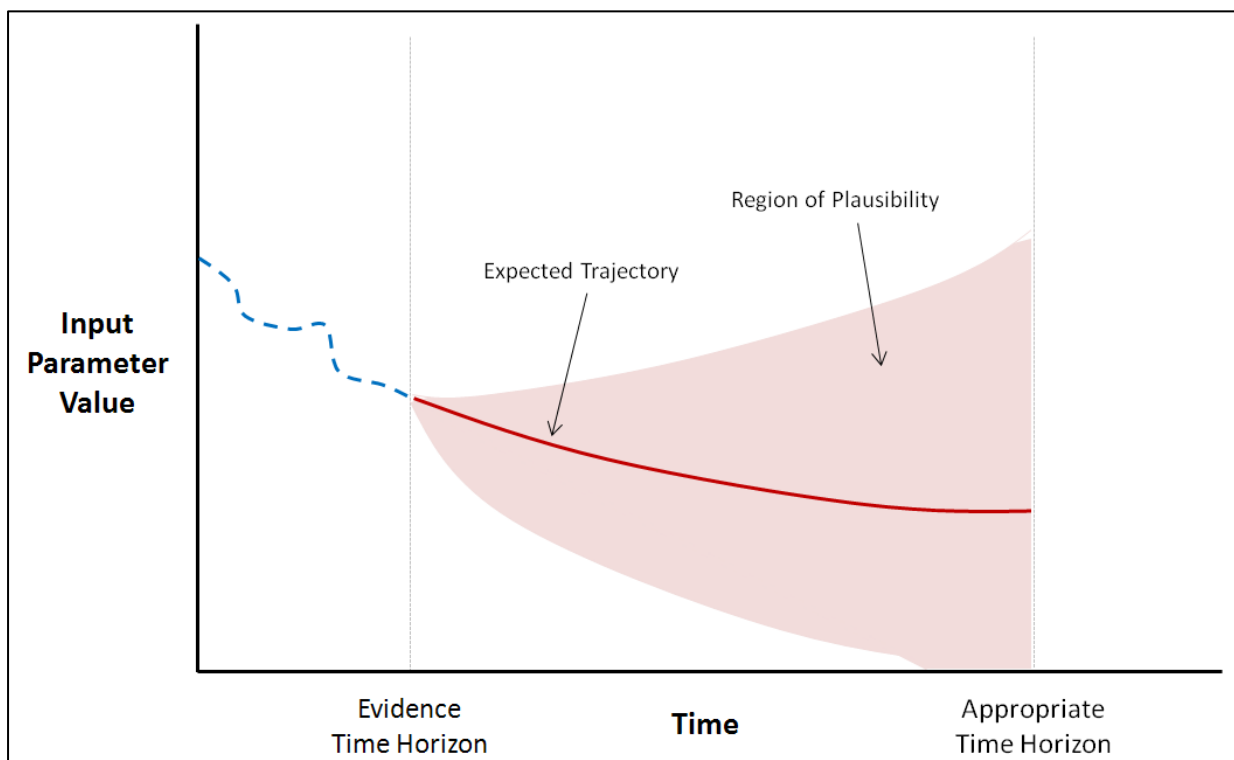
2.3.6 Continuous Parameterisation

Section 2.3.4 described a means of expressing temporal uncertainty when there exists a discrete number of broad competing scenarios. However, the temporal uncertainty in question might be better characterised by a continuous distribution around some expectation, implying an arbitrarily large number of temporal possibilities, i.e. it is deemed that there is an *area of plausibility* over the

unobserved period, rather than a number of plausible scenarios. For example, in the absence of any further direct evidence, we may wish to use an extrapolated function as our expected curve but we may want to augment this function to allow a range of curves to be possible in the PSA, thereby reflecting what is plausible over the unobserved period as well as what is currently expected.

It has been suggested that all model uncertainties can potentially be expressed as parameters (Jackson et al., 2011). We can apply this principle to temporal uncertainty. Model averaging as described above effectively parameterises the temporal uncertainty, but we could go further by explicitly introducing a ‘temporal parameter’ that dictates how the underlying parameter behaves over time. This temporal parameter can then be made probabilistic to convey temporal uncertainty in order to cover the plausible space over the unobserved period. The temporal uncertainty can then be treated essentially as another source of input parameter uncertainty (Claxton, 2008). This approach is illustrated in Figure 16.

Figure 16: Temporal uncertainty is fully parameterised so that the long-term curve may take a range of plausible trajectories



The values for p over the unobserved period using the parameterisation approach would be generated from some function of t as follows:

$$\text{For } e < t \leq r, \quad p_t = f(t)$$

$f(t)$ here is an augmented version of the function that represents the expected temporal trajectory (for example an extrapolated curve). To convey temporal uncertainty, the coefficient and/or power of t is made probabilistic. This ‘temporal parameter’ is drawn from a suitable distribution where its expected value is such that the function $f(t)$ collapses to that of the expected trajectory and its confidence interval is such that the sampled curves cover the plausible area.

2.4 The Value of Reducing Temporal Uncertainty

Expressing temporal uncertainty in the CEDM (as well as all other sources of uncertainty), facilitates calculating an estimate of the value of obtaining further information. In particular, the expected value of perfect information (EVPI) can be calculated easily once all sources of uncertainty are expressed in a probabilistic sensitivity analysis (PSA) by making use of the net benefit (NB) metric (Sculpher and Claxton, 2005). The non-parametric approach to calculating EVPI can be expressed in the following equation.

$$EVPI = E_{\theta} \max_j NB(j, \theta) - \max_j E_{\theta} NB(j, \theta)$$

For j alternative interventions with unknown parameters θ .

EVPI, as expressed in the equation above, represents the difference between net benefit (averaged over the numerous alternative realisations of costs and effect) were perfect information to be available and net benefit given currently available information (i.e. maximum of the net benefits associated with the alternative interventions).

This concept can be extended to calculate the expected value of perfect information for particular parameters (EVPPI), as expressed in the following equation.

$$EVPPI_{\varphi} = E_{\varphi} \max_j E_{\psi|\varphi} NB(j, \varphi, \psi) - \max_j E_{\theta} NB(j, \theta)$$

Where φ is the uncertain parameter of interest and ψ are the remaining uncertain parameters, i.e. $\varphi \cup \psi = \theta$.

When temporal uncertainty is expressed in a CEDM using the methods outlined in Sections 2.3.4 and 2.3.5 (i.e. model averaging or parameterisation), an uncertain parameter is effectively generated that represents the temporal trajectory of the underlying parameter of interest. When this ‘temporal parameter’ is incorporated into the PSA, it can be analysed just like any other uncertain parameter. Thus, the non-parametric method of calculating EVPPI can be employed to estimate the value of reducing the uncertainty surrounding the temporal trajectory, over the unobserved period, of a particular parameter of interest.

If, after a value of information (and specifically EVPPI) analysis, it is deemed that more evidence is required regarding the temporal nature of parameters over the unobserved period, then consideration must be given to the type of evidence that could feasibly be obtained to aid the adoption recommendation. To estimate the degree of evidence required, the expected value of sample information (EVSI) could be calculated. EVSI expresses the additional net benefits to be gained by obtaining further evidence of a specified sample size, targeting a particular uncertainty. In doing so, EVSI can help evaluate alternative research designs (Briggs et al., 2006) (McKenna, 2011). Acquiring further evidence related to the long-term temporal behaviour of a model parameter is likely to be naturally difficult. Commissioning a further RCT is unlikely to be helpful as it would involve a long wait to obtain any long-term evidence. A search for further observational evidence may prove fruitful, but may still be limited in terms of long-term outcomes. In these circumstances, eliciting the opinions of experts and quantifying them for use in the CEDM may be the most efficient course of action.

However, as discussed in Section 2.2.2.2, the evidence required to address temporal uncertainty might only be obtainable through waiting. Typically, EVPI is measured against the cost of obtaining further evidence as this is a necessary condition for further research. For circumstances where the necessary evidence can only be generated by waiting and where there are irrecoverable costs associated with a positive adoption decision, how to calculate whether the expected net benefit of approving a new technology immediately will be worth the risk of having to reverse that decision when further information is revealed?

Let's say there are the following two options: (i) immediate approval and (ii) waiting for further information. If it is assumed that sufficient further evidence will be revealed after 3 years of trial follow-up, then the expected total net benefit associated with these two scenarios are as follows:

$$E(NB(i)) = \underbrace{\sum_{t=0}^F \{max_j NB(j, \theta). (1 + d)^{-t}\}}_{\text{Expected net benefit up to point that further evidence is revealed}} + \underbrace{\sum_{t=F+1}^T \{E_{\varphi} max_j E_{\psi|\varphi} NB(j, \varphi, \psi). (1 + d)^{-t}\}}_{\text{Expected net benefit after the point that further evidence is revealed}} - \underbrace{IC}_{\text{Expected investment costs incurred}}$$

$$E(NB(ii)) = \sum_{t=0}^F \{NB(CTP, \theta). (1 + d)^{-t}\} + \sum_{t=F+1}^T \{E_{\varphi} max_j E_{\psi|\varphi} NB(j, \varphi, \psi). (1 + d)^{-t}\} - IC. \lambda. (1 + d)^{-F}$$

Where:

T = time after which evidence is no longer relevant

F = time at which further evidence is revealed

d = discount rate

CTP = current treatment practice

IC = (irrecoverable) investment costs

λ = the probability that the addition of further evidence will lead to an adoption recommendation = the proportion of times that $E_{\psi|\varphi} NB(j, \varphi, \psi)$ is greater for the new technology

The two equations above represent the net benefit expected to be gained for each scenario over the relevant lifespan of this decision problem/evidence. It is assumed that after further evidence is revealed at point F , the 'correct' decision will then be implemented. Thus the second terms in each equation above are identical and can be assumed to offset each other. What is being assessed is whether the higher net benefit being gained by immediately approving the new technology plus the investment costs incurred, outweigh the lower net benefit associated with the current treatment

plus the expected investment costs incurred (given that no investment costs may ultimately be incurred).

The 'value of waiting' could thus be expressed as: $E(NB(ii)) - E(NB(i))$.

2.5 Conclusions

The discussions and arguments in this chapter can be summarised as follows:

The problem arising when the required time horizon for CEA exceeds the evidence time horizon could and should be thought of as a problem of uncertainty.

The complexity of the modelling necessary to characterise this temporal uncertainty varies by circumstance, but in many cases, fully quantifying temporal uncertainty will be crucial if the CEA is to appropriately produce recommendations regarding technology adoption and obtaining further evidence.

Given short-term evidence, it could be assumed that (a) the short-term evidence fully explains long-term behaviour (in which extrapolation is appropriate), or (b) the short-term evidence conveys nothing of long-term behaviour (in which case a model parameter could take any logical value over the long-term). However:

- (i) It may be reasonable and realistic to assume that the short-term evidence *partly* explains the long-term, implying that extrapolation may be reasonable in order to act as an expected temporal trajectory over the long-term, but a degree of temporal uncertainty is necessary to convey the lack of direct evidence pertaining to the long-term.
- (ii) Something of what is plausible regarding long-term values is typically known: model averaging or parameterisation of temporal uncertainty may be useful tools to express expectations while allowing for other plausible temporal behaviour.

It is useful to incorporate temporal uncertainty into the value of information framework. By calculating EVPPI on one or more uncertain temporal parameters, the value of obtaining further

information specifically on the temporal behaviour of model parameters can be calculated. For the circumstance where irrecoverable costs are potentially incurred with the approval of a new technology, the framework can be extended to ascertain whether waiting for evidence that would address temporal uncertainty to be revealed through, for example, further trial follow-up would be worthwhile.

This chapter has considered the meaning and significance of temporal uncertainty and the appropriateness of alternative assumptions regarding model parameters that could be made in the face of temporal uncertainty. It also outlined hypothetical methods of expressing temporal uncertainty in CEDMs. However, the stylised examples employed pertain to a simplified world. In reality, there are numerous factors to take into account when considering how to address temporal uncertainty. The following chapter outlines the various practical issues that require consideration when addressing temporal uncertainty in cost-effectiveness decision models.

3. CHAPTER 3: KEY ANALYTICAL ISSUES IN TEMPOAL UNCERTAINTY AND REVIEW OF METHODS EMPLOYED TO DATE

3.1 Introduction

The previous chapter described the meaning and impact of temporal uncertainty in cost-effectiveness analysis (CEA). In order to move from the conceptual to the practical, this chapter outlines and discusses the key analytical issues relevant to addressing temporal uncertainty through decision modelling, with a view to highlighting the specific areas where further thought and methodological development would be most valuable. In particular, this chapter will firstly provide an overview of current methodological guidance relating to temporal uncertainty across regions, as well as a summary of recent relevant reviews and analyses. Secondly, a discussion of key analytical issues is undertaken, with particular focus on the challenge of expressing temporal uncertainty for different types of model parameter. This discussion is accompanied by a thorough review of health technology assessments (HTAs) conducted in the UK over a six year period in order to surmise and appraise the methods employed to-date. Finally, some thoughts and criticisms are offered regarding the adequacy of current methodology and the requirement for improved methodology and updated guidance.

3.2 Current Guidance and Other Relevant Reviews

3.2.1 Summary of Current Methodological Guidance

It is desirable to optimise and standardise the methods by which expected cost-effectiveness and the value of further information are calculated and the results communicated to decision makers. To these ends, guidelines are issued in a number of countries/regions. Specific guidance on the quantification of temporal uncertainty in CEA has to date been limited. Table 1 below summarises the guidance relating to temporal uncertainty in 12 countries whose HTA processes are well developed. The selection of countries is based on a comparison of guidelines conducted by Mauskopf et al. (Mauskopf et al., 2011).

Table 1: Summary of guidance relating to temporal uncertainty in 12 countries whose HTA processes are well developed

Country	Type of Guidance	Guidance Regarding Temporal Uncertainty
Australia	Formal national guidelines	<i>“The length of follow-up of participants in the trial might be less than the expected duration of treatment or expected duration of health impacts overall. Results generated in this way need to be extrapolated to the proposed duration of treatment or expected health impacts.”</i> (Pharmaceutical Benefits Advisory Committee, 2008)
Canada	Formal national guidelines	<i>“Describe the strength of the evidence for extrapolating data and assess uncertainty through a sensitivity analysis ... explain the causal relationships that are used to extrapolate” ... “Unless such an analysis is based on high quality evidence, identify it as speculative, and give appropriate caveats in the report”</i> (Canadian Agency for Drugs and Technologies in Health, 2006)
France	Informal guidelines/expert consensus	<i>“It is recommended that the time frame chosen should be long enough that all outcomes, both positive and negative, of the treatments used and evaluated be included in the study” ... “Modelling makes it possible to extrapolate clinical and economic results beyond the time horizon of the clinical Trial”</i> (College des Economistes de la Sante, 2004)
Germany	Formal national guidelines	<i>“In principle, the health economic evaluation should cover the duration of the randomized controlled trials and, as a secondary scenario, be extended beyond this time period if this is relevant for the decision maker” ... “Modelling can be carried out for the time period for which evidence on benefit and harm from clinical studies exist. In a second step, health technologies can be modelled over longer periods of time”</i> (Institute for Quality and Efficiency in Health Care (IQWiG), 2009)
Ireland	Formal national guidelines	<i>“When extrapolating beyond the period of clinical trials... inherent assumptions regarding future treatment effects and</i>

		<i>disease progression should be clearly outlined and tested as part of the sensitivity analysis” (Health Information and Quality Authority, 2010).</i>
Italy	Informal guidelines/expert consensus	There was no guidance related to temporal uncertainty found. (The Members of the Italian Group for Pharmacoeconomic Studies, 2001)
The Netherlands	Formal national guidelines	<i>“Use of modelling... to study and analyse effects and costs during a longer time horizon than that of the clinical studies.”... “Parameter values for extrapolation are obtained from clinical studies.” (College voor zorgverzekeringen Diemen, 2006)</i>
Scotland	Formal national guidelines	<i>“All structural assumptions and data inputs should be clearly documented and justified. This is particularly important in the case of modelling to extrapolate costs and health benefits over an extended time horizon. In such circumstances the results of using alternative time horizon scenarios should be reported in order to compare the implications of different assumptions for the results.” ... “Use sensitivity analysis... where there is uncertainty about the most appropriate assumption to use for extrapolation of costs and outcomes beyond trial follow-up.” (Scottish Medicines Consortium, 2007)</i>
Spain	Informal guidelines/expert consensus	<i>“Modelling techniques should be developed in different situations to extrapolate progression of the clinical outcome (i.e., survival) beyond that observed in a trial.” (Lo’pez-Bastida, 2013)</i>
Sweden	Formal national guidelines	<i>“Extrapolation must be carried out for the period outside the accessed data from clinical trials. This is then done via modelling.” (The Pharmaceutical Benefits Board, 2003)</i>
UK (England and Wales)	Formal national guidelines	<i>“When the impact of treatment beyond the results of the clinical trials is uncertain, analyses that compare several alternative scenarios reflecting different assumptions about future treatment effects should be presented. Such assumptions should include the limiting assumption of no</i>

		<i>further benefit as well as more optimistic assumptions... Assumptions used to extrapolate treatment effects should have clinical validity, be reported transparently and be clearly justified" (NICE, 2013)</i>
United States	Formal national guidelines	<i>"Modelling may be required (a) to extrapolate the progression of clinical outcomes (such as survival) beyond that observed in a trial." ... "Modelling may be required (a) to extrapolate the progression of clinical outcomes (such as survival) beyond that observed in a trial." (Academy of Managed Care Pharmacy, 2012)</i>

As Table 1 shows, most guidance documents provide some, if brief, direction regarding how to handle temporal uncertainty. Those that do, give their focus to the extrapolation of clinical trial evidence over time and where the related uncertainty is considered, a scenario analysis is typically recommended. Although this guidance is not unreasonable at a high-level, its brevity and vagueness may lead to inconsistencies in how temporal uncertainty is characterised in CEA. In this sense, supplementary guidance would be valuable regarding the modelling specifics of this issue (e.g. methods to incorporate competing long-term scenarios into the analysis and how this may differ for different parameters and different clinical contexts). All guidance documents acknowledge the inherent difficulty characterising long-term costs and effects, advocating scenario analysis as the best way to convey temporal uncertainty. Ideally however, all sources of uncertainty would be incorporated into a probabilistic sensitivity analysis. These guidelines therefore highlight the need for methods that allow temporal uncertainties to be expressed in a fully probabilistic model.

3.2.2 Recent Relevant Reviews and Analysis

A number of studies exploring and reviewing issues related to temporal uncertainty have been published in recent years.

Latimer

Latimer recently offered a guide to extrapolating patient-level data in economic evaluation using survival analysis (Latimer, 2013). His review of NICE HTAs dealing with advanced/metastatic cancer

revealed that parametric modelling was the most frequent means of extrapolation of survival outcomes (72% of studies), most analyses using ‘standard’ distributions and many not sufficiently justifying the chosen method of extrapolation. A model selection algorithm was recommended for future HTAs.

Bagust and Beale

In response to Latimer, Bagust and Beale offered a number of criticisms of the proposed model selection algorithm (Bagust and Beale, 2014). In particular, they cited a lack of access to patient-level data for many HTA researchers and questioned the appropriateness of proportional hazards¹³ modelling and the use of AIC/BIC scores. Most interestingly with regard to temporal uncertainty, they expressed concern with the method of “fitting” a statistical function to short-term data and projecting this into the long-term, outlining a number of reasons why extrapolated RCT data may badly predict long-term outcomes and ultimately arguing against the “primacy accorded to a small set of theoretical distributions” .

Guyot et al.

Guyot et al. also carried out a review of analyses within the UK HTA programme (Guyot et al., 2011). The authors asserted that estimates of efficacy and cost-effectiveness ought to be based on the same statistical analysis of the available RCT data, but found in their review that in no case was the statistical model for efficacy and CEA the same¹⁴. The authors also found that the proportional hazards assumptions was frequently employed but seldom formally justified. Finally (and importantly for this research) they found that the uncertainty in survival model choice was rarely addressed and never propagated through the cost-effectiveness model.

Davies et al.

Davies et al. compared the extrapolated survival curves of 8 year trial data to the empirical survival curves of 16 year trial data (Davies et al., 2013). They found a marked disparity between the predicted and actual curves. The hazards were approximately proportional over the first 8 years but crossed over the next 8 years, casting serious doubt on the suitability of assuming proportional hazards over the long-term.

¹³ The concept of ‘proportional hazards’ is discussed further in Section 3.4.2. In short, ‘proportional hazards’ means that the hazards of multiple groups are multiplicative for any time point t and so for example treatment groups can be characterised by hazard ratios representing treatment effects.

¹⁴ Note that in Chapter 4 of this thesis, it will be asserted that using different models to estimate baseline hazards and hazard ratios is not problematic and is in fact very useful.

Connock et al.

Connock et al. reanalysed two NICE single technology appraisals (STAs) and demonstrated that the choice of survival model to represent and extrapolate the short-term data had a significant effect on overall survival gain, and consequently expected cost-effectiveness (Connock et al., 2011). The authors highlight that without a strict and consistent process for model justification, manufacturer submissions can potentially significantly over (or under) estimate the cost-effectiveness of a health intervention.

Manca et al.

Manca et al. developed a 'wish-list' of desirable features for survival regression models when used within cost-effectiveness analysis (Manca et al., 2009). They found that flexible models like the Royston-Parmar outperform 'standard' models but concluded that beyond trial validation and uncertainty must be addressed¹⁵.

Kim and Thompson

Kim and Thompson used the example of evaluating the long-term cost-effectiveness of screening for abdominal aortic aneurysm to demonstrate the importance of validation and quantifying uncertainty in relation to extrapolation beyond the trial time horizon (Kim LG and Thompson S, 2010). They found that three alternative models when extrapolated to a lifetime horizon produced cost-effectiveness estimates ranging from £1600 to £4200 per life-year gained.

Each of these reviews/studies represents an important and useful addition to the literature regarding extrapolation/temporal uncertainty. The focus to date however, appears to have been on methods of interpreting the short-term data so as to extrapolate with justification. The focus has also been very much on survival outcomes. While efficiently exploiting the available trial data is indeed paramount and survival outcomes are frequently the key parameters, the review and discussion that takes place in this thesis chapter endeavours to give attention to the full range of model parameters that are subject to temporal uncertainty and also to methods of expressing that uncertainty as part of the analysis (as opposed to simply extrapolating).

¹⁵ 'Standard models' here broadly refers to parametric survival models that are most commonly employed in health technology assessment. Standard models are usually those that belong to the 'Generalised F' family of models, plus the Gompertz model. Royston-Parmar differs from these 'standard models' primarily in that the construct of the model is based directly on the observed data. Thus it can be flexible regarding how it fits to the nature of the empirical hazards, unlike other models which may have to be, for example, monotonically increasing/decreasing.

3.3 Key Issues When Modelling Temporal Uncertainty

Before describing the challenges of addressing temporal uncertainty for a range of model parameters, this chapter first describes a number of other key modelling issues that are pertinent to the characterisation of temporal uncertainty. From this point on, each section is supported by a review of health technology assessments (HTAs) carried out in the UK. As such, a short description of this review is first given.

As part of a project on extrapolation funded by the Medical Research Council (MRC), all health technology assessments (HTAs) from the UK HTA programme from January 2004 to October 2010 were reviewed. The reviewed HTAs included both NICE-commissioned and 'non-NICE' HTAs. The motivation for this review was the desire to understand, at a high level, the prominence and role of extrapolation in HTA. Further to this project, a more detailed review was undertaken for the benefit of this Ph.D. research. In particular, a subset of HTAs from the MRC review were selected and studied in more detail so as to document the methods employed for estimating long-term values for a range of parameter-types and the methods employed to express the related uncertainty. The title, authors and link to publication of the subset of HTAs reviewed in detail is given in Appendix 1.

3.3.1 Prevalence of Temporal Uncertainty

The MRC review analysed all NICE technology assessments (TAs) and all non-NICE reports from the UK HTA programme, with a de-novo cost-effectiveness element, published between January 2004 and October 2010. Of the 313 studies with a cost-effectiveness element, 180 (58%) were identified as featuring temporal uncertainty. A study was deemed to 'feature' temporal uncertainty if the noted time horizon of the analysis exceeded the time horizon of the primary source(s) of evidence and/or an attempt was made for any reason to extrapolate evidence over time¹⁶. A data extraction

¹⁶ To identify pertinent studies, key search terms were first used, in particular: "extrap" and "time horizon". Typically, when applied to the HTA document, these search terms quickly returned a positive or negative outcome. Where the outcome of the key term search was unclear, a 'scan read' was carried out. Seldom was a lengthy read required to ascertain the presence of temporal uncertainty.

process was then applied to each of these 180 studies where information regarding the study-specific context was recorded as well as a broad summary of the modelling approaches employed to estimate long-term parameters values. More detailed analysis (pertaining to the characterisation of temporal uncertainty) was carried out on a random sample of 64 studies (from the 180 featuring temporal uncertainty)¹⁷.

Looking at the time periods involved in the studies where there was exposure to temporal uncertainty, it was found that analysis time horizons ranged from a number of months to 100 years. In 70% of cases, the time horizon deemed appropriate was lifetime or effective lifetime. In the majority of studies, the strongest evidence used came in the form of one or more randomised controlled trials covering periods of two weeks to several years with an average of about six months. It can be assumed therefore, that very often studies are attempting to model events in an “unobserved” period many times the size of an “observed” period.

3.3.2 Clinical Context

3.3.2.1 Nature of the disease

Temporal uncertainty is likely to be pertinent to any CEA where the costs or effects attributable to alternative treatment cohorts are expected to differ over the long-term. It is tempting to think that analyses of chronic diseases are more likely to require long-term time horizons compared to acute diseases. While this may be so, it is not due to the long-term nature of the disease but the long-term differential impacts of treatment. For instance, it is perfectly feasible that alternative treatments for chronic diseases will result in different levels of health benefit for a time, before the treatment groups can be assumed to once again have the same characteristics. Conversely, it is also feasible that alternative treatments for an acute disease will result in different health benefits for the remainder of the patients’ lifetimes (certainly if there are mortality impacts). Nonetheless, the chronic/acute characteristic of a disease may aid with the quantification of temporal uncertainty. For example, if a disease is acute and any surviving patient can be assumed to revert to having the characteristics of the general population, then it will be relatively straight forward to estimate

¹⁷ A total of 64 studies for the detailed review was arrived at by employing a stopping rule broadly defined as: *Carry out detailed reviews of randomly selected HTAs (from the 180 HTAs featuring temporal uncertainty) until no further useful information is being, or is likely to be, obtained.* At 64 studies, it was deemed that a sufficient number of HTAs had been examined in detail in order to understand what approaches are being used to deal with temporal uncertainty in HTA and the proportion of time different approaches are being used.

his/her long-term survival/health benefits. Similarly, the disease area (e.g. cancer, cardiac, STD, etc.) could give an indication of how the disease could be modelled over the long-term. Recalling that a CEDM seeks to reflect long-term prognosis with and without treatment, it is clear that this must relate to an underlying biological/clinical process. The nature of the disease process then is likely to guide the development of the CEDM. For example, Tappenden et al. have documented the specific methodological issues at play when modelling cancer treatments (Tappenden et al., 2006). The disease area in question may also indicate where external sources of data could be found, how in the past such diseases have been modelled over a long-term time horizon, where to find experts that could be consulted and perhaps the extent to which data could be elicited from those experts.

Table 2: Nature of the Disease in HTAs using some form of extrapolation

Chronic/Acute	Chronic	Acute
	89%	11%

Disease Area	Cancer	Heart/Vascular	Bone/Joint	ENG	Skin	ENT	Other
	32%	13%	10%	11%	6%	6%	22%

Note: ENT = ear nose and throat, ENG = endocrine, nutritional and gastrointestinal diseases

3.3.2.2 Nature of the intervention

The expected clinical effects of a health intervention ought to give an indication as to how the related costs and QALYs could be quantified over the long-term. Factors such as whether an intervention is carried out continuously or is a once-off, whether treatment switchover or withdrawal is possible and how the intervention is carried out (e.g. drug, operation/procedure, screening) may inform the reasonableness of assumptions regarding the duration and nature of treatment effect, i.e. continued effect, ‘once-off benefit’, or a ‘rebound effect’.

The clinical characteristics of the HTAs that included some form of extrapolation are conveyed in Table 3.

Table 3: Clinical characteristics of HTAs using some form of extrapolation

Nature of Intervention	Drug	Procedure/Operation	Screening	Other
	60%	18%	10%	12%

3.3.3 Temporal Uncertainty relating to communicable diseases

Analyses of communicable diseases when model parameters interact dynamically (e.g. influenza, HIV/AIDS) pose a somewhat different set of challenges for appropriate cost-effectiveness modelling (Walker et al., 2010). In this setting, the principle sources of temporal uncertainty tend not to pertain to the trajectory over time of input parameters, but rather the temporal nature of the epidemic (i.e. the incidence and prevalence) (Pitman et al., 2012). For instance, when evaluating an influenza epidemic, evidence representing the full lifespan of the disease (a number of days) is likely to exist. However, uncertainty relating to the input parameter estimates (e.g. infection rate, recovery rate, contact rates), coupled with the non-linear feedback owing to the dynamic nature of the disease leads to temporal uncertainty regarding the predicted 'epidemic regime', i.e. a small shift in a parameter may cause a prediction of steady equilibrium to change to a prediction of widespread epidemic, with marked implications for QALYs lost and the optimal choice of intervention. Although communicable diseases present an interesting and broad array of challenges with regard to characterising temporal uncertainty, the study of communicable diseases is not within the scope of this thesis and did not feature in the HTA review.

3.3.4 Heterogeneity

Where it is required that decisions be made for a number of population 'sub-groups', it must be considered whether it is reasonable to assume that temporal assumptions or quantification of temporal uncertainty is homogeneous across sub-groups. If not, separate quantification of temporal uncertainty will have to be carried out for each sub-group and each parameter. For non-trivial models, the modelling of temporal uncertainty clearly has the potential to become cumbersome and time consuming. An initial examination of the impact of temporal uncertainty on decision making (e.g. the CINB plots outlined in Section 2.2.2) may indicate that temporal uncertainty is only of consequence for particular sub-groups.

3.3.5 Modelling Vehicles

Although the same judgements and assumptions have to be made in relation to temporal uncertainty regardless of the model structure used, the practicalities of quantifying temporal uncertainty inevitably vary by model type. The choice of model structure may well be influenced by the ability to model time. The typical model structures used in CEA are (Briggs et al., 2006):

3.3.5.1 Decision Tree

Using a decision tree involves multiplying propagating values through pathways of probabilities in order to calculate expected values for alternative options. Decision trees are simple to use and quick to construct. They have several limitations however, notably, the difficulty of incorporating a time element. They also have the potential to become complex and cumbersome if there are a lot of parameters to include. Decision trees are often used in conjunction with Markov models.

3.3.5.2 State-Transition Cohort Modelling (e.g. Markov Modelling)

The Markov model is a very common modelling approach taken in CEA. In a Markov model, complex processes are represented as sets of possible transitions (with associated transition probabilities) between health/disease states over a series of discrete time periods (although Continuous Time Markov Models can also be used). Costs and health outcomes are usually incorporated into the model as mean values ascribed to states per time period. Costs and effects are calculated by multiplying (over the full time horizon) the cost/utility values associated with each state by the time patients have spent in them and then summing across states. Markov models are frequently used for the sole purpose of extrapolation, where a treatment effect is estimated from a short-term trial and the long-term implications are estimated in the Markov model (Sculpher, 2012).

3.3.5.3 Individual Patient-level Simulation

The above types of model are generally regarded as cohort models as it is simply proportions of the patient population that they are concerned with. However, rather than 'the cohort level', it is possible and often desirable to create a model that simulates outcomes at the 'individual patient

level'. Individual patient-level simulations model the progress of individual simulated patients with heterogeneous characteristics that affect their disease progression. In discrete event simulations, each event/progression alters an individual's attributes allowing each patient a potentially unique history and set of risks (Brennan et al., 2006).

Some model structures are patently more amenable to quantifying temporal uncertainty than others, but there are difficulties associated with every model structure. For example, it is difficult to incorporate a time element into decision tree models; applying extrapolation to patient-level simulations requires strong and plentiful data; and a Markov model with many states will have many time-dependent transitions to estimate. In any model type, there are likely to be several parameters and other model components whose temporal behaviour will be uncertain.

In the HTA review, long-term outcomes were, in every case, estimated by using some form of decision modelling. The most common modelling vehicle used to account for long-term outcomes was found to be the Markov state-transition model. A Markov model was employed in 80% of HTAs that involved some extrapolation. 5% used a decision tree, 5% used a patient-level simulation and 10% used other model-types. The state-transition model structure (and Markov models in particular) will be the principal model structure in mind as temporal uncertainty relating to different model parameter-types is explored in Chapter 4.

Table 4: Modelling vehicles employed in HTA

Modelling Vehicle	Decision Tree	Markov Model	Patient-level Simulation	Other
	5%	80%	5%	10%

3.3.6 Nature of Available evidence

3.3.6.1 Trial Evidence

A typical economic evaluation is trial-based, whether a pure 'within-trial' analysis, or a 'trial plus model' analysis. The nature of the available trial data therefore can have a considerable bearing on the reliability of predicted long-term outcomes as short-term data are extrapolated into the long-

term. As the core source of evidence in an analysis, randomised controlled trial (RCT) evidence is preferable over non-randomised observational data for several reasons including the possibility of selection bias, confounding and regression to the mean associated with observational data (Torgerson, 2004). RCT data may also be at the individual patient-level, or at the aggregate level, or a combination of both. If inferences are to be made regarding outcomes in an unobserved period based on evidence in the observed trial period, then it is desirable that the data be at the patient level. The maturity of RCT data is also important. Even given a significant unobserved period, if most 'events' are captured in the observed trial period (in the case of a time-to-event variable), then there is likely to be less temporal uncertainty. Study design is a further important factor in how RCT data can be examined and potentially used to infer long-term outcomes. For example Bagust and Beale describe the influence of 'clinical protocols' on RCT outcomes both in terms of the within trial period and extrapolating beyond (Bagust and Beale, 2014).

3.3.6.2 External evidence

Often, some relevant evidence supplementary to RCT evidence is available. Although such 'external' evidence typically falls short of RCT evidence in terms of quality, it may exceed the available RCT evidence in terms of the time span it covers. Incorporating external evidence into a CEDM alongside RCT evidence can thus be a vital tool in addressing temporal uncertainty, though caution must be taken when incorporating some of these data into a CEA due to possible biases (Polsky and Basu, 2006). External evidence can come from a variety of sources, including:

- (i) Observational evidence (e.g. hospital registry data) which is often available to help inform baseline disease progression and/or long-term costs (Black, 1996).
- (ii) Patient reported outcome measures (PROMs) which are increasingly being employed to aid various types of health related research (Smith and Street, 2012). PROMs could potentially be used to inform long-term disease progression, costs and utilities.
- (iii) Mortality tables might be used to directly inform or to act as a bound on the long-term morality pertaining to a specific disease.

How any external evidence might be incorporated depends on what exactly the evidence is informing, i.e. it could be a single data point at or near the analysis time horizon, or a series of data

points over the long-term, or an indication of the temporal trajectory of the parameter after the observed period. There may be several options available regarding how the external evidence is incorporated to help in the estimation of a parameter value. For example, parameter values could be interpolated between what was observed in the trial period and what is indicated by observational data at a distal time point. Alternatively, parameter values extrapolated from the short-term RCT evidence could be combined, or averaged, with values indicated by longer-term external evidence. Furthermore, external evidence could be used not only to estimate the expected values of long-term parameters, but also, or instead:

- (i) To validate or reject extrapolated curves
- (ii) To give weight to alternative plausible assumptions
- (iii) To define a plausible region over the unobserved period, i.e. to generate bounds

3.3.6.3 Expert Elicitation

One type of external evidence that warrants particular attention in the context of addressing temporal uncertainty is expert elicitation. The opinion of clinicians and other experts can be quantified and utilised in CEA as a supplementary source of evidence (Bojke et al., 2010, Garthwaite PH, 2005). The input of experts often features in some way as a CEDM is developed. In addressing temporal uncertainty in particular, expert elicitation has the potential to play a very useful and very prominent role, ranging from simple advice regarding clinical pathways, to validation of data-based long-term estimates, to a more formal elicitation exercise in order to generate data where none existed.

How expert elicitation can be used in each of these roles to address temporal uncertainty for different parameters will be explored in further detail in Section 4.3.3.4.

In the HTA review, it was found that in most (70%) cases, some form of trial evidence was available to analysts, usually covering the time period immediately after intervention. This was often supplemented by other sources of evidence informing or validating longer-term trends. It was found that only in a minority of studies (32%) did analysts have access to individual patient-level data (IPD).

3.4 Addressing Temporal Uncertainty by Parameter-type

It was stated in the previous chapter that temporal uncertainty can potentially arise for any model input parameter. This section explores in detail the challenges of addressing temporal uncertainty for different types of model parameter. First, a brief taxonomy of the typical types of model parameter is offered.

3.4.1 A Brief Taxonomy of Parameter-types

The required outputs from a cost-effectiveness decision model (CEDM) are total costs and total health effects (e.g. QALYs) for each comparator. In most model structures, total health effects are calculated by accounting for disease progression and the health-related quality of life (HRQoL) associated with different disease conditions. Costs and HRQoL that relate to specific health states or clinical events can be considered as two parameter-types. However the characterisation of disease progression may be comprised of a number of different parameter-types. In state-transition models, disease progression is usually conveyed by the rate at which patients transition from one health state to another. Transitions between states are often determined by discrete-time transition probabilities which typically pertain to 'time-to-event' (TTE) variables. However, disease progression may also be determined by a longitudinal variable (e.g. cholesterol levels, tumour size), where either the transition to another health state is determined by the value of the longitudinal variable, or patients are assumed to stay within the same state but the health benefit associated with that state is assumed to evolve over time.

As well as the distinction between TTE variables and longitudinal variables, there is also the distinction between baseline measures and relative measures (or treatment effects). For every parameter representing disease progression, there must be a series of measures representing each competing health intervention. These are typically computed by considering a baseline and one or more treatment effects. A baseline measure of disease progression normally represents natural history or basic standard care. As a result, data are often available to inform long-term behaviour of baseline measures. CEAs may differ in their approach to comparing the clinical effectiveness of interventions. While some calculate absolute measures for each intervention and don't use any relative measure, most calculate a baseline measure and then apply treatment effects to represent the impact alternative interventions have in relation to the baseline measure. A treatment effect

therefore is a relative measure (e.g. hazard ratio, odds ratio, relative risk). Treatment effects are often the key drivers of cost-effectiveness results as they represent the comparative clinical effectiveness of treatments.

To summarise, the parameter-types that will be discussed are:

- (i) Time-to-event disease progression (baseline, treatment effect)
- (ii) Longitudinal disease progression (baseline, treatment effect)
- (iii) Health-related quality of life (HRQoL)
- (iv) Costs/resource use

3.4.2 Time-to-event Parameters

Although emphasis is given in this chapter to being mindful of the full range of parameter-types that can be exposed to temporal uncertainty in a CEDM, time-to-event parameters are examined in particular detail as they play a central role in the characterisation of disease processes. Given the prevalence of event-based models in health technology assessment (e.g. Markov models, discrete event simulation), methods to characterise long-term event rates are often paramount. Overall survival, in particular, plays a key role in the calculation of health benefit. QALYs, for example, are calculated as the product of life-years and quality of life. However, other time-to-event parameters also play a prominent role in HTA; progression-free survival (e.g. time to AIDS from HIV infection, time to cardiac event) or device failure rates (e.g. for a hip prosthesis) often represent central components of a decision model.

3.4.2.1 Overview of survival analysis

The analysis of time-to-event (TTE) data and of how TTE parameters evolve over time is referred to as survival analysis. Survival analysis thus plays a key role in the characterisation of disease processes and has featured heavily in much research to date regarding extrapolation in CEA. It is also a central

aspect of this thesis and is examined thoroughly in one of the empirical sub-chapters (Chapter 4.2). As such, an overview of the principles of survival analysis and its use in CEA is given here.

Survival analysis relates to the analysis of data in the form of times from a well-defined time origin (e.g. randomisation in a trial) until the occurrence of some particular event or end-point (e.g. death or some clinical event) (Collett, 2003). Survival analysis has been extensively employed in the medical field for some time, its key attribute being its ability to handle noninformative censoring that frequently occurs in follow-up studies (Briggs et al., 2006).

A central concept in survival analysis is that of the survivor function which expresses the probability of observing the event at or beyond time t . The survivor function can be defined as follows where T is a random variable and t is actual survival time of an individual:

$$S(t) = P(T \geq t)$$

Closely related to the survivor function is the hazard function which expresses the risk (or hazard) of the event at time t and is formally defined as follows (where $f(t)$ is the probability density function of T):

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} = \frac{f(t)}{S(t)}$$

$h(t)$ is also referred to as the hazard rate as it conveys the instantaneous rate at which events take place at time t . If the nature of the hazard function can be deduced, the related transition probabilities for use in a decision model can be calculated thusly:

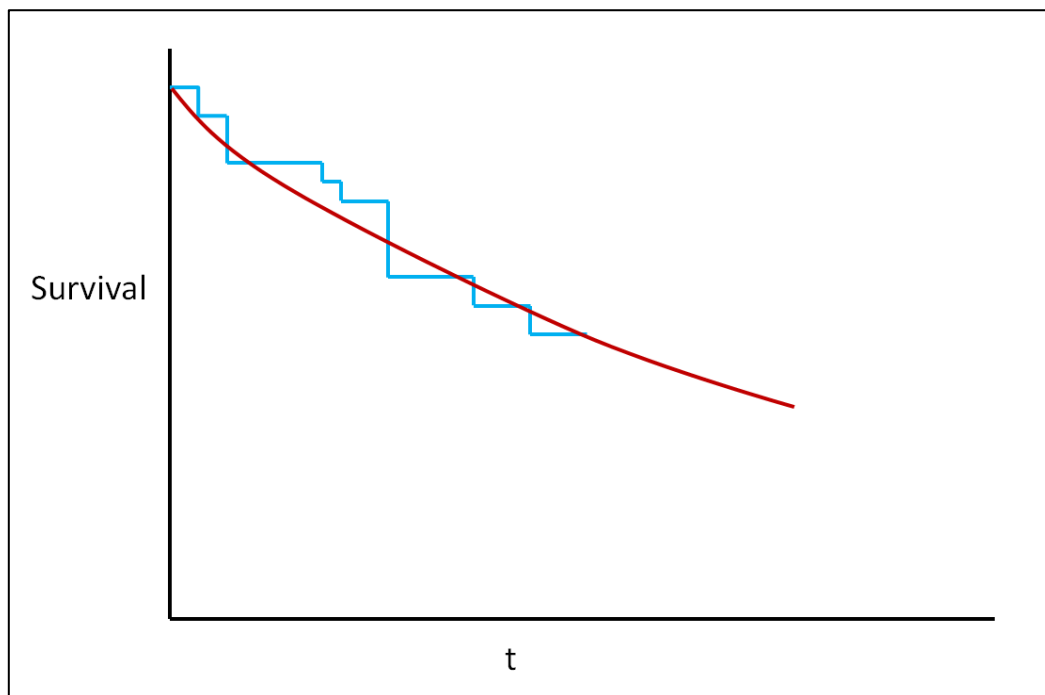
$$H(t) = \int_0^t h(u) du = -\ln[S(t)]$$

From the cumulative hazard function $H(t)$, we compute transition probabilities for chosen time intervals :

$$tp(t_u) = 1 - \exp\{H(t - u) - H(t)\}$$

Survival analysis can be used to directly interpret the available survival data as transition probabilities. However, in order to extrapolate short-term survival data beyond the evidence time horizon, a parametric distribution can be fitted to the survival data. This facilitates an estimation of the hazard function for any positive t from which the related survivor function and/or transition probabilities can be calculated up to the required time horizon. This approach is commonly taken in health technology assessment (Sculpher et al., 2006, NICE, 2013). A parametric function fit to empirical survival data (Kaplan-Meier data) is illustrated in Figure 17.

Figure 17: Illustration of a parametric function being fit to survival data in order to extrapolate beyond the data



There are a number of parametric distributions that can be used to this end (Collett, 2003). A short explanation of some of the most common distributions along with their survivor distributions is given below.

The exponential distribution is a single parameter distribution and assumes a constant hazard over time. The exponential distribution also assumes 'proportional hazards' which means the hazards of multiple groups are assumed to be multiplicative for any t and so for example treatment groups can be characterised by hazard ratios representing treatment effects.

$$\text{Exponential: } S(t) = e^{-\lambda t}$$

Where λ = a 'scale' parameter

The Weibull and Gompertz distributions are two parameters distributions which allow hazard rates to vary monotonically with time and also employ the proportional hazards assumption for multiple groups.

$$\begin{aligned} \text{Weibull: } S(t) &= \exp(-\lambda t^\gamma) \\ \text{Gompertz: } S(t) &= \exp\left\{\frac{\lambda}{\theta}(1 - e^{\theta t})\right\} \end{aligned}$$

Where λ = a 'scale' parameter

γ, θ = 'shape' parameters

The log-Normal and Log-logistic distributions allow non-monotonic hazards but are accelerated time failure (AFT) distributions as opposed to proportional hazards distributions¹⁸. Where the proportional hazards assumption is not appropriate, these distributions may be useful, although since much evidence on treatment comes in the form of hazard/odds ratios, it is often more desirable to employ a distribution with a proportional hazards metric.

$$\text{LogNormal: } S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$$

$$\text{LogLogistic: } S(t) = (1 + e^{\theta t^k})^{-1}$$

Where $\Phi(x)$ = the probability density function of the Normal distribution

μ = a location parameter

σ = a scale parameter

k = a scale parameter

θ = a shape parameter

The generalised gamma distribution is a three parameter model using the AFT metric which allows hazards to vary non-monotonically with time. It may reduce to the Exponential, Weibull and Log-Normal distributions in special cases and so may be used as either a proportional hazards model or an accelerated failure time model.

¹⁸ Note that the Weibull distribution can also be used as an AFT distribution

$$\text{Generalised Gamma: } S(t) = 1 - \Gamma_{\lambda t}(p)$$

Where $\Gamma_{\lambda t}(p)$ is the ‘incomplete gamma function’ given by $\frac{1}{\Gamma(p)} \int_0^{\lambda t} u^{p-1} e^{-u} du$

p = a shape parameter

λ = a scale parameter

Other very flexible models exist such as that developed by Royston and Parmar (Royston and Parmar, 2002). The Royston-Parmar model uses cubic splines to accurately represent the observed survival data while allowing for both a proportional hazards and proportional odds metric.

In terms of the suitability of parametric distributions, there are a number of factors to consider (Latimer, 2011, Manca et al., 2009). The validity of the proportional hazards (PH) assumption is central (Guyot et al., 2011). Often there is supplementary aggregate data in the form of hazard or odds ratios in which case the PH assumption attractive. If the PH assumption is not suitable, an AFT model could be employed or separate distributions could be fit to the different arms of the trial data. The validity of the PH assumption can be checked through visual inspection (log-log plots), goodness of fit tests and the use of time-dependent variables (Cleves, 2010). The chosen distribution must also be sufficiently flexible to characterise the nature of the observed survival data while avoiding over-specification (Jackson et al., 2010). Model fit can be judged by visual inspection (comparing the Kaplan-Meier curves to the parametric survival curves), assessing log-cumulative hazard plots, or conducting Akaike Information Criterion or Bayesian Information Criterion (AIC/BIC) tests. AIC/BIC tests in particular have been widely used in HTA to assess the relative goodness-of-fit of parametric survival distributions. These tests quantify how well parametric distributions fit the empirical data relative to other candidate distributions (i.e. an AIC/BIC score is meaningless in isolation) (Burnham and Anderson, 2004).

Of course, what is of most relevance in terms of temporal uncertainty is the validity of what the parametric function implies about parameter values after the observed period. While fitting a parametric function is a useful tool for a number of reasons, it is commonly used in order to extrapolate the survival data beyond the observed period¹⁹. Caution is warranted in this endeavour, as whatever distribution is chosen, the implicit assumption is that long-term parameter values are

¹⁹ A parametric function might also be fit in order to facilitate sub-group analysis, or to characterise parameter uncertainty, or to estimate the ‘true’ survival curve, i.e. to smooth out the sampling variation in the raw survival data

fully informed by short-term observations and that the ‘best fitting’ distribution will best predict parameter values over the unobserved period. Such an assumption may or may not reflect the clinical reality. It is nonetheless desirable to exploit the available short-term evidence when estimating long-term parameter values.

3.4.2.2 Output from HTA Review

In the HTA review, studies that featured one or more TTE parameters were flagged and examined. From these studies, data were extracted relating to the type of evidence available, the use of formal evidence synthesis and the methods employed to estimate TTE parameter values beyond the observed period. These characteristics were chosen for extraction in order to appraise the options available and the challenges existent when TTE variables require long-term characterisation, but also to observe what methods are most commonly employed.

From the randomly selected 64 HTAs, it was found that 42 (66%) included at least one time-to-event (TTE) parameter that was exposed to temporal uncertainty. Information extracted from these HTAs relating to how the parameters were modelled is summarised in Table 5.

Table 5: Details of long-term estimation of time-to-event parameters

	Number	% (Note 1)
Evidence available		
RCT (aggregate)	8	19%
RCT (individual patient-level)	6	14%
Observational (aggregate)	6	14%
Observational (individual patient-level)	3	7%
Expert opinion	0	0%
Reference to previous work but unclear what level of data	2	5%
Combination of the above	11	26%
Unclear	6	14%
Use of formal evidence synthesis	11	26%
Method to estimate long-term values (Note 2)		

Parametric model: Exponential	5	12%
Parametric model: Weibull	7	17%
Parametric model: Other, multiple or unclear	10	24%
Assumption of constant rate for unobserved period (Note 3)	15	36%
Other assumption (Note 3)	5	12%
Mortality tables / other long-term dataset	2	5%

Note 1: Percentage of the 42 HTAs where at least one time to event parameter was exposed to temporal uncertainty

Note 2: More than one TTE extrapolation was carried out for some studies

Note 3: An assumption was applied without use of parametric model fit to short-term evidence. This pertains to the base-case assumption. Alternative assumptions were sometimes explored in scenario analyses.

Significant from the results in Table 5 is that analysts do not have access to individual patient-level data (IPD) and randomised controlled trial (RCT) data to the extent that would be desirable (or that might be assumed). Working with data that are not derived from randomised trials and/or not at the patient level creates further problems of uncertainty and bias when attempting to estimate beyond the evidence period. Analysts also must frequently contend with multiple sources of relevant evidence and so must judge what evidence can be given most weight when estimating values past the evidence period. Parametric survival analysis was employed in 52% of cases, suggesting this is a significant, but not dominant, tool in estimating long-term values for time-to-event parameters. Ensuring the appropriate use of parametric models for the purposes of extrapolation is nonetheless clearly an important issue. Where parametric models were employed, the use of the proportional hazards assumption was common and an assumption of no further treatment effect was most commonly imposed. It was noted that in a few cases an assumption of limited further treatment was made. A final notable result is the frequency with which a simple assumption of a constant rate over the unobserved period was made. This, in many cases, is characterised as a conservative assumption²⁰.

How such parametric survival models are validated is the subject of much scrutiny in the methods literature (as evidenced by some of the studies described in Section 3.2.2). The fundamental concern regarding the choice of parametric function is that it may not appropriately reflect survival outcomes over the long-term (or beyond the evidence time horizon), yet extrapolating from the short-term is often understood as making best use of the available evidence. A key distinction that emerges is that

²⁰ Recall that the appropriateness and usefulness of ‘conservative assumptions’ in the context of temporal uncertainty was discussed in Section 2.2.2.1

between internal validity and external validity. Although goodness-of-fit analysis may demonstrate that a method of extrapolation is internally valid, the implied long-term estimates ought to be tested against external data (where available) to show clinical plausibility. Of course, this is arguably the crux of the issue as such data often are not available. The expression of uncertainty surrounding long-term survival estimates is thus a crucial task.

3.4.3 Longitudinal Parameters

Disease progression may also be described by non-TTE parameters. Longitudinal clinical measures are those that track continual change over time (as opposed to estimating ‘when’ some event will occur), e.g. cholesterol levels, tumour size. Although non-TTE parameter does not equate to longitudinal parameter, longitudinal is a useful category that broadly contrasts with TTE²¹. As with TTE parameters, we would expect a wider range of modelling tools to be available to analysts when data are at the patient level and/or derived from randomised trials.

From the randomly selected 64 HTAs, it was found that 13 (20%) included at least one longitudinal parameter that was exposed to temporal uncertainty. Information extracted from these HTAs relating to how the parameters were modelled is summarised in Table 6.

Table 6: Details of long-term estimation of longitudinal parameters

	Number	% (Note 4)
Evidence available		
RCT (aggregate)	2	15%
RCT (individual patient-level)	2	15%
Observational (aggregate)	2	15%
Observational (individual patient-level)	1	8%
Expert opinion	1	8%
Reference to previous work but unclear what level of data	1	8%
Combination of the above	4	31%

²¹ Note that although, a relative effect applied to a TTE parameter could be considered to be a non-TTE parameter, it was assumed that such a parameter belonged in the TTE category

Use of formal evidence synthesis	5	31%
Method to estimate long-term values		
Long-term observational dataset	1	8%
Assumption of constant value for unobserved period (Note 5)	9	69%
Other assumption (Note 5)	3	23%

Note 4: Percentage of the 20 HTAs where extrapolation of at least one longitudinal parameter was carried out

Note 5: This pertains to the base-case assumption. Alternative assumptions were sometimes explored in scenario analyses.

The primary point of note here is that the prevalence of longitudinal parameters (that are exposed to temporal uncertainty) is greater than is reflected in the methodological literature on temporal uncertainty/extrapolation. To compound this point, there were no examples found of a parametric function being used to formally extrapolate short-term data. In many cases however, there were RCT and/or patient-level data available, though it is not clear whether the data available could be used to inform the longitudinal parameter in question. As with TTE parameters, it was found that in a significant number of cases, long-term parameter values were ultimately generated through a simple assumption (usually of constant values). Such an assumption amounts to an ‘informal judgement’. These judgements were rarely validated and (as with TTE parameters) often described as ‘conservative assumptions’, made for the purposes of producing a result which would emphasise the cost-effectiveness or cost-ineffectiveness of a health technology.

3.4.4 Costs/Resource Use Parameters

Measures of cost/resource use are fundamental to CEDMs (Brouwer et al., 2001). Along with any ‘upfront’ costs that relate to alternative health interventions, costs are typically calculated in a CEDM by attributing a single cost per cycle to health states and health events. The method by which these state and event-specific costs are computed varies by disease/decision context. Typically, estimates of resource use are obtained from RCT data and the unit costs associated with each item of resource use are obtained from an external source such as hospital records (Drummond et al., 2005).

Therefore, to a large extent, the accumulation over time of costs is determined by the likelihood of experiencing health events and moving between different health states. However, these state and

event-specific estimates (both resource use and unit costs) are themselves subject to temporal uncertainty. The costs attributed to a health state or health event are often assumed to not alter over the entire period of analysis. Such an assumption may or may not be appropriate, but considering and testing its validity ought to form part of any robust analysis.

As distal time points are considered, the question of what future costs are and are not relevant becomes pertinent (van Baal et al., 2011, Meltzer, 1997). The debates surrounding related/unrelated future costs will not be added to here. However an awareness of this issue is warranted as the associated issue of temporal uncertainty is addressed.

Another 'methodological uncertainty' that relates to long-term costs (and indeed health benefits) is that of inflation. Typically, inflation is not directly tackled in HTA and is assumed to be accounted for in the 'real' discount rate employed.

In the HTA review, it was found that 58 out of 64 HTAs (91%) involved the estimation of long-term costs, i.e. in 6 HTAs, it was assumed that there were no significant costs over the long-term to estimate. The overwhelming majority of these HTAs modelled long-term costs using a state-transition model structure, i.e. the costs incurred per unit time were a product of how the patient cohort was distributed among health states, and the costs ascribed to each health state. In all HTAs that employed a state-transition structure, the costs per health state were considered not time dependant, i.e. costs per event or per time in health state were constant over time. In some cases, it was not clear how the costs per health state were populated. In one HTA (NICE TA125), a microsimulation model was employed and the costs were directly linked to measures of disease severity, namely the HAQ and PASI indexes. In no case was it assumed that costs per health state are time-dependant; that they, for example, decrease over time according to some function. The results are summarised in Table 7.

Table 7: Details of long-term estimation of cost/resource use parameters

	Number	% (Note 6)
Estimated long-term costs	58	
Method of estimating long-term costs		
State-transition Model (constant costs)	55	91%
Micro-simulation (constant costs)	1	2%
Unclear	3	5%

Note 6: % of HTAs that estimated long-term costs (58)

In terms of resource use, assuming no change over the time may be a strong assumption, as there may be factors (such as ageing) that cause significant shifts in resource use over the long-term (this issue is explored in Section 4.4.3.2.1. Any long-term uncertainty regarding unit costs is likely to relate to sudden shifts in drug prices or hospital costs. These ‘uncertain future events’ therefore are a crucial issue regarding temporal uncertainty of costs.

Uncertain future events are not model parameters themselves per se, but another consequence of a long-term time horizon; namely, the possibility of a future event that can impact the value of parameters (and in particular costs), e.g. a price shock, or a new relevant comparator emerging. Uncertain future events ought to be accounted for when conducting value of information analyses but may only be of consequence for estimating expected cost-effectiveness under specific and relatively uncommon circumstances.

The handling of uncertain future events was also investigated in the HTA review. All 180 HTAs were reviewed, as the instances of uncertain future events being accounted for were particularly sparse. Only in two HTAs was there an attempt to formally model uncertain future events (Robinson et al., 2005, Rogowski et al., 2009). These two studies formally considered the potential impact of a new comparator and the arrival of a generic version of an existing technology respectively. Robinson et al. included a further comparator as part of a sensitivity analysis justified on the grounds that the results of a major RCT for this new intervention were published during the course of undertaking the base-case analysis and the results of this RCT were already beginning to be incorporated into clinical guidelines. Rogowski et al noted that a future change was anticipated over the unobserved period,

namely the coming off patent of one of the comparators. This future event was incorporated into a value-of-information analysis.

Importantly, both studies considered events that were anticipated to occur close to the time of the evaluation and in each case the results were presented as part of a separate sensitivity analysis. Since both studies included these future events as part of a separate sensitivity analysis there was very limited information provided on the methods used and justification for these.

3.4.5 Health-related Quality of Life

Parameters representing health-related quality of life (HRQoL) are also fundamental components of CEDMs and are typically incorporated in a similar manner to costs, i.e. a value is ascribed to each health state and is generally assumed to not alter with time. In a manner that is also similar to costs, HRQoL (typically) is essentially an amalgam of two quantities. Health status is commonly recorded using an instrument such as EQ-5D or SF-36 where patients state their health status as part of a clinical trial. In order to convert these into HRQoL 'weights' for use in an economic analysis, each health status is weighted according to general population preferences (a time-trade-off based algorithm developed by Dolan et al. in the case of EQ-5D) (Dolan and Gudex, 1995, Dolan, 1997). And as with costs, both of these quantities are, in principle, subject to variation over time. It may be the case that changes in HRQoL over time are wholly captured by transitions between health states. However, the natural decline in HRQoL associated with age and/or with chronic diseases is not typically captured in transitions between health states, which calls into question the assumption of constant (HRQoL) values. If temporal decrements are to be applied to HRQoL, there is a question of what method is most appropriate (e.g. additive, multiplicative, minimum) (Ara and Wailoo, 2011). HRQoL is sometimes derived from longitudinal variables and rather than being employed to ascribe value to health states, HRQoL acts as the sole indicator of health gain/loss. HRQoL data are usually included in RCTs but may be synthesised with HRQoL data from elsewhere. For NICE HTAs, HRQoL is measured using the EQ5D instrument (Ara and Wailoo, 2012). Where HRQoL is measured by another instrument, mapping to EQ5D may be necessary (Longworth and Rowen, 2011).

In the HTA review, it was found that 61 out of the 64 HTAs (95%) included an estimation of long-term HRQoL. Three HTAs did not include estimation of long-term HRQoL due to a lack of sufficient evidence or because it was assumed life-years were a sufficient measure of health benefit. HRQoL

parameters were generally treated in the same manner as costs/resource use, where long-term HRQoL were modelled by ascribing HRQoL values to health states in a long-term state-transition model. As with costs, the vast majority assumed constant HRQoL values per health state. The results are summarised in Table 8.

Table 8: Details of long-term estimation of HRQoL parameters

	Number	% (Note 7)
Estimated long-term HRQoL	61	
Method of estimating long-term HRQoL		
State-transition Model with constant HRQoL	52	91%
Increment/decrement applied to HRQoL over time	7	2%
Unclear	2	5%

Note 7: % of HTAs that estimated long-term HRQoL (61)

In contrast to costs/resource use, there were some cases of applying a decrement to HRQoL per health state to account for change over time in health benefit not represented by progression through health states. The most common reason for this was to account for ageing within the cohort whereby a mean baseline HRQoL decrement was applied in one or more health states. There were also instances of increments/decrements being applied to HRQoL for reasons other than accounting for ageing. For example, Stevenson et al, in a study of the use of vitamin K to prevent fractures in older women, in order to model recovery, a HRQoL multiplier effect was assumed whereby a HRQoL decrement was combined multiplicatively with the general population HRQoL to provide an estimate of the HRQoL for patients in a particular health state, resulting in the absolute HRQoL increasing for a period of time (Stevenson et al., 2009). Another example can be found in a study on the cost-effectiveness of treatment for severe sepsis, where a regression analysis was used to infer the reduction of HRQoL over time and to model HRQoL beyond the last observation of the trial. It is not clear if any uncertainty in the estimated coefficients was incorporated into the model or if other functional forms were tested (Green et al., 2005).

It is clear that computing HRQoL that can be ascribed to a health state and assuming this does not vary over time is the most common method of characterising long-term HRQoL. Although there are sometimes particular reasons to model HRQoL differently, it is striking that so few studies model the

natural continuous change in HRQoL owing to ageing as this effect surely ought to be explicitly accounted for quite often in HTA.

3.5 Characterisation of Uncertainty Over Time

Recall that temporal uncertainty has been defined as relating to both estimating parameter values over the unobserved period and quantifying the related uncertainty. The above discussions primarily pertained to estimating parameter values over the observed period. This last section pertains exclusively to the quantification of uncertainty surrounding long-term estimates of parameter values.

What was sought first and foremost in this portion of the HTA review was any acknowledgement of this uncertainty (which may pertain to any of the parameter-types discussed in Section 3.4 above). It was found that in only 32 of the 64 HTAs (50%) was there any acknowledgement of, or any attempt to characterise, this uncertainty. The details are summarised in Table 9.

Table 9: Details of How Long-term Uncertainty was Characterised

	Number	% (Note 8)
Addressed Long-term Uncertainty	32	50%
Method of Addressing Uncertainty		
Scenario Analysis: Alternative Time Horizons	15	23%
Scenario Analysis: Alternative extrapolations	19	30%
Scenario Analysis: Alternative Duration of Treatment Effect	4	6%
Scenario Analysis: Alternative Long-term Costs/HRQoL	3	5%

Note 8: This is the percentage of the 64 HTAs that were reviewed.

Note 9: Some HTAs considered more than one of the aspects of temporal uncertainty listed in the table above.

Every HTA reviewed that attempted to address this uncertainty in some way did so through one-way sensitivity/scenario analyses.

In many cases (23%), the only sensitivity analysis carried out was the application of alternative time horizons. The chief reasons given for this was:

- (i) To report the cost-effectiveness results based on the trial time horizon (Main et al., 2004)
- (ii) To compare results with previous studies (Wright et al., 2006)
- (iii) To demonstrate the sensitivity of the ICER to the length of the time horizon inclusion of long-term health benefits (and possibly costs) (Robinson et al., 2005, Shyangdan et al., 2011, Wilson et al., 2005, Paulden et al., 2010)

Examining the sensitivity of the ICER to the length of the time horizon was the most common reason given to conduct this particular sensitivity analysis. Many HTAs showed significant sensitivity to the time horizon applied. This ought to have firstly, confirmed the importance of including long-term health benefits and costs, but secondly, demonstrated the importance of estimating long-term outcomes with care and characterising the uncertainty surrounding the long-term values of model input parameters. However, in the majority (60%) of cases where a time horizon sensitivity analysis was carried out, no uncertainty analysis was conducted regarding the estimated long-term values of parameters.

In a relatively large proportion of HTAs (30%), there was some attempt to express uncertainty regarding how evidence was extrapolated from the short-term in order to infer long-term parameter values. There was much inconsistency in how this was done. For example, Morgan et al constructed two models, one that assumed that the costs and effects seen up to 6 months would continue into the future, and one that assumed that costs and effect would exhibit an exponential decay (Morgan et al., 2004). Another example is the HTA carried out by Black et al who explored the impact of uncertainty regarding long-term disease progression by varying a transition probability by +/-50% (Black et al., 2009). A number of HTAs attempted to characterise the uncertainty around extrapolating TTE outcomes by producing cost-effectiveness results for alternative parametric fits to the short-term TTE data (Dundar et al., 2009, Fox et al., 2007, Bond et al., 2009). For example, the manufacturer's submission in TA137 showed results for both Weibull and Log-logistic survival models. These models were chosen based on their 'goodness-of-fit' to the short-term data. However, as the evidence review group in this case point out, both functions imply generous life expectancy and it may be that neither is appropriate to characterise long-term survival. This example suggests that applying alternative parametric functions is unlikely to be a sufficient means of

expressing uncertainty with regard to the estimation of the long-term values of a parameter (Boland et al., 2009).

Uncertainty regarding the relative effect of treatment was most commonly characterised by altering the treatment effect duration after the observed period. If a treatment effect is shown to exist (and not alter over time) up to the evidence time horizon, then clearly, a difficult judgement must be made regarding the nature of treatment effect over the long-term. Representing this uncertainty by assuming the treatment effect continues at a constant magnitude for a period of time is perhaps an oversimplification, as the effect is more likely to dissipate over time, or there could even be a rebound effect (a possibility not incorporated into any of the analyses reviewed). However, like other scenario analyses, the purpose of applying alternative treatment effect durations was simply to test the sensitivity of the adoption decision (or the mean ICER) to the nature of long-term treatment effect.

In three cases, there was uncertainty expressed regarding the estimates of long-term costs/utilities (Clar et al., 2005, Main et al., 2010, Morgan et al., 2004). For example, Main et al simulated a 50% increase and decrease in supportive care and drug administration costs in order to show that the ICER does not change significantly (Main et al., 2010).

Although a scenario analysis goes some way to expressing the (impact of) uncertainty over the long-term and avoids imposing assumptions regarding the relative likelihood of alternative scenarios, it carries significant limitations. In presenting a base-case and a number of alternative scenarios, decision-makers are effectively left to weight all scenarios, which may lead to (i) too much weight given to the base-case and (ii) inappropriate calculation of an overall expected cost-effectiveness.

For example, in the case of Sorafenib for advanced hepatocellular carcinoma, the manufacturer's submission examined a range of survival distributions to extrapolate overall survival but ultimately employed a single distribution (Log-Normal) based on superior AIC (Connock et al., 2010). However, as Connock et al. illustrate in the ERG report, the extrapolated Log-Normal function implied very optimistic survival relating to Sorafenib which may not be clinically valid. The Log-Normal distribution alone did not provide a fair estimate of overall survival, whereas multiple distributions weighted according to their clinical validity may have provided a more suitable estimate. Sorafenib was ultimately not recommended for use by NICE.

Another (perhaps more worrying) example is that of Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. As part of an appeal against NICE's approval of Trastuzumab, analysts representing Newbury and Community Primary Care Trust highlighted that optimistic assumptions regarding primary efficacy are extrapolated beyond the trial period but that adverse events observed in the trial are not extrapolated into the long-term. Whereas the significant effects of alternative temporal assumptions were demonstrated in a sensitivity analysis, it was the base-case analysis (with temporal assumptions favouring Trastuzumab) that was used as the basis for reimbursement decision. This point was ultimately left out of the appeal for reasons of "simplicity and efficiency". The appeal was ultimately rejected and the decision upheld (Ward et al., 2009, Trust, 2006, NICE, 2006).

These examples highlight the need to firstly validate temporal assumptions in health technology assessments but also the need to make every effort to incorporate those alternative valid assumptions into the base-case analysis in order to provide a fair estimate of expected cost-effectiveness.

3.6 Conclusions & Discussion

Through the above examination of the key analytical issues in addressing uncertainty, the existent methods guidance and the methods employed in HTA to date, a number of points of note emerge.

Temporal uncertainty is undoubtedly an issue common in HTA, as evidenced by its prevalence in the HTAs reviewed. How often temporal uncertainty is a pivotal issue (i.e. could impact decision making) is less clear. A number of HTAs produced results for a range of time horizons, seemingly to demonstrate the importance (or otherwise) of long-term outcomes. A means of demonstrating the significance of temporal uncertainty is to be welcomed and perhaps ought to be required. However, a more sophisticated and reliable method is warranted (such as calculating cumulative net benefit over time as outlined in Section 2.2.2.1). A revelation that temporal uncertainty impacts decision-making should be followed by a thoughtful and thorough modelling of temporal uncertainty.

The endeavour made regarding extrapolation/temporal uncertainty in HTA would appear to be generally deficient. The many instances of temporal uncertainty being so diverse, coupled with the dearth of clear guidance often leads those developing CEDMs to either not fully confront temporal

uncertainty problems, or to employ very simple assumptions to estimate long-term outcomes. There seems to be a disparity between the relative consideration given to how best to tackle temporal uncertainty and the impact that temporal uncertainty can potentially have on the cost-effectiveness results. Even when data are lacking or weak, as it often is, it was found that there is often scope for a more thoughtful consideration of modelling options.

Relatedly, the presentation of how temporal uncertainty is addressed is often poor and the techniques used less than transparent. Therefore, a requirement for an explicit consideration of how temporal uncertainty was approached in an analysis may be productive.

There is a disproportionate focus in methods guidance and in the methods literature on extrapolating survival outcomes and treatment effect (in the methods literature there is particular focus on using parametric survival analysis to extrapolate survival outcomes). Although survival (and other time-to-event) outcomes are of obvious importance in HTA, there are limitations to the benefits of fitting parametric functions to the short-term data and directly extrapolating over time into the unobserved period. In particular, there are issues of: justification of distribution choice, clinical validity and expression of uncertainty. Moreover, there are other parameter-types that are deserving of attention when it comes to estimation beyond the evidence period (e.g. the assumption of constant HRQoL over time is often not suitable). There is a need to analyse the broader problem of a lack of long-term evidence and to develop techniques for sensibly plugging the evidence gap and quantifying the related uncertainty.

The 'base-case' assumptions regarding the post-trial behaviour of model parameters are typically (and perhaps sometimes not unreasonably) very simple assumptions, often of no change over time. While *some* assumptions must be imposed in order to produce an estimate of cost-effectiveness, there is often worryingly little attention paid to the (un)suitability of these assumptions.

Crucially, the uncertainty regarding the temporal assumptions imposed or the suitability of the modelling approach used is not sufficiently captured in HTAs. There is a danger that varying time horizons is seen as 'ticking the box' of exploring uncertainty pertaining to the long-term. Scenario analyses (which can potentially go some way to expressing temporal uncertainty) are sometimes employed. However, to allow the decision-maker to make a fully informed decision regarding technology adoption and/or to deduce that further evidence relating to long-term outcomes is required, temporal uncertainty must be incorporated into a single probabilistic model.

It is hoped this review will strengthen the argument for giving temporal uncertainty greater attention in the assessment of health technologies and will help build towards more robust guidance on temporal uncertainty in future methods guides.

This chapter has endeavoured to outline the array of analytical issues relevant to the characterisation of temporal uncertainty and to express the 'state of play' in how temporal uncertainty is being addressed. In the following chapter (Chapter 4), many of the issues raised are explored and tackled further through the thorough re-analysis of a CEDM.

4. THESIS CHAPTER 4: METHODS TO ADDRESS TEMPORAL UNCERTAINTY

4.1 Chapter 4.1: Overview

4.1.1 Introduction

The purpose of Chapter 4 is to address, using a motivating example, a range of analytic challenges that arise when there exists a dearth of long-term data in a cost-effectiveness decision model (CEDM).

This overview sub-chapter sets the scene by explaining the overall structure of Chapter 4, introducing the motivating example that will be used throughout, outlining the issues of temporal uncertainty that will be addressed in the remaining sub-chapters and finally focusing on an important parameter relationship and how this can be dealt with.

4.1.2 Rationale Behind Structure of Chapter 4

For the sake of clarity and narrative, the empirical section of the thesis comprises one overarching chapter. As discussed in Chapter 3, there is a range of parameter-types that are typically exposed to temporal uncertainty. Because methods are required to appropriately address temporal uncertainty regarding each of these parameter-types and because the interactions and cumulative effects of these parameters are also crucial factors, a chapter structure where inter-related sub-chapters build towards a comprehensive methodology would best facilitate a thorough analysis. Thus, there are six 'sub-chapters' within this chapter, with four of these sub-chapters focusing on one particular parameter-type (or group of parameter types) where there is an issue of temporal uncertainty. These are (i) baseline risk, (ii) treatment effect and (iii) costs/resource use and (iv) HRQoL. As well as this overview sub-chapter, there is also an overall results and discussion sub-chapter that compares the original CEDM to the 'updated' CEDM and discusses some key outcomes and findings.

Each of central four sub-chapters broadly follows the same structure (introduction, available evidence, methods, results) and employs the same motivating example (RITA-3). Parameter interactions can then be explored and overall conclusions can be drawn in the final sub-chapter. The efforts at characterising temporal uncertainty in the RITA-3 model are not exhaustive. All issues of

temporal uncertainty in the RITA-3 model are outlined in this overview section, but not all are analysed in detail in the following sub-chapters. It is envisioned however, that the issues of temporal uncertainty that are addressed in detail broadly represent the key issues of temporal uncertainty facing analysts in health technology assessment (HTA). In short, the RITA-3 model is used as a vehicle for exploring and developing methods for addressing temporal uncertainty, the need for which has been articulated in the previous two chapters.

4.1.3 Introduction to the Motivating Example: The RITA-3 Model

The RITA-3 cost-effectiveness decision model (CEDM) represents a useful vehicle through which to explore issues of temporal uncertainty. This case study comprises an event-based model structure, individual patient-level data from a randomised controlled trial, a long-term analysis time horizon and a narrowly defined patient population. These features present challenges for estimating long-term outcomes which are typical of those arising in HTA.

It should be noted at this point that the RITA-3 model is employed in this Ph.D. research solely for the purposes of developing the methods employed in healthcare decision modelling. Some of the assumptions and analyses in this chapter are stylised (though the final 'updated model' is based only on real evidence). Furthermore it is likely that the RITA-3 trial itself is largely obsolete as relevant clinical procedures (e.g. the insertion of stents) have advanced somewhat since the RITA-3 trial took place (Vardi et al., 2005). Therefore, the results of this re-analysis are not intended for use in clinical decision making.

4.1.4 Background to Disease and Decision Problem

Patients with non-ST-elevation acute coronary syndrome (NSTEMI-ACS) face a significant risk of mortality and cardiovascular events. It is expected that an early interventional strategy (routine angiography followed by revascularisation if clinically indicated) will represent a lower risk of death/a cardiovascular event compared to a conservative strategy (ischaemia or symptom-driven angiography), but also a higher cost to the health system. There is uncertainty regarding whether implementing the early interventional strategy represents good value for money from a health system's point of view. A decision problem arises therefore, concerning whether the early

interventional strategy or the conservative strategy should be recommended for patients presenting with NSTEMI-ACS.

It is assumed that the ICER threshold for this decision problem is £20,000 per QALY.

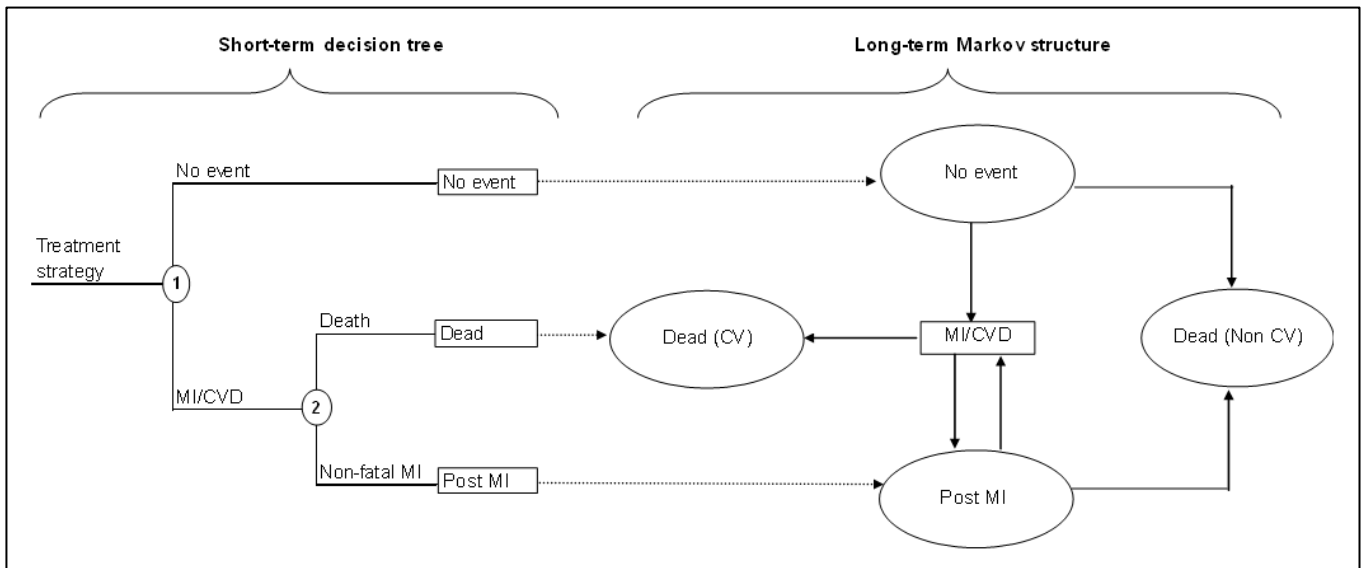
4.1.4.1 The RITA-3 trial

The Randomised Intervention Treatment of Angina (RITA-3) trial was a prospective, randomized multicentre trial with parallel groups, enrolling 1810 patients from 45 hospitals in England and Scotland (Fox et al., 2005). Patients were said to be eligible if they presented with cardiac pain associated with electrocardiographic or previous arteriographic evidence of coronary artery disease, or an elevated serum cardiac marker (Fox et al., 2002).

4.1.4.2 A Decision Model

A cost-effectiveness decision model (CEDM) was developed by Henriksson et al. in 2008 based predominantly on the individual patient-level data (IPD) from the RITA-3 trial (Henriksson et al., 2008). The CEDM is composed of a short-term tree structure (assumed to be instantaneous in time) and a long-term Markov structure. A series of regression equations are used to estimate the transition probabilities between Markov states. Costs and QALYs per Markov state (and for the index hospitalisation period) are estimated using standard OLS regressions. The CEDM is probabilistic and is written in Stata and Excel. The model structure is depicted in Figure 18 below.

Figure 18: Structure of RITA-3 CEDM



4.1.5 Original Model Results

4.1.5.1 Cost-effectiveness

Patients were divided into quartiles of risk. Because of the much higher event rate in the fourth (uppermost) quartile, this quartile was further divided in two (risk groups 4a and 4b). Each risk group was represented by a particular risk profile (See Appendix 2). Table 10 below outlines the cost-effectiveness results for 5 risk groups.

Table 10: Cost-effectiveness Results from Original CEDM

Risk group	Incremental Cost (£)	Incremental QALY	Mean ICER	Adopt/reject early interventional (EI)
Risk group 1	4,885	0.091	53,760	Reject
Risk group 2	4,898	0.213	22,949	Reject
Risk group 3	6,029	0.285	21,186	Reject
Risk group 4a	6,538	0.547	11,957	Adopt
Risk group 4b	6,530	0.512	12,750	Adopt

With an ICER threshold of £20,000, the results in Table 10 suggest that for risk groups 4a and 4b, the early interventional treatment should be adopted based on cost-effectiveness, whereas for risk groups 1, 2 and 3, the conservative treatment should be retained. Risk group 3 is associated with a mean ICER very close to that of the willingness to pay threshold (£20,000). It is expected therefore, that uncertainty around the model inputs will have the greatest impact on the decision uncertainty associated with risk group 3.

4.1.5.2 Uncertainty

Table 11 below summarises the impact of uncertainty on the outputs of the CEDM for each of the 5 risk groups, showing the probability of cost-effectiveness at a threshold of £20,000/QALY and the expected value of perfect information (EVPI) at both the patient and population level. For the calculation of EVPI/population, an annual UK incidence rate for NSTEMI-ACS of 59,756 was obtained from the Office for National statistics (Office for National Statistics, 2012). It was assumed that evidence pertinent to this decision problem would remain relevant for 10 years and that the value of this evidence could be discounted at a rate of 3.5% per annum. It was further assumed that the incident population could be divided proportionally among the risk groups, i.e. 25%, 25%, 25%, 12.5%, 12.5% for risk groups 1, 2, 3, 4a, 4b respectively.

Table 11: Summary of Effect of Uncertainty on for each Risk Group

Risk group	Prob(EI cost-effective) at £20,000/QALY	EVPI/patient (£)	EVPI/population (£)
Risk group 1	0.009	1.92	349,068
Risk group 2	0.328	367.70	66,850,066
Risk group 3	0.420	475.15	86,385,121
Risk group 4a	0.945	61.13	5,556,900
Risk group 4b	0.924	107.96	9,813,888

For the results outlined above, the original temporal assumptions (described in detail in Table 12 below) were applied. It is important to note that some of these temporal assumptions were explicitly 'conservative', i.e. they openly underestimate the effectiveness of the early interventional strategy in order to be confident in its cost-effectiveness for higher risk groups. As temporal uncertainty is more carefully modelled in subsequent sub-chapters, both the level of uncertainty

surrounding the model inputs and the expected ICER are likely to shift, both of these impacting decision uncertainty.

These same results will be presented at the close of every sub-chapter, i.e. after each issue of temporal uncertainty has been addressed, the impact of incorporating this temporal uncertainty is conveyed for each risk group. However, for simplicity, some analysis carried out within sub-chapters will be illustrated for just one risk group (risk group 3).²²

4.1.6 Temporal uncertainty in the RITA-3 model

4.1.6.1 Time horizon mismatch

The primary sources of data (data from RITA-3 and other relevant clinical trials) pertain to a 5 year time horizon or less. The appropriate analysis time horizon (the time over which costs and effects are expected to differ between the two strategies) however is circa 60 years. Since there are mortality effects related to the competing treatment strategies, it is necessary to impose a lifetime time horizon. 60 years is assumed to be the time horizon after which effectively all patients are deceased.

It is this time horizon mismatch that gives rise to temporal uncertainty within the CEDM.

4.1.6.2 Will this matter?

In this section, it will be considered whether the temporal uncertainties that arise due to this time horizon mismatch could affect the recommendations that the decision model exists to inform.

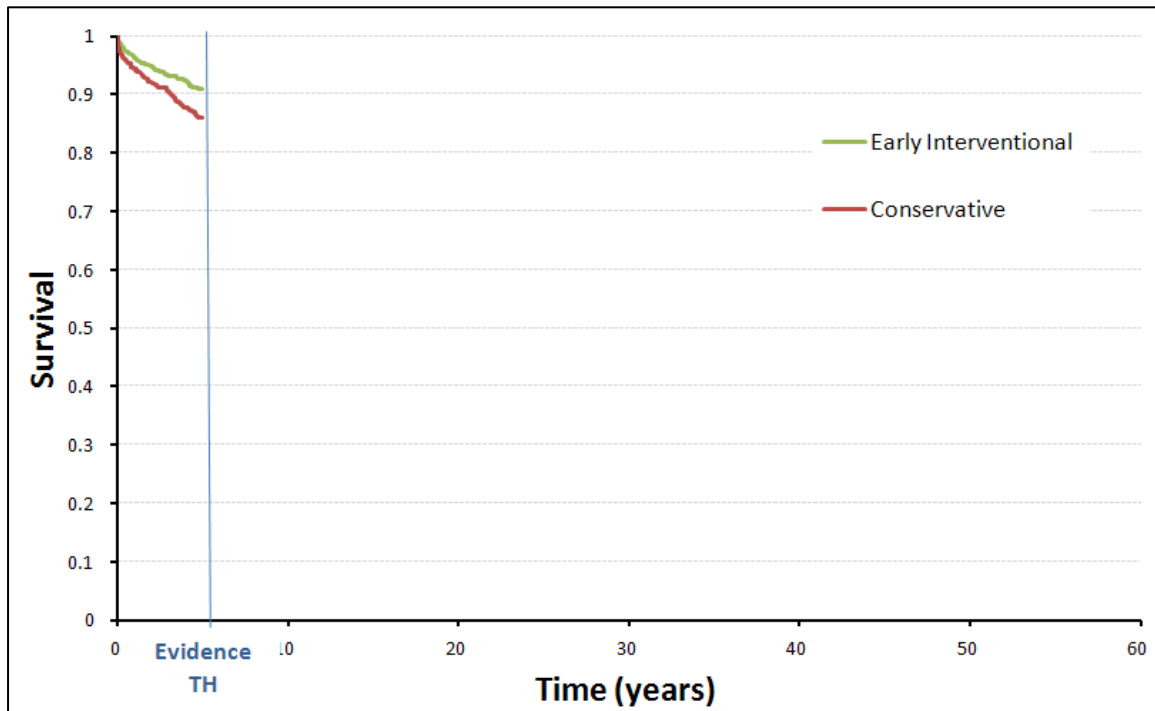
4.1.6.2.1 Observed Period vs. Unobserved Period

To initially gauge the expected impact of temporal uncertainty on the cost-effectiveness results, we can simply consider the disparity between the duration of the observed period (5 years) and the duration of the unobserved period (55 years). We would expect the existence of such a long unobserved period relative to the observed period to potentially lead to substantial temporal

²² To be clear, temporal uncertainty is incorporated into the model results cumulatively. Thus the updated results for one sub-chapter will act as the base-case results for the subsequent sub-chapter.

uncertainty. Also important is the number of events of interest that have occurred within the observed period (i.e. the 'maturity' of the data). Consider, for example, the survival curves in Figure 19 that pertain to patients experiencing a first composite event (myocardial infarction or cardiovascular related death).

Figure 19: Available survival data (Kaplan-Meier curves) against full model time horizon



Not only does this convey the extent of the survival 'space' yet to be filled, it also shows that just 11% of patients have experienced an event within the observed period. As such, these data could be described as immature. However, it must also be noted that outcomes attributable to earlier periods are of more value. In this model, approximately 31% of all costs and QALYs are attributable to the observed period (using, for now, the temporal assumptions applied in the original model). The reason for this is that there are more patients alive or in better health states over earlier periods and costs and QALYs accrued are discounted to a lesser degree.

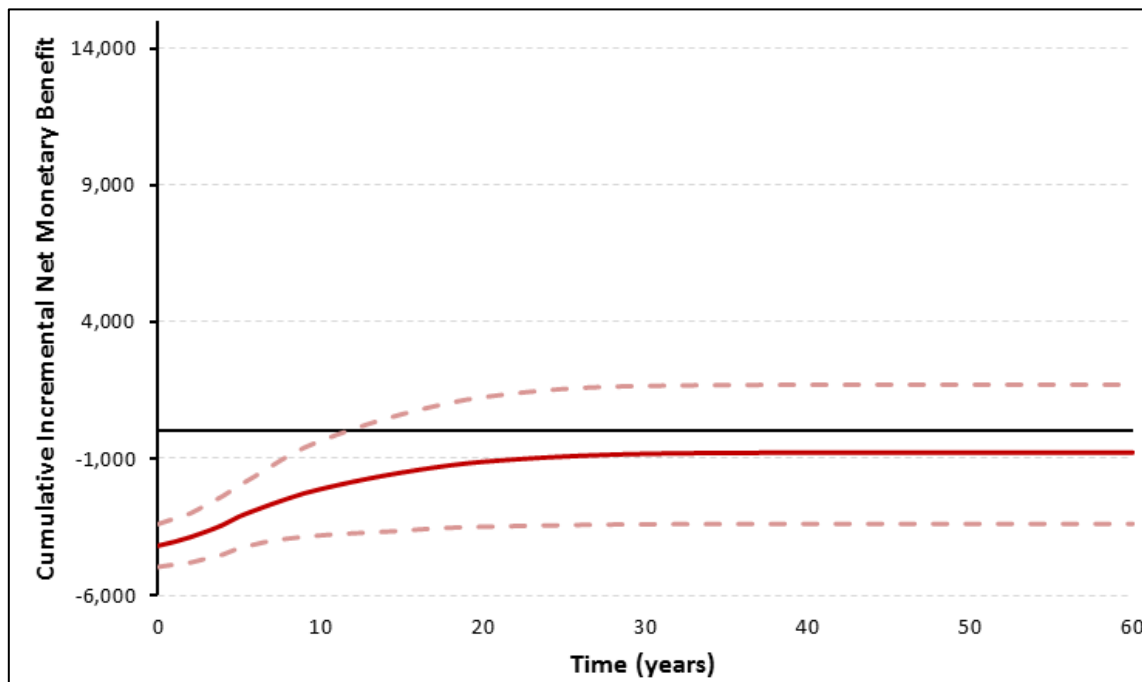
Although this information is an indicator of the extent of the temporal uncertainty that may exist, what ultimately is of consequence is the impact that temporal uncertainty has on the decisions the analysis is designed to inform, i.e. whether or not to adopt a new health intervention, and whether

or not to seek further evidence. We can appraise the role that temporal assumptions may play in generating decision recommendations by considering how cost-effectiveness evolves over time.

4.1.6.2.2 Cost-effectiveness Over Time

If it was the case that the early interventional strategy was found to be cost-effective at the evidence time horizon, we could be confident that this strategy would also be cost-effective at the full time horizon (as per the discussion in Section 2.2.2.1). The reason for this is that the early interventional strategy is the more expensive strategy (with high up-front costs). Cost-effectiveness at the evidence time horizon therefore implies significant health gains over the observed period. Since most of the health gains are survival related and since we don't expect any rebound effect or significant future costs, the early interventional strategy would be expected to continue to accrue more QALYs relative to its comparator over the unobserved period and at no significant extra cost, therefore remaining cost-effective. This in turn would imply that any set of plausible temporal assumptions would not be expected to affect the adoption recommendation. Conversely, if the early interventional strategy was found to be cost-ineffective at the evidence time horizon, then there would be scope for the temporal assumptions to determine the cost-effectiveness at the full time horizon and therefore the adoption recommendation. To observe this, we can calculate the cumulative incremental net health benefit over time (assuming a cost-effectiveness threshold of £20,000).

Figure 20: Cumulative Incremental Net Health Benefit over time for risk group 3 showing mean CINB over time and 5th and 95th percentiles



It can be seen in Figure 20 that net health benefits (for the early interventional strategy) accrue over the observed period but do not reach the point where the early interventional strategy is deemed to be cost-effective (i.e. where cumulative incremental net health benefit > 0). Although, we expect net health benefits to continue to accrue over the unobserved period, it is not certain whether these additional net health benefits will be sufficient for the strategy to be deemed cost-effective (as can be seen from three hypothetical realisations depicted in the graph). In other words, since the early interventional strategy relies on the net health benefits accrued over the unobserved period in order to be deemed cost-effective, it is possible that the adoption decision recommended by the CEDM will depend on the temporal assumptions that are imposed. In short, temporal uncertainty directly impacts the adoption recommendation.

4.1.6.3 What are the issues of temporal uncertainty in RITA-3?

Having established that temporal uncertainty may be a crucial element in this analysis, it is now considered where exactly there exists temporal uncertainty in the CEDM. Essentially all of the key components of the long-term Markov structure (see Figure 18) are subject to some kind of temporal uncertainty, i.e. the transition probabilities between health states, the costs and HRQoL attributed to those states and the assumptions implicit in the overall model structure. Details about each

pertinent model component and the temporal assumptions made in the original analysis are provided in the table below (Table 12).

Table 12: Issues of Temporal Uncertainty in RITA-3 CEDM

Model Component	Where in Model?	Specific Parameter	Variable type	Data availability	Temporal assumption
Transition Probabilities	Transitions between the 'no event' state and the 'MI/CVD' state	Baseline measure (conservative arm of trial)	Time-to-event (TTE) variable. Hazard rates are calculated and transformed into transition probabilities	IPD from RCT (RITA 3) for 5 years	Hazard rates were assumed to stay constant over the unobserved period except for a 10 yearly increase to represent increasing risk with age
		Treatment effect (in order to characterise the 'intervention' arm of the trial from the conservative arm)	Relative effect in the form of a hazard ratio	IPD from RCT (RITA 3) for 5 years and aggregate data from 7 other trials	Pooled treatment effect was applied over trial period. Hazard ratio was then assumed to be 1 after the trial period, i.e. no difference in hazards was assumed
	Transitions between the 'Post MI state' and the 'MI/CVD' state	Baseline measure (conservative arm of trial)	TTE variable. Hazard rates are calculated and transformed into transition probabilities	IPD from RCT (RITA 3) for 5 years	It was assumed that the probability of suffering a second event in the model was the same as the probability of suffering a first (but with the covariate for 'previous event' switched to 1) for 5 years then (if no further event) the hazard reverts to the same as those in the 'no event' state
		Treatment effect (in order to characterise the 'intervention' arm of the trial from the conservative arm)	Relative effect in the form of a hazard ratio	IPD from RCT (RITA 3) for 5 years and aggregate data from 7 other trials	It was assumed that there is no difference between arms in terms of the probability of suffering another event
	Transitions between the 'MI/CVD' state and the 'Dead CV' and 'Post MI' states	Event type (logistic regression). Same for both comparators (i.e. no treatment effect)	Odds ratio. Transition probabilities are then calculated from this.	IPD from RCT (RITA 3) for 5 years	Odds ratio was assumed to stay constant over unobserved period
HRQoL	Composite event & 'No Event' and 'Post MI' health states	Health-related quality of life weight	Factor between 0 and 1	IPD from RCT (RITA 3) for 5 years	HRQoL weights per health state and health event were assumed to stay constant over unobserved period and were assumed equal for both cohorts
Costs	Composite event & 'No Event' and 'Post MI' health states	Resource use and unit costs	Pound Sterling	IPD from RCT (RITA 3) for 1 year	Costs per health state and health event were assumed to stay constant over the unobserved period and were assumed equal for both cohorts
Structure	Entirety of Markov model	n/a	n/a	IPD from RCT (RITA 3) for 5 years	It was assumed that the model structure reflected reality over the long-term sufficiently well for the purposes of the analysis

What Table 12 demonstrates is that when temporal uncertainty is explicitly taken into account, it becomes apparent that there are a number of parameters and model components whose long-term behaviour ought to be examined. In other words, temporal uncertainty pertains to far more than survival curves.

Transition Probabilities

Transition probabilities are the drivers of disease progression in this event-based model. Any analysis of survival curves, treatment effects, multiple cardiovascular events etc. must ultimately be translated into transition probabilities. As Table 12 illustrates, transition probabilities can be broken down into more specific model parameters. In the RITA-3 model, the transitions between each of the key health states merit consideration, with a baseline measure (which represents the conservative treatment strategy) and treatment effect (which when applied to the baseline measure represents the early interventional strategy) for each transition²³. Although there are significant temporal uncertainties relating to each transition, the key transition in this CEDM is the transition from the 'No Event' health state to the 'MI/CVD' health state; that is, the post-intervention risk of experiencing a first composite event (myocardial infarction or cardiovascular related death). The effect of alternative treatment strategies on this risk is the principle outcome of the RITA-3 trial and determines the subsequent pathways in the CEDM. Hence the 'No Event' to 'MI/CVD' component of the model will form the focus of much of this re-analysis.

Costs and Resource Use

There are costs associated with two health states in the long-term portion of the model – there are no costs associated with death, and the 'MI/CVD' health state is assumed to be instantaneous. For both the 'No Event' health state and the 'Post MI' health state, there are constant costs assumed (the same amount for both treatments) for the second and subsequent years. There is also a cost associated with experiencing an MI. The temporal uncertainty relating to costs pertains to the reasonableness of these assumptions. It is plausible that other factors could come into play that may influence the cost of patients inhabiting these health states.

Health-related Quality of Life

As with costs, there are HRQoL 'weights' associated with both the 'No Event' health state and the 'Post MI' health state in the second year after intervention and subsequent years. These weights are assumed to be constant over time and the same for both treatments in the base-case. Again, there is

²³ In Section 4.1.7 the possibility of using absolute measure for both the conservative and early interventional treatments is discussed.

temporal uncertainty relating to these assumptions. It is likely that there is a reduction in HRQoL associated with aging and there could be other factors to consider over the long-term.

Structure

The structure of the Markov portion of the CEDM implies further assumptions about progression over the long-term. The Markov assumption plays only a minor role in this model as there is only one path to the key 'No Event' health state and transitions from this health state can easily be made dependant on the time spent in the health state. The 'Post MI' state is effectively 5 states (with each state representing an additional year alive after a myocardial infarction associated with decreasing risk of another composite event) and so the Markov assumption is circumvented in this part of the model also (for 5 cycles/ years)²⁴. The appropriateness and sufficiency of this time dependency could be questioned. More generally, perhaps more health states are required to reflect the complexity of long-term disease progression and perhaps other long-term clinical events need to be explicitly taken into account.

Although all the potential areas of temporal uncertainty in the RITA-3 model have been outlined above, the scope for this case study is to analyse in detail the temporal uncertainty pertaining to key parameters. In doing so, this chapter will seek to cover the central analytic issues regarding temporal uncertainty. The subsequent sub-chapters will thus focus on:

- (i) Baseline risk of a first composite event (Chapter 4.2)
- (ii) Treatment effect pertaining to a first composite event (Chapter 4.3)
- (iii) Costs and resource use (Chapter 4.4)
- (iv) Health-related Quality of Life (Chapter 4.5)

²⁴ Recall that the 'Markov assumption' means that each model state is memoryless, i.e. when patients are in a health state, it is of no consequence how long those patients have been in that state nor what health states they were previously in.

4.1.7 Parameter Relationships

An issue that ought to be addressed before focus is given to individual parameters is the relationship between some of these parameters. It is inevitable that some parameters will be inextricably linked to others. Certainly in terms of conveying the impact of temporal uncertainty, the effect of a temporal assumption regarding one parameter may be very sensitive to the temporal assumption regarding another. The full picture of (and impact on decision-making of) temporal uncertainty will only become clear after temporal uncertainty has been expressed for each of the key parameters. It will be necessary therefore, in the subsequent sub-chapters of this chapter, to explore how the parameter under discussion relates to other key parameters. It could be argued that some parameters ought to be modelled in tandem. However, it is suggested here that it is more beneficial to focus on one parameter-type at a time, while referring intermittently to the relationship with other parameters, so as to develop methodology specific to that parameter-type.

There is one parameter relationship that warrants particular consideration from the outset; that is the relationship between baseline risk and treatment effect (in terms of a first composite event) and how these two parameters ultimately generate transition probabilities for each of the treatment strategies. The estimation of these parameters is central to the analysis (indeed they both have substantial sub-chapters dedicated to them) and are obviously closely related. It is important therefore that the approach taken to modelling these closely related parameters is explained and justified.

4.1.7.1 Modelling baseline risk and treatment effect

In the Markov (long-term) portion of the model, it is the transition probabilities that (almost solely) differentiate the two treatment strategies (the higher costs for the early interventional strategy occur in the index hospitalisation period which is assumed instantaneous in time). As outlined earlier, the key transition is that from the 'No Event' state to the 'MI/CVD' state. It must be considered what approach to take in order to estimate transition probabilities for both the conservative and early interventional arms of the model given the nature of the evidence available and given the need to characterise temporal uncertainty. This task of estimating transition probabilities for multiple comparators is a common scenario in HTA and is an especially important issue in the context of a need to model beyond the evidence period.

The primary evidence available to inform these parameters comes in the form of 5 year individual patient-level data (IPD) from the RITA3 trial. Additional information on treatment effect in the form of aggregate trial data (specifically odds ratios) is also available. In Section 3.4.2.1, it was outlined how transition probabilities can be derived from the hazard functions calculated from analysis of the survival data. There are a number of alternative modelling approaches that can be employed to produce these hazard functions from the available data for both comparators. It is suggested here that there are five key criteria to consider when choosing a modelling approach.

- (i) Flexibility regarding treatment effect. As we characterise treatment effect over the observed period, it is desirable to allow our estimates to vary with time. Calculating a single estimate for treatment effect (and consequently assuming no change over the unobserved period) may not be appropriate, but importantly, it also makes it difficult to judge how the treatment effect parameter might behave over the unobserved period.
- (ii) Flexibility regarding baseline risk. It is similarly desirable to ensure that baseline risk is modelled with sufficient flexibility to reflect the complexity of the disease. Simple assumptions such as a constant or monotonically decreasing baseline risk may be sufficient, but there ought to be scope to assume a more complex baseline risk.
- (iii) Explicitness regarding treatment effect. It is desirable to be explicit regarding what is being assumed about the treatment effect, for the sake of clarity about the relative effectiveness of an intervention, but also as it is scenarios regarding future treatment effect that are highlighted in the current NICE guidance on extrapolation (NICE, 2013).
- (iv) Consideration/incorporation of all relevant evidence. The consideration of all relevant evidence is a key principle of economic evaluation (Sculpher et al., 2006, Philips et al., 2006). As a result, any modelling approach ought not to preclude the use of any relevant evidence. Although it is debatable as to what can be defined as relevant evidence, in this case study, the immediately relevant evidence comprises the individual patient-level data from the RITA-3 trial and the estimates of treatment effect from other published trials.

- (v) Quantification of correlation between parameters. When employing a survival regression to compute transition probabilities, the availability of individual patient-level data (IPD) allows us to incorporate the correlations between statistical parameters by constructing a Cholesky decomposition²⁵ (Briggs et al., 2006). Correlation between parameters is relevant for running a probabilistic sensitivity analysis. It should be incorporated where possible although it may have negligible impact on the model results. The 5-year IPD available makes it possible to incorporate the correlation between baseline risk and treatment effect. It may be more difficult however to legislate for correlation as temporal changes beyond the observed period are modelled.

Three approaches to using the RITA-3 data in order to produce hazard functions (and ultimately transition probabilities) for both comparators are outlined below. The merits of each are discussed in terms of the criteria listed above.

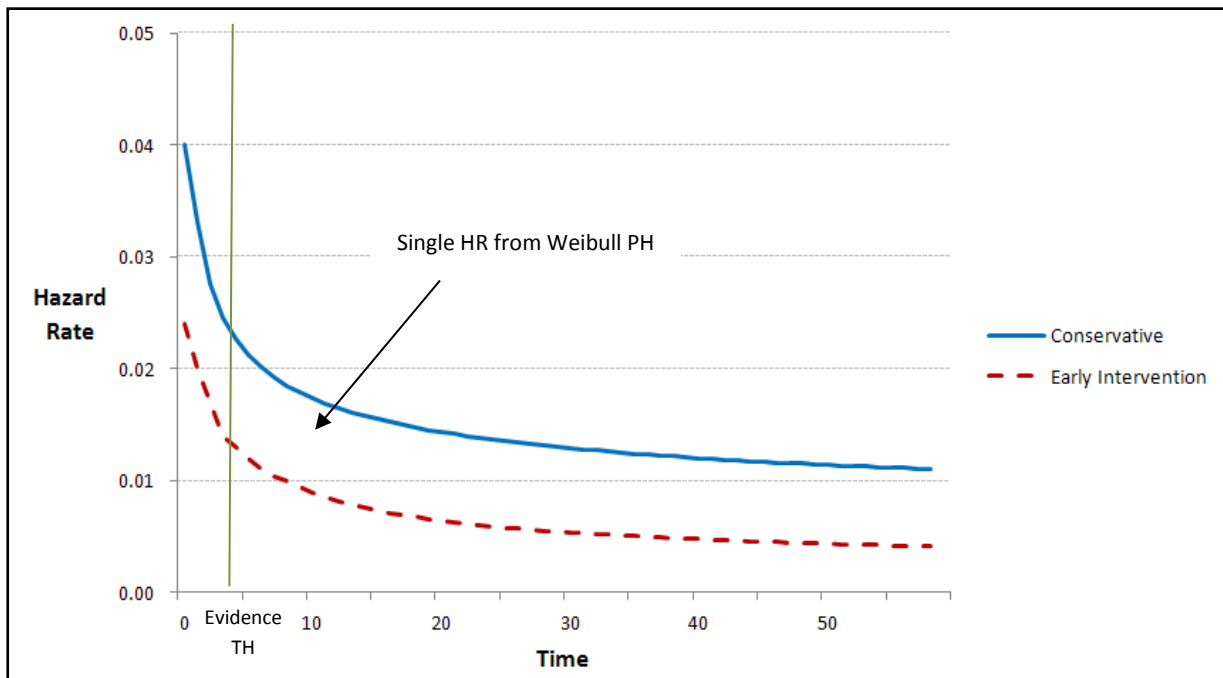
Approach 1:

A common modelling approach in this situation is to assume proportional hazards. This assumes that the effect of treatment is multiplicative with respect to baseline risk and that this effect is independent of time. For the RITA-3 case study, a proportional hazards (PH) survival function can be fit to the survival data in order to simultaneously produce: (i) a baseline hazard function which represents the conservative arm and (ii) a single hazard ratio representing treatment effect which when multiplied by the baseline hazard function produces the hazard function pertaining to the early interventional arm. Separate judgements can then be made regarding the behaviour of both the baseline hazard and the treatment effect over the unobserved period. This is what was done in the original version of this CEDM where a Weibull PH model was fit to the RITA-3 data. The treatment effect applied to the observed period was the hazard ratio from the Weibull PH model, although a pooled treatment effect comprised of the hazard ratio from the Weibull model and the odds ratios from the external data was applied in a sensitivity analysis. Although this approach is straightforward, simply assuming proportional hazards and estimating both parameters in one regression equation is restrictive with regard to extrapolating in order to estimate temporal behaviour over the unobserved period. Though maybe not unreasonable, we are constrained to

²⁵ Note the distinction between model input parameters (parameters that are input directly into the decision model, e.g. transition probabilities, costs, utilities) and statistical parameters (parameters that contribute to a regression equation that may help generate a model input parameter, e.g. age, sex, smoker). Parameters like treatment effect and baseline risk could be placed into either category. In this instance, we are concerned about the correlation between these two when thought of as statistical parameters, but generally in this chapter and elsewhere they are discussed as model input parameters.

assuming that both hazard functions follow a Weibull distribution, while the PH model produces a single estimate of treatment effect for the observed period without estimating how treatment effect might evolve with time. This approach does incorporate estimates of correlation between baseline and treatment effect. It is also reasonably flexible with regard to baseline risk as not only can standard PH model be employed, but also more flexible models like the Royston-Parmar flexible parametric model. Approach 1 is depicted in the Figure 21 below. (Note that the extrapolated hazard curves depicted are hypothetical. In fact, the original model assumed constant post-trial hazards and no further treatment effect.)

Figure 21: The ‘proportional hazards’ approach to modelling this component of the observed period



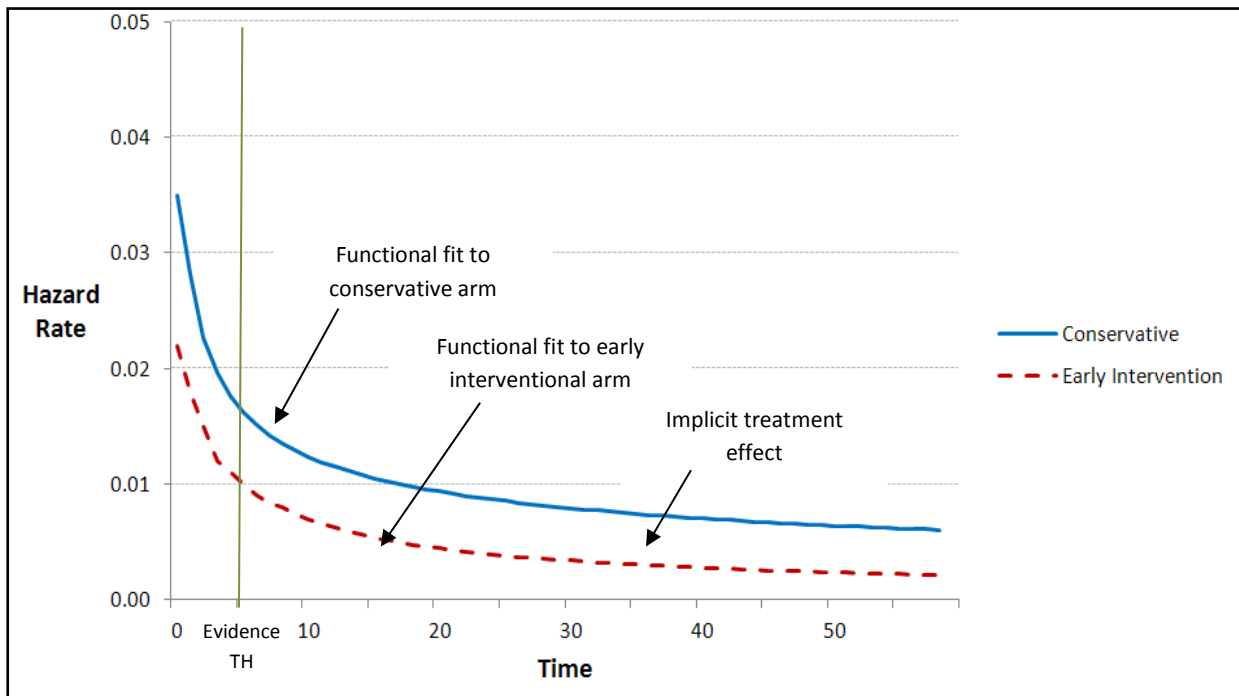
Note: In the original model, assumptions of constant hazards and no further treatment were made when extrapolating. Note that this graph does not account for the change in hazards when the age covariate is updated

Approach 2

An alternative approach that might be less restrictive and more conducive to extrapolating evidence over time would be to fit separate survival distributions to each arm of the trial data (for example Weibull and Log-Normal). No relationship would then be assumed between the arms. Each arm could be independently extrapolated over the unobserved period, either assuming a continuation of the survival distributions, or making alternative assumptions regarding the behaviour of the hazards. Since this approach does not consider a relative effect, but separate absolute risks for each arm, it is

not necessary to only select survival distributions that use a PH metric. This approach could arguably be thought of as very flexible in terms of both baseline risk (as distributions using non-PH metrics can be used) and treatment effect (as treatment is allowed implicitly to change over time). However this approach also implies that we will not use the supplementary aggregate data available as they pertain to treatment effect specifically. It also involves making an implicit, rather than explicit, judgement regarding the magnitude of treatment effect and how this evolves over time. Furthermore, this approach makes it more difficult to incorporate correlation between (in this case) the absolute hazards (assuming that distributions are fit separately to the trial arms). This ‘independent arms’ approach is depicted in Figure 22.

Figure 22: The ‘independent arms’ approach to modelling this component of the decision model

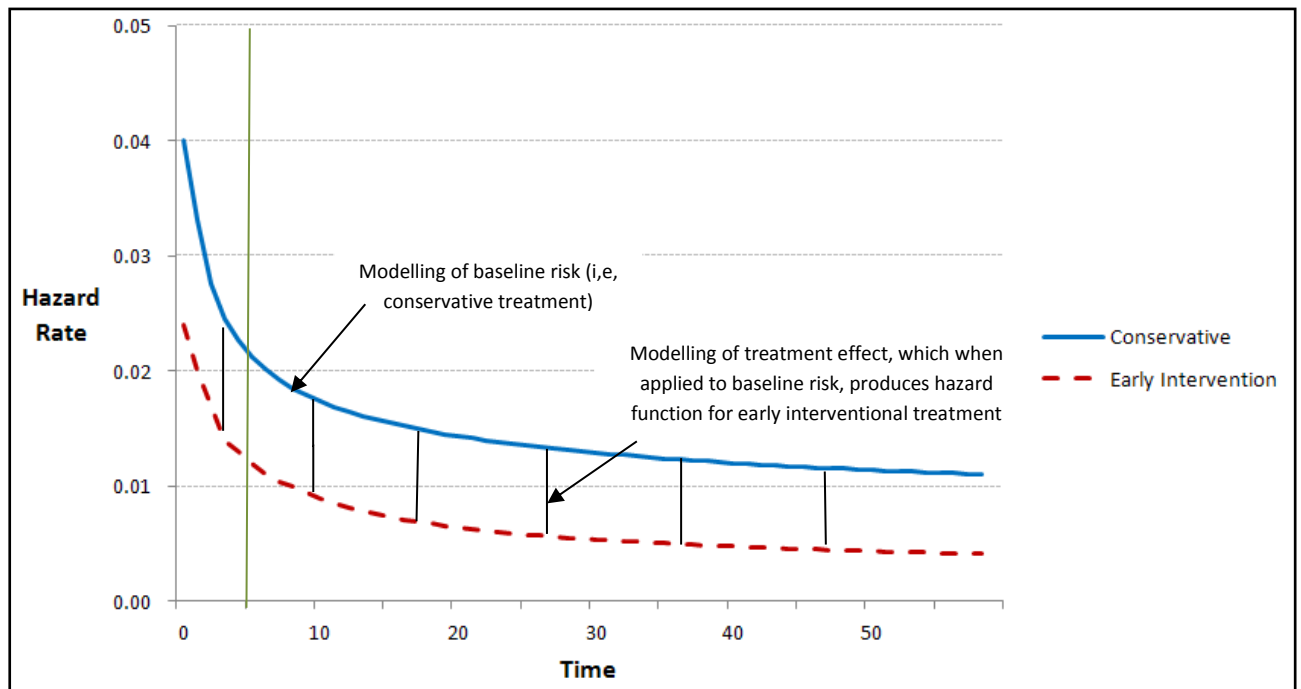


Approach 3

A third approach would be to focus on baseline risk and treatment effect separately, essentially analysing the IPD twice in order to estimate these two parameters. The baseline risk (i.e. the conservative arm of the trial) would be modelled in relation to the observed period from which alternative long-term scenarios would be generated based on the temporal trend observed in the trial period, or on external evidence relating to natural history, or on other clinically plausible scenarios. The treatment effect over the observed period would be analysed and modelled while ignoring the nature of the baseline. This would allow for a time-dependent hazard ratio and the

inclusion of hazard ratios from other evidence sources if appropriate. Future temporal behaviour can then, to the greatest extent possible, be based on the temporal trends observed. Even if the function used to represent baseline risk implies a different treatment effect to the one used in the decision model, I do not consider this contradiction as problematic as of course a model is 'correct' in so far as it is useful. This approach, like Approach 1, is restricted to using only PH distributions to characterise baseline risk. However, with the availability of an array of PH distributions including very flexible models like the Royston-Parmar model, this approach would appear to ensure sufficient flexibility. Also, like Approach 1, this approach could incorporate the correlation between baseline risk and treatment effect (it would seem reasonable to use the correlation estimate from a PH model including treatment effect even if a different estimate of treatment effect is ultimately used). Note that an extension to the Royston-Parmar model has been developed which allows both a flexible baseline hazard and a time-dependent treatment effect, and jointly estimates both. Although, this precludes the use of other estimates of treatment effect, it may be a useful model to employ under some circumstances. The approach of considering baseline risk and treatment effect separately is depicted in the Figure 23 below.

Figure 23: Approach of modelling baseline risk and treatment effect separately



Choosing the most appropriate approach clearly requires judgment and there may be more criteria to consider than those outlined above. It is also not a case of choosing the approach that satisfies

the most criteria as some ought to carry more weight than others. However, judging these alternative approaches by the criteria described above is a useful process.

Table 13: How the alternative approaches fare in terms of the three criteria

Criterion	Approach 1	Approach 2	Approach 3
Flexibility re treatment effect		✓	✓
Explicit re treatment effect	✓		✓
Incorporate all relevant evidence	✓		✓
Flexibility re baseline risk	✓	✓	✓
Quantification of correlation	~✓		~✓

For this analysis, Approach 3 incorporates, and best exploits, the relevant evidence available while also allowing explicit and flexible modelling regarding both baseline risk and treatment effect. It is suggested here that modelling baseline risk and treatment effect separately best facilitates extrapolation and ultimately appropriate characterisation of temporal uncertainty in this circumstance. Furthermore, in terms of addressing temporal uncertainty, baseline risk and treatment effect pose quite different challenges and so it is useful, for the purposes of this re-analysis, to dedicate separate sub-chapters to them.

The remaining sub-chapters in Chapter 4 will now explore in detail the key issues of temporal uncertainty in RITA-3 and seek to appraise and develop methodology with a view to constructing a comprehensive methodology for appropriately addressing temporal uncertainty in CEA.

4.2 Chapter 4.2: Baseline Risk

4.2.1 Introduction

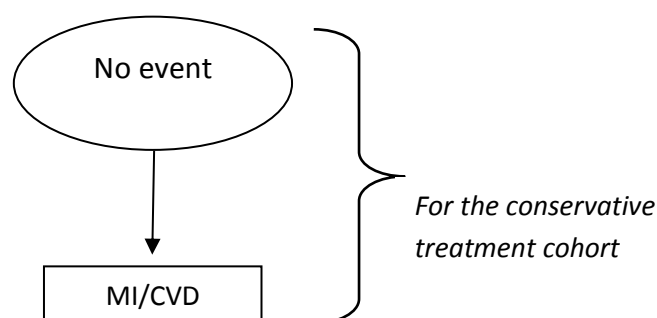
Following the overview given in Chapter 4.1, the purpose of this chapter is to address, in detail, the temporal uncertainty relating to the first of the key parameters in the RITA-3 model: baseline risk of a first composite event (myocardial infarction or cardiovascular-related death).

In particular, this chapter endeavours to explore: the usefulness and limitations of employing survival analysis to extrapolate a time-to-event variable; the issue of accounting for the effects aging when estimating baseline risk over the long-term; and crucially, methods that appropriately express uncertainty for estimates of baseline risk beyond the observed period.

4.2.1.1 The parameter

The model parameter under analysis is the baseline risk of experiencing a first composite event after intervention. Baseline, in this instance, refers to patients subject to the conservative treatment strategy. Baseline risk therefore will be converted into annual transition probabilities that determine the proportion of the 'conservative cohort' that transition from the 'No Event' health state to the 'MI/CVD' health state as illustrated in Figure 24 below. A treatment effect (the derivation of which is the subject of Chapter 4.3) will then be applied to this baseline risk in order to determine the risk, and ultimately the transition probabilities, for patients subject to the early interventional treatment strategy.

Figure 24: Baseline risk refers to the transition between the 'No Event' health state and the 'MI/CVD' health state for patients receiving the conservative treatment strategy



4.2.1.2 Challenges regarding estimating baseline risk over the long-term

The condition in question, non-ST-elevation acute coronary syndrome (NSTEMI-ACS), is a collective term for unstable angina and non-ST elevated acute myocardial infarction and relates to a spectrum of diseases that involves an imbalance of supply and demand of oxygen available to the myocardium (Anderson et al., 2007). Patients presenting with these symptoms represent a heterogeneous group with a wide variety of clinical outcomes (Grech and Ramsdale, 2003). As such, it is not straightforward to model the long-term prognosis for a patient presenting with NSTEMI-ACS.

To compound this difficulty, the risk that requires characterisation is that of a 'composite event', which is defined in this study as a cardiovascular death or myocardial infarction (Henriksson et al., 2008). Even if it were feasible to obtain long-term mortality data for this particular patient population, it is unlikely that long-term data on the prevalence of such composite events could be obtained (to say nothing of heterogeneous risk between risk groups). It may be possible however, to assess the validity of the implied mortality from the model. If nothing else, the implied long-term mortality can be compared to that of the general population.

Age is recorded in the RITA-3 trial by categorizing patients into 'age groups' (group 0=50's, group 1=60's, group 2=70's, group 3=80's, group 4=90+). Although it is possible to assess the impact of the 'age group' covariate on the hazard rate over the observed period, it is more difficult to assess the effect of continuous aging as patients remain in their age groups for the duration of the trial (5 years). There are essentially two factors to contend with as we consider the behaviour of hazard rates over time: (i) the change in risk associated with event-free survival and (ii) the change in risk associated with ageing. This combined effect of ageing and event-free survival is not adequately captured in the RITA-3 data. This shortcoming is indicative of the fact that RCTs are seldom designed with estimation beyond the trial period in mind. In the original CEDM developed by Henriksson et al. (Henriksson et al., 2008), the change in risk associated with age over the unobserved period was modelled by updating the age covariate in the Weibull PH model every ten years (as patients are assumed to transfer into the next age group), while there was assumed to be no further change in hazard rates (after five years) associated with time since hospitalisation. The appropriateness of these assumptions will be central as we characterise temporal uncertainty regarding baseline risk.

4.2.2 Available evidence

4.2.2.1 Evidence from RITA-3

The key data available to inform baseline risk come in the form of 5 year individual patient-level data (IPD) from the RITA-3 trial. These data are time-to-event or survival data which means that parametric survival analysis may be employed to characterise risk over the observed period and potentially extrapolate beyond, as described in Section 3.4.2.1.

4.2.2.2 Further evidence

As part of this re-analysis, a search for further relevant evidence was carried out.

Firstly, data were collected pertaining to UK population cardiovascular-related mortality (British Heart Foundation Health Promotion Research Group, 2010, Office of National Statistics, 2011). These data was collected to serve two purposes: (i) to act as a lower-bound to the cardiovascular mortality rate implied by the model and (ii) to give an indication of the change in risk of a cardiovascular death with age.

Secondly, a search of UK health technology assessments (HTAs) was carried out to take into account previous assumptions made regarding long-term event risk for patients that presented with NSTEMI-ACS. Six full HTAs containing the term “non-ST” in the text were found. The content of these studies is discussed in Section 4.2.3.4.5. Some other studies outside of the HTA literature are also included in the discussion of long-term prognosis of patients with NSTEMI-ACS.

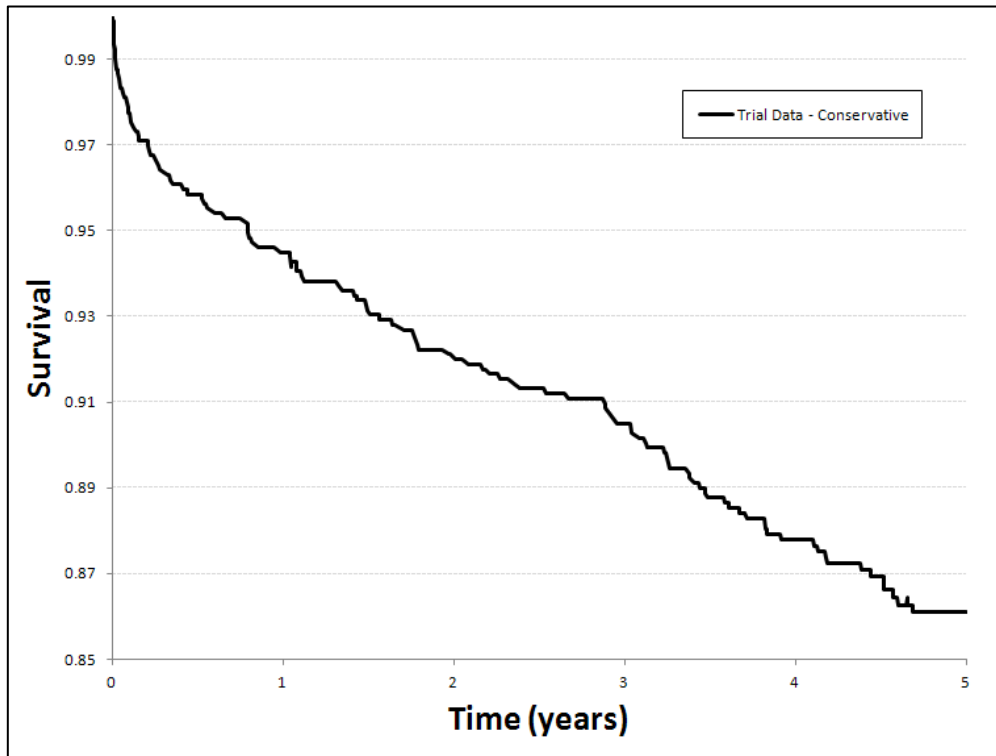
4.2.3 Analysis

4.2.3.1 Analysing and interpreting the survival data

It is first necessary to compute hazard rates (and then transition probabilities) for the observed period, not only because these inputs are required in the CEDM, but also because characterisation of baseline risk over the observed period may partially inform our characterisation of baseline risk over the unobserved period. To these ends, the available individual patient-level data (IPD) from the RITA-3 trial which pertain to the first 5 years after intervention can be analysed.

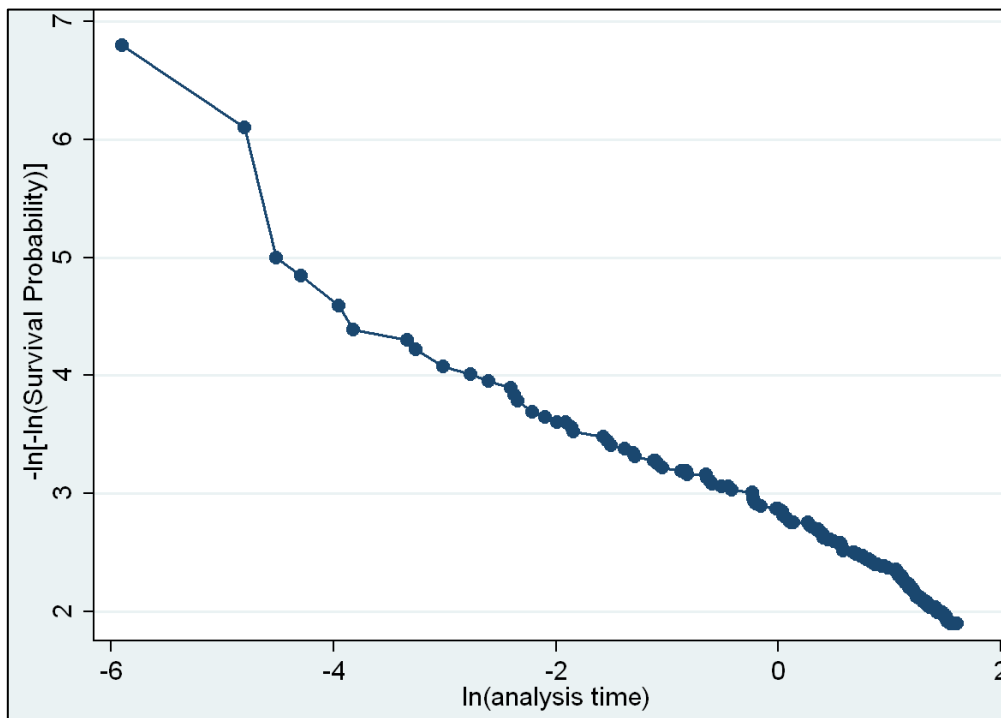
Figure 25 shows the survival data, as a Kaplan-Meier curve, from the RITA-3 trial for those given the 'conservative' treatment (i.e. baseline risk).

Figure 25: The baseline survival data from RITA-3



This graph indicates that the decrease in survival is quite rapid over the first year and then relatively stable for the remaining years. The nature of baseline risk in relation to time can be further investigated by constructing “-log-log” plots, where the negative log of the cumulative hazard function is plotted against the log of time (Cleves, 2008).

Figure 26: “-Log-Log” plots of the baseline survival data



Like the Kaplan-Meier curve, this plot indicates that the behaviour of the hazard rate alters at a particular point over the observed period. An approximately straight line on the –log-log plot would indicate that the survival data would be well represented by a Weibull distribution. A slope of approximately -1 would indicate that the data would be well represented by an exponential distribution (Kleinbaum and Klein, 2005)²⁶.

This above plot therefore suggests that:

- (i) The data may be adequately represented by a Weibull distribution, although hazard behaviour seems only to become stable after an initial period of higher hazards, or
- (ii) The latter portion of the data may be well represented by a Weibull or even Exponential distribution where the hazard rate appears to monotonically decrease at a steady rate. But to also capture the initial less stable behaviour of hazards, a more flexible modelling approach may be desirable.

²⁶ Note that similar plots can be constructed to test for the suitability of other common distributions.

4.2.3.2 Fitting parametric functions to the survival data

It is desirable to compute hazard rates by fitting a parametric function to the available survival data. Recall from Section 3.4.2.1 that there may be a number of reasons to do this, of which extrapolation beyond the observed period is just one.

As part of this re-analysis, the following parametric distributions (each of which were discussed in Section 3.4.2.1) were fit to the survival data:

- (i) Exponential (PH)
- (ii) Weibull (PH/AFT)
- (iii) Gompertz (PH)
- (iv) Royston-Parmar (PH)
- (v) Log-Normal (AFT)
- (vi) Log-Logistic (AFT)
- (vii) Generalised Gamma (AFT)

With the exception of the Royston-Parmar and Gompertz models, each of these distributions is a special case of the Generalised F distribution (Cox, 2008). Four of the distributions (Exponential, Weibull, Gompertz and Royston-Parmar) use the proportional hazards metric while three (Log-Normal, Log-Logistic and Generalised Gamma) use the accelerated failure time metric²⁷.

The survival regressions were carried out in Stata version 11 (StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.). Table 14 shows the statistical fit (AIC and BIC) to the data for each distribution. In terms of the 'best' statistical fit, the Royston-Parmar model achieves the best AIC score, while the Weibull distribution achieves the best BIC score. The reason

²⁷ Note that the Weibull model can also use the accelerated failure time metric and the Log-Logistic model can also use a proportional odds metric (Collett, 2003).

that the Weibull outperforms the Royston-Parmar in terms of BIC is that the Royston-Parmar has several more parameters than the Weibull and BIC penalises complexity more than AIC (Cox et al., 2006). It has been suggested that AIC is a preferable adequacy measure in these circumstances as it has superior predictive validity (Jackson C et al., 2009).

Table 14: AIC/BIC scores for each parametric distribution

Parametric distribution	AIC	BIC
Weibull (PH)	1739.422	1805.071*
Exponential (PH)	1812.143	1872.322
Gompertz (PH)	1786.519	1852.168
Royston-Parmar (PH)	1736.351*	1812.942
Log-logistic (AFT)	1743.217	1808.066
Log-Normal (AFT)	1755.757	1821.406
Generalised Gamma (AFT)	1739.966	1811.086

Note: A lower AIC/BIC score indicates a better statistical fit. * indicates the ‘best fit’ for each measure

Let us also consider each distribution’s visual fit to the empirical survival data and the implied yearly transition probabilities over the observed period²⁸. Note the immaturity of the survival data at the evidence time horizon.

²⁸ Transition probabilities were calculated from the hazard rates as per the formulae shown in Section 3.4.2.1. These calculations were more cumbersome for some distributions (Generalised Gamma, Royston-Parmar, Log-Logistic) than for others (Exponential, Weibull).

Figure 27: Empirical survival data (Kaplan-Meier curves) against the survivor functions of the parametric distributions

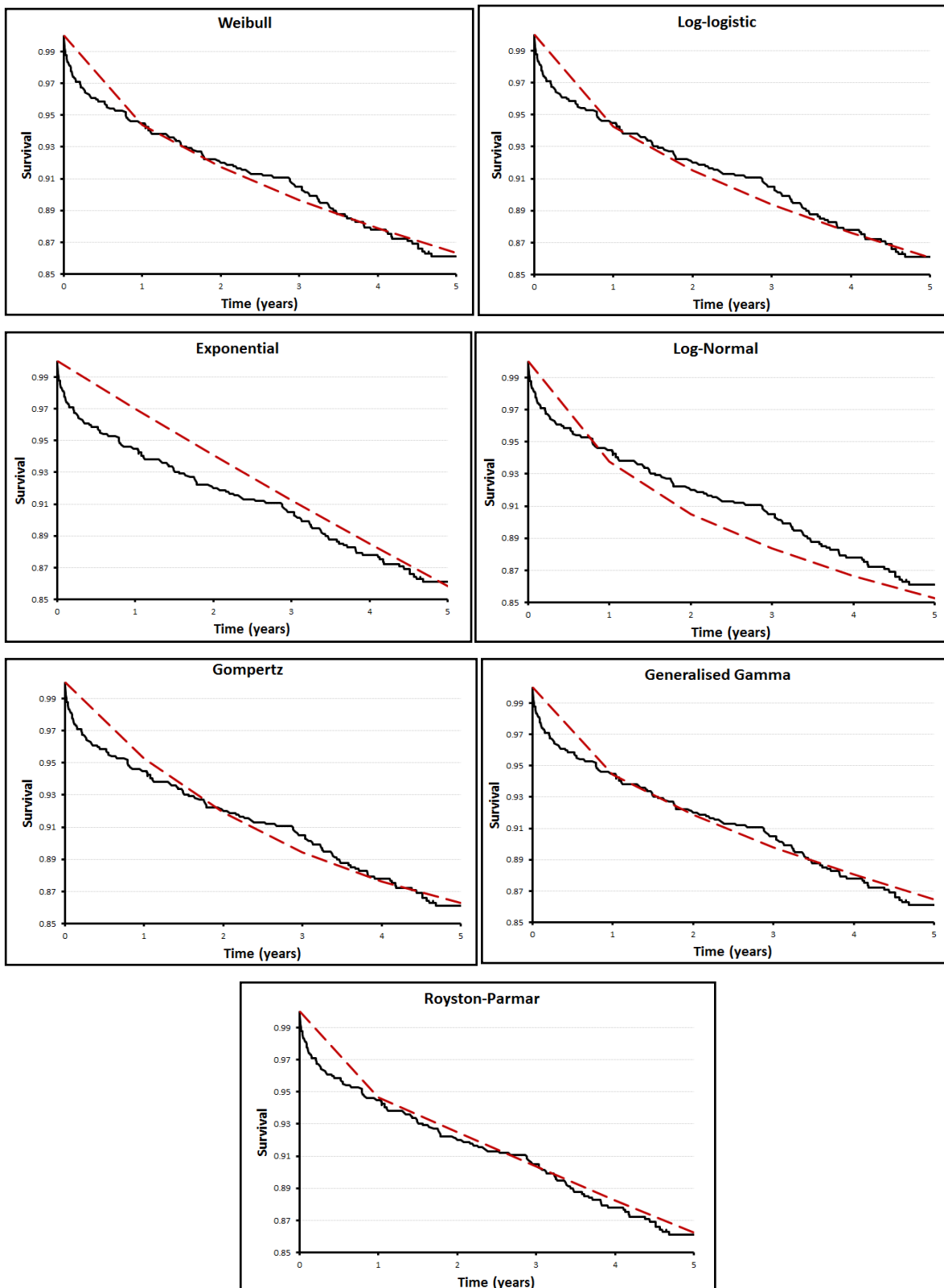
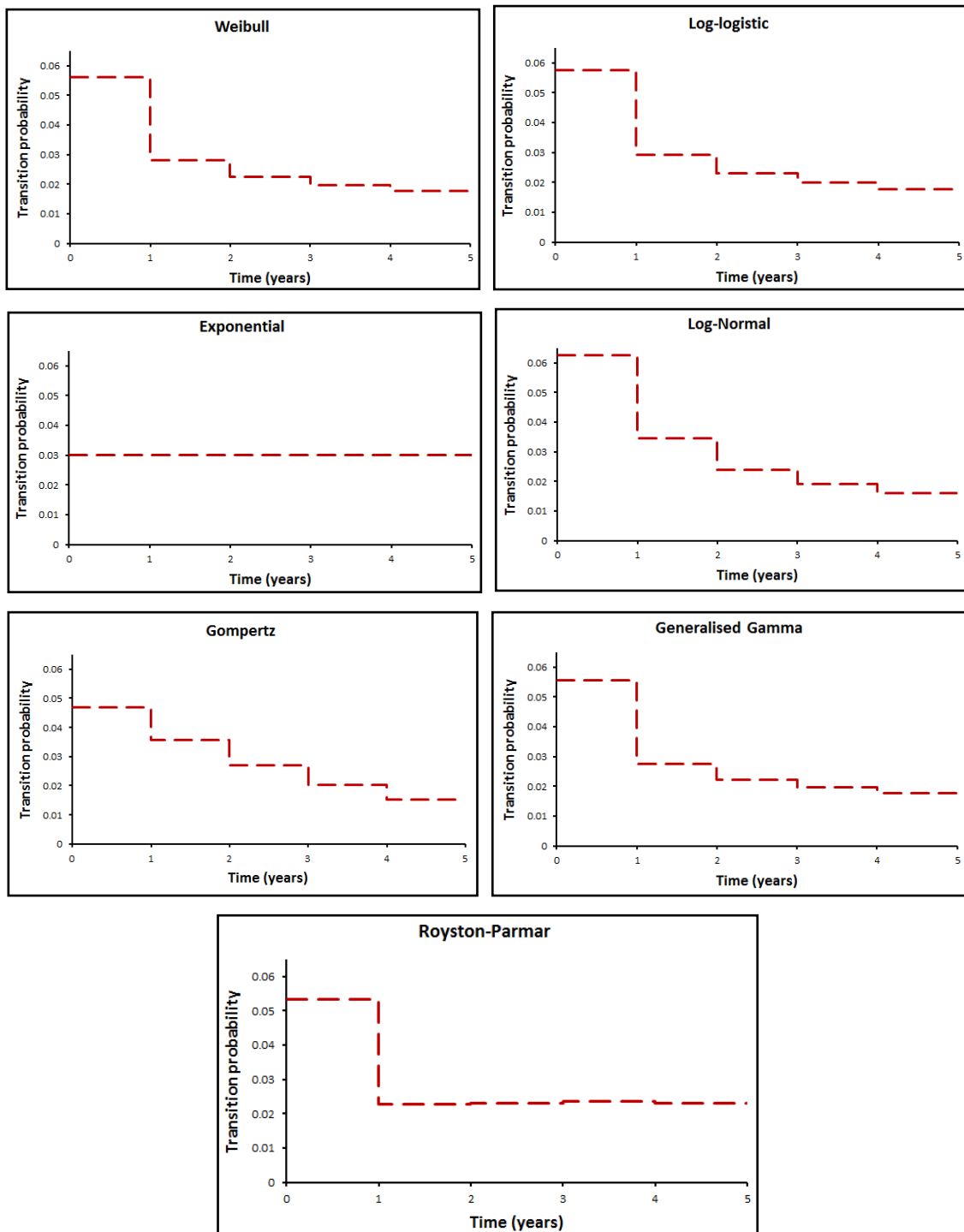


Figure 28: The resultant yearly transition probabilities over the observed period (5 years) for each distribution



There are four notable features of these plots.

First is the unsuitability of the exponential distribution for characterising the observed period. The single parameter characteristic of the exponential distribution would seem to render it too restrictive for these survival data. As well as the exponential survival curve not fitting well to the Kaplan-Meier curve, the results of the Weibull regression (mirrored in the results of other regressions) demonstrated that a declining hazard with time over the full observed period was highly probable (i.e. the shape parameter of the Weibull function was less than 1 with statistical significance indicating a decline in hazard rate over time), thus it would seem that an assumption of constant hazards over the observed period (as imposed by the Exponential distribution) would be inappropriate.

Second is that most of the distributions (all but the Exponential and the Royston-Parmar) indicate a monotonically decreasing hazard. Each of these distributions is tending towards constant hazards as time elapses (i.e. each is represented by a convex hazard curve). A plausible clinical explanation for declining hazards is that as time elapses from time of hospitalisation without experiencing an adverse clinical event, the probability of a patient experiencing such an event decreases.

Third is the flexibility of the Royston-Parmar model and the suggestion from fitting this model that hazards are high for the first year and then drop to a steady level for the remainder of the observed period. It seems clinically credible that there may be an initial 'hazardous period' of one year after hospitalisation/intervention, after which the hazard rate levels off and becomes approximately constant. Such a feature would not be captured by the other less flexible models. A clinical explanation for the slight increase in hazards during the fourth year as indicated by the Royston-Parmar model is not immediately apparent. In this instance, the Royston-Parmar model may be guilty of 'overfitting' (Harrell et al., 1985).

Fourth is that arguably none of these functions fit the data particularly well. It is highly likely that were we to attempt to fit a multi-part Weibull, or a Royston-Parmar with more carefully chosen splines²⁹, a better fit would be achieved. However, it is unlikely that an improved fit would impact the cost-effectiveness results significantly. The difference in mean ICER between the two "well-fitting models" (Weibull and Royston-Parmar) is relatively minor as we shall see below.

²⁹ For example, two or three splines to represent the first year plus a single spline to represent the subsequent four years would likely represent a well-fitting Royston-Parmar model

It can be concluded from exploring this range of distributions that there are three broad competing assumptions that could be made regarding the nature of baseline risk over the observed period.

- (i) The hazard rate stays constant over the entire period (as represented by the exponential distribution), though this can be all but ruled out.
- (ii) The hazard rate steadily decreases over the observed period (best represented by the Weibull distribution).
- (iii) The hazard rate is high for the first year but then decreases and remains approximately constant over the next four years (as represented by the Royston-Parmar distribution).

4.2.3.3 Model averaging over parametric functions

Instead of attempting to fit increasingly complex parametric models, let's assume one of the above functions appropriately characterises baseline risk over the observed period. Because none of the AFT distributions has the best statistical fit and we have a strong preference for distributions that use the proportional hazards metric (so hazard ratios from other trials may be incorporated when it comes to estimating treatment effect), it is assumed that uncertainty regarding baseline risk in the observed period (over and above the parameter uncertainty) pertains only to which PH distribution is "correct". This kind of structural uncertainty can be expressed through model averaging.

Bayesian model averaging involves ascribing relative weights to a number of alternative models or model assumptions and averaging across each model to produce a posterior predictive distribution that represents the expected model outcome as well as the related uncertainty (Leamer, 1978).

To incorporate the Weibull, Royston-Parmar, Gompertz and the Exponential distributions into the base-case decision model, a distribution can be selected for each Monte Carlo simulation according to a vector of probabilities $p = (p_W, p_{RP}, p_E, p_G)$ where $\sum p_i = 1$.

$$Survival\ distribution = \begin{cases} Weibull, & \text{with probability } p_W \\ Royston - Parmar, & \text{with probability } p_{RP} \\ Exponential, & \text{with probability } p_E \\ Gompertz, & \text{with probability } p_G \end{cases}$$

To generate these probabilities, an adequacy measure is required. Jackson et al. describe how posterior model probabilities can be derived from an adequacy measure such as AIC by using the following formula (Jackson C et al., 2009):

$$p(M_k|x) = \frac{1}{1 + \sum_{r \neq k} \exp\{-\frac{a_r - a_k}{2}\}}$$

where $p(M_k|x)$ is the probability assigned to model k given data x and a_r is the adequacy measure associated with model r .

Applying this formula to the AIC results in Table 14, the following probabilities are obtained:

$$\begin{aligned} p_W &= 0.1772 \\ p_{RP} &= 0.8228 \\ p_E &\cong 0.0 \\ p_G &\cong 0.0 \end{aligned}$$

As expected, the ill-fitting Exponential and Gompertz distributions are given a weight of approximately zero, so they are effectively excluded. The Royston-Parmar characterisation of the survival data is given the highest probability, suggesting that the hazard rate is high for the first year after hospitalisation and then constant for the remaining four years of the observed period. The parameter ' p ' can now be incorporated into the probabilistic sensitivity analysis and treated like any other probabilistic parameter, i.e. for each simulation, the Weibull distribution (and therefore its associated transition probabilities) is selected with probability 0.1772 and the Royston-Parmar with probability 0.8228.³⁰

By incorporating alternative characterisations of the short-term survival data, the uncertainty around how these data should be interpreted is characterised and the expected nature of baseline risk over the observed period is appropriately expressed. However, as these alternative functions are 'extrapolated', to what extent is temporal uncertainty (uncertainty regarding what happens beyond

³⁰ These probabilities represent a discrete distribution that expressed the uncertainty around which functional fit ought to be used. It is not suggested that there is any need for further uncertainty to be expressed around these probabilities (or the underlying adequacy measure).

the observed period) being captured? Note that in characterising baseline risk over time, there are 3 sources of uncertainty being quantified: (i) the parameter uncertainty surrounding the values of the statistical parameters that make up a parametric function representing baseline risk over the observed period, (ii) the structural uncertainty surrounding which parametric function best represents baseline risk over the observed period³¹ and (iii) the uncertainty surrounding how baseline risk will evolve over the unobserved period (temporal uncertainty).

4.2.3.4 Moving beyond the Observed Period

4.2.3.4.1 Extrapolating the parametric functions

Frequently in HTAs, survival curves are extrapolated directly from parametric functions fit to the short-term trial data. Some process of validation may then be undertaken to choose the most 'plausible' long-term curve. Figure 29 below illustrates the long-term survival curves that result from extrapolating each of the parametric distributions discussed in Section 4.2.3.2. Note that the Gompertz and Exponential distributions represent the most disparate outcomes, while the AFT distributions offer similar outcomes to the Weibull but with slightly poorer statistical fit to the short-term data

³¹ To be clear, there is a discrete distribution representing the uncertainty surrounding which parametric survival function should be used, i.e. an array of probabilities that sum to one. There is not uncertainty expressed regarding any of these probabilities, i.e. for every (Monte Carlo) simulation, each survival function is given the same probability value, based on the AIC score calculated for that survival function.

Figure 29: The long-term survival curves extrapolated directly from the parametric functions described in section 7.1.2

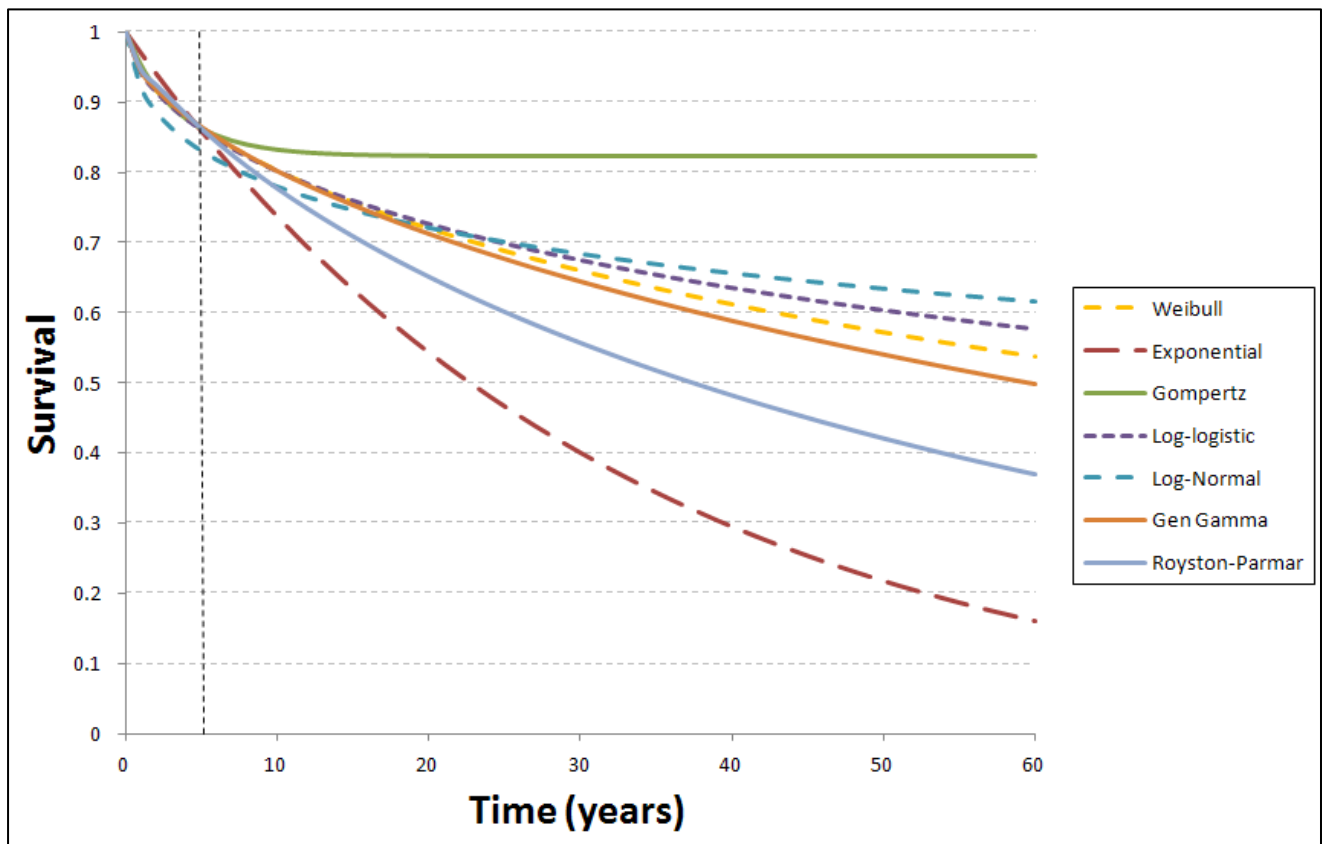


Figure 29 depicts disparate and in some cases manifestly implausible survival curves. The Gompertz survival curve, for example, implies cardiovascular mortality that is less than that of the general population which would not be expected for this cohort of patients that have presented with NSTEMI-ACS. In fact, all of these curves appear overly optimistic. This is because in every case (except the Exponential), hazard rates continue to decline over time leading to optimistic long-term survival. However it is likely that ageing will result in increased risk over the long-term. This fact renders direct extrapolation from the short-term data inherently unreliable.

Extrapolating directly from short-term data (usually through parametric modelling) is a common technique in health technology assessment. Such a technique may be manifestly inappropriate as is the case here. However, the primary issue in this example is the very long (in absolute calendar time) unobserved period. Age effects and other long-term factors are often the main drivers of change in parameter values over a period of this length, but these factors are not (directly) captured by the short-term data. Simple extrapolation from the short-term data therefore, is likely to be useful when

the unobserved period is short enough so that the effects of ageing and other long-term factors are negligible, but of limited use when the unobserved period is of significant length.

In order to characterise the change in baseline risk over the long-term, there are three factors that ought to be accounted for.

- (i) Event-free survival
- (ii) Period risk
- (iii) Ageing

4.2.3.4.2 Event-free survival

Event-free survival here means time having passed without a patient experiencing the event of interest assuming the patient does not age, i.e. it relates to (the dissipation of) the effect of intervention rather than the effect of ageing or any other effect relating to the passage of time. Event-free survival would be expected to influence the risk of a patient experiencing a composite event in the future. It is not clear whether this would be a positive or negative effect. For example, an event-free passing of time could indicate the good health of a patient, or it is possible that the benefits of intervention are dissipating and baseline risk is thus increasing. This effect can be assumed to be captured (to at least some extent) by the trial follow-up. In the RITA-3 example, baseline risk was observed to decline as time since intervention increased (assuming the 5 year time frame was too short for the effects of ageing to be captured). Indications from the analysis of the survival data were that baseline hazards were tending towards constant as time since intervention increased. In the case of the Royston-Parmar model (which had the best statistical fit), hazards were approximately constant from year 2 onwards.

Based on this available evidence, an expectation of no further change in risk due to event-free survival would appear reasonable (though such an assumption ought to be validated by a clinical expert). Without the means to posit an alternative assumption (with current evidence), it is not clear how uncertainty might be expressed regarding this assumption. Yet uncertainty is undoubtedly present at this stage. In this case it may be prudent to incorporate general stochasticity to express such uncertainty. Such a step is common in fields such as macroeconomic forecasting, but has not, to date, been employed for economic evaluation of healthcare. The incorporation of stochasticity

could be effected by modelling the parameter's development over the unobserved period as a simple Weiner process of the form:

$$W_{t+1} = W_t + W_s$$

where the increment W_s is Gaussian with mean zero and variance s (Horrocks and Thompson, 2004).

Such a modelling approach may not aptly represent the nature of this uncertainty however. A more sophisticated modelling approach that incorporates a stochastic drift, such as a variation of the Lee-Carter method might be more suitable (Lee, 2000). It is unlikely however, that employing such (relatively) complex methods would be a worthwhile endeavour (in this case at least), as this is not even the principal source of temporal uncertainty for this parameter. This matter will be revisited in Section 4.2.3.5 (expressing uncertainty).

4.2.3.4.3 Period risk

Period risk refers to the potential change over the long-term of medical care and population lifestyles, leading to changes (most likely improvements) in health outcomes.

In the RITA-3 example, the analysis time horizon is 60 years. It is unlikely that the risk of experiencing a composite event will be the same for patients today and patients 50 years from today, even if their characteristics are identical (including age and time since intervention). It is important to recall however, that these most distal parameter values are likely to be relatively insignificant for the reasons given in Section 4.1.6.2.1.

Period risk is similar to the well-documented issue of longevity risk which is typically the concern of actuaries. Longevity risk refers specifically to changes over the long-term in life expectancy and the related impact on pensions, life assurance, etc. (Antolin, 2007). Longevity risk, in fact, is what the Lee-Carter method (mentioned above) was designed to address. Such extrapolative stochastic models, or variants thereof, may be useful in quantifying period risk in health economic decision making.

4.2.3.4.4 Adjusting for age

Both of the effects described above will be relatively insignificant in terms of modelling the long-term change in baseline risk in comparison to the effect of ageing.

Consideration of the effect of ageing is often required in analyses where the time horizon is long-term. This effect is something that can rarely be inferred from trial data. Typically, patients will not age sufficiently during trial follow-up for the effects of aging to be captured in the analysis. Indeed, in the RITA-3 example, only an 'age group' encompassing ten years was used to record the effect of age and so patients did not advance in 'age-group' during trial follow-up.

Impact of age would be expected to have the greatest impact of baseline risk over the long-term and therefore is the greatest source of temporal uncertainty

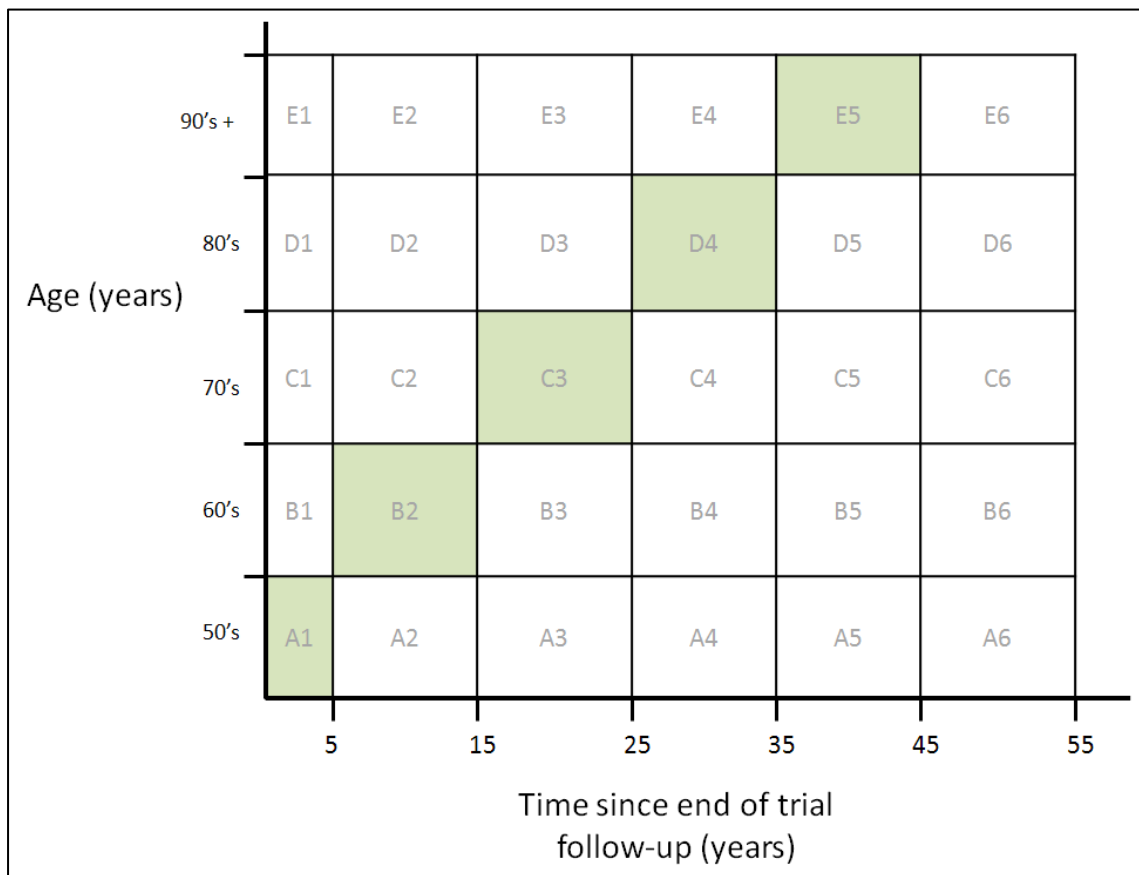
As patients age, they are expected to become more at risk of adverse clinical events. Generally this increase in risk becomes apparent only over the long-term, i.e. much longer than the follow up of a typical trial. For example, the age of patients in risk group 3 in the RITA-3 trial is assumed to be 52³². So as the cohort moves into old age (e.g. 20-30 years after intervention), it can be assumed that their risk of experiencing a composite event has increased. Often in health technology assessments (as was observed in Chapter 3), increases in baseline risk owing to age effects are not considered and assumptions of constant hazards (or constant transition probabilities) are imposed.

4.2.3.4.5 Alternative Temporal Scenarios

Event-free survival and period risk pertain to change over calendar time, as opposed change with age. The issue of simultaneously modelling age effects and other effects owing to the passage of calendar time can be helpfully illustrated using a Lexis diagram where event rates are displayed by age and by calendar time (Carstensen, 2007). For the RITA-3 example, it is useful to consider calendar time starting at the end of trial follow-up (5 years after intervention) and age starting at the age (52) of the cohort at the end of trial follow-up.

³² For the purposes of clarity and simplicity, only risk group 3 will be used to illustrate the analysis for the remainder of this sub-chapter

Figure 30: Simple Lexis diagram pertaining to the RITA-3 example



Note: The shaded cells (A1, B2, C3, D4, E5) represent where hazard rates are required to be estimated

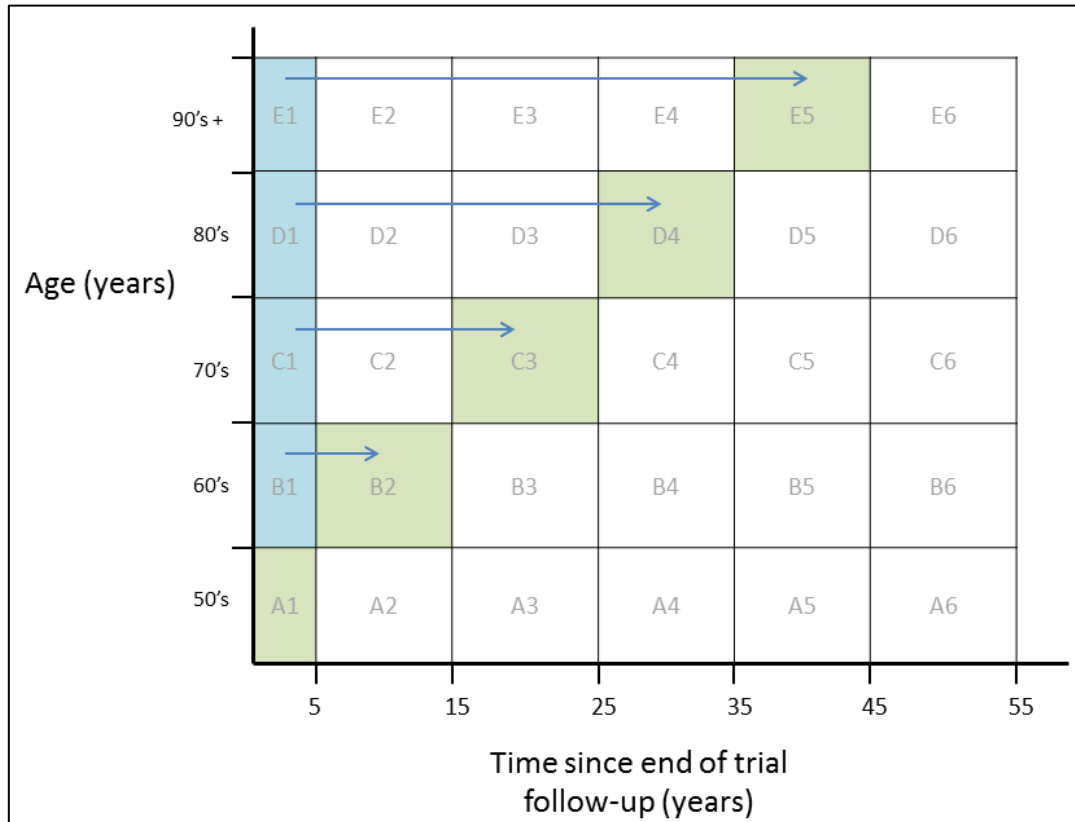
There are a number of assumptions that could be made regarding the behaviour of hazard rates over the unobserved period.

Assumption 1:

The assumption made in the original CEDM regarding age was that age could be modelled by updating the 'age group' covariate in the survival regression equation every 10 years. This, in many respects, appears a reasonable assumption, especially if we accept that hazards were tending towards being constant at the end of the observed period. This approach implies that we expect the impact of age not to alter over the long-term. For example, the risk associated with an 82 year old patient 30 years after intervention is assumed to be the same as the risk associated with an 82 year old patient 5 years after intervention. In this scenario the hazard rates over the unobserved period would be determined by (i) the hazard rate at 5 years (end of trial follow-up) and (ii) the age co-efficient in the survival regression model (recall that we are model averaging over both the Weibull

and Royston-Parmar models). The hazard rates we would use are conveyed in the Lexis diagram in Figure 31. In effect, we are assuming $B2 = B1$, $C3 = C1$, $D4 = D1$, $E5 = E1$.

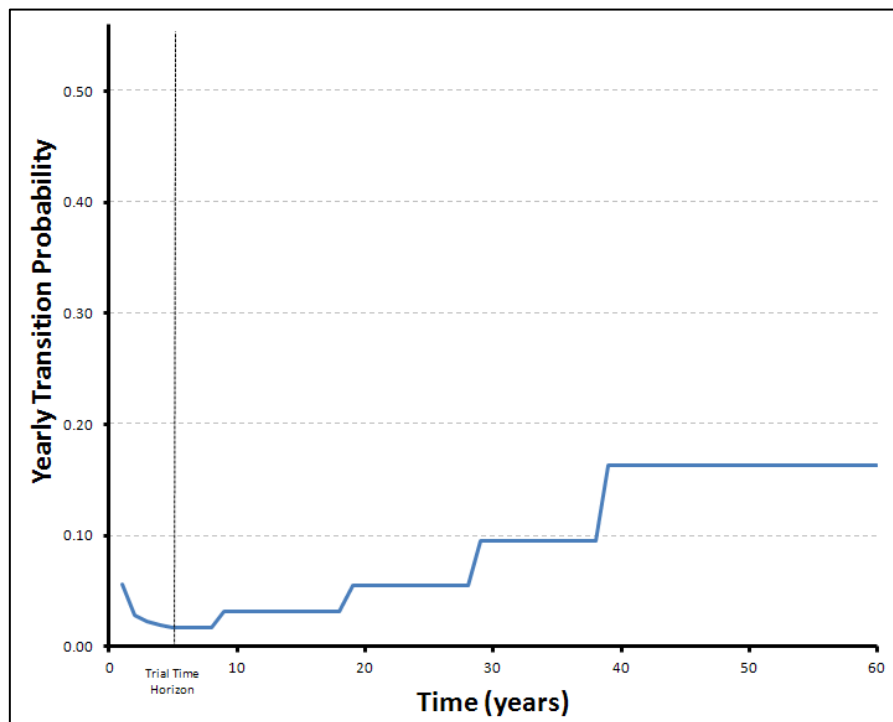
Figure 31: Lexis diagram conveying rates used under assumption that age can be modelled by applying the within trial age effect



The resultant transition probabilities are illustrated in Figure 32 where the transition probability increases in 'steps' over time, as the age-group covariate is updated³³.

³³ Note that because the transition probabilities in question are quite small, they closely resemble the related hazard rates. Transition probabilities (calculated as per the equations shown in section 5) as opposed to hazard rates will be shown in most of the illustrations as it is transition probabilities that are ultimately required for use in the CEDM.

Figure 32: Transition probabilities resulting from Assumption 1 (for risk group 3)



Assumption 2:

There is uncertainty regarding whether the age effect observed during trial follow-up (i.e. the effect on baseline risk of the age of a patient during the period following intervention) can be used to represent the effect of aging over the long-term. A second possible assumption was obtained from searching previous health technology assessments in this disease area (NSTE-ACS).

In order to find alternative plausible assumptions regarding the long-term effect of aging, a brief review of health technology assessments relating to NSTE-ACS was undertaken. A search of documents including the term “non-ST” was carried out on the UK health technology assessment database (www.hta.ac.uk)³⁴. Six full HTA documents were returned. Of these, one only assessed the short-term impacts of interventions to NSTE-ACS (Robinson et al., 2005), one pertained to a somewhat different disease area (occlusive vascular events) (Jones et al., 2004), one pertained specifically to non-ST-elevation acute myocardial infarction and contained long-term estimates on survival alone using extrapolation from short-term trial evidence (Simpson et al., 2011) and one was a systematic review which documented efforts at predicting long-term risk of a composite outcome,

³⁴ A Google Scholar search was also carried out using the search term “non-ST elevation long-term”. The results contained reports of clinical trials which considered outcomes 5 to 7 years after intervention, no further.

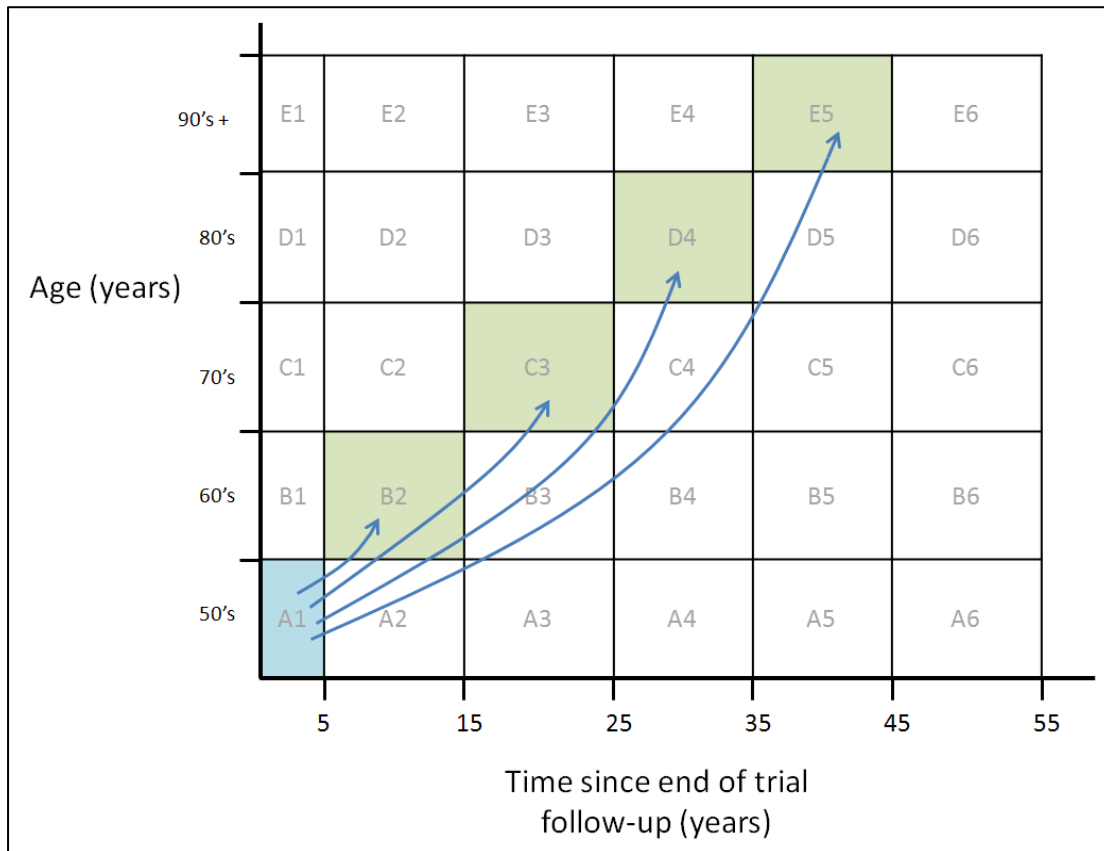
though these again were primarily based on extrapolation from the short-term without any explicit accounting for age, or incorporation of any external evidence that would be applicable to the RITA-3 example.

Two HTAs however pertained specifically to NSTEMI-ACS and attempted to estimate the long-term risk of a composite event similar to that defined in the RITA-3 decision model by incorporating some external evidence. These studies were related and based the long-term transition probabilities from a 'no event' health state to a 'composite event' health state in a model constructed for NICE in 2002 which compared alternative management strategies for the use of glycoprotein IIb/IIIa antagonists in NSTEMI-ACS (Palmer et al., 2002). This model incorporated evidence from the Nottingham Heart Attack Registry (NHAR) to ascertain the change in baseline risk over time³⁵. An Exponential distribution was found to be the best fit to these data and so an assumption of constant hazards over time was imposed. The data were not particularly long-term (5 to 7 years) and arguably did not account for the effect of aging. However it was judged by the modellers that the assumption of constant transition probabilities was reasonable based on this best available evidence. Effectively, this approach assumes that longer survival time (time without experiencing a composite event) is associated with decreasing risk and that this effect approximately offsets the increasing risk over time associated with age.

This assumption, in light of the reasoning outlined earlier would seem to generate an optimistic scenario. Assuming constant hazards/transition probabilities (i.e. no increase in risk with age) would seem to be a lower-bound to the behaviour of baseline risk over the long-term, i.e. this scenario is just about plausible, but we would certainly not expect anything more extreme (lower) than this. In terms of the Lexis diagram, the implications of this assumption are illustrated in Figure 33.

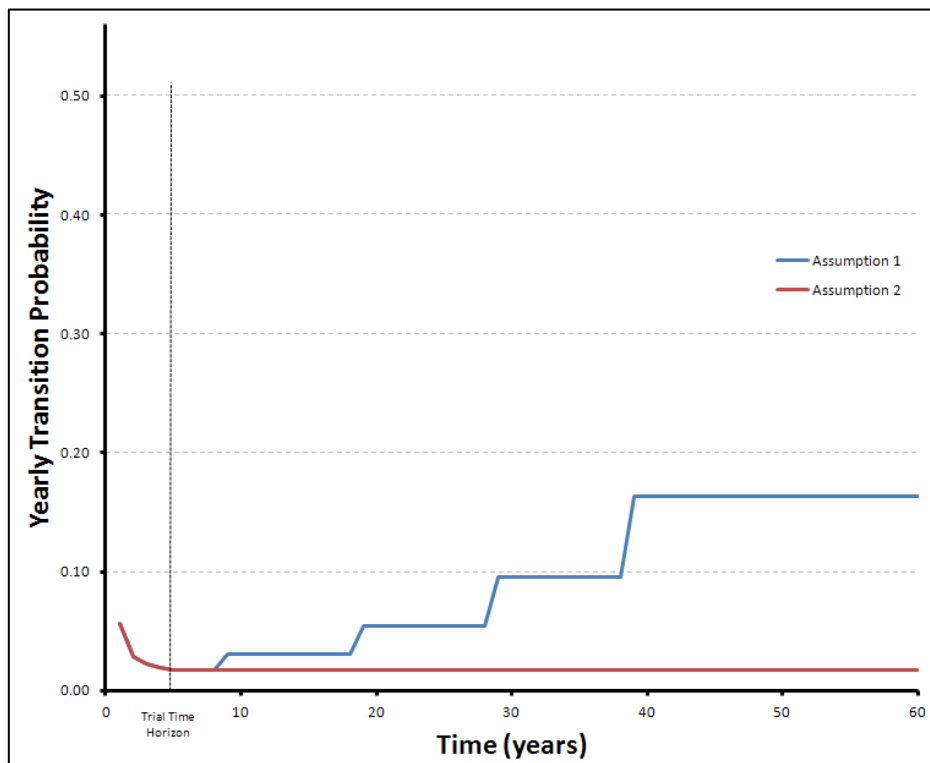
³⁵ These data also served to validate the transition probabilities imposed over the short-term, i.e. when the parameters of the survival regression equation were changed to reflect the characteristics of the mean patient in the NHAR data, the resultant short-term transition probabilities were similar. The key difference between the RITA-3 data and the NHAR data is the apparent decline in hazards over time. This may be explained by the NHAR data not capturing the initial 'hazardous period' immediately after hospitalisation.

Figure 33: Lexis diagram illustrating Assumption 2 where the hazard rates used over the long-term are the same as those used at the end of trial follow-up, i.e. the effects of aging and time without experiencing an event are assumed to cancel each other out



The resultant transition probabilities are illustrated in Figure 34 alongside the transition probabilities representing Assumption 1.

Figure 34: Transition probabilities resulting from Assumptions 1 and 2 (for risk group 3)



Assumption 3:

What is essentially required is an estimate of the effect of age on baseline risk that is not influenced by proximity to hospitalisation/intervention, i.e. an age effect for those having once presented with NSTEMI-ACS, were given a treatment similar to the conservative treatment in RITA-3, have not experienced an event and can be assumed to no longer benefit from the initial treatment.

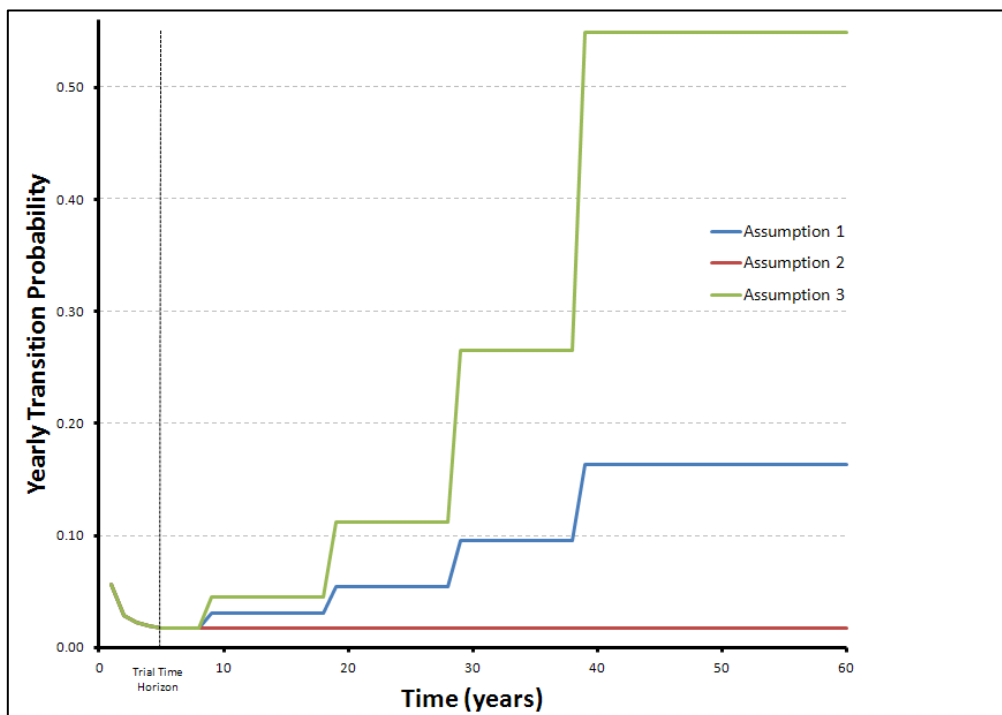
Obtaining data to inform such an estimate is difficult as the data need to be relatively long-term and the nature of standard treatment may have changed over time. But because an age effect is essentially what is required, it may be informative to analyse population mortality data and observe the effect of age on disease-specific mortality. From this the likely change in risk of experiencing a composite event related specifically to age could be inferred. Table 15 shows male death rates from coronary heart disease per age group. The data are taken from UK coronary heart disease statistics 2010 edition (British Heart Foundation Health Promotion Research Group, 2010).

Table 15: Death rates from coronary heart disease for males in 2008

Age Group	Number of deaths in 2008 per 100,000 population
45 – 54	67
55 – 64	175
65 – 74	443

The risk in one age group is approximately 2.6 that of the previous age group. This can be interpreted as a hazard ratio and applied to the ‘age group’ covariate in the survival regression. This represents an age effect somewhat stronger than that calculated from the RITA-3 data. The transition probabilities produced by Assumption 3 are arguably too pessimistic. These could be thought of as representing the upper-bound of what is plausible regarding baseline risk over the unobserved period. The transition probabilities representing Assumption 3 are illustrated in Figure 35 alongside the transition probabilities representing Assumption 1 and Assumption 2.

Figure 35: Transition probabilities resulting from Assumptions 1, 2 and 3 (for risk group 3)



Note that although the alternative curves in Figure 35 become quite disparate as we approach the full (60 year) time horizon, the values ascribed to transition probabilities over the final 20 years have relatively little impact due to discounting and fewer patients being event-free at distal time points.

Let us observe the impact of these alternative assumptions regarding baseline risk on the cost-effectiveness results (mean ICER), assuming all other temporal assumptions are as per the original CEDM and contrast this with the impact of alternative survival distributions.

Table 16: Cost-effectiveness results (mean ICER) for alternative assumptions regarding baseline risk and alternative parametric survival functions

	Assumption 1 (risk changes with age as per RITA-3 trial)	Assumption 2 (no change with age)	Assumption 3 (risk changes with age as per general population disease-related mortality)
Model averaged (as per Section 4.2.3.3)	22,238	20,160	24,052
Weibull	21,186	18,982	22,930
Exponential	22,672	20,098	24,607
Royston-Parmar	22,106	20,352	24,139
Gompertz	23,678	21,401	25,462
Log-logistic	21,266	19,360	22,771
Log-Normal	22,722	20,946	24,650
Generalised Gamma	21,275	19,385	22,843

Note that the choice of parametric function used to interpret the short-term has only a small, though not insignificant, impact on the mean ICER. At this point, the assumption imposed regarding long-term change in baseline risk also has a relatively minor impact (though it is somewhat more significant than the choice of parametric function) on the mean ICER. Recall that a conservative assumption of no treatment effect over the unobserved period is being applied. When less conservative assumptions regarding long-term treatment effect are imposed, the impact of the long-term assumption regarding baseline risk will have greater impact, as will be shown in Section 4.3.

4.2.3.5 Expressing uncertainty

In Section 4.3.2.4.5 above, three alternative scenarios are posited regarding the behaviour of baseline risk over the unobserved period. It is assumed that these alternative scenarios, though heavily focusing on the effect of age, account for the three relevant sources of temporal uncertainty outlined in section 4.2.3.4.2, 4.2.3.4.3 and 4.2.3.4.4 (event-free survival, period risk, age risk). These scenarios, represent alternative cost-effectiveness results (mean ICERs). At this point (before other sources of temporal uncertainty are characterised), the uncertainty surrounding long-term baseline risk does not impact the adoption recommendation for any of the risk groups. This is demonstrated in Table 17, where the shaded cells highlight ICERs which effect a positive adoption decision.

Table 17: Mean ICER for each risk group for each assumption regarding long-term baseline risk

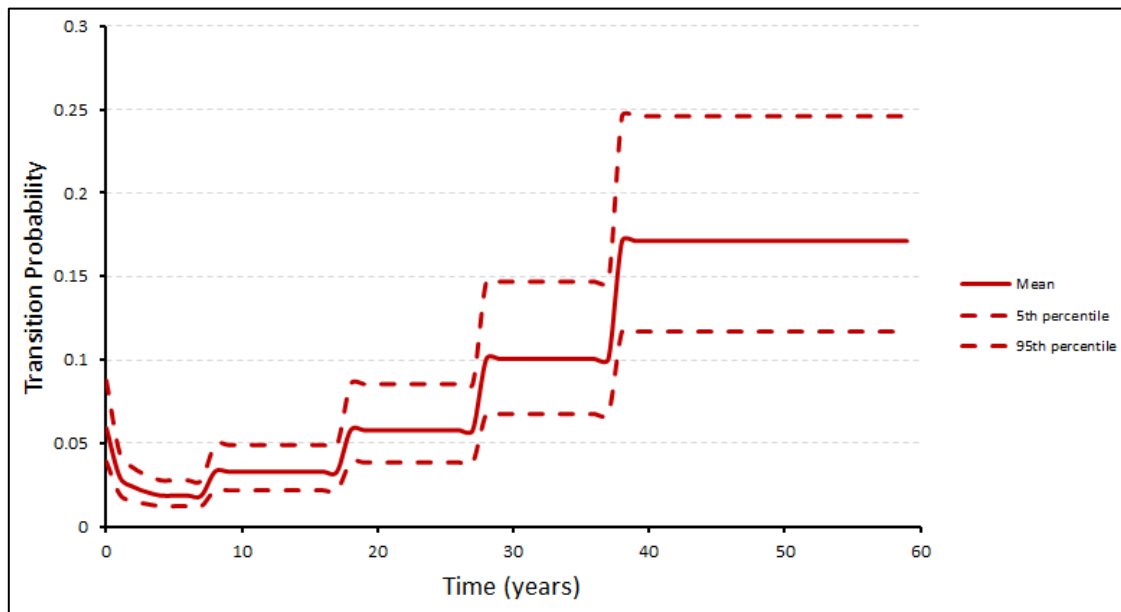
	Assumption 1 (risk changes with age as per RITA-3 trial)	Assumption 2 (no change with age)	Assumption 3 (risk changes with age as per general population disease-related mortality)
Risk Group 1	53,272	51,711	55,486
Risk Group 2	23,698	21,622	25,904
Risk Group 3	22,238	20,160	24,052
Risk Group 4a	12,540	11,566	13,764
Risk Group 4b	13,462	11,671	16,258

As things stand, therefore, temporal uncertainty relating to baseline risk alone does not directly lead to changes in adoption recommendation. However, because this source of temporal uncertainty may combine with other sources to effect decision uncertainty and because like all sources of uncertainty, it is desirable, where possible, to fully characterise the uncertainty in a probabilistic sensitivity analysis (PSA), we now endeavour to express temporal uncertainty as a single uncertain parameter.

When a parametric function with uncertain statistical parameters is extrapolated beyond the evidence time horizon, increasing uncertainty over time is naturally expressed. Figure 36 below illustrates the baseline transition probabilities as well as the 5th and 95th percentiles from the PSA,

resulting from a Weibull extrapolation of the RITA-3 data, assuming the change over time with age is equal to the change with age observed in the RITA-3 trial.

Figure 36: Baseline transition probabilities and percentiles from Weibull extrapolation of the RITA-3 data



Although extrapolating the uncertainty associated with a parametric function into the unobserved period could conceivably be used as a proxy for expressing uncertainty over the long-term, it is not a true representation of temporal uncertainty, i.e. this does not reflect our lack of knowledge pertaining to the unobserved period, only the uncertainty pertaining to how well the parametric function fits the short-term data. Such an approach is, in fact, likely to underestimate temporal uncertainty. A more appropriate approach is to characterise, using reasonable upper and lower bounds, the plausible space around the expected trajectory from extrapolated short-term evidence.

The optimistic and pessimistic scenarios outlined in Section 4.2.3.4.5 are well suited to act as such bounds. Although all 3 scenarios could be incorporated into the PSA using model averaging, the resultant characterisation of temporal uncertainty (effectively a discrete distribution) would not well represent the nature of the uncertainty. This uncertainty is such that there is a plausible range over the unobserved period, i.e. baseline risk could take an arbitrarily large number of temporal trajectories over the unobserved period within certain bounds. As such, a continuous distribution would seem a more suitable means to express the uncertainty.

Each of the assumptions discussed in Section 4.2.3.4.5 were executed by augmenting the co-efficient for age in the survival regression equation. This parameter is thus the vehicle through which temporal uncertainty can be expressed.

The equation for hazards over the unobserved period is effectively of the form:

$$h(t) = c * \exp(\beta_{Age} * x_{Age})$$

where:

$h(t)$ = the hazard rate at time t

c = a constant < 0.02

β_{Age} = the coefficient for age

$$x_{Age} = \begin{cases} 0, & \text{if } 50 \leq 52 + t < 60 \\ 1, & \text{if } 60 \leq 52 + t < 70 \\ 2, & \text{if } 70 \leq 52 + t < 80 \\ 3, & \text{if } 80 \leq 52 + t < 90 \\ 4, & \text{if } 90 \leq 52 + t \end{cases}$$

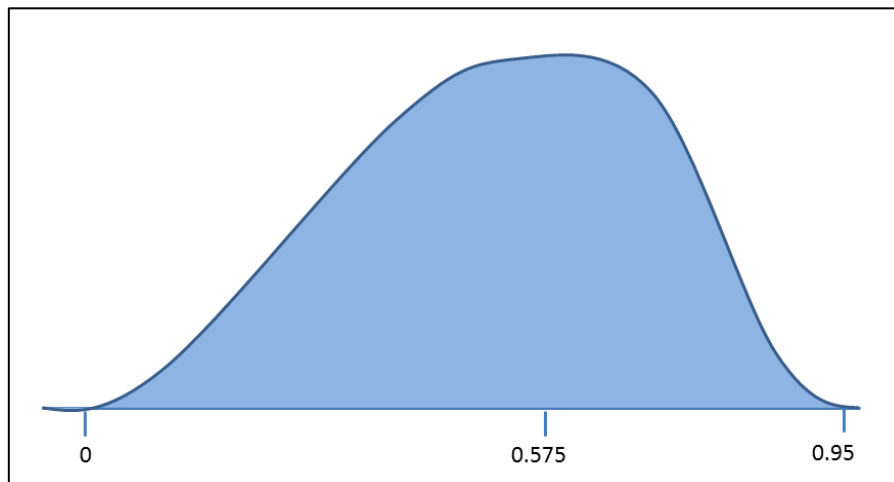
Of course, there is already a distribution around β_{Age} generated from the probabilistic parametric fit to the short-term data. Hence there is a degree of uncertainty already expressed regarding the trajectory of baseline risk over the unobserved period. As stated above however, what this distribution represents is uncertainty pertaining to the observed period, rather than uncertainty regarding the nature of baseline risk over the unobserved period. Thus a distribution around β_{Age} from the evidence time horizon (5 years) onwards that conveys what can be believed with current information to be the plausible space over the unobserved period as well as a reasonable expected trajectory.

Let us characterise β_{Age} such that the expected value corresponds to Assumption 1 where the effect of aging over time is assumed to correspond to the effect of age within the observed period. This assumption represents current rational beliefs based on the best available evidence. Let upper and lower extreme values of β_{Age} correspond to Assumptions 2 and 3 (the optimistic and pessimistic

assumptions respectively) so that the long-term transition probabilities fall within the plausible region represented by these assumptions.

In particular, let the expected value for β_{Age} equal 0.575 as generated from the Weibull distribution³⁶. The values for β_{Age} that correspond to Assumptions 2 and 3 are 0 and 0.95 respectively. In order to allow 0 and 0.95 to broadly represent confidence intervals, a slightly left-skewed distribution to be fitted around β_{Age} is required as Figure 37 illustrates.

Figure 37: Desired distribution for β_{Age}



This distribution ought to have three characteristics pertaining to the mean and the cumulative distribution function:

$$\begin{aligned}\mu &= 0.575 \\ F(0.005) &\cong 0 \\ F(0.995) &\cong 0.95\end{aligned}$$

A distribution with more parameters may perform better in terms of exactly meeting these criteria, but a simpler Beta distribution does well to approximate. By using the method of moments, a suitable α and β can be obtained for the Beta distribution.

³⁶ The same process is undertaken in relation to the Royston-Parmer distribution. The particular distribution for β_{Age} used corresponds to the survival distribution (Weibull or Royston-Parmer) used in the simulation, though both (versions of the distribution for β_{Age}) are very similar.

$$\alpha = \mu * \left(\frac{\mu(1 - \mu)}{\sigma^2} - 1 \right)$$

$$\beta = \alpha * \frac{(1 - \mu)}{\mu}$$

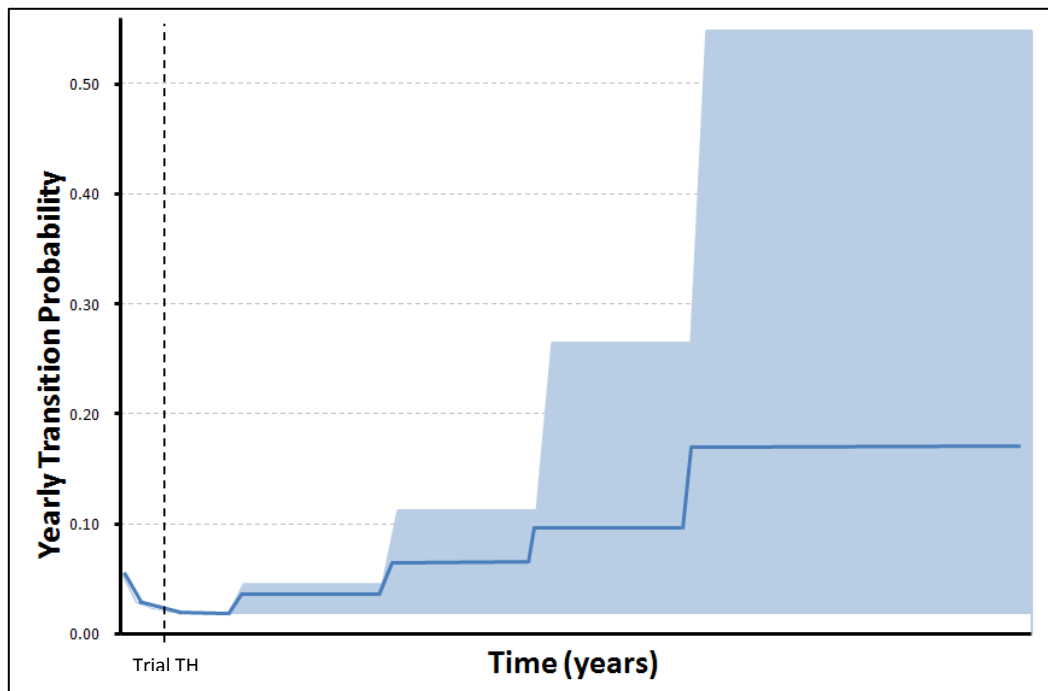
Values of σ are sampled until $F(0.005)$ and $F(0.995)$ approximate the desired values. The values generated for α and β are:

$$\alpha = 2.235$$

$$\beta = 1.652$$

By applying this distribution to β_{Age} in the probabilistic sensitivity analysis, temporal uncertainty around long-term baseline risk (transition probabilities) is expressed as illustrated in Figure 38. The dark blue line represents the expected temporal trajectory of baseline risk, while the shaded area represents the plausible region.

Figure 38: Yearly transition probabilities when temporal uncertainty is expressed regarding baseline risk



The temporal uncertainty relating to baseline risk can now be considered to be appropriately expressed in the CEDM given current information.

A relevant issue to this characterisation of temporal uncertainty is that of the potential role of expert elicitation. Formal expert elicitation might have acted as an alternative means of characterising the uncertainty around the expected progression of baseline risk over the unobserved period. Given that a data-driven approach was employed, an appropriate role for expert opinion would be to validate this characterisation of the temporal uncertainty.

4.2.4 Results

The CEDM results are now shown below given the following two updates.

- (i) Uncertainty surrounding appropriate parametric fit to the short-term data now characterised through model averaging
- (ii) Temporal uncertainty around baseline over unobserved period now characterised through Beta distribution applied to age parameter.

4.2.4.1 Cost-effectiveness

Table 18: Cost-effectiveness Results after temporal uncertainty has been addressed for baseline risk. Compare with Table 10 to observe the change in the results.

Risk group	Incremental Cost (£)	Incremental QALY	Mean ICER	Adopt/reject early interventional (EI)
Risk group 1	4892	0.0877	55,782	Reject
Risk group 2	4841	0.1841	26,302	Reject
Risk group 3	5874	0.2537	23,149	Reject
Risk group 4a	6259	0.4702	13,311	Adopt
Risk group 4b	6183	0.4288	14,419	Adopt

The mean ICER has increased slightly for each risk group. This is for two reasons. First, baseline risk over the observed period is now characterised by a model averaging of the Weibull and Royston-

Parmar parametric functions instead of the Weibull function alone. Since the Royston-Parmar fit implies a higher ICER, the model averaged fit does also. Second, the CEDM still assumes that there is no treatment effect after the observed period, but for the patients who make to beyond 5 years without experiencing a composite event (of whom there are more in the early interventional cohort) their long-term baseline risk is more uncertain (though the expected trajectory is still the same). The non-linearity of the model means that the greater uncertainty impacts the mean ICER, increasing it by a small degree.

4.2.4.2 Uncertainty

Table 19 below summarises the current impact of uncertainty on the outputs of the CEDM for each of the 5 risk groups.

Table 19: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for baseline risk. Compare with Table 11 to observe the change in the results.

Risk group	Prob(EI cost-effective) at £20,000/QALY	EVPI/patient (£)	EVPI/population (£)
Risk group 1	0.001	0.38	86,358
Risk group 2	0.181	133.54	30,348,046
Risk group 3	0.270	266.10	60,473,379
Risk group 4a	0.907	110.62	12,569,645
Risk group 4b	0.872	133.53	15,172,887

In the original CEDM results, both risk groups 2 and 3 has significant error probabilities. With the modifications to the CEDM, the error probabilities for these risk groups are smaller, i.e. there can be more confidence in the decisions being recommended. Thus there is less value in obtaining further information to inform these decisions. However, the error probabilities for the higher risk groups (4a and 4b) have risen slightly leading to higher estimates of value of information.

4.3 Chapter 4.3: Treatment Effect

4.3.1 Introduction

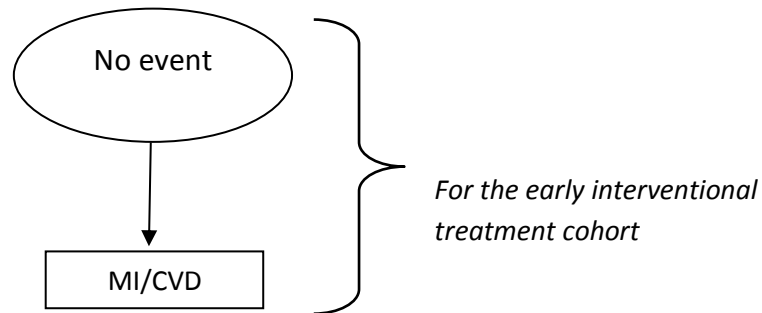
The purpose of this sub-chapter is to address temporal uncertainty relating to another key parameter in the RITA-3 model: the treatment effect pertaining to a first composite event (myocardial infarction or cardiovascular-related death), i.e. the measure representing the effect of the early interventional treatment that when applied to baseline risk expresses the risk of a first composite event for patients receiving the early interventional treatment. In doing so, this chapter will address, more generally, the characterisation of temporal uncertainty pertaining to treatment effects.

In particular, this chapter endeavours to explore: (i) methods to interpret short-term trial data in order to observe trends in treatment effect parameters that can potentially be extrapolated beyond the observed period; (ii) the sensitivity of the mean ICER to alternative assumptions regarding long-term treatment effect and how this relates to uncertainty regarding long-term baseline risk; (iii) the expression of uncertainty when there are essentially no data pertaining to long-term treatment effect; and (iv) the use of expert elicitation to inform long-term treatment effect.

4.3.1.1 The parameter

The model input parameter under analysis is the relative measure representing the effect of the early interventional treatment, which when applied to baseline risk expresses the risk of a first composite event for patients receiving the early interventional treatment. Like the previous sub-chapter (4.2), the ultimate aim is to compute transition probabilities pertaining to the transition from the 'No Event' health state to the 'MI/CVD' health state, in this case, transition probabilities representing the early interventional treatment.

Figure 39: Treatment effect is applied to baseline risk in order to compute the risk of transitioning from the 'No Event' health state to the 'MI/CVD' health state for patients receiving the early interventional treatment strategy



It will be desirable to compute a hazard ratio to represent the treatment effect - as this is a time-to-event analysis and the measure ought to be compatible with the hazard rate calculated to represent baseline risk. The data for the full patient population in RITA-3 are used to calculate the hazard ratio, which is then applied to each of the risk groups. Note that it is common in cost-effectiveness studies in the cardiac field to assume the same treatment effect for patients with different risk profiles (Briggs et al., 2007, Mihaylova et al., 2006) and indeed this was the approach taken in the original Henriksson analysis.

The question of sustained health benefit owing to the early interventional treatment is fundamental to the decision problem at hand. The early interventional treatment strategy had already been shown to improve health outcomes at 1 year, but at a higher cost than a conservative treatment strategy (Kim et al., 2005). The RITA-3 CEDM exists to account for all relevant health outcomes and costs in order to make a judgement regarding costs-effectiveness. This of course depends on a number of components in the CEDM, as outlined in Chapter 4.1. However, the long-term effect of treatment on the risk of a first composite event is undoubtedly a core element of the analysis.

4.3.1.2 What do we mean by treatment effect as we move further from randomisation?

An interesting aspect of analysing the evolution of treatment effect over time is that thought must be given to what is meant by treatment effect as more distal time periods are considered. Treatment effect in health technology assessment is usually defined as the effect attributable only to a certain

treatment (Facey, 2006). It can also be taken to refer to the mean difference between two treatment cohorts of interest. At the point of randomisation, these definitions can be assumed to mean the same thing, as the only difference between the two cohorts is assumed to be the treatments they are receiving. However, at any time point after randomisation, the two cohorts have been exposed to different risks and now have different characteristics. Any difference in outcomes cannot now be attributable purely to the different treatments. In other words, a kind of temporal selection bias creeps into the analysis as time moves forward. Bagust and Beale have referred to this phenomenon as “progressive survivor bias” (Bagust and Beale, 2014). In RITA-3 for example, a group of more susceptible patients in the conservative cohort will have experienced a composite event by the end of the observed period. If we assume that there is an equivalent susceptible group among the early interventional cohort that *would* have experienced a composite event had they been subject to the conservative treatment, then it is clear that the early interventional cohort is now more ‘susceptible’ on average. It is possible that these susceptible patients may simply experience composite events later than they would have, and so it may be that the early interventional treatment simply delays the adverse event. In other words, there may be a rebound effect (Drummond et al., 2005).

Such difficulties regarding the interpretation of treatment effects (and in particular hazard ratios) over time have been reported and discussed elsewhere (Hernan, 2010, van Walraven et al., 2004). However, for cost-effectiveness analysis, our concern pertains simply to being clear about what we want to quantify as we look to the long-term. At any time point after randomisation, it is not the effect attributable only to a treatment that is of interest, but the mean difference between the two cohorts as they stand. This is important not just for clarity, but for characterising the trajectory of treatment effect after the observed period, especially if the nature of the treatment effect parameter requires explanation for the purposes of employing expert opinion.

4.3.2 Available evidence

4.3.2.1 Evidence from RITA-3

The key data available to inform treatment effect come in the form of 5 year individual patient-level data (IPD) from the RITA-3 trial. These data are time-to-event (or survival) data which means that parametric survival analysis may be employed to characterise treatment effect over the observed

period and potentially extrapolate beyond, as described in Section 3.4.2.1. The patients participating in the RITA-3 trial were randomised into conservative and early interventional cohorts so that the total differences in outcomes could be attributed to the different treatment strategies.

4.3.2.2 Further evidence

There exist a number of other trials that compare conservative and early interventional strategies for the treatment of patients with NSTEMI-ACS (Spacek et al., 2002, Anderson et al., 1995, Lagerqvist et al., 2006, de Winter et al., 2005, Boden et al., 1998, McCullough et al., 1998, Cannon et al., 2001). From these, alternative estimates of treatment effect (in the form of odds ratios) can be obtained. Each of the trials had follow-up no longer than that of RITA-3.

4.3.3 Analysis

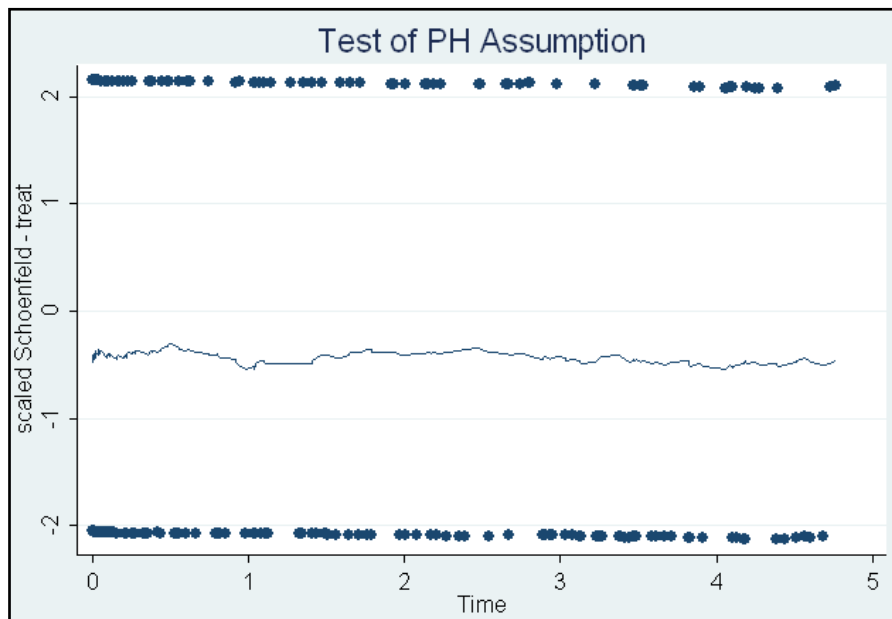
4.3.3.1 Analysing and interpreting the survival data

4.3.3.1.1 The proportional hazards assumption

As discussed in Section 3.4.2.1, it is common (and to a large extent desirable) to assume proportional hazards when analysing survival data. To recap, ‘proportional hazards’ means that the hazards of multiple groups are assumed to be multiplicative for any time-point t and so for example, treatment groups can be characterised by single hazard ratios representing treatment effects. A useful first step in estimating treatment effect is to test the validity of a proportional hazards assumption. Two methods of testing for proportional hazards are outlined below and applied to the RITA-3 example.

Schoenfeld residuals can be used to test for proportional hazards. A smooth function of time is fit to residuals to test for a relationship. They do not involve an estimated hazard function and so are useful in examining only the relative hazards between treatment groups (Schoenfeld, 1982). A slope of approximately zero in the resultant graph would indicate that proportional hazards could reasonably be assumed. This test was carried out in the RITA-3 example using the `estat phtest` command in Stata (Cleves, 2010). The graphical result of the test is shown in Figure 40 below

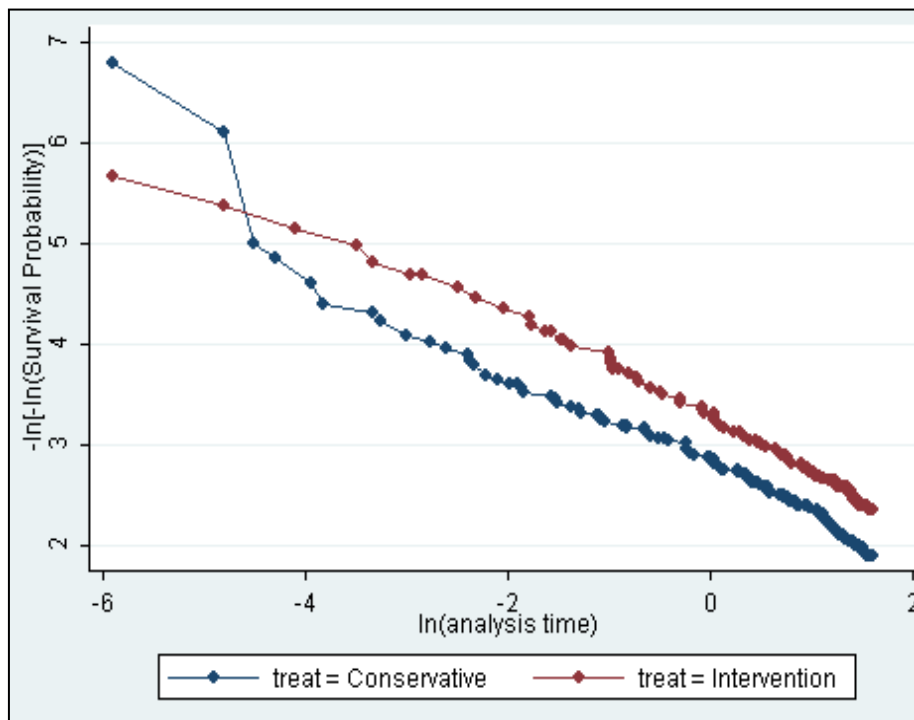
Figure 40: Schoenfeld Residuals for alternative treatment strategies in RITA-3



The graph shows a function with slope of approximately zero, supporting an assumption of proportional hazards.

Another useful test is to generate -Log-Log curves and look for approximate parallelism (Kay, 1977). Under proportional hazards, the plot of $-\ln(-\ln(S(t)))$ against $\ln(t)$ for each treatment group should be roughly parallel. This test was carried out in the RITA-3 example using the `stphplot` command in Stata (Cleves, 2010). The graphical result of the test is shown in Figure 41 below.

Figure 41: -Log-Log curves for alternative treatment strategies in RITA-3

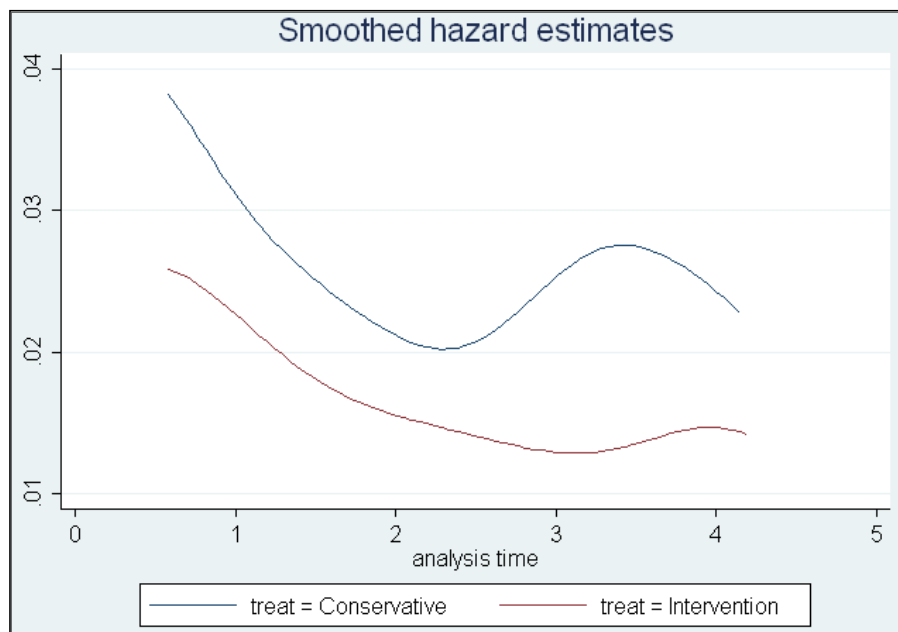


Although the curves cross at the left-hand side of the graph, this refers to approximately the first week of the 5 years under analysis and so should not be considered consequential³⁷. The -Log-Log curves are approximately parallel, thus further supporting an assumption of proportional hazards.

These tests simply seek to demonstrate approximate proportionality of hazards in order to justify a PH assumption. For the purposes of characterising the observed period, it is rational in this case to assume proportional hazards. For the purposes of extrapolation however, there may still be some use in analysing more closely the hazard ratio over the observed period in order to look for evidence of any temporal trend. Graphical estimates of the empirical hazard functions for both arms of the trial are shown in Figure 42 below. These plots were generated for the RITA-3 example using the `stcurve`, `hazard` command in Stata (Cleves, 2010).

³⁷ This may be indicative of higher mortality experienced by patients in the early interventional cohort immediately after treatment.

Figure 42: Empirical hazard curves showing treatment effect over the observed period



Although because of right-censoring, the curves representing the latter stages of the observed period are less reliable, there is a suggestion of a non-constant hazard ratio over this period.

A change in hazard ratio over time can be further investigated by creating a piecewise Cox model with time-dependant covariates added to the ‘treatment’ covariate in order to produce a hazard ratio for each year in the observed period³⁸. The model was implemented in Stata using the `stsplit` command (Cleves, 2010).

Table 20: Results of Piecewise Cox Regression

Time Period (Year)	Estimated Hazard Ratio
1	0.647
2	0.671
3	0.96
4	0.444
5	0.604

The hazard ratio seems to become somewhat unstable over the third and fourth years as depicted in Figure 4 and supported by the results of the piecewise Cox model. However, it must be noted that

³⁸ Note that such a model can be problematic in terms of fewer data pertaining to later time periods.

the time-dependant covariates that give rise to the period-specific hazard ratios are not statistically significant. As such, it would be inappropriate to extrapolate based on these estimates. Nonetheless, this analysis suggests that we ought not preclude scenarios where the treatment effect rises, falls, or stays constant over the unobserved period.

4.3.3.1.2 Estimating treatment effect for the observed period

Assuming proportional hazards, a hazard ratio is required to characterise treatment effect over the observed period (and to use as a starting point for treatment effect over the unobserved period). In the original model, a Weibull proportional hazards model was fit to the survival data and from this a hazard ratio was calculated:

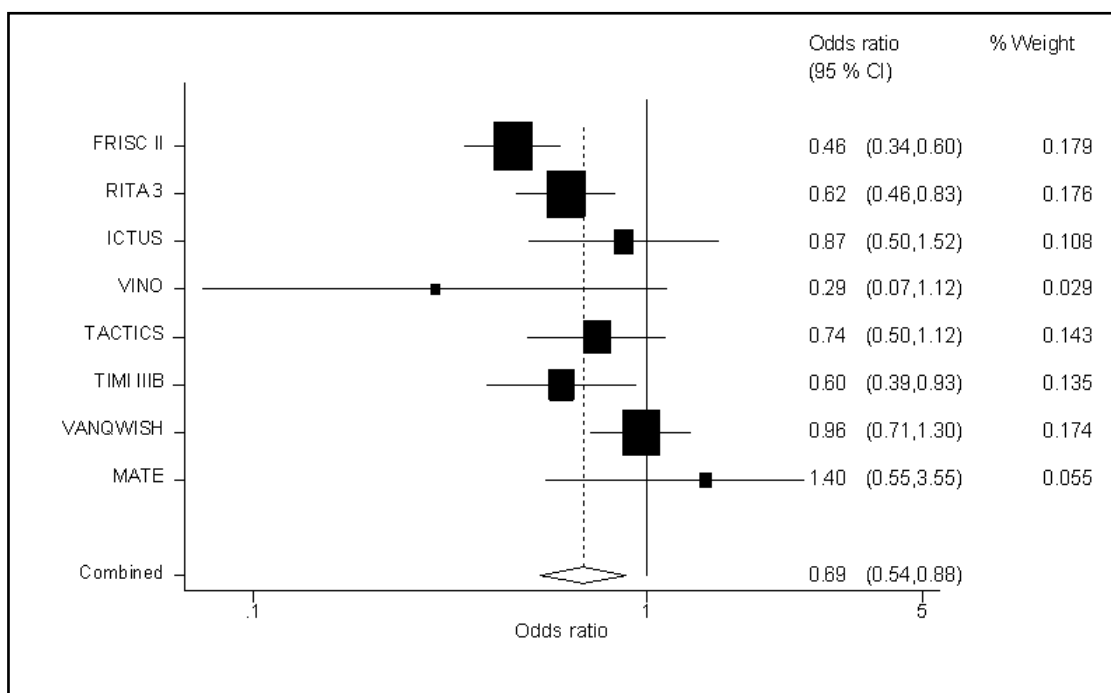
Model	Hazard Ratio	95% CI	
Weibull PH	0.620	0.464	0.830

Recall from Section 4.1.7.1 that it is advantageous to calculate the hazard ratio independent of the calculation for baseline hazards. The Cox proportional hazards model allows estimation of a hazard ratio without specifying a baseline hazard (Cox, 1972). This produces a slightly greater treatment effect than that of the Weibull PH model.

Model	Hazard Ratio	95% CI	
Cox PH	0.618	0.462	0.826

However, there are treatment effects available from multiple trials. A meta-analysis can be conducted on all relevant estimates of treatment effect to produce a pooled estimate. The results of this meta-analysis are shown in Figure 43.

Figure 43: Forest Plot of meta-analysis of treatment effect carried out by Henriksson et al.



Note: The analysis includes the estimated treatment effect from RITA3 and treatment effects from 7 other trials (Spacek et al., 2002, Anderson et al., 1995, Lagerqvist et al., 2006, de Winter et al., 2005, Boden et al., 1998, McCullough et al., 1998, Cannon et al., 2001)

This pooled odds ratio which we will interpret as a hazard ratio reflects all of the relevant evidence regarding treatment effect in the short-term period after intervention³⁹.

Model	Hazard Ratio	95% CI	
Pooled	0.69	0.54	0.88

The pooled estimate would seem to best represent the use of ‘all relevant evidence’, although there may be questions regarding the applicability of some of the other trials to this analysis. A quick sensitivity analysis (using only risk group 3) demonstrates that, all else being equal, the inclusion of estimates from other trials has a not insignificant impact on the estimate of cost-effectiveness.

³⁹ In this case study, an odds ratio can be used to approximate a hazard ratio because the risk of the event in question is very low and a hazard ratio tends towards an odds ratio as the underlying probability tends towards zero (Symons and Moore, 2002).

Model	Mean ICER
Weibull PH	23,149
Cox PH	22,690
Pooled	26,846

It is not within the scope of this reanalysis to assess the suitability of the other trials. Thus, as per the original analysis, it will be assumed here that the Weibull fit from the RITA-3 trial produces an appropriate estimate of the hazard ratio for the observed period, though it is important to note that applying the pooled estimate continues to represent a useful sensitivity analysis. The remaining sections of this sub-chapter focus on the nature of treatment effect after the observed period.

4.3.3.2 Moving beyond the observed period.

From the analysis of the short-term survival data, there is little indication of how the hazard ratio might evolve over the unobserved period. The IPD from RITA-3 did not suggest a temporal trend and only in one of the reports from the other trials was there a suggestion of how treatment effect might alter with time. This was the report of the FRISC-II trial where there was a suggestion that treatment effect over the first two years of follow-up was noticeably greater than that of the following three years, though questions of generalisability regarding this study should be taken into account (Lagerqvist et al., 2006). Since the proportional hazards assumption was valid for the observed period, it could be assumed that this holds for the unobserved period, though Davies et al. have warned of the dangers of such an assumption (Davies et al., 2013). What is advocated by many methods guidelines (including the NICE methods guidance) in this situation is to carry out analyses that compare alternative scenarios reflecting different assumptions about future treatment effects and that such assumptions should include the limiting assumption of no further benefit as well as more optimistic assumptions. In the original cost-effectiveness analysis by Henriksson et al. a 'conservative assumption' of no further treatment effect was made in the base-case. Let us compare this assumption with alternative assumptions that could be made regarding the nature of treatment over the unobserved period. Addressing the temporal uncertainty of treatment effect can begin by observing the impact of these alternative assumptions.

4.3.3.2.1 Scenario analysis and model averaging

What the NICE methods guidance suggests, in particular, is to compare scenarios where (i) the treatment effect over the unobserved period is nil; (ii) the treatment effect continues at the same level as during the observed period; and (iii) the treatment effect diminishes over time (NICE, 2013). Applying this principle to the RITA-3 model, the following alternative scenarios can be posited.

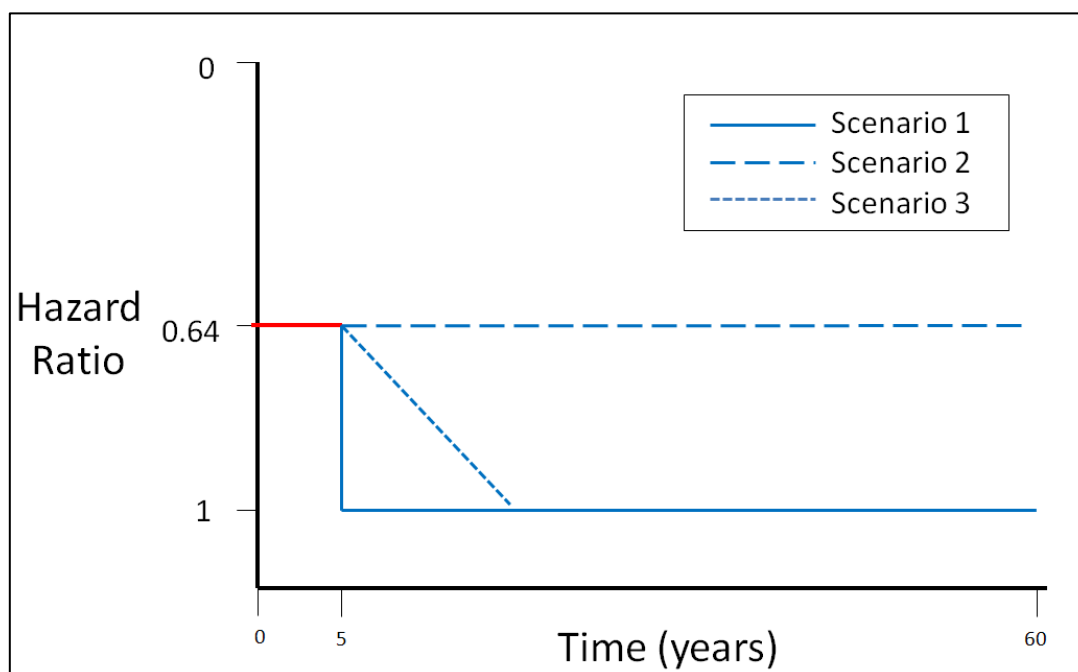
Scenario 1: The treatment effect is nil after the observed period. This scenario is represented by assuming the hazard ratio equals 1 throughout the unobserved period (this was the assumption used in the original analysis).

Scenario 2: The treatment effect continues as observed in the trial. This scenario is represented by assuming the hazard ratio calculated for the observed period (mean = 0.62) holds for the duration of the unobserved period.

Scenario 3: The treatment effect slowly dissipates over the unobserved period. This scenario is represented by assuming the hazard ratio increases linearly (tends towards 1 and does not exceed 1) from the beginning of the unobserved period. The rate of increase reflects that suggested by the FRISC-II report: 1.05/year.

The three scenarios are illustrated in Figure 44.

Figure 44: Illustration of the alternative temporal trajectories of the hazard ratio according to the three scenarios posited



Applying each of these scenarios to the RITA-3 model, the following cost-effectiveness results are obtained (for risk group 3).

Table 21: Results (for risk group 3) of sensitivity analysis regarding behaviour of treatment effect (hazard ratio) over unobserved period

Scenario	Incremental Cost (£)	Incremental QALY	Mean ICER (£)	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
1	5874	0.2537	23,149	0.270	Reject
2	8735	0.7915	11,037	0.999	Adopt
3	6663	0.4075	16,353	0.806	Adopt

It is clear that the adoption recommendation is dependent on the assumption made regarding the long-term behaviour of treatment effect. An assumption of no further treatment returns a mean ICER just above the threshold of £20,000 per QALY. It is likely therefore that a scenario where even a relatively small amount of treatment effect is assumed over the unobserved period would return a mean ICER below £20,000, and therefore a positive cost-effectiveness result for the early interventional treatment strategy. In fact, it can be deduced from the decision model that it would

take a scenario of less than 2 years of further treatment effect (assuming treatment effect stays at the same level) for the early interventional strategy to be deemed cost-effective (as things stand)⁴⁰.

It is interesting at this point to note the joint impact of the assumptions made regarding treatment effect and baseline risk by conducting a two-way sensitivity analysis. Applying each of the three scenarios for treatment effect described above along with the three scenarios described for baseline risk described in Section 4.2.3.4.5, the results in Table 22 are obtained. A shaded cell implies an adoption recommendation for the early interventional treatment strategy.

Table 22: Results (for risk group 3) in terms of mean ICERs (£) for two-way sensitivity analysis regarding behaviour of treatment effect and baseline risk over unobserved period

		Treatment effect over unobserved period		
		None	Declines over time	Continues as per trial
Effect of age on baseline risk	No effect	20,160	15,621	11,621
	As per RITA-3 trial	22,238	16,411	10,934
	As per population	24,052	17,008	11,112

It is clear that the assumption regarding treatment effect has greater influence on the mean ICER, again with either assumption involving continued treatment effect over the unobserved period returning a recommendation to adopt the early interventional treatment strategy. The extent of that influence however, is largely dictated by the assumption made regarding baseline risk, with greater long-term baseline risk leading to a greater impact for treatment effect. It is notable also that the impact of baseline risk alters as greater treatment effect over the unobserved period is assumed. This is because there are two forces at work. First, the smaller the baseline risk over unobserved period, the greater the benefits from within trial treatment effect. But second, the smaller the baseline risk over the unobserved period, the smaller the impact of any *post-trial* treatment effect. Therefore, as greater treatment effect over the unobserved period is assumed, the influence of the latter force comes to the fore. Evidently however, if substantial baseline risk is assumed, the negation of the benefit from treatment over the observed period is enough to surpass the benefit of

⁴⁰ Note that the temporal uncertainty pertaining to other input parameters is yet to be addressed.

greater benefit gained over the unobserved period, even under an assumption of continued treatment effect.

If the scenarios for treatment effect, when applied in the model, had resulted in the same decision recommendation, then this scenario analysis would simply act to further endorse the recommendation. However, this is not the case in the RITA-3 example (for risk group 3 at least). Although, the scenario analysis demonstrates the importance of the treatment effect assumption and to some extent conveys the uncertainty that currently surrounds long-term treatment effect, the CEDM is not producing a decision recommendation, thus leaving the decision-maker to implicitly weight these scenarios in order to make a judgement.

Like the uncertainty relating to the choice of parametric distribution (in Section 4.2.3.3), this issue could be considered as one of structural uncertainty, and addressed through model averaging. As described previously, model averaging must occur for each individual simulation, effectively parameterising the structural uncertainty. Without any directly relevant evidence with which to weight the alternative scenarios, the scenarios could, in the first instance, be given equal weighting. The vector of probabilities that characterises the uncertainty relating to treatment effect behaviour over the unobserved period would be:

$$p = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

This produces the following cost-effectiveness results for risk group 3.

Model averaged with equal weighting	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
	7,098	0.4851	14,632	0.6993	Adopt

The mean ICER is now well below the threshold of £20,000 per QALY and adoption of the early interventional strategy is recommended. As an alternative, the ‘middle’ scenario of gradual decline in treatment could be given most weight, with the more optimistic/pessimistic scenarios given equal lesser weight.

$$p = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$$

This produces the following cost-effectiveness results for risk group 3.

Model averaged: 25%, 25%, 50%	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
	7,002	0.4685	14,944	0.7313 ⁴¹	Adopt

These weights also result in a mean ICER below the threshold of £20,000 per QALY and a recommendation to adopt the early interventional strategy.

However, model averaging over alternative posted scenarios is a deficient course of action under these circumstances for a number of reasons:

- (i) With the currently available evidence, there is no adequacy measure available with which to weight the alternative scenarios. The weightings employed therefore are necessarily arbitrary.
- (ii) The scenarios broadly represent optimistic, pessimistic and ‘middle-ground’ assumptions. However, there is no cause to believe that any are plausible scenarios, or that together they represent the range of plausible parameter values. They were generated simply to test the sensitivity of the ICER to the assumption regarding long-term treatment effect. It may in fact be more meaningful to discuss this uncertainty in terms of the duration of the treatment effect (as observed in the trial period) over the unobserved period.
- (iii) There is no reason to believe that the nature of the uncertainty is discrete, i.e. there could be an arbitrarily large number of plausible scenarios, effectively rendering the uncertainty continuous in nature.

⁴¹ Note that since there is ‘less’ uncertainty with these weightings, the probability of cost-effectiveness is greater even though the mean ICER is slightly closer to the threshold.

4.3.3.3 The need for further evidence

It is apparent from the scenario analysis that the evidence currently available is not sufficient to characterise treatment effect over the long-term considering its impact on the cost-effectiveness results. This can be shown more formally by conducting a value of information analysis; in particular by calculating the expected value of perfect information for a parameter (EVPPI) where the parameter in question is the vector of probabilities that determines the temporal trajectory of treatment effect over the unobserved period. However, as explained in the section above, it is difficult given current evidence, even to accurately quantify the uncertainty surrounding treatment effect over the unobserved period. Nonetheless, calculating EVPPI with the current, 'naïve' characterisation of temporal uncertainty ought to convey the difficulty in making an informed judgement based on cost-effectiveness and the appropriateness of obtaining further evidence. The results of an EVPPI analysis are given below assuming uncertainty can be appropriately characterised by ascribing equal weight to each of the 3 scenarios described above, i.e. each have equal probability of being 'correct'. The vector of probabilities p therefore can be written:

$$p = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

The non-parametric computation methods for EVPPI is employed, i.e.

$$EVPPI = E_p \max_j E_{\psi|p} NB(j, p, \psi) - \max_j E_{\theta} NB(j, \theta)^{42}$$

Where:

θ = all uncertain parameters

p = uncertain parameter of interest (nature of long-term treatment effect)

ψ = all uncertain parameters besides parameter of interest

j = alternative interventions

⁴² The uncertainty around p is described by a discrete distribution (the three alternatives scenarios with equal probability of 1/3). The 'temporal parameter' p is then treated like any other uncertain parameter in an EVPPI. Thus, for the left-hand side of the EVPPI equation, a value of p is drawn randomly from the distribution given, a PSA is run, and the greatest net benefit (of the two interventions) is recorded. This is done a number of times and then an average is taken.

Executing this calculation for the 5 risk groups produces the following results:

Table 23: EVPPI for temporal uncertainty relating to treatment effect, given current model settings

	Risk group 1	Risk group 2	Risk group 3	Risk group 4a	Risk group 4b
$E_p \max_j E_{\psi p} NB(j, p, \psi)$	227,062	242,329	132,227	122,005	90,828
$\max_j E_{\theta} NB(j, \theta)$	225,297	241,891	130,151	122,005	90,828
EVPPI/patient	734	438	315	0	0
EVPPI/population	166,807,440	99,539,045	71,586,299	0	0

The EVPPI calculations show that, given the current characterisation of this temporal uncertainty, there is likely to be value in obtaining further evidence in order to inform the adoption recommendations for risk groups 1, 2 and 3. Incorporating more optimistic scenarios of long-term treatment effect has meant there is now a not insignificant probability that it would be cost-effective for the lower risk groups to receive the early interventional treatment strategy. Although allowing for the possibility of continued treatment effect over the long-term is potentially too optimistic, explicitly imposing only a conservative assumption of no continued treatment is likely to be too pessimistic. What this analysis demonstrates is that the scale of the uncertainty existent in the CEDM with the currently available evidence is too vast to make a well-informed decision based on cost-effectiveness and that obtaining further evidence at a reasonable cost is likely to be worthwhile.

However, by the nature of the parameter, there exists no direct relevant evidence pertaining to the treatment effect of a new health intervention beyond the follow-up of a relevant trial. It is sometimes possible to infer long-term behaviour of treatment effect from long-term evidence of comparable health interventions, or from observational cases carried out before the trial⁴³. In the circumstance where there are no existent data with which to characterise the long-term nature of a parameter, the uncertainty around which (as naively characterised) is shown to impact the adoption

⁴³ A search for such evidence did not return any useful results in the RITA-3 example.

decision, it is desirable to incorporate into the model information that best expresses current clinical belief.

In previous sub-chapter, the potential role of expert elicitation was discussed. In particular, it was suggested that the opinion of clinical experts could be usefully employed to validate estimates of long-term baseline risk or perhaps to characterise the uncertainty around long-term estimates. In the situation presented in this sub-chapter however, a more prominent role for expert elicitation may be warranted due to the absence of relevant data and the demonstrable importance of the parameter. Although expert elicitation increasingly features as a source of evidence in economic evaluation, there does not, at present, exist a methodology to employ expert elicitation specifically to inform the (post-trial) temporal trajectory of model parameters. Such a methodology is thus posited and developed here.

4.3.3.4 Employing expert elicitation to characterise long-term behaviour of treatment effect

4.3.3.4.1 Expert elicitation in health technology assessment

Formal elicitation is the process of interpreting, as a probability distribution, a person's knowledge and beliefs about an uncertain quantity (Garthwaite PH, 2005). A person whose knowledge is to be elicited is typically referred to as an 'expert', which can be assumed to simply mean that this person's knowledge and judgement is, on some level, worth having (O'Hagan et al., 2006). Expert elicitation has been employed in a number of fields including HTA. Its use in HTA has been sporadic to date (as found in the HTA review in Chapter 3). Although, preference should typically be given first to the 'gold standard' of randomised controlled trials (Charlton, 1991) as sources of evidence for relative effects, then to forms of observational evidence, formal expert elicitation is likely to be a useful and appropriate endeavour in HTA under a number of circumstances.

- (i) When there are no data to inform the value of an input parameter. Expert opinion is a legitimate source of information where there exist no data to inform particular parameters, or where the data existent are not suitable (Philips et al., 2006)

- (ii) When there is no means to ascribe weights to alternative plausible scenarios; or more generally, no way to quantify the uncertainty surrounding a parameter value of model assumption in order to appropriately estimate the value of obtaining further evidence
- (iii) When it is desirable to use informative priors as part of a Bayesian process (Bojke et al., 2010)
- (iv) When there is a need to evaluate the cost-effectiveness of healthcare interventions that have not used randomised studies of efficacy to inform the licensing process, e.g. medical diagnostics (Sullivan and Payne, 2011)

4.3.3.4.2 Why formal elicitation is a suitable means to inform post-trial treatment effect

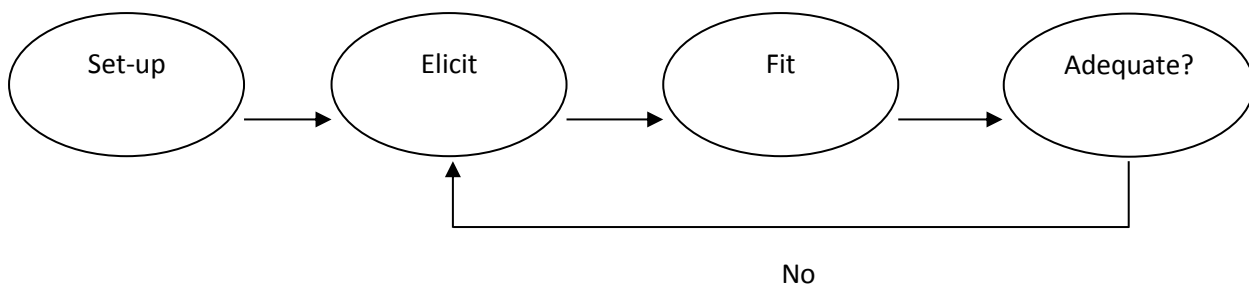
The problem to be addressed in the RITA-3 example is a common one. There does not exist direct evidence to inform treatment effect beyond the time horizon of the relevant trial(s). This problem is an example of the first circumstance outlined above where expert elicitation may be useful and appropriate. In particular, the parameter for which there are no data available to inform can be considered to be the underlying temporal parameter that dictates the temporal trajectory of treatment effect over the unobserved period. It can be assumed that there are no data to inform this temporal parameter if the short-term evidence is deemed unable or unsuitable to inform the temporal behaviour of treatment effect over the long-term, which is true for the RITA-3 example. Moreover, it has been demonstrated that the adoption decision is dependent on how treatment effect is characterised over the unobserved period. As formal expert elicitation is generally relatively parsimonious with both time and money and can be designed to obtain the specific data required, it represents an efficient means of generating the required data so as to produce an informed adoption recommendation (for the immediate future) and the characterisation of uncertainty for the purposes of value of information analysis and future decisions.

4.3.3.4.3 Appropriate execution of expert elicitation

Before positing methods to employ expert elicitation for the purposes of characterising post-trial treatment effect in particular, it is first important to consider the appropriate use of expert

elicitation more generally. There are four key steps to the elicitation process, as outlined by Garthwaite et al. and illustrated in Figure 45.

Figure 45: The elicitation process – based on a diagram from Garthwaite et al. (Garthwaite PH, 2005)



(i) Set-up

Once it has been established that expert elicitation is an appropriate means of obtaining further data, a suitable preparation for the elicitation must take place. This, in short, involves: selecting and training the experts; creating a timeline for the exercise; and importantly, identifying what exactly is to be elicited.

It is important to aim from the outset to elicit probability distributions rather than point estimates (O'Hagan et al., 2006). The key reason for this, as Sculpher et al. outline is that there will inevitably be uncertainty between and within expert opinions. This uncertainty ought to be reflected in the analysis, rather than forcing a consensus (Sculpher et al., 2000).

There are, of course, potential biases associated with eliciting the opinions of individuals (Garthwaite PH, 2005, Kahneman and Slovic, 1982). As well as a careful selection of experts, an obvious tool to reduce bias and to better quantify the uncertainty among experts, is to elicit from as many experts as is feasible (though this must be balanced with the timeliness and expense of the exercise as well as ensuring the experts have the relevant expertise).

(ii) The elicitation process

The crux of the exercise is the elicitation process itself. There have been several methods proposed to elicit data from experts (Jenkinson, 2005, Chaloner and Duncan, 1983, Gavasakar, 1988). There are two aspects for which there is broad consensus: first, group or discussion based elicitation have been deemed inappropriate due to the potential bias toward dominant individuals and group pressure for conformity (Fischer, 1978); second, it is paramount that the expert understands the nature of the parameter he/she is being asked to estimate and also has some understanding of basic statistical concepts like mean and variance.

Among the types of judgement that can be asked of experts are the fixed interval and variable interval, where experts are asked for probabilities and quantiles respectively (Oakley, 2010). A commonly used method in HTA is the histogram approach where individual experts place a number of crosses on a frequency chart with each cross representing a percentage of the distribution of the uncertainty quantity (Van Noortwijk et al., 1992). The visual aspect of this method allows easy understanding and expression of quantitative judgements (for perhaps non-quantitative-minded experts), while output can be quantified with relative ease and with minimal scope for misinterpretation.

(iii) Fitting a distribution

The output from the elicitation exercise must be quantified and synthesised for use in the decision model. Assuming it is reasonably straightforward to quantify the output from individual elicitation, as with the histogram approach, there are a number of synthesising techniques that could be employed to produce a single distribution that may then represent the uncertain parameter for use in a decision model.

Bojke et al. outline and apply four alternative approaches (Bojke et al., 2010): linear pooling without weighting; linear pooling with weighting; random effects meta-analysis without weighting; random effects meta-analysis with weighting.

'Weighting' here refers to the notion of applying differential weights to the estimates of individual experts. If deemed appropriate, differential weights can be generated through calibration, i.e. by taking a parameter for which RCT evidence is available and comparing the estimates of the experts with the RCT data so as to obtain a sense of the reliability of

the different experts. Differential weighting is a contentious issue among elicitation advocates as it is not clear whether the relative suitability of an expert to estimate an unknown parameter can be reliably inferred through calibration (Cooke, 1991).⁴⁴

Linear pooling involves aggregating experts' estimates using simple linear combinations of the form $\theta = \sum_i w_i * i(\theta)$ where θ is the unknown parameter and w_i is the weight of expert i in order to produce a 'super distribution' expressing all estimates (Bojke et al., 2010). As an alternative to linear pooling (which assumes no relationship between experts' distributions), a random effects meta-analysis can be carried out whereby the expert judgements are treated like data which are combined with non-informative priors to produce a posterior estimate of the unknown parameter. This approach incorporates both the within and between expert variation (Bojke et al., 2010).

It is also possible, and arguably desirable, to fit a smooth parametric function rather than directly inputting the discrete distribution formed when the experts' estimates are combined (Leal et al., 2007).

(iv) Assessing adequacy

The final step is to validate, or to in some way, assess the adequacy of the elicitation. This requirement pertains to (a) an elicitation exercise capturing an expert's 'true' beliefs and (b) elicitation output representing an accurate estimate of the unknown parameter.

There are theoretical and psychological issues relating to the idea of obtaining a representation of an expert's 'true' beliefs (O'Hagan, 1988, Winkler, 1967). However, to ensure the output of an elicitation accurately expresses an expert's judgement, Garthwaite recommends that the internal consistency of the expert's statements are tested by running 'tests of coherence' and allowing the expert to revise some of his/her statements (Garthwaite PH, 2005).

Assessing the adequacy of output from an elicitation exercise is also clearly problematic. If there existed data pertaining to the unknown parameter with which to validate the judgements of experts, these data would almost certainly be better used from the

⁴⁴ Note that in the application carried out by Bojke et al., no discernible difference was recorded between using equal weights and using differential weights.

outset, possibly in conjunction with elicitation, or possibly negating the need for elicitation. As data become available over time, then the estimates from elicitation can be calibrated and refined. O'Hagan outlines these calibration methods in some detail (O'Hagan et al., 2006). For a decision that must be made in the immediate future however, the reliability of elicited evidence can be, to some extent, assessed by including a 'seed question', i.e. by comparing elicited output for a known parameter with (ideally) RCT evidence for that parameter (Bojke et al., 2010). Of course, this approach has its limitations, primarily that the 'known' parameter may be relatively easy to estimate compared to the 'unknown' parameter.

If the elicitation is deemed to be invalid, the elicitation process must be repeated, or possibly abandoned, as Figure 45 suggests.

4.3.3.4.4 Methods to employ elicitation specifically to inform post-trial treatment effect

In developing methods for the use of formal elicitation specifically in order to estimate post-trial treatment effect, it is steps (ii) and (iv) from the list above that come into focus. While step (i) (set-up) and step (iii) (fitting a distribution) are crucial elements in the elicitation process, they are issues that must be addressed in any formal elicitation. It is in step (ii) (the elicitation process) and step (iv) (assessing adequacy) where questions arise as to what is the appropriate approach when seeking to estimate post-trial treatment effect in particular.

4.3.3.4.5 The Elicitation Process

Much of what has been recommended in the literature and described under 'step (ii)' above will still hold true for the use of formal elicitation to estimate post-trial treatment effect, e.g. the inappropriateness of group-based elicitation exercises and the merits of the histogram method for the elicitation exercise itself. However, a number of more specific issues arise.

Defining and explaining the problem

The problem of how treatment effect evolves over time could be characterised in a number of different ways. Since clinical experts have found it challenging to express beliefs on some mathematical quantities such as coefficients (Kadane and Wolfson, 1998), careful consideration must be given to how to frame the question posed to experts, both for the purposes of

appropriately filling the gap in the decision model and for the purposes of ensuring the responses from the experts best reflect their understanding of the treatment and disease. Furthermore, the clinical nature of the patient cohorts must be carefully explained to the experts⁴⁵. This is especially important for garnering the insights of experts regarding how one cohort is expected to change *over time* with respect to another.

Quantifying predicted changes over time

A typical task of elicitation is to quantify an unknown parameter. However the specific task of quantifying the change in a parameter over time presents particular challenges. As per the discussion of temporal uncertainty in general (see Chapter 2), change over time in health economic models can be characterised in a number of different ways and when employing elicitation, care must be taken to ensure there can be a good understanding of the quantity being expressed (indeed the use of expert elicitation is a further factor to consider as we choose which modelling approach is most appropriate). Graphical techniques can be used to allow experts to express change over time, including non-linear change. For survival data in particular, elicitation methods outcomes have been developed in the field of engineering (Jager and Bertsche, 2004, Campodónico and Singpurwalla, 1994), but such techniques have not been employed in health economic modelling to date.

Relative or absolute measures?

With regard to treatment effect in particular, there is the option to quantify not the relative measure (i.e. treatment effect itself), but the absolute measure (in the case of RITA-3, the hazards/probabilities of the early interventional treatment cohort). The latter may be a more intuitive measure for the experts to quantify. However, it is arguably more relevant to estimate the measure for treatment group relative to baseline, i.e. the most useful input from experts may relate to how one cohort ‘differs’ from another as time elapses.

Weight scenarios or generate scenarios?

In some cases, elicitation has been carried out by asking experts to weight (or ascribe probabilities to) pre-determined alternative scenarios (Bojke et al., 2010). If it is indeed appropriate to characterise the uncertainty as discrete scenarios (as was done in the initial characterisation in 3.2.1) then at the very least these scenarios ought to be validated by the experts. It would seem more appropriate however, for the scenarios themselves to be generated by the experts – even if they

⁴⁵ A detailed breakdown of the clinical nature of the cohorts and the nature of the uncertain quantity is given in Appendix 3 where a stylised elicitation exercise for the RITA-3 example is outlined.

were to take the general form of 'optimistic', 'pessimistic' and 'middle-ground' scenarios. In any case, weighting discrete scenarios may not be an appropriate characterisation of the uncertainty.

Expressing the trial evidence

There is a general question of how to incorporate the short-term trial evidence. Assuming for now, that it is desirable for the experts' judgements to be partly based upon the observed outcomes in the short-term data, a mechanism that usefully and clearly expresses the outcome of the trial analysis is warranted.

With these issues in mind, a number of alternative approaches could be considered.

(i) Eliciting the temporal trend of treatment effect directly

One option would be to elicit directly on the temporal trend of the treatment effect. This could be done explicitly by estimating treatment effect as a single relative effect, or implicitly by estimating the absolute risk associated with the treatment cohort.

The relative effect approach would effectively involve eliciting on an unknown 'temporal parameter' that dictates the change in value over time of treatment effect after the observed period. In other words, treatment effect would be some function of the temporal parameter. Framing the problem this way is attractive as the missing information (how treatment effect evolves over time) is represented by a single unknown parameter, which is also useful for value of information analysis. However, this temporal parameter may be difficult for clinical experts to correctly interpret and then quantify; it would be especially complicated to allow for anything other than a linear change over time.

Instead of quantifying a treatment effect (i.e. a relative effect), the absolute risk for the treatment cohort could be quantified. For the RITA-3 example, it was argued in Section 4.1.7.1 that it would be preferable to characterise baseline risk and then treatment effect as opposed to absolute risks. However, given that the available supplementary evidence in the form of hazard/odds ratios was not useful for characterising long-term treatment effect, the option to consider the absolute risk for patients in the early interventional cohort may once again become valid. A common elicitation technique that would be suitable for this approach is to mark a number of time-points over the unobserved period and simply ask the experts, what proportion of the cohort they would expect to

have experienced the event by each time point (Soares et al., 2011). From this, hazard rates or transition probabilities can be computed, albeit forming a piecewise linear curve. Estimating event occurrence for a single cohort of patients would arguably be easier for clinicians (as opposed estimating treatment effect which could be a difficult concept to understand especially where two or more cohorts are evolving over time). However, it may be more important to elicit the judgments of experts regarding the treatment group in light of what has been observed (and assumed) regarding the control/baseline group. If this approach is to be considered therefore, thought must be given to how the information and assumptions regarding baseline risk can be conveyed clearly.

(ii) Eliciting on the duration of treatment effect

Thinking again in terms of a relative effect, a simpler approach would be to elicit the experts' judgements regarding the *duration* of the treatment effect from the trial time horizon onwards. This approach would involve an assumption that the magnitude of treatment effect remains as it was in the observed period and it is simply the longevity of this effect that is uncertain. The elicitation could be carried out by asking the experts to estimate a distribution for a parameter (for the RITA-3 example, this parameter would be bounded by 0 and 55, representing the timespan in years over the unobserved period). This approach was developed and employed successfully in a health technology evaluating enhanced external counterpulsation for the treatment of stable angina and heart failure (McKenna et al., 2009). This approach would be relatively intuitive for the clinical experts (as long as the meaning of treatment effect in this context is clearly explained). The approach would however, simplify the issue somewhat, in effect trading accuracy for clarity. Nonetheless, this characterisation of the uncertainty ought to be sufficient to fill the evidence gap.

(iii) Eliciting weights for alternative pre-defined scenarios

A further option would be to discretise the uncertainty and ask the experts to simply ascribe weights to alternative scenarios. The temporal uncertainty in question may not be accurately represented by discrete alternative scenarios but nonetheless this may be a useful simplification. As stated above, the choice of scenarios would need to be validated by, if not generated by, the experts. The most straightforward execution of this approach would be to weight the scenarios that have already been generated (based on the guideline of considering optimistic, pessimistic and middle-ground scenarios). This is arguably the simplest approach for the elicitation process (both for the modellers and the experts) but it may be too restrictive.

4.3.3.4.6 Assessing Adequacy

Assessing adequacy in terms of obtaining an expert's 'true' belief does not propose a substantially different challenge for treatment effect over time than for any other parameter. It is simply imperative that the experts understand the concepts of treatment effect and change in cohorts over time. It would be sensible to apply the internal consistency tests recommended by Garthwaite.

Assessing adequacy in terms of the reliability and accuracy of the elicited evidence is clearly challenging. The use of one or more seed questions is desirable and plausible bounds on the magnitude on treatment effect (something the judgements of clinical experts would be expected to be well within) ought to be applied. After these checks however, the elicited evidence must simply be considered to be the 'best' currently available evidence.

It was not within the scope of this thesis to carry out a real life elicitation exercise. However, a stylised elicitation exercise was constructed and applied to the RITA-3 example. This exercise and results are outlined in Appendix 3.

4.3.4 Results

It is preferable to carry out the re-analysis of the RITA-3 CEDM based on existent evidence. Therefore, rather than employing the stylised elicitation, the results given in this section are based on the 'naïve' characterisation of temporal uncertainty outlined in section 4.3.3.2.1, i.e. three broad alternative scenarios given equal weighting. Given the available evidence, this characterisation best reflects current knowledge and expresses considerable uncertainty as is appropriate. However, were a decision to be based on this analysis, it ought to be strongly conveyed to the decision maker that the characterisation of temporal uncertainty regarding treatment effect over the long-term is naïve and as such this uncertainty may be under (or over) estimated. In practice, in such a scenario (where the mean ICER is close to the threshold and as a result there is significant decision uncertainty), there would certainly be a need for some expert input to better characterise this particular source of uncertainty that plainly has such sway over the adoption decision.

4.3.4.1 Cost-effectiveness

Table 24: Cost-effectiveness Results after temporal uncertainty has been addressed for treatment effect. Compare with Table 18 to observe the change in the results.

Risk group	Incremental Cost (£)	Incremental QALY	Mean ICER	Adopt/reject early interventional (EI)
Risk group 1	5197	0.2141	24,227	Reject
Risk group 2	5132	0.3718	13,804	Adopt
Risk group 3	7,098	0.4851	14,632	Adopt
Risk group 4a	7080	0.7043	10,052	Adopt
Risk group 4b	6989	0.6481	10,784	Adopt

The characterisation of temporal uncertainty which incorporates alternative scenarios regarding the nature of treatment effect over the unobserved period has replaced the ‘conservative’ assumption that had been in place where no further treatment effect was assumed. As a result, the relatively pessimistic mean ICERs have been replaced with mean ICERs which, as far as possible given the evidence immediately available, reflect current expectations. All but one risk group (risk group 1) now returns a positive adoption decision⁴⁶.

⁴⁶ Note that the risk groups are ordered by risk of experiencing a first composite event. This does not necessarily equate to order of resultant costs, QALYs, or cost-effectiveness. For instance, costs vary by risk group characteristics and higher age leads to fewer expected life years.

4.3.4.2 Uncertainty and Value of Information

Table 25 below summarises the current impact of uncertainty on the outputs of the CEDM for each of the 5 risk groups.

Table 25: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for treatment effect. Compare with Table 19 to observe the change in the results.

Risk group	Prob(EI cost-effective) at £20,000/QALY	EVPI/patient (£)	EVPI/population (£)
Risk group 1	0.31	848.16	192,714,863
Risk group 2	0.64	476.62	108,174,852
Risk group 3	0.69	423.19	96,130,173
Risk group 4a	0.96	47.25	5,340,565
Risk group 4b	0.96	44.85	4,999,678

The early interventional treatment for both risk groups 2 and 3 is now likely to be cost-effective, though there would be considerable value in obtaining further evidence. The shift downwards of all mean ICERs has meant that there is now a significant probability that the early interventional strategy may also be cost-effective for risk group 1 (the lowest risk group). It is also in relation to risk group 1 that there would be most value in obtaining further information.

4.4 Chapter 4.4: Costs

4.4.1 Introduction

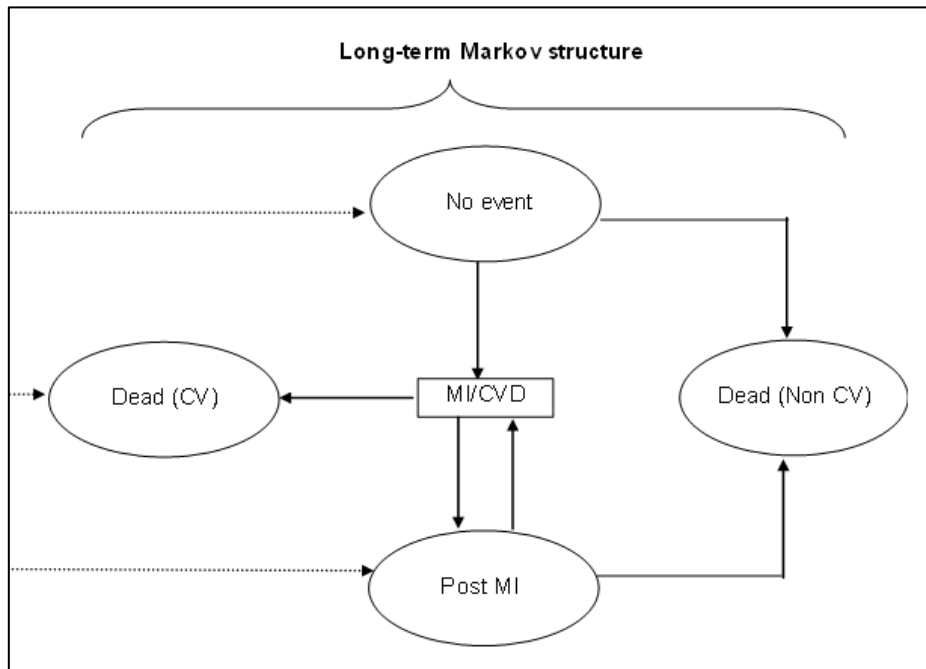
The purpose of this sub-chapter is to address temporal uncertainty relating to costs in the RITA-3 cost-effectiveness decision model (CEDM). In doing so, this chapter endeavours to address, more generally, the characterisation of temporal uncertainty pertaining to costs in CEDMs. First, a series of sensitivity analyses are carried out in order to test the robustness of the cost-effectiveness results to alternative assumptions regarding how costs evolve over time. Second, a number of factors that may impact long-term costs are discussed and temporal uncertainty is expressed. These factors are: the impact of age when estimating long-term costs, accounting for inflation when estimating long-term costs and accounting for uncertain future events when estimating long-term costs. The latter two issues will be for illustration and discussion only and re-analysis will not be incorporated into the updated CEDM.

4.4.1.1 The parameters

Recall that the long-term portion of the RITA-3 CEDM is represented by a state transition structure, where 'events' cause proportions of the patient cohort(s) to migrate to and from health states. Both experiencing an event and spending a cycle in a health state has an associated cost. Such a cost is calculated as a product of unit costs and resource use. In Chapter 3, it was discussed how, in an event-based model, costs accumulate via patients moving between and residing in health states over time. In this sense, cumulative costs are driven, to a large extent, by the event rates assumed in the model. The uncertainty regarding future costs therefore has, to some degree, been addressed through the analysis of event rates carried out in Chapters 4.1, 4.2 and 4.3. However, there will inevitably also be uncertainty relating to the evolution over time of the resource use and unit costs associated with health states and health events. It is these parameters that are the focus of this sub-chapter. In particular, the parameters under analysis are:

- The costs associated with a composite health event, i.e. transitioning to the 'MI/CVD' health state
- The costs associated with residing in the 'No event' health state and the 'Post MI' health state

Figure 46: Costs are associated with each health state in the long-term Markov portion of the CEDM



Note: Costs are assumed to be zero for the death states and the MI/CVD state is assumed to be instantaneous in time and thus also has zero costs.

4.4.2 Available evidence

4.4.2.1 Evidence from RITA-3

Detailed resource use data were collected as part of RITA-3 over the 5 years of trial follow-up. The details of the items of resource use were outlined and analysed in detail in a study examining the costs of an early intervention versus conservative strategy in NSTEMI-ACS (Epstein et al., 2008). Briefly, the items of resource use are categorised into: angiogram, percutaneous coronary interventions, coronary bypass surgery, myocardial infarction, key cardiac medications and other costs. Items variously pertain to: intervention (conservative and early interventional), the no event state in the first year after intervention (conservative and early interventional), the no event health state for second and subsequent years, the post MI state in the first year and the post MI state for the second and subsequent years.

4.4.2.2 Other evidence

Unit cost data can be derived from RITA-3's predecessor, RITA-2 (1997). The unit costs of pertinent consumables were recorded in a survey of five centres carried out as part of RITA-2. These unit costs were updated to current price levels as part of the RITA-3 cost analysis (Epstein et al., 2008). Further unit costs were derived from reference costs, PSSRU and the British National Formulary (BNF) (Sculpher et al., 2002, Health, 2004).

4.4.3 Analysis

4.4.3.1 One-way Sensitivity Analyses

4.4.3.1.1 Costs associated with a composite health event, i.e. transitioning from the 'No event' health state to the 'MI/CVD' health state

In the original analysis, standard OLS regressions were employed to determine each required mean cost, for each risk group, based on the unit cost and resource use data available. Table 26 shows the estimated mean cost per relevant co-variate.

Table 26: Estimated mean cost per relevant co-variate

Covariate	Coefficient	95 % CI
First year after non-fatal MI	5,467	3,890 to 7,044
Treat	-1,106	-1,562 to -650
Male	586	111 to 1,061
Angina	1,034	550 to 1,518
Previous MI	724	210 to 1,239
Constant	2,735	2,249 to 3,220

The mean cost associated with a non-fatal MI, i.e. the cost incurred for the first year after a non-fatal MI for risk group 3 was calculated as £9,440. No difference was assumed between the interventions and this cost was assumed to not change over time.

With this simple assumption of a constant mean cost, the following cost-effectiveness results are obtained for risk group 3.

Table 27: Cost-effectiveness results for assumption of constant costs

Constant costs for first year after MI	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
	7,098	0.4851	14,632	0.6913	Adopt

The robustness of the simple temporal assumption of constant costs can be assessed by applying alternative assumptions of increasing costs over time and decreasing costs over time.

Table 28: Cost-effectiveness results for alternative assumptions regarding costs associated with a composite event

Δ Costs	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Constant costs	7,098	0.4851	14,632	0.6913	Adopt
↑ 2%/annum	7,113	0.4851	14,662	0.6820	Adopt
↑ 5%/annum	7,320	0.4851	15,089	0.6741	Adopt
↓ 2%/annum	7,082	0.4851	14,599	0.7010	Adopt
↓ 5%/annum	7,072	0.4851	14,578	0.7087	Adopt

While it is clear that the temporal assumption regarding costs incurred during the first year after a non-fatal MI does have an effect on the ICER, all else being equal, the effect is not significant. The adoption decision could therefore be said to be robust against this particular uncertainty.

4.4.3.1.2 Costs associated with residing in the 'No event' health state and the 'Post MI' health state

The costs associated with the 'No event' health state and the 'Post MI' health state were also derived, in the original analysis, from the simple OLS regression depicted in Table 26. The cost associated with the 'Post MI' health state (for the second and subsequent years) was assumed to

equal to the cost associated with the 'No event' health state (for the second and subsequent years). This is because it was assumed the cost of revascularisation – the biggest contributor to mean annual costs - would be the same in the two strategies from this point onwards. Again a temporal assumption of constant costs over time was employed. The mean cost associated with both the 'No event' and 'Post MI' health states for risk group 3 was £3973.

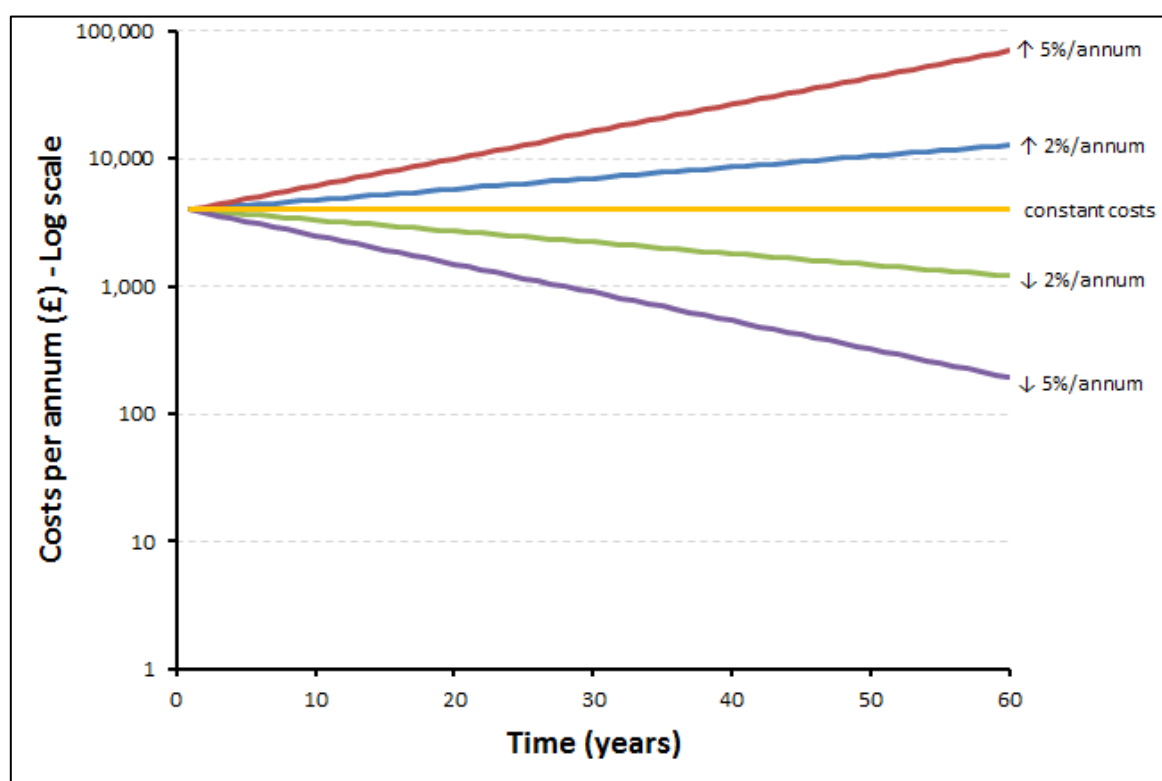
The robustness of the simple temporal assumption of constant costs over time can again be assessed by applying alternative assumptions of increasing costs over time and decreasing costs over time. Table 29 below shows the cost-effectiveness results for alternative assumptions relating to costs for risk group 3.

Table 29: Cost-effectiveness results for alternative assumptions regarding the change over time of costs per health state for risk group 3

Δ Costs	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Constant costs	7,098	0.4851	14,632	0.6913	Adopt
↑ 2%/annum	9,509	0.4851	19,602	0.4762	Adopt
↑ 5%/annum	15,852	0.4851	32,678	0.0104	Reject
↓ 2%/annum	5,305	0.4851	10,935	0.8210	Adopt
↓ 5%/annum	3,645	0.4851	7,513	0.9056	Adopt

It is plain that the mean ICER is far more sensitive to the temporal assumption made regarding the long-term costs per health state compared to the temporal assumption regarding the cost of experiencing a non-fatal MI. This is because higher costs associated with health states over the long-term would imply more expensive medical management of surviving patients, of whom there are more in the early interventional cohort. Although these rates of increase/decrease are extreme, the potential impact on the ICER ought to prompt a consideration of what factors may cause these costs to in practice either increase or decrease over time. There is, in fact likely to be upward pressure on resource use and downward pressure on unit costs over time as will be discussed in sections 3.4 and 3.6 respectively. The trajectory over time of these health state costs are illustrated in Figure 47 below.

Figure 47: The trajectory over time of costs for risk group 3 according to 5 alternative assumptions



4.4.3.2 Expressing Temporal Uncertainty

In the previous section, one-way sensitivity analyses were carried out on the key temporal assumptions related to costs in order to appraise their impact on the mean ICER and ultimately the adoption recommendation. A number of factors that may cause costs to alter over the long-term are now investigated. Where appropriate, temporal uncertainty is quantified and incorporated into the updated model.

4.4.3.2.1 Accounting for age when estimating long-term costs

Similar to HRQoL, long-term costs per health state are likely to be affected by ageing. Specifically, it is likely that resource use will increase with age. In the original RITA-3 CEDM, constant costs (for both the 'No event' and 'Post MI' health states) were assumed for second and subsequent years (following intervention and MI respectively). However, detailed analysis by Epstein et al. on the cost outcomes of the RITA-3 trial reveals that age is a predictor of increased mean costs (Epstein et al., 2008). In particular, it was found, through a multivariate analysis, that mean costs increase by £737

for every 10 years for patients over 60 years of age. 95% confidence intervals were also given (£342, £1147). This age-related increase can be incorporated into the model and the uncertainty surrounding the magnitude of the increase can be expressed through a gamma distribution based on the confidence intervals given.

$$10 \text{ yearly increase in costs per health state} \sim \text{Gamma}(12.92, 57.02)$$

The application of this uncertain age-related increase in costs yields the cost-effectiveness results shown in Table 30.

Table 30: Cost-effectiveness results for risk group 3 when no age-related increase in costs is assumed versus when an uncertain 10 yearly increase in costs is assumed

Age-related increase in costs	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
No increase	7,098	0.4851	14,632	0.6913	Adopt
Uncertain 10 yearly increase	7,582	0.4851	15,630	0.6890	Adopt

The results show a slight upward shift in the ICER leading to a decreased probability of cost-effectiveness. This is explained by the increased costs of medical management of those surviving patients over the long-term (of whom there are more in the early interventional cohort).

4.4.3.2.2 Accounting for inflation when estimating long-term costs and health effects

It has been observed (in Chapter 3) that assuming constant costs per health state over time is a common practice in HTA and indeed that is the case in the original RITA-3 CEDM. However, leaving aside resource use for now, it is almost certain that unit costs – the prices of various clinical equipment and procedures – will alter over time. It may be that some items become relatively cheap due to improvements in technology or clinical practice, but costs will almost certainly come under upward pressure due to inflation.

There are two options regarding handling inflation in CEA, given that discounting must also be accounted for due to time preference (Drummond et al., 2005, Kumaranayake, 2000):

- (i) Inflate all future costs by predicted inflation rate and apply a larger discount rate that accounts for general inflation (an 'inflation-adjusted' discount rate)
- (ii) Do not inflate future costs and use smaller discount rate, i.e. 'real' discount rate. This was assumed to be the case in the original RITA-3 analysis and indeed most CEAs, as it is the approach recommended by NICE

NICE recommends a 'real' discount rate of 3.5% for both costs and health effects (i.e. taking account of inflation). The 'real' discount rate used in health economic evaluations in the UK is taken from the Green Book of the UK Treasury, which is in principle based on the social time preference of the general population (Treasury, 2003). Although a wealth of literature and debate exists on how to estimate a suitable discount rate, or whether differential discounting (between costs and QALYs) is suitable (Gravelle and Smith, 2001, Claxton et al., 2011b, Claxton et al., 2006, Brouwer et al., 2005), it suffices to state in this research, that both the absolute social discount rate and future rates of inflation (and thus the 'real' discount rate) are subject to considerable uncertainty. Generally, this 'methodological uncertainty' is set to one side in HTA. However, the NICE guidance suggests that for assessments with long-term time horizons, a sensitivity analysis using discount rates of 1.5% for both costs and health effects may be presented alongside the base-case analysis (NICE, 2013). It has been suggested that this may not even be appropriate as a sensitivity analysis (HEDS, 2011).

It is, at the very least, interesting to note the sensitivity of the mean ICER and decision recommendations to the discount rate(s) – and implicit inflation rates - used. Let us first perform the sensitivity analysis suggested by NICE⁴⁷.

⁴⁷ Note that the assumption made here is that it is inappropriate to apply differential discount rates to costs and health effects. Therefore whatever rate we apply to costs, we will also apply to health effects.

Table 31: Cost-effectiveness results for alternative assumptions regarding discount rates

Discount rate for costs & benefits	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
3.5%*	7,582	0.4851	15,630	0.6890	Adopt
1.5%*	8,865	0.6697	13,237	0.8130	Adopt

*incorporating the temporal uncertainty quantified thus far in chapter 4.4

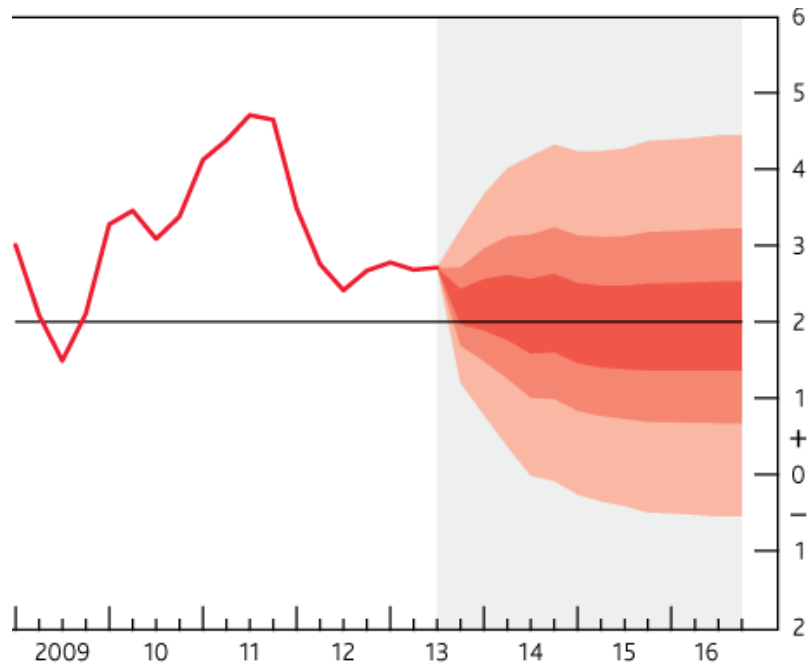
The effect of this change in discount rate is not insignificant but not dramatic. The more favourable result for the early interventional strategy is a consequence of the long-term health gains associated with this strategy ‘mattering more’ when a lower discount rate is applied (recall that the majority of costs are incurred upfront when intervention takes place).

Instead of applying a ‘real’ social discount rate (i.e. the latter of the two approaches for addressing outlined above), let us employ the former option, applying an ‘inflation-adjusted’ discount rate of 5%, coupled with an uncertain inflation rate. An outcome x in year y therefore is subject to the following adjustment where d is the discount rate and i is the inflation rate:

$$x \rightarrow \frac{x(1+i)^y}{(1+d)^y}$$

Estimating long-term inflation is, unsurprisingly, highly difficult. The graph in Figure 48, taken from a recent Bank of England inflation report, depicts consumer price index (CPI) inflation projection based on market rate expectations and ongoing central bank asset purchases (Bank of England, 2013). The confidence intervals shown demonstrate the vast uncertainty inherent in predicting inflation rates even 3 years into the future. It is separate matter again as to whether the CPI index would be an appropriate inflation benchmark for health costs and effects, but what this estimation conveys is the scale of the uncertainty present when valuing health costs and health effects so far into the future.

Figure 48: CPI inflation projection: percentage increase in prices on a year earlier (Bank of England, 2013)



Let us take the latest long-term inflation estimates from PricewaterhouseCoopers (which gives expected inflation until the year 2020) (PricewaterhouseCoopers, 2014). Let us then apply normal distributions to the predicted rates with 90% confidence intervals of $\pm 2.5\%$ for years 1 to 3. Let us further assume that these confidence intervals rise to $\pm 4\%$ for the subsequent years⁴⁸. It is assumed that, after 5 years, there is a change in the inflation rate every 5 years. Figure 49 below depicts the assumed mean inflation rate, the 90% confidence intervals and an example of a simulated inflation rate curve.

⁴⁸ Note that this analysis is purely for illustrative purposes. The base year for the original CEA is in fact 2008, but it was felt it would be more interesting and informative to use the latest inflation data

Figure 49: Simulated long-term inflation rate

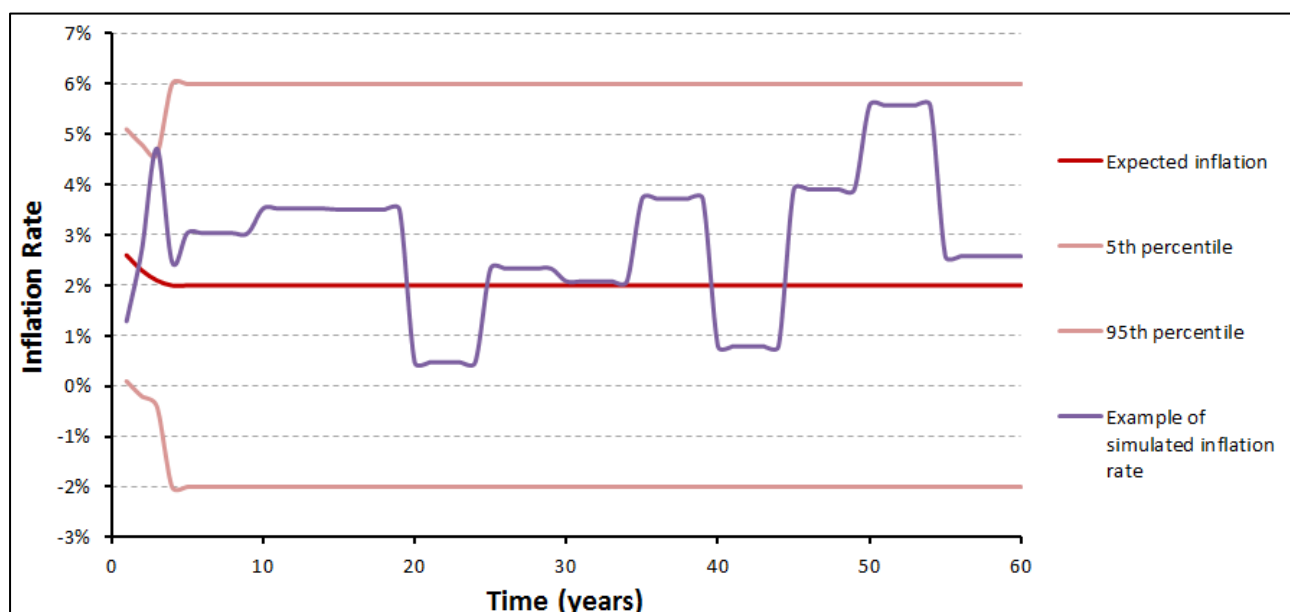


Table 32 below shows the CE results when this ‘inflation adjusted’ discount rates along uncertain inflation approach is taken, versus the standard ‘real’ discount rates approach.

Table 32: Cost-effectiveness results when this ‘inflation adjusted’ discount rates along uncertain inflation approach is taken, versus the standard ‘real’ discount rates approach

Discount rate for costs & benefits	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
‘real’ discount rates 3.5%*	7,582	0.4851	15,630	0.6890	Adopt
‘inflation adjusted’ discount rates + uncertain inflation*	7,980	0.5386	14,649	0.7656	Adopt

*incorporating the temporal uncertainty quantified thus far in chapter 4.4

The effect on the mean ICER (and ultimately the adoption recommendation) is small. And although, further parameter uncertainty has been incorporated, the effect on overall decision uncertainty is also small.

Whatever the inflation curve assumed, the same rates will be applied to both costs and health effects for both treatment strategies. Only therefore, if there are significantly greater health benefits for over the long-term (or indeed significantly greater costs) for one treatment would there be expected to be a notable impact on an ICER. It may be cautiously concluded therefore, that in many CEAs, the choice of approach to accounting for inflation will not meaningfully impact the outcome of the analysis, unless the estimated inflation rate differs wildly from that implied by the 'real' discount rate and there is significant temporal disparity between costs and health effects.

Certain 'methodological uncertainties' such as discount rates, growth over time of CE threshold, etc. are generally thought of as a value judgements that are made by the relevant decision-making authority and outlined in local methodological guidance. For the purposes of producing an updated RITA-3 CEDM in this research, the principle of taking methodological guidance at face value and not incorporating methodological uncertainty into the CE analysis will be adhered to. Thus the uncertainty surrounding inflation is not included in the overall results.

4.4.3.2.3 Accounting for uncertain future events when estimating long-term costs

Although the task of accounting for inflation, to a large extent, is married to the task of employing sound and appropriate economic theory, there are often more tangible reasons to consider changes in unit costs in future time periods. Possible sudden future changes in unit costs forms the subject matter of this final investigation. Uncertain future events (UFE) can have a significant bearing on adoption decision, but only under particular circumstances, while their effect on future decisions, and consequently the value of further evidence is to be expected. The extent of these impacts is tested for RITA-3 using two simulated UFEs.

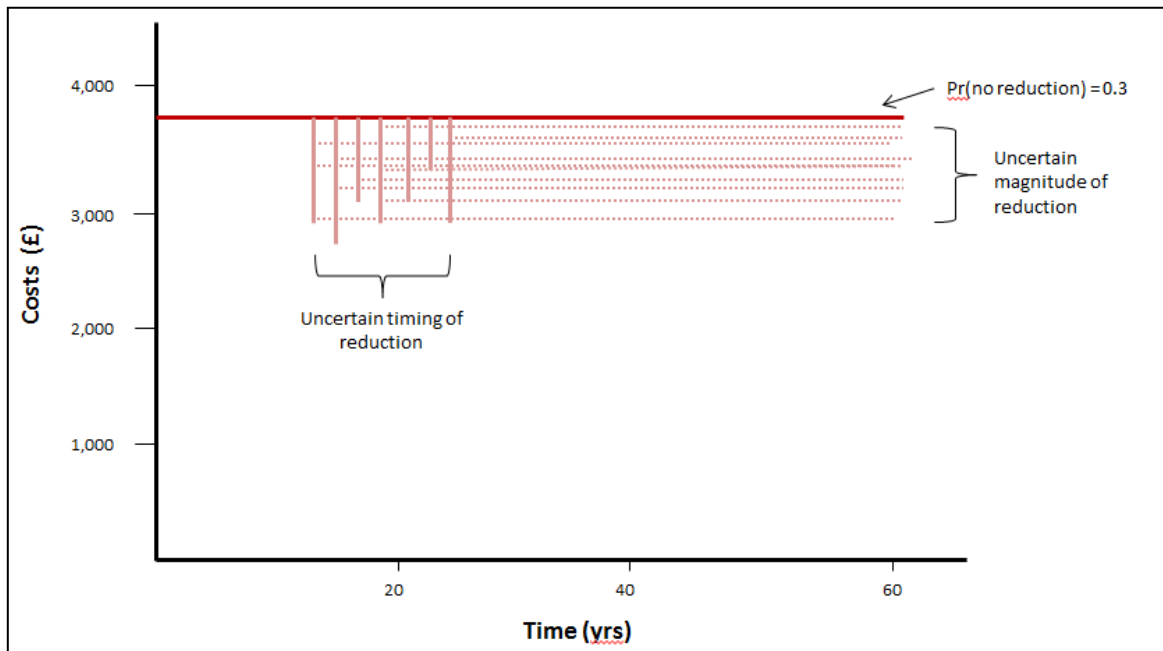
4.4.3.2.3.1 An Uncertain Reduction in Unit Cost of Cardiac Medication

Let us assume that there is an anticipated reduction in the unit cost of a key long-term cardiac medication and that as a result the annual cost associated with both the 'No Event' health state and the 'Post MI' health state is expected to reduce some time in the future. There are potentially three uncertainties at play:

- (i) Uncertainty regarding the occurrence of the cost reduction (there may also be second-order uncertainty, i.e. uncertainty regarding the value of the probability of a cost reduction, but it is assumed here that there is none)
- (ii) Uncertainty regarding the magnitude of the cost reduction
- (iii) Uncertainty regarding the timing of the cost reduction

Let us assume that the uncertainty regarding the occurrence of the cost reduction can be expressed by a single probability 0.7, that the uncertainty regarding the magnitude of any reduction can be expressed as follows: $reduction\ r \sim Gamma(10, 100)$ and that the uncertainty regarding the timing of any cost reduction can be expressed as follows: $timing\ t \sim Normal(20, 2)$. It may, in some circumstances, make more sense to express a single distribution for the magnitude of the reduction with a mass point at zero rather than separate estimates for the probability of a reduction and the magnitude of a reduction, but for this example it is assumed that having separate estimates is suitable. It is also assumed that these distributions are independent. The graph in Figure 50 illustrates this temporal uncertainty.

Figure 50: Temporal uncertainty relating to uncertain future cost reduction



The CE results in Table 33 show the contrasting results for the model with the above simulated UFE and without.

Table 33: Cost-effectiveness results with and without UFE

Discount rate for costs & benefits	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
No UFE*	7,582	0.4851	15,630	0.6890	Adopt
UFE regarding annual costs*	6,793	0.4851	14,003	0.7126	Adopt

*incorporating the temporal uncertainty quantified thus far in chapter 4

There is a notable impact on the mean ICER, though not significant. The change in assumption is a positive one for the early interventional strategy as there is a greater proportion of patients in the ‘No event’ health state, leading to reduced overall costs when a reduction in annual costs is applied.

For this UFE example, the future change, though only becoming manifest over the long-term, affects the patients treated today and for whom an adoption recommendation is being made today.

However, a UFE may also affect a parameter that will not be relevant to the population of patients treated today, but to future incident populations.

4.4.3.2.3.2 An Uncertain Increase in the Cost of Angiography

Putting to one side the fact that such a future event is unlikely in reality, let us simulate an anticipated increase in the cost of an early invasive angiography. Assume again that probability of the price shift is 0.7, that the magnitude of the increase can be expressed as $increase\ i \sim Gamma(50, 100)$ and that the timing of the price shift, if it occurs, is certain to be 10 years from the present year.

The UFE will not affect today’s incident population as it is assumed that all relevant (early interventional) angiographies will have taken place by the time the price change occurs. The impact of such an event relates solely to future decisions. The adoption recommendation for these future incident populations can be made at the relevant future date, i.e. a new decision can be made (albeit with some caveats), as discussed in Chapter 2. What is relevant today however, is the impact of the value of obtaining further information (Vol).

The calculation of Vol will be carried out in detail in the final sub-chapter (Chapter 4.6), but for now let us make some observations in relation to this UFE. The expected value of perfect information per patient (EVPI/patient) for risk group 3 given the current model specifications is £357.58 EVPI/patient given the UFE regarding the cost of angiography is £402.29, implying the estimates for EVPI/population as given below.

	EVPI/total population
Without UFE re cost of angiography	£81,262,949
With UFE re cost of angiography	£91,423,658

This distinction between the two types of UFE simulated in this section is important. Anticipated future changes ought to be accounted for in cost-effectiveness decision modelling, but such changes may not always be relevant for today's adoption recommendation. They will however, always be relevant for value of information analysis (assuming the change occurs within the timeframe when evidence is deemed still relevant). More generally this distinction illustrates the explanation given in Chapter 2 that different time horizons and different decisions are at play as uncertainty over time is accounted for.

4.4.4 Results

The results after temporal uncertainty has been expressed for costs are outlined below. Only the update regarding age-related cost increase is included.

4.4.4.1 Cost-effectiveness

Table 34: Cost-effectiveness Results after temporal uncertainty has been addressed for costs. Compare with Table 24 to observe the change in the results.

Risk group	Incremental Cost (£)	Incremental QALY	Mean ICER	Adopt/reject early interventional (EI)
Risk group 1	5,435	0.2055	26,455	Reject
Risk group 2	5,370	0.3783	14,197	Adopt
Risk group 3	7,582	0.4851	15,630	Adopt
Risk group 4a	7,438	0.7099	10,478	Adopt
Risk group 4b	7,251	0.6463	11,220	Adopt

All ICERs have increased slightly after this update for the reason outlined in Section 4.4.3.2.1. This expression of temporal uncertainty has, like in the previous sub-chapter, replaced a ‘conservative’ assumption. This assumption however, unlike other conservative assumptions, was conservative regarding the magnitude of long-term costs, not conservative regarding the effectiveness of a health intervention. Thus, this update has caused a pessimistic shift in results in relation to the cost-effectiveness of the early interventional strategy.

4.4.4.2 Uncertainty and Value of Information

Table 35 below summarises the current impact of uncertainty on the outputs of the CEDM for each of the 5 risk groups.

Table 35: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for costs. Compare with Table 25 to observe the change in the results.

Risk group	Prob(EI cost-effective) at £20,000/QALY	EVPI/patient (£)	EVPI/population (£)
Risk group 1	0.274	624.35	141,888,590
Risk group 2	0.638	510.87	116,099,342
Risk group 3	0.689	398.24	90,503,263
Risk group 4a	0.974	28.81	3,273,653
Risk group 4b	0.948	58.94	6,697,296

4.5 Chapter 4.5: Health-related Quality of Life

4.5.1 Introduction

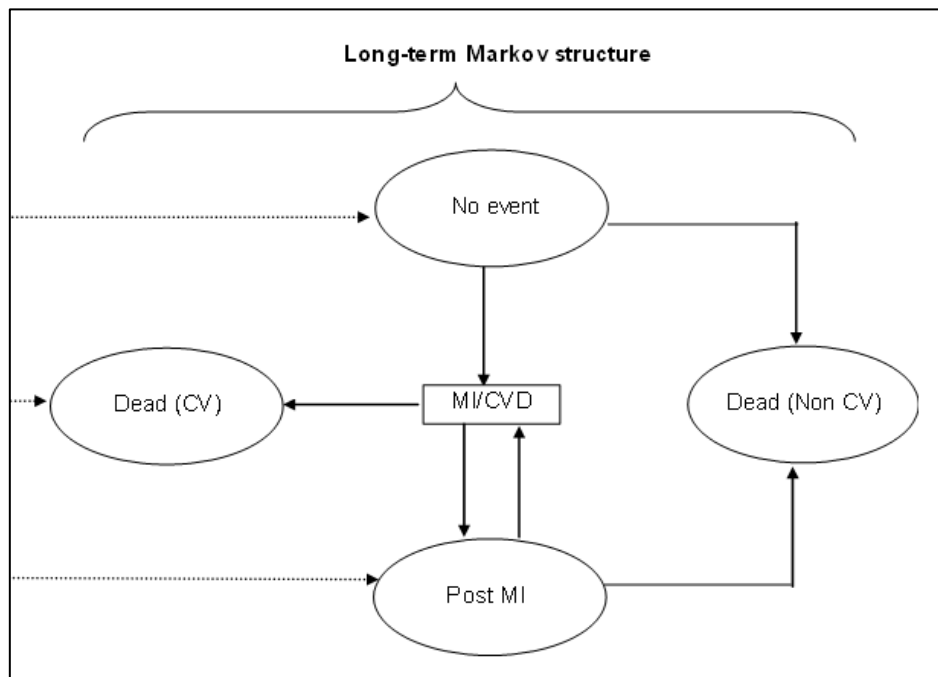
The purpose of this sub-chapter is to address temporal uncertainty relating to health-related quality of life (HRQoL) in the RITA-3 cost-effectiveness decision model (CEDM). In doing so, this chapter endeavours to address, more generally, the characterisation of temporal uncertainty pertaining to HRQoL in CEDMs. First, a series of sensitivity analyses are carried out in order to test the robustness of the cost-effectiveness results to alternative assumptions regarding how HRQoL evolves over time. Second, two factors that may impact long-term HRQoL are discussed and temporal uncertainty is expressed and incorporated into the updated CEDM. These factors are: a possible treatment effect regarding HRQoL and the impact of age when estimating long-term HRQoL.

4.5.1.1 The parameters

Recall that the long-term portion of the RITA-3 CEDM is represented by a state transition structure, where 'events' cause proportions of the patient cohort(s) to migrate to and from health states. Time spent in a health state is associated with the accrual of HRQoL. The occurrence of an event may also be associated with a 'one-off' hike or a drop in HRQoL. As with costs therefore, cumulative HRQoL is, to a large extent, driven by the event rates assumed in the model. However, there is also uncertainty related to how the state-related and event-related HRQoL might evolve over time. It is these parameters that are the focus of this sub-chapter. In particular, the parameters under analysis are:

- (i) The HRQoL associated with a composite health event, i.e. transitioning to the 'MI/CVD' health state
- (ii) The HRQoL associated with residing in the 'No event' health state and the 'Post MI' health state

Figure 51: Unit costs and utilities are associated with each health state in the long-term Markov portion of the CEDM



Note: HRQoL weights are assumed to be zero for the death states and the MI/CVD state is assumed to be instantaneous in time and thus also has zero costs and HRQoL.

4.5.2 Available evidence

4.5.2.1 Evidence from RITA-3

HRQoL data were collected as part of RITA-3 at randomisation, at 4 months, at one year and yearly thereafter. The data were recorded using EQ-5D. HRQoL weights were calculated through employing preferences from the UK population (Brooks, 1996, Dolan, 1997).

4.5.3 Analysis

4.5.3.1 One-way Sensitivity Analyses

4.5.3.1.1 The HRQoL associated with a composite health event, i.e. transitioning to the 'MI/CVD' health state

In the original analysis, a standard OLS regression was employed to determine mean HRQoL at randomisation (baseline), for each risk group, based on the EQ-5D data collected as part of RITA-3. Baseline was calculated to be 0.58097 for patients in risk group 3. Changes after randomisation were then determined using a panel data approach. In particular, a GLS random-effects model was fitted where a covariate for experiencing a ‘recent’ (within one year to the time of the follow-up interview) MI was generated, along with binary covariates representing whether the HRQoL measure was taken at 4 months or 12 months. The resulting coefficients are shown in Table 36.

Table 36: Estimated gain in health-related quality of life (Henriksson et al., 2008)

Covariate	Coefficient	Standard error	95 % CI
D4*	0.0441	0.0133	
D4t*	0.0384	0.0168	0.0054 to 0.0714
D12*	0.0383	0.0076	0.0234 to 0.0533
D12t*	0.0177	0.0154	-0.0126 to 0.0480
Previous MI	-0.0097	0.0156	-0.0404 to 0.0209
Current MI	-0.0353	0.0220	-0.0784 to 0.0078
Constant	0.0442	0.0126	0.0195 to 0.0689
Between patient standard error (σ_u)		0.295	
Within patient standard error (σ_e)		0.183	
Fraction of variance due to u_i (ρ)		0.722	

*These covariates represent the change in utility (from baseline) at follow-up of 4 months or 12 months as indicated.

Those indicated with a ‘t’ represent the gain in HRQoL in the early interventional strategy over and above that of the conservative strategy

The temporal assumption regarding the impact of a non-fatal MI was that HRQoL for the first year following a non-fatal MI would equal baseline (0.58097) plus the ‘current MI’ decrement (-0.0353) and this figure would not change over time. It was also assumed that this would be the same for both treatment cohorts.

Let us appraise the robustness of the temporal assumption by applying linear increments and decrements.

Table 37: Cost-effectiveness results for alternative assumptions regarding the change over time of HRQoL associated with a composite event for risk group 3

Δ HRQoL	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Constant HRQoL	7,582	0.4851	15,630	0.6890	Adopt
↑ 2%/annum (upperbound 0.9)	7,582	0.4771	15,892	0.6802	Adopt
↓ 2%/annum (lowerbound 0)	7,582	0.4965	15,270	0.7214	Adopt

As with the temporal assumption regarding costs incurred during the first year after a non-fatal MI, the impact of alternative temporal assumptions regarding HRQoL during the first year after a non-fatal MI is relatively minor.

4.5.3.1.2 The HRQoL associated with residing in the ‘No event’ health state and the ‘Post MI’ health state

For the HRQoL associated with residing in the ‘No event’ and ‘Post MI’ health states (every year except the first year after a non-fatal MI), the original model assumed, based on a separate detailed analysis carried out by Kim et al., that for patients in the ‘No event’ health state, the overall change in HRQoL from baseline observed at one year (i.e. ignoring intermediary recordings) is maintained for the remainder of their lives and that this change is the same for both conservative and early interventional cohorts (Kim et al., 2005). Similarly, for the Post MI state, the HRQoL was assumed to remain the same from the second year onwards (and that the same HRQoL weight exists for both treatment strategies). Thus, as with costs, assumptions of constant HRQoL per health states were applied.

Let us test the sensitivity of the cost-effectiveness results to the temporal assumption relating HRQoL per health state by applying alternative temporal assumptions of increasing and decreasing HRQoL over time.

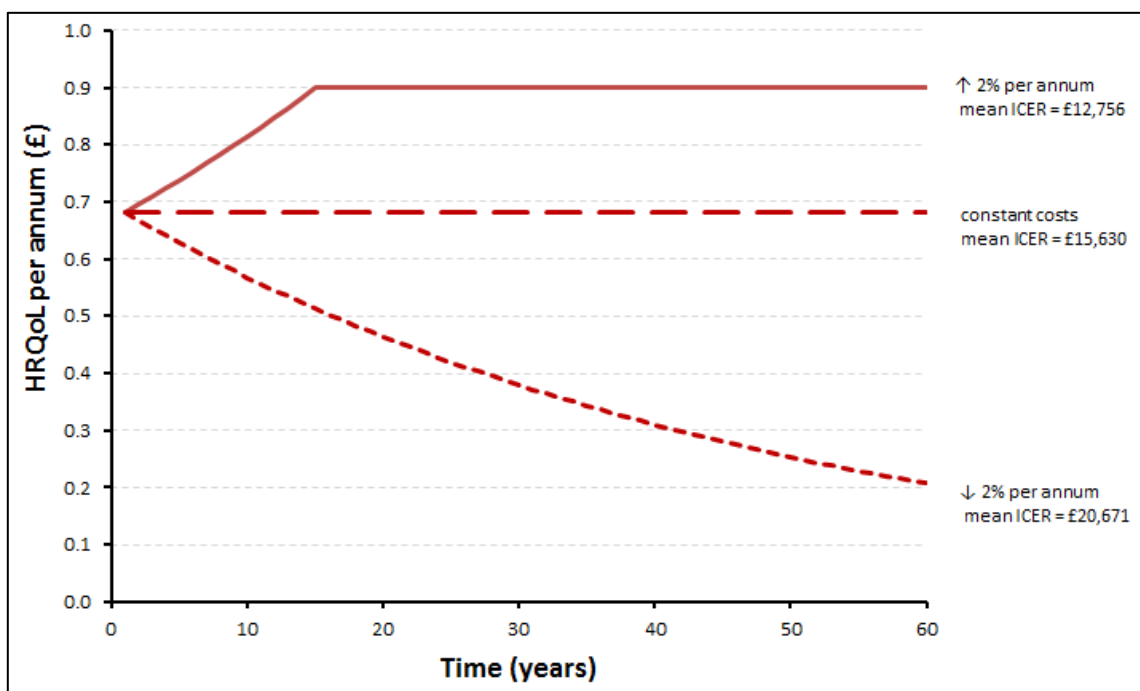
Table 38: Cost-effectiveness results for alternative assumptions regarding the change over time of HRQoL per health state for risk group 3

Δ HRQoL	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Constant HRQoL	7,582	0.4851	15,630	0.6890	Adopt
↑ 2%/annum (upperbound 0.9)	7,582	0.5944	12,756	0.8231	Adopt
↓ 2%/annum (lowerbound 0)	7,582	0.3668	20,671	0.5334	Reject

As one would expect, the early interventional treatment becomes more favourable with rising HRQoL over time, as patients who survive for longer (of whom there are more in the early interventional cohort) accrue more health benefits. Given the sensitivity of the mean ICER (and indeed the adoption decision) to this temporal assumption, we are prompted to consider what factors may cause HRQoL to change over time.

The trajectories over time of HRQoL of the alternative assumptions are illustrated in Figure 52 below.

Figure 52: The trajectory over time of HRQoL per health state for risk group 3 according to 3 alternative assumptions



Let us also address the assumption that there is no difference in HRQoL between the two treatment cohorts in the ‘No event’ health state after the first year. The assumption of parity in HRQoL between the two treatment strategies in the ‘No event’ health state in the second and subsequent years was deemed conservative in the original model report (M Henriksson et al., 2007). This was a similar approach to that taken in the analysis of long-term composite event rates (see Chapter 4.3). The assumption is arguably reasonable (i.e. not conservative), as the early interventional treatment strategy is designed to benefit patients through the prevention of composite cardiac events, and thus the HRQoL associated with patients in the ‘No event’ health state could be expected to be the same for both treatment groups after sufficient time since intervention has elapsed. Nonetheless, some alternative scenarios could reasonably be assumed. In particular:

- (i) Instead of assuming equal long-term HRQoL, a ‘treatment effect’ could be assumed for the second and subsequent years in the ‘No event’ health state. This can be done by including the treatment interaction term at 12 months (12t in Table 36 above) for the calculation of HRQoL for the early interventional strategy.
- (ii) An assumption of differential HRQoL could be assumed, but for a limited amount of time (5 years). The implication of this scenario is that it takes somewhat longer after intervention for patients in the ‘No event’ health state to experience the same HRQoL.

As with other issues of temporal uncertainty, these scenarios broadly represent optimistic, pessimistic and ‘middle’ assumptions. The results of these scenarios are shown in Table 39.

Table 39: Cost-effectiveness results for risk group 3 for alternative assumptions regarding differential long-term HRQoL in the ‘No event’ health state

Δ HRQoL	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Equal HRQoL	7,582	0.4851	15,630	0.6890	Adopt
Differential HRQoL for remaining years	7,582	0.6908	10,976	0.8965	Adopt
Differential HRQoL for 5 years	7,582	0.5703	13,295	0.8413	Adopt

The results in Table 39 demonstrate that assuming differential HRQoL in the ‘No event’ health state up to the model time horizon has a notable effect on the mean ICER, though it does not alter the adoption recommendation for risk group 3. This suggests that the relative HRQoL for patients remaining in the ‘No event’ health state over the long-term is an important element of the model, even though the majority of both cohorts leave the ‘No event’ health state over the short to medium term (there is roughly a third of the conservative cohort in the ‘No event’ health state by year 20).

4.5.3.2 Expressing Temporal Uncertainty

In the previous section, one-way sensitivity analyses were carried out on the key temporal assumptions related to HRQoL in order to appraise their impact on the mean ICER and ultimately the adoption recommendation. Two factors that may impact the temporal nature of HRQoL are now investigated. Where appropriate, temporal uncertainty is quantified and incorporated into the updated model.

4.5.3.2.1 Differential long-term HRQoL in the ‘No event’ health state

Rather than applying an assumption of no difference (with the knowledge that it is perhaps conservative), it is desirable to express temporal uncertainty by averaging across the scenarios outlined in Table 39. Assuming that each of the scenarios described in Table 39 is equally likely (though these scenarios, in practice, ought to be informed and validated by expert opinion), temporal uncertainty by model averaging for this assumption can be expressed as:

$$Assumption = \begin{cases} equal\ HRQoL\ with\ probability\ \frac{1}{3} \\ differential\ HRQoL\ for\ remaining\ years\ with\ probability\ \frac{1}{3} \\ differential\ HRQoL\ for\ 5\ years\ with\ probability\ \frac{1}{3} \end{cases}$$

The results of this model averaged assumption are shown in Table 40.

Table 40: Cost-effectiveness results for model-averaged characterisation of long-term relative HRQoL in 'No event' health state

Δ HRQoL	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Temporal uncertainty - Model averaged ($1/3, 1/3, 1/3$)	7,582	0.5705	13,290	0.8035	Adopt

4.5.3.2.2 Accounting for age and event-free survival when estimating long-term HRQoL

In estimating long-term HRQoL, a similar problem to the one that arose in Chapter 4.2 is found, where the impact of age at baseline (i.e. at the point of randomisation following NSTEMI-ACS) may not equate to the effect of age over the long-term. Indeed in RITA-3, a counterintuitive result is produced by the regression analysis where age has a positive effect on HRQoL rather than negative. This is shown in Table 41 where 'agegroup' is included as a covariate in the regression to calculate HRQoL.

Table 41: Co-variables in calculation of HRQoL including age

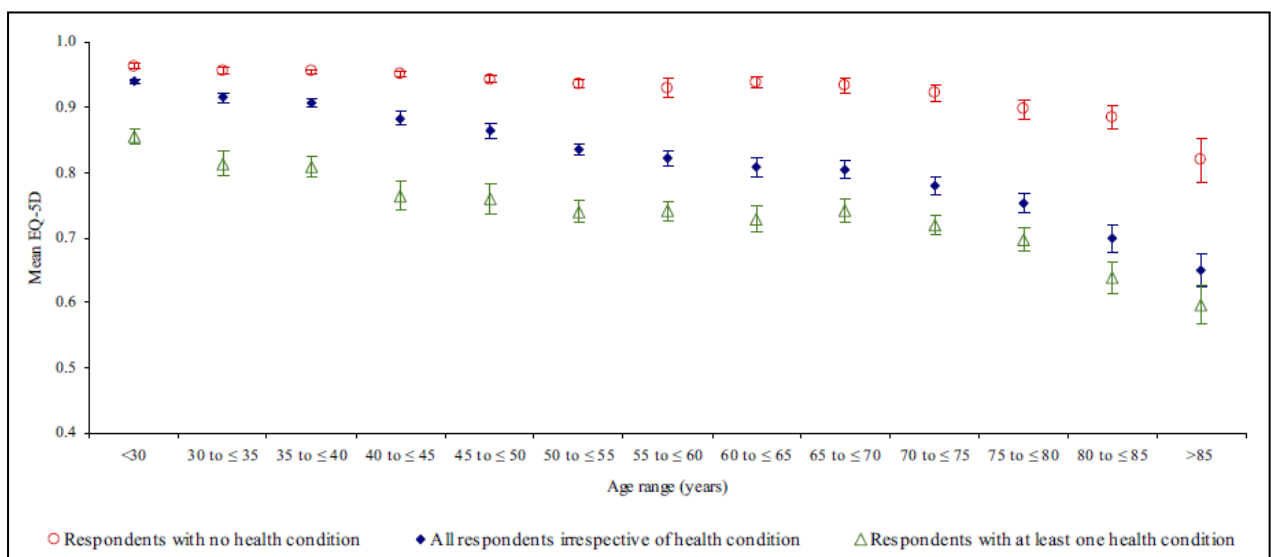
eq5d0	Coef.	Std. Err.	[95% Conf. Interval]	
agegrp0	0.021178	.0084933	0.00452	0.037836
diabmell	-0.05241	.0208927	-0.09339	-0.01144
prevmi	-0.04932	.0163219	-0.08133	-0.01731
stdepres	-0.07118	.0149127	-0.10042	-0.04193
agrade	-0.07457	.0150338	-0.10406	-0.04509
sex	0.074819	.0148092	0.045774	0.103864
_cons	0.675985	.0160855	0.644437	0.707533

This may relate to the unreliability of self-reported HRQoL so close to a traumatic medical event, i.e. those in lower age groups in times of stress may report lower HRQoL than those in higher age groups experiencing the same stress.

Even if the trial data are deemed to be uninformative or unreliable, it is prudent to consider the long-term effect of ageing on HRQoL per health state, rather than assuming constant HRQoL over time. A number of studies have examined the natural evolution of HRQoL that comes with the ageing process (Busschbach et al., 1993, Pliskin, 1980), but consideration of the impact of ageing on disease specific health states is rare. One enlightening study by Ara and Brazier examined pooled health survey data in order to assess its appropriateness for use as baseline HRQoL data (Ara and Brazier, 2011). Among their findings was that HRQoL declines (with statistical significance) with age and that data from the general population can be used, in many but not all, circumstances to approximate baseline utility.

Another useful output of this study was an estimate of the trajectory with age of HRQoL (in particular, EQ-5D). This was estimated for cohorts with no health condition, with at least one health condition and for the general population irrespective of health status. These estimations are depicted in Figure 53.

Figure 53: Trajectory of HRQoL with age from Ara and Brazier



Looking at these trajectories (and concentrating in particular on those with at least one health condition), it is clear that there exists a decrement in HRQoL with age and it follows that some estimate of that decrement ought to be incorporated into any analysis with a long-term time horizon

involving HRQoL. It is also notable however, that from age group 50 to ≤ 55 to age group 70 to ≤ 75, HRQoL is relatively stable (note that the ages of patients representing risk groups 1 to 5 are 45, 52, 52, 62 and 65 respectively). It may be therefore that the decline in HRQoL that is apparent in old age does not meaningfully impact the cost-effectiveness results in RITA-3. Let us test this hypothesis by comparing an assumption of constant HRQoL over time with an assumption of a decline in HRQoL at the rate suggested by Ara and Brazier, only from age 75 onwards, for the proportion of the cohort residing in the ‘No event’ health state. It is assumed that this cohort can be represented by ‘respondents with at least one health condition’ in the Ara and Brazier paper. This description would seem to fit the RITA-3 cohort. For example, HRQoL at age 52 for these respondents (0.5975) is similar to that estimated in RITA-3 for risk group 3 (0.5712).

Table 42: Cost-effectiveness results for an assumption of constant HRQoL vs. the inclusion of an end-of-life decrement

Δ HRQoL	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Constant HRQoL*	7,582	0.5705	13,290	0.8035	Adopt
End of life decrement*	7,582	0.5682	13,343	0.7833	Adopt

*TU has been expressed for HRQoL differential

The difference is moderate (as expected since more extreme alternative scenarios were applied in Table 38, with little impact). Nonetheless, it would seem more appropriate to apply decrement associated with age than to not.

More challenging is estimating the change over time in HRQoL for patients in the ‘Post MI’ health state specifically. These patients have a significantly lower HRQoL in light of having experienced a myocardial infarction. The model assumes that this lower HRQoL is sustained irrespective of subsequent event-free survival. However, it may be more reasonable to assume that for patients who have survived a prolonged period of time without experiencing a further MI, a recovery of in terms of HRQoL has occurred. In other words, there is an upward pressure on this initial estimate of HRQoL associated with event-free survival, as well as a longer-term downward pressure associated with ageing. It is desirable therefore, to express uncertainty regarding the possible recovery in HRQoL after a period of event-free survival as well as uncertainty regarding the impact of ageing for patients in the ‘Post MI’ health state.

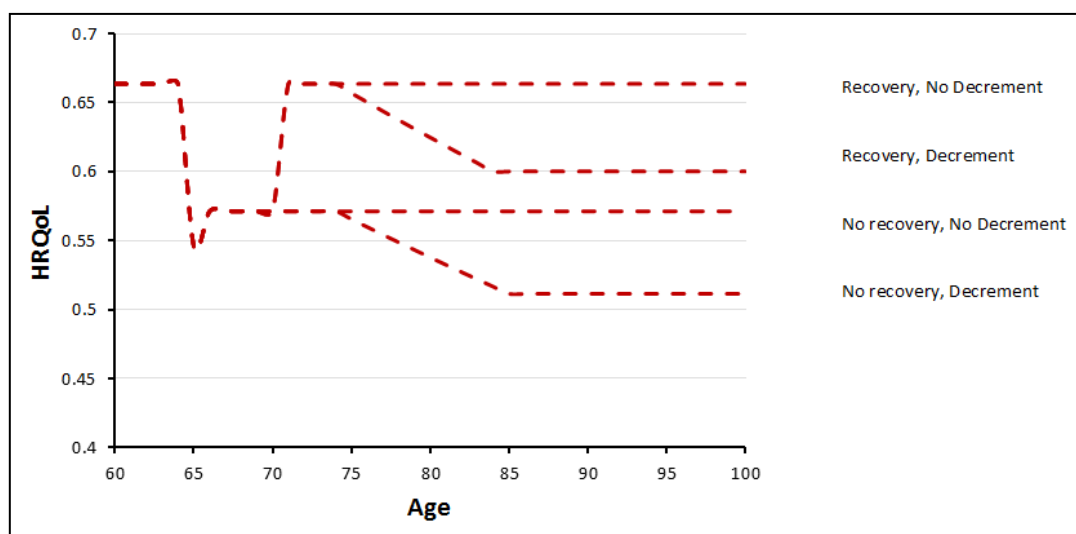
It is important to note at this point, that the 'Post MI' health state is in fact made up of 4 tunnel states representing the first 4 years after a non-fatal MI and a fifth state which represents patients 5 years or more after a non-fatal MI. Let us first appraise the impact of alternative assumptions regarding change in HRQoL over time in the 'Post MI' health state. It is assumed that a recovery in HRQoL can be represented by a return to the 'No event' HRQoL (for the conservative treatment) once a patient reaches the '5 years since MI' health state and it is assumed that an end-of-life decline in HRQoL can be represented by the percentage decrement in HRQoL shown in Table 43 for respondents with at least one health condition. Note that the decrement is applied to both the 'No event' health state and the 'Post MI' health state.

Table 43: Mean ICERs for risk group 3 for alternative assumptions regarding change in HRQoL in the 'Post MI' health state

	Recovery in HRQoL at 5 years	No recovery in HRQoL at 5 years
No end of life decrement in HRQoL	13,677	13,290
End-of-life decrement in HRQoL	13,759	13,343

Again, the impact of the end-of-life decrement is minimal as is the impact of the recovery assumption. Nonetheless, it is appropriate to incorporate both sources of temporal uncertainty. An example of both sources of temporal uncertainty being expressed in combination is illustrated in Figure 54. Here a patient on the conservative treatment experiences an MI at age 60.

Figure 54: Expression of temporal uncertainty in HRQoL for patient (in risk group 3) who experiences an MI aged 60



Again assuming that each point of uncertainty is represented by equally weighted alternative scenarios, temporal uncertainty can be expressed, through model averaging, for HRQoL in place of a conservative temporal assumption. The results (as compared to assumptions of no recovery and no end-of-life decrement) are shown in Table 44.

Table 44: Cost-effectiveness results for an assumption of constant HRQoL vs. the inclusion of an end-of-life decrement and post MI recovery

Δ HRQoL	Incremental Cost (£)	Incremental QALY	Mean ICER	Prob(EI cost-effective) at £20,000/QALY	Adopt/reject early interventional (EI)
Constant HRQoL (no recovery, no EOL decrement)*	7,582	0.5705	13,290	0.8035	Adopt
Temporal Uncertainty for Post MI recovery and EOL decrement*	7,582	0.5622	13,486	0.7935	Adopt

*incorporating the temporal uncertainty quantified thus far in chapter 4.5

4.5.4 Results

4.5.4.1 Cost-effectiveness

Table 45: Cost-effectiveness Results after temporal uncertainty has been addressed for HRQoL. Compare with Table 34 to observe the change in the results.

Risk group	Incremental Cost (£)	Incremental QALY	Mean ICER	Adopt/reject early interventional (EI)
Risk group 1	5,435	0.3286	16,539	Adopt
Risk group 2	5,370	0.4613	11,641	Adopt
Risk group 3	7,582	0.5622	13,486	Adopt
Risk group 4a	7,438	0.7564	9,833	Adopt
Risk group 4b	7,251	0.6812	10,644	Adopt

All ICERs have decreased somewhat after these updates to the CEDM. Most notably, a positive adoption recommendation is now being given for risk group 1. This is primarily explained by the ICER of risk group 1 being especially sensitive to the assumption regarding differential long-term HRQoL in the 'No event' health state, as a larger proportion of risk group 1 patients reside in this health state over the long-term. Under an assumption of equal HRQoL in the 'No event' health state, a mean ICER of £26,529 is returned. As with the previous sub-chapters, here a characterisation of temporal uncertainty replaces a 'conservative' assumption with regard to the cost-effectiveness of the early interventional treatment, leading to a more optimistic mean ICER.

4.5.4.2 Uncertainty and Value of Information

Table 46 summarises the current impact of uncertainty on the outputs of the CEDM for each of the 5 risk groups.

Table 46: Summary of Effect of Uncertainty for each Risk Group after temporal uncertainty has been addressed for HRQoL. Compare with Table 35 to observe the change in the results.

Risk group	Prob(EI cost-effective) at £20,000/QALY	EVPI/patient (£)	EVPI/population (£)
Risk group 1	0.505	1,119.22	254,301,806
Risk group 2	0.764	400.79	90,903,237
Risk group 3	0.793	357.58	81,131,139
Risk group 4a	0.970	40.70	4,545,162
Risk group 4b	0.969	31.56	3,522,500

Given the decrease in ICERs for all risk groups, it is not surprising to see that the probability of cost-effectiveness for all risk groups has increased. Risk groups 4a and 4b are now almost certain to be cost-effective and risk group 1 is now more likely than not to be cost effective. Again this is explained by a 'conservative' assumption regarding the change in HRQoL over time being replaced by a 'neutral' expression of temporal uncertainty. EVPI for group 1 has naturally increased as the distribution of ICERs now centres closer to the cost-effectiveness threshold. EVPI for risk groups 2, 3 and 4b has decreased as their ICERs have moved further from the threshold making the adoption decision less 'uncertain'. However, interestingly, EVPI for risk group 4a has increased despite the associated distribution of ICERs being centred further from the threshold. This is due to the greater dispersion of ICERs (due to further temporal uncertainty being expressed) outweighing the shift in where the ICERs are centred in terms of the impact on EVPI.

4.6 Chapter 4.6: Overall Results and Discussion

4.6.1 Introduction

In this last sub-chapter, the original CEDM is compared to the 'updated' version. The overall impact of accounting for temporal uncertainty is observed as well as the impact of individual temporal parameters and their interactions. Key outcomes and aspects of the methodology are then discussed.

4.6.2 Original CEDM vs. CEDM Updated for Temporal Uncertainty

4.6.2.1 Summary of Model Updates

Table 47: Summary of model updates

Model Component	Temporal parameter	Chapter where analysed	Original Assumption	Method of expressing temporal uncertainty	Evidence used to inform temporal trajectory
Transition Probabilities	Baseline risk regarding 1 st composite event after 5 years	4.2	Constant baseline risk with 10 yearly update of age parameter	Parameterisation using continuous (beta) distribution	RITA-3 IPD, related HTAs, population mortality tables
	Treatment effect regarding 1 st composite event after 5 years	4.3	No further treatment effect	Model averaging over alternative assumptions	RITA-3 IPD, alternative scenarios in NICE guidance
HRQoL	Long-term treatment effect re HRQoL in 'No Event' state	4.5	No treatment effect	Model averaging over alternative assumptions	RITA-3 IPD, alternative assumptions
	Recovery in HRQoL for patients in 'Post MI' state	4.5	No recovery in HRQoL	Model averaging over alternative assumptions	RITA-3 IPD, alternative assumptions
	Age-related (end of life) decrement	4.5	No age-related end of life decrement	Model averaging over alternative assumptions	RITA-3 IPD, Aggregate data on HRQoL decrement with age
Costs	Age-related increase in costs	4.4	No age-related increase in costs	Parameterisation using continuous (gamma) distribution	RITA-3 IPD, RITA-2 resource use information, PSSRU, BNF, analysis by Epstein et al.

4.6.2.2 Cost-effectiveness Results

Table 48: Cost-effectiveness results for original CEDM vs. updated CEDM

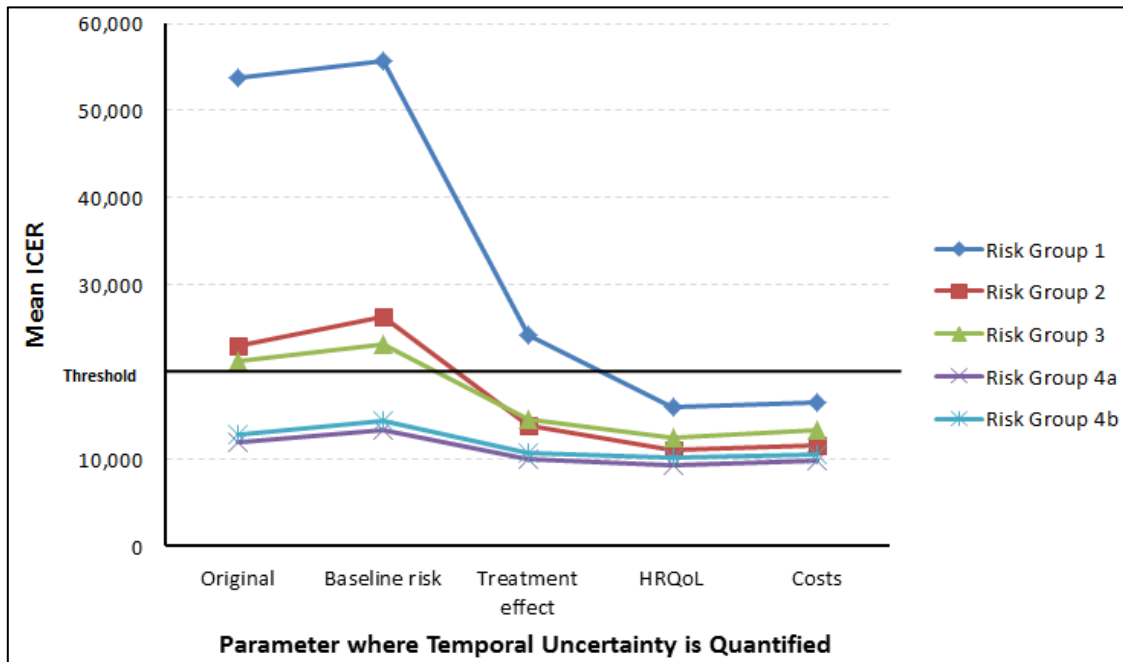
Risk Group	Mean ICER (adoption decision)	
	Original CEDM	Updated CEDM
1	53,760 (reject)	16,539 (adopt)
2	22,949 (reject)	11,641 (adopt)
3	21,186 (reject)	13,486 (adopt)
4a	11,957 (adopt)	9,833 (adopt)
4b	12,750 (adopt)	10,644 (adopt)

Note: a shaded cell indicates a positive adoption recommendation

The results in the Table 48 show that the re-analysis where temporal uncertainty is more closely examined and quantified leads to significantly altered mean ICERs and adoption recommendations. The principle reason for this is that the re-analysis did not simply consist of fitting distributions around point estimates to express uncertainty; rather it incorporated and weighted alternative temporal scenarios which expressed expectations and uncertainty given the currently available evidence, replacing the single, often ‘conservative’ assumptions of the original CEDM. The main cause of the shift is the change in the modelling of treatment effect over the long-term.

Figure 55 below illustrates how the mean ICERs for each risk group evolved with each model update, i.e. each new characterisation of temporal uncertainty (the updates relating to HRQoL have been grouped into one). It is clear that characterisation of temporal uncertainty regarding treatment effect has the greatest impact, especially on the mean ICER of risk group 1.

Figure 55: Evolution of Mean ICER for each characterisation of temporal uncertainty for each risk group



4.6.2.3 Uncertainty

The cost-effectiveness planes for original and updated CEDMs for each risk group are illustrated in Figure 56.

Figure 56: Cost-effectiveness planes for original and updated CEDMs

Risk Group	CE Plane for Original CEDM	CE Plane for Updated CEDM
1		
2		
3		
4a		
4b		

The cost-effectiveness planes in Figure 56 representing the original and updated CEDMs show the vast increase in uncertainty in costs and health effects associated with the updated CEDM. The result is unsurprising as more uncertainty regarding the inputs into the CEDM is being expressed. The change in decision uncertainty however owes more to how the ICERs for each risk group have shifted in relation to the ICER threshold (assumed to be £20,000/QALY). Table 49 below gives the probabilities of cost-effectiveness for each risk group for the original and updated CEDMs.

Table 49: Probability that the early interventional strategy is cost-effective for the original and updated CEDM for each risk group

Risk group	Prob(EI cost-effective) at £20,000/QALY	
	Original CEDM	Updated CEDM
Risk group 1	0.009	0.505
Risk group 2	0.328	0.764
Risk group 3	0.420	0.793
Risk group 4a	0.945	0.970
Risk group 4b	0.924	0.969

For risk groups 4a and 4b, there is far less decision uncertainty as it is now highly likely that the early interventional strategy is cost-effective for these risk groups. For risk groups 1, 2 and 3, the early interventional strategy is now also likely to be cost-effective, but for these risk groups there is a significant error probability, in particular for risk group 1. It is therefore expected that there may be value in obtaining further information for these lower risk groups.

4.6.2.4 Value of Information

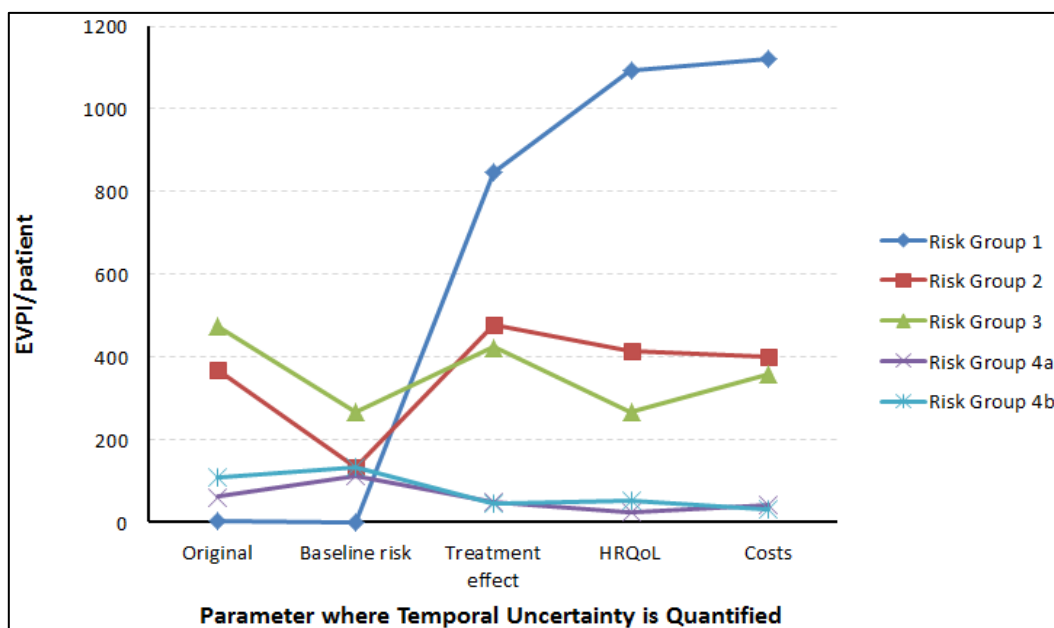
The principal motivations for expressing and incorporating temporal uncertainty into the CEDM were to calculate the true expected cost-effectiveness given currently available evidence and to assess the need for further evidence given the current decision uncertainty. Because each source of temporal uncertainty has been expressed in the CEDM through an uncertain temporal parameter, it is also possible to estimate the value of obtaining further information specifically in relation to the temporal trajectory of key model parameters. Table 50 below outlines the expected value of perfect information (EVPI) per patient and per population for both the original and updated CEDMs for each risk group.

Table 50: EVPI per patient and per population for each risk group

Risk Group	EVPI/patient (£)		EVPI/population (£)	
	Original CEDM	Updated CEDM	Original CEDM	Updated CEDM
1	1.92	1,119.22	349,068	254,301,806
2	367.70	400.79	66,850,066	90,903,237
3	475.15	357.58	86,385,121	81,131,139
4a	61.13	40.70	5,556,900	4,545,162
4b	107.96	31.56	9,813,888	3,522,500

The change in EVPI as temporal uncertainty is characterised for more parameters is illustrated in Figure 57 below.

Figure 57: Change in EVPI as each source of temporal uncertainty was addressed



Again the biggest impact is that of the characterisation of treatment effect and the risk group most affected by this is risk group 1. What is of particular interest however, is the value of obtaining further evidence pertaining to the specific uncertain temporal parameters. The calculation of EVPI/patient for each temporal parameter ought to (i) express the relative significance of each source of temporal uncertainty and (ii) demonstrate the usefulness of parameterising temporal

uncertainty as the resultant temporal parameter can be examined and it can be determined whether obtaining further information specific to the temporal trajectory of the underlying model parameter would be worthwhile. Because of the replacement of conservative assumptions with distributions that incorporate all plausible scenarios and that reflect our uncertainty, the mean ICER has moved far enough from the cost-effectiveness threshold to render EVPPI negligible or zero for many temporal parameters for many risk groups.

Table 51 below shows the expected value of perfect information for a specific temporal parameter (EVPPI) per patient for each characterised source of temporal uncertainty and for each risk group.

Table 51: EVPPI/patient (£) for all uncertain temporal parameters and for each individual uncertain temporal parameter for each risk group

Risk Group	All temporal parameters	Baseline risk regarding 1 st composite event after 5 years	Treatment effect regarding 1 st composite event after 5 years	Long-term treatment effect re HRQoL in 'No Event' state	Recovery in HRQoL for patients in 'Post MI' state	Age-related (end of life) decrement	Age-related increase in costs
1	977.42	~0	127.97	515.71	~0	~0	~0
2	180.04	~0	~0	~0	~0	~0	~0
3	133.05	~0	~0	~0	~0	~0	~0
4a	~0	~0	~0	~0	~0	~0	~0
4b	~0	~0	~0	~0	~0	~0	~0

The results in Table 51 suggest that there is meaningful value in obtaining further information with regards to the temporal behaviour of parameters only for risk groups 1, 2 and 3. For risk group 1 in particular, further information on the nature of treatment effect pertaining to a first composite event and pertaining to HRQoL in the 'No event' health state would be of value.

The analysis in Chapter 4.3 suggested that the uncertainty around treatment effect regarding a first composite event would impact decision making for risk groups 2 and 3, thus we would expect there to be value in obtaining further information regarding this temporal parameter. However, the results in Table 51 show that, when all sources of temporal uncertainty are accounted for, there is in fact negligible value in obtaining further information solely on this temporal parameter as all assumptions regarding the temporal behaviour of treatment effect return a positive adoption decision for risk groups 2 and 3. In other words, it was only when a number of conservative

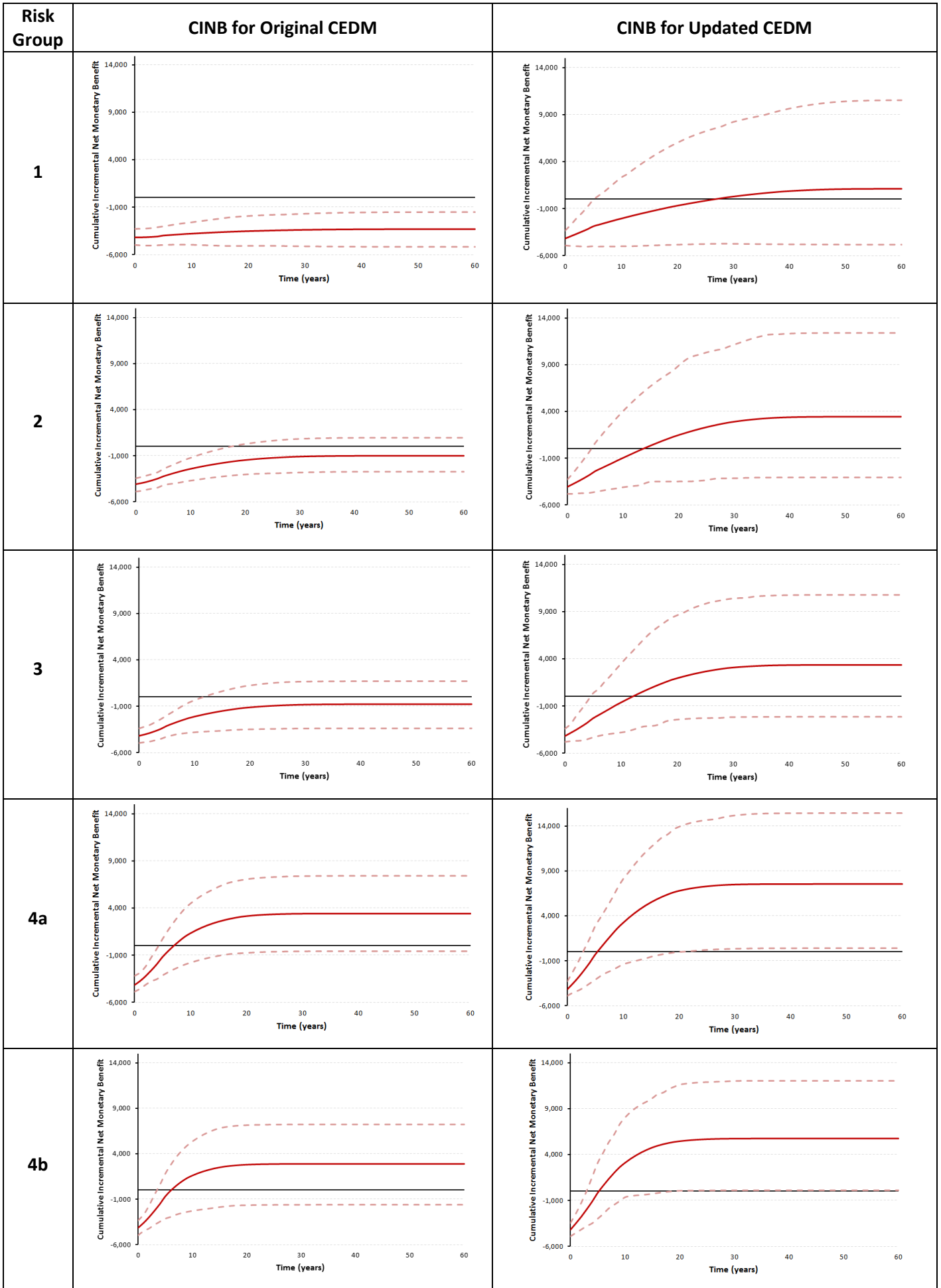
assumptions were together applied (as in the original CEDM) that a negative adoption decision could be returned for risk groups 2 and 3. Thus a 'different' adoption decision is returned in a proportion of scenarios where 'all temporal parameters' are varied in risk groups 2 and 3 leading to the positive EVPI. In terms of obtaining further information on temporal parameters, the results in Table 51 show that in many cases, further information on individual temporal parameters will not be sufficient, it will likely be necessary to obtain information on a number of temporal parameters, as well as other model parameters.

4.6.2.5 Cost-effectiveness Over Time

In the context of understanding the effects of uncertainty over time in CEDMs, it is useful to observe the impact of the model updates on the cumulative incremental net benefit (CINB) curves for each risk group.

The CINB curves in Figure 58 illustrate both how expected cost-effectiveness over time has shifted for each risk group and how increasing uncertainty over time has been expressed with the model updates. The solid curves represent mean cumulative incremental net monetary benefit and the dashed lines represent 5th and 95th percentiles.

Figure 58: CINB for original and updated CEDMs



4.6.3 The value of waiting

Given current evidence and assumptions, the early interventional treatment strategy should be adopted for all risk groups. For risk group 1 however, there is considerable decision uncertainty. Let us assume that there would be an irrecoverable investment cost of £2,000,000 associated with implementing the early interventional treatment strategy. Let us further assume that after 3 years of further trial follow-up, the extent of long-term treatment effect regarding HRQoL in the ‘no event’ health state can be known with certainty.

Using the formulae derived in Section 2.4, we can calculate whether it would be worthwhile implementing the early interventional strategy immediately given the risk that the decision may have to be reversed when further evidence reveals that the early interventional treatment strategy is in fact not cost-effective and the investment costs will be lost. Say that $E(NB(IA))$ = the expected net benefit (up to the point of further evidence revelation) associated with immediate approval and $E(NB(W))$ = the expected net benefit (up to the point of further evidence revelation) associated with waiting until further evidence is revealed to make an adoption decision.

$$E(NB(IA)) = \sum_{t=0}^2 \{ \max_j NB(j, \theta) \cdot (1.035)^{-t} \} - 2.10^6$$

$$E(NB(W)) = \sum_{t=0}^2 \{ NB(\text{conservative strategy}, \theta) \cdot (1.035)^{-t} \} - \lambda \cdot 2.10^6 \cdot (1.035)^{-2}$$

$$\text{Value of waiting} = E(NB(W)) - E(NB(IA)) = \text{£}3,663,673 - \text{£}3,016,791 = \text{£}646,882$$

Because the value of waiting is positive, it is better to wait for the further evidence rather than adopt the early interventional strategy immediately and risk the possibility of incurring irrecoverable costs.

It can also be calculated that, under the given assumptions, the amount that the investments costs would have to be under in order for immediate adoption to be the better option is £287,141

In practice, the investment costs should be relatively small for risk group 1, as given the results above, it would be expected that the early interventional treatment strategy is immediately

approved for all other risk groups and hence the bulk of the investment costs would already be incurred.

4.6.4 Findings and Discussion

This sub-chapter concludes with an overview of the purpose and findings of the preceding sub-chapters, followed by a discussion of some key issues that have emerged over the course of the empirical analysis in Chapter 4. More general issues relating to temporal uncertainty which may also have stemmed from this empirical chapter are discussed in the Chapter 5.

This principle purpose of Chapter 4.1 was to set the scene for the empirical portion of this thesis (Chapter 4) by introducing and undertaking some initial analysis on the RITA-3 model. It was found that:

- (i) Temporal uncertainty had the potential to significantly influence recommendations regarding adoption and obtaining further evidence.
- (ii) There was a range of issues of temporal uncertainty pertaining to the RITA-3 CEDM, principally relating to the temporal trajectories of model input parameters.
- (iii) While there was a number of factors to consider when using TTE data to compute transition probabilities for multiple comparators, it is advantageous (in this instance, if not generally) to consider the temporal trajectories of baseline risk and treatment effect(s) separately.

The focus of Chapter 4.2 was long-term baseline risk associated with a first composite event. It was found that:

- (i) Survival analysis can be used to characterise and potentially extrapolate short-term survival data, but its usefulness can be limited, e.g. the long-term effect of age may not be captured, extrapolating a short-term trend may be a strong assumption.
- (ii) A key issue in the characterisation of long-term outcomes is simultaneous modelling of effects of age, event-free survival and period risk.
- (iii) It is possible and often appropriate to express temporal uncertainty by fitting a suitable continuous distribution around the term that determines parameter's relationship with time.

The focus of chapter 4.3 was long-term treatment effect regarding a first composite event. It was found that:

- (i) The assumption regarding long-term treatment is central to the RITA-3 CEDM and most likely many CEDMs. The magnitude of the impact can be influenced by other uncertainties. The requirement for more evidence regarding long-term treatment can be ascertained by incorporating temporal uncertainty into the EVPI framework and calculating EVPPi for the temporal parameter in question.
- (ii) There is still much use in conducting a simple scenario analysis. It is advantageous to demonstrate, at an early stage of the modelling process, that the evidence available regarding long-term treatment effect is not sufficient to make a confident adoption decision or even to characterise the uncertainty surrounding that decision. In this sense, a simple scenario analysis may adequately convey this information to the decision maker and set in motion the process of obtaining further information.
- (iii) Though a number of factors must be taken into account and there exist a series of challenges associated with eliciting evidence on how a parameter will evolve over time, there are a number of valid approaches. The most appropriate will depend on the particulars of the analysis but the approach of quantifying the duration of treatment effect is likely to represent a good balance of comprehensibility and accuracy.

The focus of Chapters 4.4 and 4.5 was the potential change over time of costs and health-related quality of life. It was found that:

- (i) As with other types of temporal uncertainty, there were many instances of temporal uncertainty regarding costs and HRQoL that did not have a significant bearing on decision-making (in particular on the mean ICER).
- (ii) Similar to transition probabilities, the uncertainty regarding whether a treatment effect existed over the long-term had a notable impact on the mean ICER.
- (iii) Though costs and HRQoL per health state are often assumed to stay constant over the long-term, there are a number of factors that would likely cause these state-specific values to increase or decrease over time (ageing, price drop, etc.). These factors ought to be considered routinely in HTA, as they can (as demonstrated in this sub-chapter) impact decision-making for some risk groups.

The re-analysis in Chapter 4 has negated the need for a number of 'conservative' assumptions. As a result, the mean ICER is less pessimistic and can be said to more faithfully represent expectations and uncertainties regarding cost-effectiveness given the evidence currently available. Expressing decision uncertainty is a key element of this approach. The expected cost-effectiveness may be a 'neutral' estimation but if it is partially based on weak or scant evidence then it is paramount that decision uncertainty is expressed so as to allow for decisions beyond those of adopt or don't adopt. It is important to note however that decisions that involve delaying the adoption recommendation are only relevant if there are irrecoverable costs associated with implementing a new strategy or if it would be difficult or contentious to reverse a decision in the future.

This re-analysis has made the adoption decision less uncertain for most risk groups. Where there would be most value in obtaining further evidence however is regarding treatment effect for a composite event and treatment effect for long-term HRQoL. As was discussed in Chapter 4.3, good quality data regarding treatment effect may be difficult to obtain. However, eliciting the opinion of experts would be a very useful early avenue to reduce decision uncertainty.

The interactions and relationships between key parameters can be central as uncertainty over time is quantified. Quantifying one source of temporal uncertainty may not seem significant in terms of the effect on ICER but in conjunction with the quantification of another related source of temporal uncertainty it may become significant. This was seen in the relationship between baseline risk and treatment effect (for a first composite event), as demonstrated in the two-way sensitivity analysis in Chapter 4.3. As a consequence of these relationships, the calculation of EVPI and EVPPI only becomes meaningful once all key sources of uncertainty are quantified.

Relatedly, it was notable that even though the impact of uncertainty regarding long-term baseline risk and long-term differential HRQoL had a sizeable impact on the mean ICER, there was zero EVPPI associated with the temporal parameters at a threshold of £20,000 for all risk groups except risk group 1. This is because each of the scenarios posited returned a positive adoption recommendation for the early interventional strategy. The same was true for recovery in HRQoL in the 'Post MI' health state and end-of-life HRQoL decrement, though these results are less surprising.

Note that simple equal weighting of scenarios was employed a number of times when implementing model averaging. This is an estimate of uncertainty and should be used primarily to ascertain whether this uncertainty meaningfully impacts the adoption decision and whether there is value in obtaining further information regarding temporal trajectory of parameters. For adoption decisions,

where more information is necessary the scenarios should be weighted by further evidence or input from relevant experts.

For the risk of experiencing a first composite event, it is important to note that three sources of uncertainty (all related to how this risk changes over time) are simultaneously quantified: the parameter uncertainty, the uncertainty regarding which survival distribution best represents the short-term evidence, and the temporal uncertainty regarding how this risk might further evolve as it moves beyond the observed period.

The CINB curves in figure 58 demonstrate that outcomes over the very long-term, in fact, have little influence over the decision-making. In particular, it seems that uncertainty regarding only outcomes up to approximately the 30 year time point held sway over the adoption decision. What the CINB curves also demonstrate is the relationship between uncertainty over outcomes (i.e. costs and health benefits) and uncertainty over the adoption decision. It can be seen with regard to risk groups 2 and 3, for example, that uncertainty around outcomes is greatly increased due to the expression of temporal uncertainty, but because these changes have also caused the mean ICER to shift further from the ICER threshold, the amount of decision uncertainty associated with these risk groups has greatly reduced.

5. THESIS CHAPTER 5: CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

This final chapter comprises: a number of conclusions and discussion points given the research, analysis and results in the previous chapters, an outline of the limitations of this thesis and a number of recommendations for Health Technology Assessment.

5.1 Conclusions

5.1.1 The Significance of Temporal Uncertainty

The issue that arises due to a disparity between the evidence time horizon and the appropriate analysis time horizon in cost-effectiveness analysis (CEA) is most accurately and most usefully thought of as an issue of uncertainty. This ‘temporal uncertainty’ pertains predominantly to estimating expectations and uncertainties regarding the long-term temporal trajectories of input parameters in cost-effectiveness decision models (CEDMs). Whatever the scale of temporal uncertainty, it is of importance in CEA only in so far as it influences the related decisions that CEA is employed to inform: whether or not to adopt a health technology and whether or not to obtain further evidence. The relationship between temporal uncertainty and decision making can be helpfully illustrated by calculating and illustrating cumulative incremental net benefit over time (CINB). While temporal uncertainty can be a significant contributor to the uncertainty around whether the adoption of a new health technology is cost-effective for the present incident population, it also relates to the value of the (at least partial) resolution over time of uncertainty in order to make coverage decisions regarding both present and future incident populations.

5.1.2 State-of-play in Health Technology Assessment

Temporal uncertainty is a prevalent issue in Health Technology Assessment (HTA). However, the steps taken to ascertain the significance of temporal uncertainty for decision making and to address temporal uncertainty within decision models have been inconsistent and inadequate to date. Temporal uncertainty both in the HTA literature and in the published methods literature appears to be most strongly associated with survival parameters and the use of extrapolation techniques to utilise the short-term evidence. Temporal uncertainty is however pertinent to a range of model

parameters including survival parameters, longitudinal parameters, cost parameters and health-related quality of life parameters. In general, relatively simple approaches have been taken to plug the gap between the evidence time horizon and analysis time horizon when estimating model input parameters, e.g. no change over time. Such simple approaches may or may not be reasonable as base-case assumptions. There is however, a striking lack of analysis pertaining to the suitability of these assumptions and the uncertainty surrounding them.

5.1.3 Short-term Evidence and Extrapolation

CEAs are often comprised of short-term RCT evidence and a decision model where the available RCT evidence is utilised, to at least some degree, to inform the model assumptions. For some long-term parameter estimations, the trend observed in the short-term RCT evidence is explicitly 'extrapolated' over time. The extrapolation of short-term evidence over time to inform what happens over the long-term (in particular long-term parameter values) is, though a useful tool, a somewhat dangerous one. Although a modicum of uncertainty over time can be expressed through uncertain parametric distributions or the fitting of alternative distributions, true temporal uncertainty is not captured, i.e. the uncertainties expressed relate to the interpretation of the short-term evidence and not the 'unknowns' of the unobserved period. Any method of extrapolation contains an overarching assumption that the values or outcomes of the short-term can in some way inform those of the long-term. Extrapolation is likely to be most appropriate when data are mature or the unobserved period is short. In other circumstances, the extrapolated short-term evidence may be taken into account, but so too must a number of uncertainties relating to the long-term, such as the effect of age, period risk and the effect of long-term event-free survival.

Whether or not extrapolating evidence into the long-term is appropriate, it is often difficult to know how to sensibly interpret the available short-term evidence. If there is not a clear temporal trend that could logically be extended beyond the evidence time horizon, the usefulness of short-term evidence for the purposes of extrapolation is limited. For example, it was found regarding treatment effect in the RITA-3 CEDM that although proportional hazards could be assumed to hold over the observed period, there was little indication of if, or how long, a treatment effect could be assumed to continue.

5.1.4 Expressing Temporal Uncertainty

The approach of expressing temporal uncertainty through scenario analyses is not sufficient, primarily because it is implicit that decision makers must ascribe the necessary relative weights in order to estimate the expected cost-effectiveness of a new technology. Where an expression of temporal uncertainty for model input parameters is warranted, it is desirable to characterise the temporal uncertainty as an uncertain 'temporal parameter'. This temporal parameter can then be incorporated into a probabilistic analysis in order to express an unbiased estimate of expected cost-effectiveness, to calculate the value of obtaining further information specifically on the temporal nature of one or more model input parameters and to calculate the value of waiting for sources of temporal uncertainty to resolve over time before making an adoption decision given the existence of irrecoverable costs. Depending on the nature of the temporal uncertainty, this can be achieved through model averaging or continuous parameterisation.

The challenge of appropriately expressing temporal uncertainty varies by context and by parameter-type. For baseline disease progression, there is likely to be some form of external evidence available with which to inform, validate or bound long-term estimates. It may be reasonable to base the expected temporal trajectory of a parameter on what was observed over the short-term, however supplementary evidence along with other valid scenarios ought to form part of the expression of uncertainty over time. Where short-term estimates (which are themselves uncertain) are extrapolated to inform long-term estimates, there are in fact three sources of uncertainty that ought to be accounted for: the parameter uncertainty associated with the statistical parameters of the distributional fit to the short-term data, the uncertainty regarding which distributional fit best represents the short-term data and the temporal uncertainty associated with moving beyond the observed period. For treatment effect parameters, long-term evidence is much less likely to be available. The implications of the short-term evidence can again play a role. However, broad alternative scenarios (such as: optimistic, pessimistic and intermediate) can form an expression of temporal uncertainty. Suitable relative weights should be ascribed to these scenarios, but where that is not possible, a disinterested equal weighting can be applied in the first instance. Temporal uncertainty related to costs and health-related quality of life (HRQoL) is typically overlooked in HTA. However, a number of long-term factors such as the impact of aging and uncertain future events have the potential to significantly alter expected cost-effectiveness and ought to be considered explicitly in decision modelling.

For a range of contexts and parameter-types, incorporating the elicited opinion of experts is likely to be a very useful resource in the characterisation of temporal uncertainty. For many situations, especially where there exists some external evidence with which to characterise temporal uncertainty, expert opinion can be employed to validate the expression of temporal uncertainty and the implications for health outcomes. For other, even more uncertain situations, and in particular the expression of temporal uncertainty for treatment effects, a more formal role for expert elicitation is appropriate. There is value in extending methods of expert elicitation for use in HTA for the specific purposes of expressing temporal uncertainty relating to treatment effects.

Often in HTA, in place of expressing temporal uncertainty, assumptions are imposed that are explicitly 'conservative', generally meaning that they knowingly underestimate the cost-effectiveness of a new health technology. This can be a useful approach; in particular it can convey that a new technology is cost-effective *even* when conservative assumptions are imposed. However this approach implies that (i) expected cost-effectiveness is an underestimation (e.g. mean ICER is higher than true expectation) and (ii) analysis of uncertainty and value of information is all but meaningless since decision uncertainty is not faithfully expressed. Therefore in cases where a technology is rejected based on an analysis containing conservative assumptions, it is imperative that those assumptions be replaced with an unbiased expression of temporal uncertainty.

5.1.5 Effect of Expressing Temporal Uncertainty

The replacement of conservative assumptions with expressions of temporal uncertainty can cause a downward shift in the mean ICER and potentially a change in decision from one of rejection to one of adoption. There is also likely to be an inflation of the uncertainty around costs and health effects. The change in decision uncertainty however, will be a product of both the shift in expected cost-effectiveness and the inflation in uncertainty around costs and health effects.

The successful expression of temporal uncertainty should lead to a fair estimate, given current knowledge, of decision uncertainty. As a result, the value of obtaining further information can be estimated as well as the value of obtaining further information on the temporal nature of one or more model parameters. The impact of specific instances of temporal uncertainty can only be appraised once all temporal uncertainty (and uncertainty in general) has been characterised. It may transpire that only one source of temporal uncertainty ultimately influences the adoption decision. The calculation of CINB over time (post expression of temporal uncertainty) will also give an

indication as to the time it is expected to take in order for uncertainty to be resolved with further follow-up, as opposed to procuring supplementary evidence.

5.2 Limitations

This RITA-3 re-analysis was not exhaustive. There were other sources of temporal uncertainty in the RITA-3 case study that were not addressed such as: the uncertainty surrounding the long-term suitability of the logistic regression determining whether a patient who experienced a composite event moved to the death state or the 'Post MI' state, the structural uncertainty relating to the composition of the state-based model and its suitability to the long-term clinical pathways in this decision problem.

Moreover, there were many issues of temporal uncertainty that could did not arise in the RITA-3 motivating example, in particular, temporal uncertainty relating to longitudinal parameters, an area which has been overlooked to date.

A key issue that also did not arise in the RITA-3 example was that of the availability of long-term observational data which can be directly employed to address temporal uncertainty. There would be much value in exploring how best to utilise such evidence where it is available. For instance, it may be desirable to execute a kind of temporal evidence synthesis of the long-term observational evidence with the short-term RCT evidence.

The RITA-3 example represents a particular, though common, model structure. Although the methodological challenges of quantifying temporal uncertainty will be similar, there may be different challenges to implementing these methods using other model structures, such as patient-level simulations. More generally, there are other issues in CEA that relate to the disparity between the evidence and analysis time horizons that were not explored in this thesis. These include normative issues such as long-term discount rates and end-of-life QALYs, the broader role of time in CEA, e.g. optimal cycle lengths, the impact of uncertain future events which was touched upon but whose impact in CEA were not fully appraised.

For parameters where there existed a number of broad alternative temporal trajectories, an 'equal weighting' approach was used in order to incorporate each scenario into the probabilistic analysis and express temporal uncertainty. Although it is argued that this approach is legitimate in the face of a paucity of long-term evidence, its requirement ought to be rare. The weighting of alternative scenarios should be informed by evidence where

possible. If observational evidence is not available then expert elicitation ought to be employed. The choice of scenarios was, for the most part, based upon current NICE guidance which recommends applying optimistic, pessimistic and intermediate scenarios in cases of structural uncertainty.

5.3 Recommendations

From the considerations and analysis in this thesis, it is recommended that the issue of temporal uncertainty be explicitly and systematically considered in future HTAs. In particular, it is recommended to:

- (i) Appraise and demonstrate the significance of temporal uncertainty on decision making through the calculation of cumulative incremental net benefit and through scenario analyses.
- (ii) Identify the particular model parameters where there is exposure to temporal uncertainty.
- (iii) Characterise temporal uncertainty for each pertinent model parameter by obtaining all relevant evidence (including extrapolated short-term evidence) and expressing uncertainty through a single uncertain parameter – either using a discrete or continuous distribution. All assumptions relating to behaviour of parameters over long-term should be externally validated (by external data if possible, or by clinical experts if necessary).
- (iv) Define a standard set of alternative scenarios that should be used in cases where there are no data to inform long-term values of parameters
- (v) Calculate expected cost-effectiveness for all relevant risk or sub groups and determine whether there is value in obtaining further information related to the temporal trajectory of one or more model parameters

5.4 Related Issues and Potential Further Research

This thesis included an analysis of the value of waiting for further evidence on the assumption that sufficient further evidence would become available at a specific time point. A valuable extension of this would be incorporate the uncertainty regarding ‘when’ that time point would be, taking account

of the alternative implications of further evidence. Such a framework ought to work by combining the concepts of the value of waiting and the value of sample information.

A pertinent issue in the study of temporal uncertainty is how to best incorporate relevant long-term external data when it is available. For instance, if there exists a single data point to inform a parameter's temporal trajectory over the unobserved period, how exactly ought this data point to be incorporated? The external data point could be used to validate long-term parameter values extrapolated from short-term evidence, or it could be synthesised with the short-term evidence in order to generate the expected parameter values in between. In the latter case, is it reasonable to use a simple linear interpolation? Or should some relative weighting be employed to strike a balance between the long-term parameter value implied by the external data point and the trend apparent in the short-term evidence?

An important step in the process of addressing temporal uncertainty in order to inform appropriate decision-making in healthcare resource allocation is identifying pertinent further evidence when the need for which is indicated by the analysis recommend above. Evidence that is likely to help reduce any temporal uncertainty existent in a CEDM is unlikely to come in the form of the typical 'new primary research' that might be commissioned in circumstances of uncertainty, e.g. a new randomised control trial. Thus there has been an emphasis in this thesis on alternative modes of obtaining further evidence such as 'the value of waiting' and expert elicitation. More research would be useful on what types of obtainable evidence would be most relevant for a variety of model parameters that might be exposed to temporal uncertainty. For instance, it may be the case, that for some parameters, certain (carefully designed) retrospective observational studies could be valuable.

The issue of period risk, as discussed in chapter 4.2 is pertinent when discussing future decisions. There is often the assumption that evidence can only remain relevant for so long, but even within that timeframe the change in medical care and patient characteristics may significantly alter the decision problem. For example, if standard medical care improved to the extent that the early interventional treatment strategy was no longer a cost-effective approach in 10 years' time, could the positive adoption decision made today be reversed? Would this be ethical? Is this something that should be taken into account as today's adoption decision is considered?

A number of alternative approaches to eliciting evidence to inform temporal trajectories were discussed and one approach was implemented using a stylised elicitation output. It would be interesting and worthwhile if these alternative approaches were implemented in a real elicitation setting so as to note any difficulties with any of the approaches and any differences in the outcomes.

Temporal uncertainty in the context of communicable diseases poses a range of challenges beyond those addressed in this thesis which were in a context of non-communicable diseases. In general, research to marry the methodologies of infectious disease modelling and the economic evaluation of healthcare interventions is warranted. A notable expected feature of temporal uncertainty related to communicable diseases is the multi-modality of the distribution of ICERs. This is because the uncertainty around an input parameter in an infectious disease model can be the difference between a steady-state equilibrium and an epidemic.

The parameters analysed in the RITA-3 example that conveyed effectiveness were of a time-to-event nature. Although the issues associated with temporal change in a parameter and the relevance of extrapolated evidence are equally pertinent, there would be much value in conducting similar research on extrapolation in relation to longitudinal parameters and expressing the related temporal uncertainty.

Finally, research comparing the predicted long-term parameter values from extrapolated short-term evidence with the empirical parameters values eventually revealed would be of value. Such an investigation has already been carried out by Davies et al. where it was found that continued proportional hazards predicted from the short-term evidence did not become manifest after further evidence was revealed (Davies et al., 2013). Further research like this is warranted in order to better understand when and why extrapolated evidence is likely to be misleading.

APPENDICES

Appendix 1: Details of HTAs Reviewed

The title, authors and link to publication of the 64 HTAs reviewed in detail.

Report title	Author	Publication URL
Psychological treatment for insomnia in the regulation of long-term hypnotic drug use	K Morgan, S Dixon, N Mathers, J Thompson and M Tomeny	http://www.hta.ac.uk/fullmono/mon808.pdf
Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling	M Robinson, S Palmer, M Sculpher, Z Philips, L Ginnelly, A Bowens, S Golder, K Alfakih, A Bakhai, C Packham, N Cooper, K Abrams, A Eastwood, A Pearman, M Flather, D Gray and A Hall	http://www.hta.ac.uk/fullmono/mon927.pdf
Outcomes of electrically stimulated gracilis neosphincter surgery	T Tillin, M Chambers and R Feldman	http://www.hta.ac.uk/fullmono/mon928.pdf
Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling	R Garside, M Pitt, M Somerville, K Stein, A Price and N Gilbert	http://www.hta.ac.uk/fullmono/mon1008.pdf
Health benefits of antiviral therapy for mild chronic hepatitis C: randomised control trial and economic evaluation	M Wright, R Grieve, J Roberts, J Main and HC Thomas	http://www.hta.ac.uk/fullmono/mon1021.pdf
The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review	M Connock, A Burls, E Frew, A Fry-Smith, A Juarez-Garcia, C McCabe, A Wailoo, K Abrams, N Cooper, A Sutton, A O'Hagan and D Moore	http://www.hta.ac.uk/fullmono/mon1024.pdf
Epidemiological, social, diagnostic, and economic evaluation of population screening for genital chlamydial infection	N Low, A McCarthy, J Macleod, C Salisbury, R Campbell, TE Roberts, P Horner, S Skidmore, JAC Sterne, E Sanford, F Ibrahim, A Holloway, R Patel, PM Barton, SM Robinson, N Mills,	http://www.hta.ac.uk/fullmono/mon1108.pdf

	A Graham, A Herring, EO Caul, G Davey Smith, FDR Hobbs, JDC Ross and M Egger	
'Cut down to quit' with nicotine replacement therapies in smoking cessation: a systematic review of effectiveness and economic analysis	D Wang, M Connock, P Barton, A Fry-Smith, P Aveyard and D Moore	http://www.hta.ac.uk/project/1542.asp
Blood glucose self-monitoring in type 2 diabetes: a randomised controlled trial	AJ Farmer, AN Wade, DP French, J Simon, P Yudkin, A Gray, A Craven, L Goyder, RR Holman, D Mant, A-L Kinmonth and HAW Neil	http://www.hta.ac.uk/fullmono/mon1315.pdf
The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis	W Rogowski, J Burch, S Palmer, C Craigs, S Golder and N Woolacott	http://www.hta.ac.uk/fullmono/mon1331.pdf
The clinical effectiveness of glucosamine and chondroitin supplements in slowing or arresting progression of osteoarthritis of the knee: a systematic review and economic evaluation	Black C, Clar C, Henderson R, MacEachern C, McNamee P, Quayyum Z, et al.	http://www.hta.ac.uk/fullmono/mon1352.pdf
The clinical and cost effectiveness of testing for cytochrome P450 polymorphisms in patients treated with antipsychotics: a systematic review and economic evaluation	Fleeman N, McLeod C, Bagust A, Beale S, Boland A, Dundar Y, et al.	http://www.hta.ac.uk/fullmono/mon1403.pdf
Antenatal screening for haemoglobinopathies in primary care: a cohort study and cluster randomised trial to inform a simulation model. The Screening for Haemoglobinopathies in First Trimester (SHIFT)	Dormandy E, Bryan S, Gulliford MC, Roberts TE, Ades AE, Calnan M, et al.	http://www.hta.ac.uk/fullmono/mon1420.pdf

trial		
The safety and effectiveness of different methods of ear wax removal: a systematic review and economic evaluation	Clegg A, Loveman E, Gospodarevskaya E, Harris P, Bird A, <i>et al</i>	http://www.hta.ac.uk/fullmono/mon1428.pdf
Sugammadex for the reversal of muscle relaxation in general anaesthesia: a systematic review	Chambers D, Paulden M, Paton F, Heirs M, Duffy S, <i>et al</i>	http://www.hta.ac.uk/fullmono/mon1439.pdf
Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation	C Main, S Palmer, S Griffin, L Jones, V Orton, M Sculpher, R Henderson, C Sudlow, N Hawkins and R Riemsma	http://www.hta.ac.uk/fullmono/mon840.pdf
The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation	R Garside, K Stein, E Castelnuovo, M Pitt, D Ashcroft, P Dimmock and L Payne	http://www.hta.ac.uk/fullmono/mon929.pdf
Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation	C Clar, E Cummins, L McIntyre, S Thomas, J Lamb, L Bain, P Jobanputra and N Waugh	http://www.hta.ac.uk/fullmono/mon947.pdf
Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation	C Main, L Bojke, S Griffin, G Norman, M Barbieri, L Mather, D Stark, S Palmer and R Riemsma	http://www.hta.ac.uk/fullmono/mon1009.pdf

Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation	J Shepherd, J Jones, A Takeda, P Davidson and A Price	http://www.hta.ac.uk/fullmono/mon1028.pdf
A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer	R Collins, E Fenwick, R Trowman, R Perard, G Norman, K Light, A Birtle, S Palmer and R Riemsma	http://www.hta.ac.uk/fullmono/mon1102.pdf
Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation	N Woolacott, Y Bravo Vergel, N Hawkins, A Kainth, Z Khadjesari, K Misso, K Light, C Asseburg, S Palmer, K Claxton, I Bruce, M Sculpher and R Riemsma	http://www.hta.ac.uk/fullmono/mon1031.pdf
The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease	E Loveman, C Green, J Kirby, A Takeda, J Picot, E Payne and A Clegg	http://www.hta.ac.uk/fullmono/mon1001.pdf
The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model	M Fox, S Mealing, R Anderson, J Dean, K Stein, A Price and RS Taylor	http://www.hta.ac.uk/fullmono/mon1147.pdf
Natalizumab for multiple sclerosis	Pentag	http://www.hta.ac.uk/erg/reports/1647.pdf
Stapled haemorrhoidectomy (haemorrhoidopexy) for the treatment of haemorrhoids: a systematic review and economic evaluation	J Burch, D Epstein, A Baba-Akbari, H Weatherly, D Fox, S Golder, D Jayne, M Drummond and N Woolacott	http://www.hta.ac.uk/fullmono/mon1208.pdf
Omalizumab for severe persistent allergic asthma	Jones J, Shepherd J, Hartwell D et al	http://www.hta.ac.uk/erg/reports/1649.pdf

Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma	LRiG	http://www.hta.ac.uk/erg/reports/1654.pdf
Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis	C McDaid, S Griffin, H Weatherly, K Durée, M van der Burgt, S van Hout, J Akers, RJO Davies, M Sculpher and M Westwood	http://www.hta.ac.uk/fullmono/mon1304.pdf
Rimonabant for the treatment of obese or overweight patients	J Burch, G Norman, J Glanville, C McKenna, S Palmer, M Sculpher, N Woolcott	http://www.hta.ac.uk/erg/reports/1691.pdf
The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model	M Bond, S Mealing, R Anderson, J Elston, G Weiner, RS Taylor, M Hoyle, Z Liu, A Price and K Stein	http://www.hta.ac.uk/fullmono/mon1344.pdf
Bevacizumab, sorafenib tosylate and sunitinib for renal cell carcinoma: a systematic review and economic evaluation	J Thompson Coon, M Hoyle, C Green, Z Liu, K Welch, T Moxham and K Stein	http://www.hta.ac.uk/fullmono/mon1402.pdf
The clinical and cost-effectiveness of rituximab for the first line treatment of chronic lymphocytic leukaemia: an evidence review of the submission from Roche	Caroline Main, Martin Pitt, Tiffany Moxham, Ken Stein	http://www.hta.ac.uk/erg/reports/1785.pdf
Ustekinumab for the treatment of moderate to severe psoriasis	SHTAC	http://www.hta.ac.uk/erg/reports/1819.pdf
Prasugrel for the treatment of acute coronary artery syndromes with percutaneous coronary intervention	Janette Greenhalgh, Adrian Bagust, Angela Boland, Carlos Martin Saborido, Nigel Fleeman, Claire McLeod, Yenal Dundar, Rumona Dickson, Chris Proudlove, Ruwanthi Kolamunnage-Dona, Michael Fisher	http://www.hta.ac.uk/erg/reports/1852.pdf

Recombinant human growth hormone for the treatment of growth disorders in children: a systematic review and economic evaluation	A Takeda, K Cooper, A Bird, L Baxter, GK Frampton, E Gospodarevskaya, K Welch and J Bryant	http://www.hta.ac.uk/fullmono/mon1442.pdf
Omalizumab for the treatment of severe persistent allergic asthma in children aged 6-11	Walker S, Burch J, McKenna C, Wright K, Griffin S, Woolacott N	http://www.hta.ac.uk/erg/reports/2212.pdf
Liraglutide for the treatment of Type 2 diabetes	Ewen Cummins, Pamela Royle, Deepson Shyangdan, Norman Waugh	http://www.hta.ac.uk/erg/reports/2157.pdf
Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK	BH Willis, P Barton, P Pearmain, S Bryan and C Hyde	http://www.hta.ac.uk/fullmono/mon913.pdf
A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over: The SAFE study	FDR Hobbs, DA Fitzmaurice, S Jowett, J Mant, E Murray, S Bryan, J Raftery, M Davies and G Lip	http://www.hta.ac.uk/fullmono/mon940.pdf
Randomised clinical trial, observational study and assessment of cost-effectiveness of the treatment of varicose veins (REACTIV trial)	Michaels	http://www.hta.ac.uk/fullmono/mon1013.pdf
Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy	Williams	http://www.hta.ac.uk/fullmono/mon1034.pdf
Taxanes for the adjuvant treatment of early breast cancer: systematic review and economic evaluation	S Ward, E Simpson, S Davis, D Hind, A Rees and A Wilkinson	http://www.hta.ac.uk/project/1516.asp

Enhanced external counterpulsation for stable angina or heart failure: a systematic review and economic evaluation	C McKenna, C McDaid, S Suekarran, N Hawkins, K Claxton, K Light, M Chester, J Cleland, N Woolacott and M Sculpher	http://www.hta.ac.uk/fullmono/mon1324.pdf
Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care	J Mant, J Doust, A Roalfe, P Barton, MR Cowie, P Glasziou, D Mant, RJ McManus, R Holder, J Deeks, K Fletcher, M Qume, S Sohanpal, S Sanders and FDR Hobbs	http://www.hta.ac.uk/fullmono/mon1332.pdf
Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer	Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths T R L, et al.	http://www.hta.ac.uk/fullmono/mon1404.pdf
A systematic review and economic evaluation of the clinical effectiveness and cost-effectiveness of aldosterone antagonists for postmyocardial infarction heart failure	McKenna C, Burch J, Suekarran S, Walker S, Bakhai A, et al.	http://www.hta.ac.uk/fullmono/mon1424.pdf
Newer agents for blood glucose control in type 2 diabetes: systematic review and economic evaluation	Waugh N, Cummins E, Royle P, Clar C, Marien M, <i>et al</i>	http://www.hta.ac.uk/fullmono/mon1436.pdf
Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation	J Wilby, A Kainth, N Hawkins et al	http://www.hta.ac.uk/fullmono/mon915.pdf
Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation	J Wilson, M Connock, F Song, G Yao, A Fry-Smith, J Raftery and D Peake	http://www.hta.ac.uk/fullmono/mon925.pdf

Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation	L Jones, S Griffin, S Palmer, C Main, V Orton, M Sculpher, C Sudlow, R Henderson, N Hawkins and R Riemsma	http://www.hta.ac.uk/fullmono/mon838.pdf
A systematic review and economic model of the clinical and cost-effectiveness of immunosuppressive therapy for renal transplantation in children	G Yao, E Albon, Y Adi, D Milford, S Bayliss, A Ready, J Raftery and RS Taylor	http://www.hta.ac.uk/fullmono/mon1049.pdf
Etanercept and efalizumab for the treatment of psoriasis: a systematic review	N Woolacott, N Hawkins, A Mason, A Kainth, Z Khadjesari, Y Bravo Vergel, K Misso, K Light, R Chalmers, M Sculpher and R Riemsma	http://www.hta.ac.uk/fullmono/mon1046.pdf
The clinical and cost effectiveness of inhaled insulin in diabetes mellitus: a systematic review and economic evaluation	C Black, E Cummins, P Royle, S Philip and N Waugh	http://www.hta.ac.uk/fullmono/mon1133.pdf
Adalimumab for the treatment of psoriasis	S Erhorn, V Young, S Brent, A Dyker, W Horsley, K Macfarlane, J Martin, B Reddy, S White, S Thomas, Y Bravo Vergel, S Palmer	http://www.hta.ac.uk/erg/reports/1646.pdf
Bortezomib for the treatment of multiple myeloma patients	C Green, J Bryant, A Takeda, K Cooper, A Clegg, A Smith, M Stephens	http://www.hta.ac.uk/erg/reports/1637.pdf
Infliximab for psoriasis	Emma Loveman, David Turner, Debbie Hartwell, Keith Cooper, Andy Clegg	http://www.hta.ac.uk/erg/reports/1650.pdf
Infliximab for the sub-acute manifestations of ulcerative colitis	Chris Hyde, Stirling Bryan, Ariadna Juarez-Garcia, Lazaros Andronis, Anne Fry-Smith	http://www.hta.ac.uk/erg/reports/1690.pdf
adalimumab for the treatment of psoriasis	David Turner, Joanna Picot, Keith Cooper,	http://www.hta.ac.uk/erg/reports/1692.pdf

	Emma Loveman	
Cetuximab for recurrent and/or metastatic squamous-cell carcinoma of the head and neck	J Greenhalgh, A Bagust, A Boland, N Fleeman, C McLeod, Y Dundar, C Proudlove and R Shaw	http://www.hta.ac.uk/erg/reports/1778.pdf
Alitretinoin for the treatment of severe chronic hand eczema.	Mike Paulden, Mark Rodgers, Susan Griffin, Russell Slack, Steven Duffy, John R Ingram, Nerys Woolacott, Mark Sculpher	http://www.hta.ac.uk/erg/reports/1798.pdf
Trabectedin for the treatment of advanced metastatic soft tissue sarcoma	P Harris, K Welch and A Clegg	http://www.hta.ac.uk/erg/reports/1994.pdf
Dronedarone within its licensed indication for the treatment of atrial fibrillation and atrial flutter	McKenna C, Maund E, Sarowar M, Fox D, Stevenson M, Pepper C, Palmer S, Woolacott N	http://www.hta.ac.uk/erg/reports/2153.pdf
Eltrombopag for the treatment of chronic idiopathic (immune) thrombocytopenic purpura	Dwayne Boyers, Xueli Jia, Mark Crowther, David Jenkinson, Cynthia Fraser, Graham Mowatt	http://www.hta.ac.uk/erg/reports/2163.pdf

Appendix 2: RITA-3 Cohort by Risk Profile

	Risk group 1	Risk group 2	Risk group 3	Risk group 4	Risk group 5
Age	45	52	52	61	66
Diabetes	No	No	No	No	Yes
Previous myocardial infarction	No	No	Yes	Yes	Yes
Smoker	No	Yes	No	Yes	No
Pulse (beats per minute)	72	82	82	87	97
ST depression	No	No	Yes	Yes	Yes
Angina	Yes	No	Yes	No	No
Sex	Female	Male	Male	Male	Male
Left bundle branch block	No	No	No	No	No

Appendix 3: Application of Expert Elicitation Methodology to RITA-3 example (stylised elicitation)

It was not within the scope of this thesis to carry out an elicitation exercise. However, it will be instructive to apply some of the concepts discussed above by carrying out a stylised elicitation. Therefore, let us say we want to employ formal elicitation to characterise the temporal uncertainty regarding treatment effect in the RITA-3 example.

From the four options outlined above for conducting the elicitation process, Option 2 (eliciting judgements regarding the duration of treatment effect) represents a good balance of modelling simplicity, comprehensibility (for the clinical experts) and reasonable accuracy.

To carry out the elicitation, a thorough explanation of the decision problem, the RITA-3 trial and the precise clinical nature of the patients cohorts must be given to the clinical experts. Following this, the elicitation exercise itself can be presented.

(i) Clinical Explanation of Uncertain Quantity

For the sake of clarity as we endeavour to elicit information from clinical experts, it is valuable to consider more carefully the clinical nature of the patients in each of the cohorts in RITA-3 and communicate this to the experts.

In the RITA-3 decision problem, the early interventional cohort consists of those who were, following NSTEMI-ACS, given a routine angiography followed by revascularisation if clinically indicated, whereas patients in the conservative cohort were given ischaemia or symptom-driven angiography. It can be assumed that a number of the more 'at risk' patients in the conservative cohort were given an angiography and if necessary, revascularisation at some stage during the observed period (although this information was not itself observed in the RITA-3 trial). Also it can be assumed that a proportion of the conservative cohort will have experienced a composite event where they would not have done under the early interventional treatment strategy.

Ultimately, what is required is simply the difference in risk between the average patient in the conservative cohort and the average patient in the early interventional cohort. The clinical factors that inform this risk differential include:

1. The proportion of the conservative cohort (that didn't experience a composite event) that had a revascularisation and the extent to which this differs from that same proportion in the early interventional cohort
2. The longevity of the benefit of revascularisation, i.e. when, or in what sense, the benefit of revascularisation 'wears off'
3. When the 'hazardous period' after first presenting with NSTEMI-ACS be said to have passed, i.e. whether there will be a rebound effect when revascularisation 'wears off' or whether the hazardous period having passed allows us to consider the event-free survivors from the early interventional cohort as equivalent to those in the conservative cohort.
4. Other elements such as possible different post-treatment clinical pathways (pertaining to hospital care and self-care) that may not be captured in the trial data

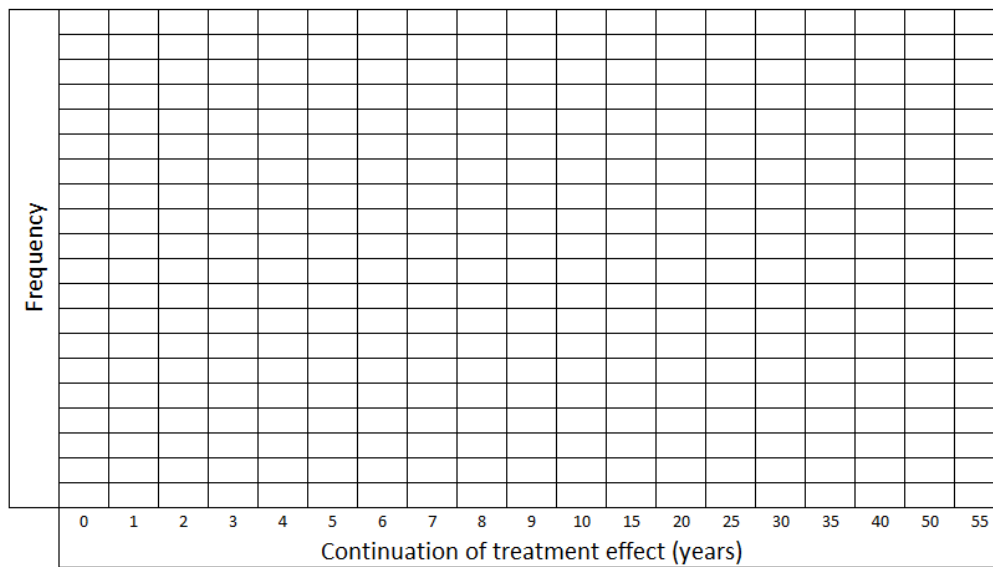
(ii) Elicitation Exercise

A question such as the following could be posed to the experts:

“The RITA-3 trial has shown that for 5 years after presenting with NSTEMI-ACS, an early interventional strategy results in a lower risk of a first composite event to patients compared to a conservative treatment strategy. For how long would you expect this ‘treatment effect’ to continue?”

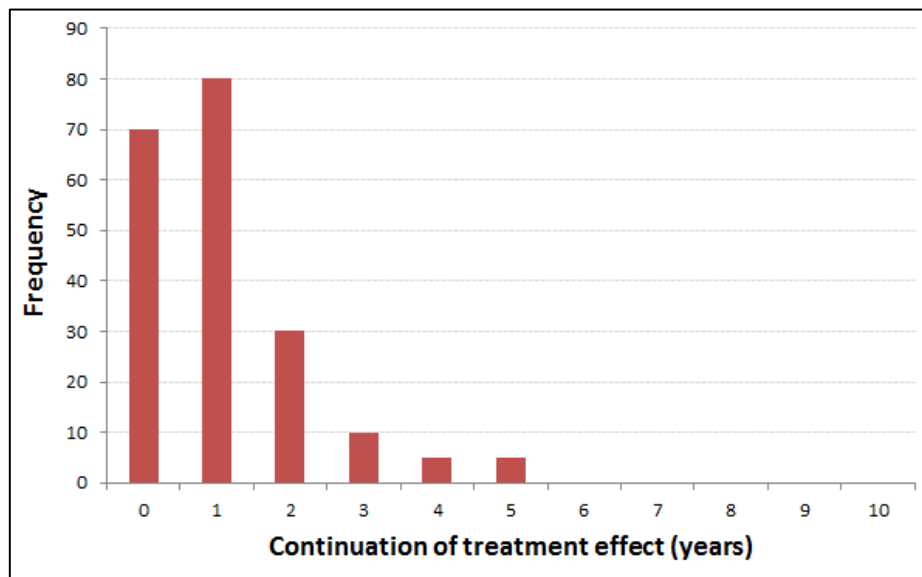
As per the histogram method, the experts would be asked to answer this question by placing 20 crosses on a chart such as the one in Figure 59.

Figure 59: Frequency chart to be filled in by experts



Let us assume that there were 10 experts who responded and that there was a seed question that demonstrated every expert was reliable. If equal weighting is given to the experts and their responses are combined through linear pooling, the result is a ‘super distribution’ such as the one illustrated in Figure 60.

Figure 60: Resultant ‘super-distribution’ from elicitation



Commonly at this stage, a smooth parametric distribution is fit to the experts’ ‘super-distribution’. For the distribution in Figure 60 above, a beta or possibly a gamma distribution may be suitable. It is also possible however, to directly input the discrete ‘super distribution’ into the decision model. Since the decision model is constructed in terms of yearly transition probabilities, it is more straightforward to employ this direct approach rather than modelling precise treatment effect durations. Using the direct approach, the stylised elicitation output as depicted in Figure 60 can be parameterised as follows.

For a random variable $X \sim U(0,1)$, the parameter D , which represents the duration of treatment effect from the 5 year time-point, is distributed as:

$$D = \begin{cases} 0 & \text{if } 0 \leq X < 0.35 \\ 1 & \text{if } 0.35 \leq X < 0.75 \\ 2 & \text{if } 0.75 \leq X < 0.9 \\ 3 & \text{if } 0.9 \leq X < 0.95 \\ 4 & \text{if } 0.95 \leq X < 0.975 \\ 5 & \text{if } 0.975 \leq X < 1 \end{cases}$$

Temporal uncertainty can therefore be neatly represented by the uncertain parameter D . Therefore, as with baseline risk in chapter 4.2, EVPI analysis can be easily employed to ascertain how much value there would be in knowing for certain the duration of treatment effect.

REFERENCES

- Health Technology Assessment UK* [Online]. Available: <http://www.hta.ac.uk/search/results.asp?Action=Search&QU=non-st&go=Go&PAGE=1> [Accessed November 2012].
1997. Coronary angioplasty versus medical therapy for angina: the second Randomised Intervention Treatment of Angina (RITA-2) trial. RITA-2 trial participants. *Lancet*, 350, 461-8.
- ACADEMY OF MANAGED CARE PHARMACY. 2012. *A Format for Submission of Clinical and Economic Evidence of Pharmaceuticals in Support of Formulary Consideration* [Online]. Available: <http://www.amcp.org/practice-resources/amcp-format-formulary-submissions.pdf> 2013].
- ANDERSON, H. V., CANNON, C. P., STONE, P. H., WILLIAMS, D. O., MCCABE, C. H., KNATTERUD, G. L., THOMPSON, B., WILLERSON, J. T. & BRAUNWALD, E. 1995. One-year results of the Thrombolysis in Myocardial Infarction (TIMI) IIIB clinical trial. A randomized comparison of tissue-type plasminogen activator versus placebo and early invasive versus early conservative strategies in unstable angina and non-Q wave myocardial infarction. *J Am Coll Cardiol*, 26, 1643-50.
- ANDERSON, J. L., ADAMS, C. D., ANTMAN, E. M., BRIDGES, C. R., CALIFF, R. M., CASEY, D. E., JR., CHAVEY, W. E., 2ND, FESMIRE, F. M., HOCHMAN, J. S., LEVIN, T. N., LINCOFF, A. M., PETERSON, E. D., THEROUX, P., WENGER, N. K., WRIGHT, R. S., SMITH, S. C., JR., JACOBS, A. K., HALPERIN, J. L., HUNT, S. A., KRUMHOLZ, H. M., KUSHNER, F. G., LYTLE, B. W., NISHIMURA, R., ORNATO, J. P., PAGE, R. L. & RIEGEL, B. 2007. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients With Unstable Angina/Non ST-Elevation Myocardial Infarction): developed in collaboration with the American College of Emergency Physicians, the Society for Cardiovascular Angiography and Interventions, and the Society of Thoracic Surgeons: endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation and the Society for Academic Emergency Medicine. *Circulation*, 116, e148-304.
- ANTOLIN, P. 2007. Longevity risk and private pensions.
- ARA, R. & BRAZIER, J. E. 2011. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value Health*, 14, 539-45.
- ARA, R. & WAILOO, A. 2011. NICE DSU Technical Support Document 12: the use of health state utility values in decision models. *Sheffield, UK: School of Health and Related Research, University of Sheffield*.
- ARA, R. & WAILOO, A. 2012. Using health state utility values in models exploring the cost-effectiveness of health technologies. *Value Health*, 15, 971-4.
- BAGUST, A. & BEALE, S. 2014. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Med Decis Making*, 34, 343-51.
- BANK OF ENGLAND. 2013. *Overview of the Inflation Report November 2013* [Online]. Available: <http://www.bankofengland.co.uk/publications/Pages/inflationreport/infrep.aspx> [Accessed January 2014].
- BILCKE, J., BEUTELS, P., BRISSON, M. & JIT, M. 2011. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. *Med Decis Making*, 31, 675-92.
- BLACK, C., CLAR, C., HENDERSON, R., MACEACHERN, C., MCNAMEE, P., QUAYYUM, Z., ROYLE, P. & THOMAS, S. 2009. The clinical effectiveness of glucosamine and chondroitin supplements in slowing or arresting progression of osteoarthritis of the knee: a systematic review and economic evaluation. *Health Technol Assess*, 13, 1-148.

- BLACK, N. 1996. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*, 312, 1215-8.
- BODEN, W. E., O'ROURKE, R. A., CRAWFORD, M. H., BLAUSTEIN, A. S., DEEDWANIA, P. C., ZOBLE, R. G., WEXLER, L. F., KLEIGER, R. E., PEPINE, C. J., FERRY, D. R., CHOW, B. K. & LAVORI, P. W. 1998. Outcomes in patients with acute non-Q-wave myocardial infarction randomly assigned to an invasive as compared with a conservative management strategy. Veterans Affairs Non-Q-Wave Infarction Strategies in Hospital (VANQWISH) Trial Investigators. *N Engl J Med*, 338, 1785-92.
- BOJKE, L., CLAXTON, K., BRAVO-VERGEL, Y., SCULPHER, M., PALMER, S. & ABRAMS, K. 2010. Eliciting distributions to populate decision analytic models. *Value Health*, 13, 557-64.
- BOJKE, L., CLAXTON, K., SCULPHER, M. & PALMER, S. 2009. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. *Value Health*.
- BOLAND, A., BAGUST, A., HOCKENHULL, J., DAVIS, H., CHU, P. & DICKSON, R. 2009. Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma. *Health Technol Assess*, 13 Suppl 2, 41-8.
- BOND, M., MEALING, S., ANDERSON, R., ELSTON, J., WEINER, G., TAYLOR, R. S., HOYLE, M., LIU, Z., PRICE, A. & STEIN, K. 2009. The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model. *Health Technol Assess*, 13, 1-330.
- BRENNAN, A., CHICK, S. E. & DAVIES, R. 2006. A taxonomy of model structures for economic evaluation of health technologies. *Health economics*, 15, 1295-1310.
- BRIGGS, A., CLARK, T., WOLSTENHOLME, J. & CLARKE, P. 2003. Missing... presumed at random: cost-analysis of incomplete data. *Health Economics*, 12, 377-92.
- BRIGGS, A., CLAXTON, K. & SCULPHER, M. J. 2006. *Decision modelling for health economic evaluation*, Oxford, Oxford University Press.
- BRIGGS, A., MIHAYLOVA, B., SCULPHER, M., HALL, A., WOLSTENHOLME, J., SIMOONS, M., DECKERS, J., FERRARI, R., REMME, W. J., BERTRAND, M. & FOX, K. 2007. Cost effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study. *Heart*, 93, 1081-6.
- BRIGGS, A. H. 2000. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*, 17, 479-500.
- BRIGGS, A. H., WEINSTEIN, M. C., FENWICK, E. A., KARNON, J., SCULPHER, M. J., PALTIEL, A. D. & FORCE, I.-S. M. G. R. P. T. 2012. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health*, 15, 835-842.
- BRITISH HEART FOUNDATION HEALTH PROMOTION RESEARCH GROUP 2010. Coronary Heart Disease Statistics 2010 edition. Oxford: British Heart Foundation Health Promotion Research Group.
- BROOKS, R. 1996. EuroQol: the current state of play. *Health Policy*, 37, 53-72.
- BROUWER, W., RUTTEN, F. & KOOPMANSCHAP, M. 2001. Costing in Economic Evaluations. In: MCGUIRE, M. D. A. A. (ed.) *Economic evaluation in health care: merging theory and practice*. Oxford: Oxford University Press.
- BROUWER, W. B., NIESSEN, L. W., POSTMA, M. J. & RUTTEN, F. F. 2005. Need for differential discounting of costs and health effects in cost effectiveness analyses. *BMJ*, 331, 446-8.
- BURNHAM, K. P. & ANDERSON, D. R. 2004. Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, 33, 261-304.
- BURTON, A. & ALTMAN, D. G. 2004. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *British Journal of Cancer*, 91, 4-8.
- BUSSCHBACH, J. J., HESSING, D. J. & DE CHARRO, F. T. 1993. The utility of health at different stages in life: a quantitative approach. *Soc Sci Med*, 37, 153-8.
- CAMPODÓNICO, S. & SINGPURWALLA, N. D. 1994. A Bayesian analysis of the logarithmic-Poisson execution time model based on expert opinion and failure data. *Software Engineering, IEEE Transactions on*, 20, 677-683.

- CANADIAN AGENCY FOR DRUGS AND TECHNOLOGIES IN HEALTH 2006. Guidelines for the economic evaluation of health technologies: Canada [3rd Edition]. Ottawa: Canadian Agency for Drugs and Technologies in Health.
- CANNON, C. P., WEINTRAUB, W. S., DEMOPOULOS, L. A., VICARI, R., FREY, M. J., LAKKIS, N., NEUMANN, F. J., ROBERTSON, D. H., DELUCCA, P. T., DIBATTISTE, P. M., GIBSON, C. M. & BRAUNWALD, E. 2001. Comparison of early invasive and conservative strategies in patients with unstable coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. *N Engl J Med*, 344, 1879-87.
- CARSTENSEN, B. 2007. Age-period-cohort models for the Lexis diagram. *Stat Med*, 26, 3018-45.
- CHALONER, K. & DUNCAN, G. 1983. Assessment of a beta prior distribution: PM elicitation. *Journal of the Royal Statistical Society*, 32, 174-180.
- CHARLTON, B. G. 1991. Medical practice and the double-blind, randomized controlled trial. *Br J Gen Pract*, 41, 355-6.
- CLAR, C., CUMMINS, E., MCINTYRE, L., THOMAS, S., LAMB, J., BAIN, L., JOBANPUTRA, P. & WAUGH, N. 2005. Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation. *Health Technol Assess*, 9, iii-iv, ix-x, 1-82.
- CLAXTON, K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*, 18, 341-64.
- CLAXTON, K. 2008. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics*, 26, 781-98.
- CLAXTON, K., MARTIN, S., SOARES, M., RICE, N., SPACKMAN, E., HINDE, S., DEVLIN, N., SMITH, P. C. & SCULPHER, M. 2013. Methods for the estimation of the NICE cost effectiveness threshold. *York: University of York*.
- CLAXTON, K., PALMER, S., LONGWORTH, L., BOJKE, L., GRIFFIN, S., MCKENNA, C., SOARES, M., SPACKMAN, E. & YOUN, J. 2011a. Uncertainty, evidence and irrecoverable costs: Informing approval, pricing and research decisions for health technologies. *CHE Research Paper 69*.
- CLAXTON, K., PAULDEN, M., GRAVELLE, H., BROUWER, W. & CULYER, A. J. 2011b. Discounting and decision making in the economic evaluation of health-care technologies. *Health Economics*, 20, 2-15.
- CLAXTON, K., SCULPHER, M., CULYER, A., MCCABE, C., BRIGGS, A., AKEHURST, R., BUXTON, M. & BRAZIER, J. 2006. Discounting and cost-effectiveness in NICE - stepping back to sort out a confusion. *Health Economics*, 15, 1-4.
- CLAXTON, K., SCULPHER, M. & DRUMMOND, M. 2002. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet*, 360, 711-5.
- CLAXTON, K., SCULPHER, M., MCCABE, C., BRIGGS, A., AKEHURST, R., BUXTON, M., BRAZIER, J. & O'HAGAN, T. 2005. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics*, 14, 339-47.
- CLEVES, M. A. 2008. *An introduction to survival analysis using Stata*, Texas, StataCorp LP.
- CLEVES, M. A. 2010. *An introduction to survival analysis using Stata*, College Station, Tex., Stata Press.
- COLLEGE DES ECONOMISTES DE LA SANTE. 2004. *French guidelines for the economic evaluation of health care technologies* [Online]. [Accessed 2013].
- COLLEGE VOOR ZORGVERZEKERINGEN DIEMEN. 2006. *Guidelines for pharmacoeconomic research, updated version* [Online]. Available: <http://www.ispor.org/PEguidelines/source/HTAGuidelinesNLupdated2006.pdf> [2013].
- COLLETT, D. 2003. *Modelling survival data in medical research*, Boca Raton, Fla. ; London, Chapman & Hall/CRC.
- CONNOCK, M., HYDE, C. & MOORE, D. 2011. Cautions Regarding the Fitting and Interpretation of Survival Curves: Examples from NICE Single Technology Appraisals of Drugs for Cancer. *Pharmacoeconomics*, 29, 827-37.

- CONNOCK, M., ROUND, J., BAYLISS, S., TUBEUF, S., GREENHELD, W. & MOORE, D. 2010. Sorafenib for the treatment of advanced hepatocellular carcinoma. *Health Technol Assess*, 14 Suppl 1, 17-21.
- COOKE, R. M. 1991. Experts in uncertainty: opinion and subjective probability in science.
- COX, C. 2008. The generalized F distribution: an umbrella for parametric survival analysis. *Stat Med*, 27, 4301-12.
- COX, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187-220.
- COX, G. M., GIBBONS, J. M., WOOD, A. T. A., CRAIGON, J., RAMSDEN, S. J. & CROUT, N. M. J. 2006. Towards the systematic simplification of mechanistic models. *Ecological Modelling*, 198, 240-246.
- CULYER, A., MCCABE, C., BRIGGS, A., CLAXTON, K., BUXTON, M., AKEHURST, R., SCULPHER, M. & BRAZIER, J. 2007. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *J Health Serv Res Policy*, 12, 56-8.
- CULYER, A. J. 1971. The nature of the commodity "health care" and its efficient allocation. *Oxford Economics Papers*, 23, 189-211.
- DAVIES, C., BRIGGS, A., LORGELLY, P., GARELLIC, G. & MALCHAU, H. 2013. The "Hazards" of Extrapolating Survival Curves. *Med Decis Making*.
- DE WINTER, R. J., WINDHAUSEN, F., CORNEL, J. H., DUNSELMAN, P. H., JANUS, C. L., BENDERMACHER, P. E., MICHELS, H. R., SANDERS, G. T., TIJSSEN, J. G. & VERHEUGT, F. W. 2005. Early invasive versus selectively invasive management for acute coronary syndromes. *N Engl J Med*, 353, 1095-104.
- DOLAN, P. 1997. Modeling valuations for EuroQol health states. *Med Care*, 35, 1095-108.
- DOLAN, P. & GUXEX, C. 1995. Time preference, duration and health state valuations. *Health economics*, 4, 289-299.
- DRUMMOND, M. F., SCULPHER, M. J., TORRANCE, G. W., O'BRIEN, B. J. & STODDART, G. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed.: Oxford University Press.
- DUNDAR, Y., BAGUST, A., HOUNSOME, J., MCLEOD, C., BOLAND, A., DAVIS, H., WALLEY, T. & DICKSON, R. 2009. Rituximab for the first-line treatment of stage III/IV follicular non-Hodgkin's lymphoma. *Health Technol Assess*, 13 Suppl 1, 23-8.
- ECKERMAN, S. & WILLAN, A. R. 2007. Expected value of information and decision making in HTA. *Health Economics*, 16, 195-209.
- EPSTEIN, D. M., SCULPHER, M. J., CLAYTON, T. C., HENDERSON, R. A., POCOCK, S. J., BUXTON, M. J. & FOX, K. A. 2008. Costs of an early intervention versus a conservative strategy in acute coronary syndrome. *Int J Cardiol*, 127, 240-6.
- FACEY, K. 2006. *INAHTA Health Technology Assessment (HTA) Glossary* [Online]. Available: <http://www.inahta.org/GO-DIRECT-TO/Members/> [Accessed August 2013].
- FISCHER, R. G. 1978. The Delphi method: a description, review and criticism. *Journal of academic librarianship*, 4, 64-70.
- FOLLAND, S., STANO, M. & GOODMAN, A. C. 2004. *The economics of health and health care*, Upper Saddle River, NJ, Pearson Prentice Hall.
- FOX, K. A., POOLE-WILSON, P., CLAYTON, T. C., HENDERSON, R. A., SHAW, T. R., WHEATLEY, D. J., KNIGHT, R. & POCOCK, S. J. 2005. 5-year outcome of an interventional strategy in non-ST-elevation acute coronary syndrome: the British Heart Foundation RITA 3 randomised trial. *Lancet*, 366, 914-20.
- FOX, K. A., POOLE-WILSON, P. A., HENDERSON, R. A., CLAYTON, T. C., CHAMBERLAIN, D. A., SHAW, T. R., WHEATLEY, D. J. & POCOCK, S. J. 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, 743-51.

- FOX, M., MEALING, S., ANDERSON, R., DEAN, J., STEIN, K., PRICE, A. & TAYLOR, R. S. 2007. The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model. *Health Technol Assess*, 11, iii-iv, ix-248.
- GARTHWAITE PH, K. J., O'HAGAN A. 2005. Statistical methods for eliciting probability distributions. *J Am Stat Assoc*, 680-700.
- GAVASAKAR, U. 1988. A comparison of two elicitation methods for a prior distribution for a binomial parameter. *Management Science*, 34, 784-790.
- GRAVELLE, H. & SMITH, D. 2001. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Economics*, 10, 587-99.
- GREAT BRITAIN. DEPT. OF HEALTH 1998. *A first class service : quality in the new NHS*, [London], Department of Health.
- GRECH, E. D. & RAMSDALE, D. R. 2003. Acute coronary syndrome: unstable angina and non-ST segment elevation myocardial infarction. *BMJ*, 326, 1259-61.
- GREEN, C., DINNES, J., TAKEDA, A., SHEPHERD, J., HARTWELL, D., CAVE, C., PAYNE, E. & CUTHBERTSON, B. H. 2005. Clinical effectiveness and cost-effectiveness of drotrecogin alfa (activated) (Xigris) for the treatment of severe sepsis in adults: a systematic review and economic evaluation. *Health Technol Assess*, 9, 1-126, iii-iv.
- GRIFFIN, S. C., CLAXTON, K. P., PALMER, S. J. & SCULPHER, M. J. 2011. Dangerous omissions: the consequences of ignoring decision uncertainty. *Health Economics*, 20, 212-24.
- GUYOT, P., WELTON, N. J., OUWENS, M. J. & ADES, A. E. 2011. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. *Value Health*, 14, 640-6.
- HARRELL, F. E., JR., LEE, K. L., MATCHAR, D. B. & REICHERT, T. A. 1985. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep*, 69, 1071-77.
- HEALTH, D. O. 2004. NHS reference costs 2003 and National tariff 2004.
- HEALTH INFORMATION AND QUALITY AUTHORITY 2010. Guidelines for the Economic Evaluation of Health Technologies in Ireland.
- HEDS. 2011. NICE changes discount rate methods. Available from: <http://scharrheds.blogspot.ie/2011/08/nice-changes-discount-rate-methods.html> [2014].
- HENRIKSSON, M., EPSTEIN, D. M., PALMER, S. J., SCULPHER, M. J., CLAYTON, T. C., POCOCK, S. J., HENDERSON, R. A., BUXTON, M. J. & FOX, K. A. 2008. The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. *Heart*, 94, 717-23.
- HERNAN, M. A. 2010. The hazards of hazard ratios. *Epidemiology*, 21, 13-5.
- HORROCKS, J. & THOMPSON, M. E. 2004. Modeling event times with multiple outcomes using the Wiener process with drift. *Lifetime Data Anal*, 10, 29-49.
- INSTITUTE FOR QUALITY AND EFFICIENCY IN HEALTH CARE (IQWiG). 2009. *General Methods for the Assessment of the Relation of Benefits to Costs* [Online]. Available: http://www.ispor.org/PEguidelines/source/Germany_AssessmentoftheRelationofBenefitstoCosts_En.pdf [Accessed 2013].
- JACKSON C, THOMPSON S & SHARPLES L 2009. Accounting for uncertainty in health economic decision models using model averaging. *Journal of the Royal Statistical Society*, 172.
- JACKSON, C. H., BOJKE, L., THOMPSON, S. G., CLAXTON, K. & SHARPLES, L. D. 2011. A Framework for Addressing Structural Uncertainty in Decision Models. *Med Decis Making*.
- JACKSON, C. H., SHARPLES, L. D. & THOMPSON, S. G. 2010. Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. *Int J Biostat*, 6, Article 34.
- JAGER, P. & BERTSCHE, B. A new approach to gathering failure behavior information about mechanical components based on expert knowledge. Reliability and Maintainability, 2004 Annual Symposium-RAMS, 2004. IEEE, 90-95.

- JENKINSON, D. 2005. *The Elicitation of Probabilities - A Review of the Statistical Literature* [Online]. Available: <http://www.sheffield.ac.uk/beep/publications.html> 2012].
- JONES, L., GRIFFIN, S., PALMER, S., MAIN, C., ORTON, V., SCULPHER, M., SUDLOW, C., HENDERSON, R., HAWKINS, N. & RIEMSMA, R. 2004. Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation. *Health Technol Assess*, 8, iii-iv, 1-196.
- KADANE, J. & WOLFSON, L. J. 1998. Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47, 3-19.
- KAHNEMAN, D. & SLOVIC, P. 1982. A. Tversky, 1982, Judgment Under Uncertainty: Heuristics and Biases. *CUP, Cambridge*.
- KAY, R. 1977. Proportional Hazard Regression Models and the Analysis of Censored Survival Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26, 227-237.
- KEYNES, J. M. & FEINSTEIN, C. H. 1921. *A treatise on probability*, London, Macmillan and co., limited.
- KIM, J., HENDERSON, R. A., POCOCK, S. J., CLAYTON, T., SCULPHER, M. J. & FOX, K. A. 2005. Health-related quality of life after interventional or conservative strategy in patients with unstable angina or non-ST-segment elevation myocardial infarction: one-year results of the third Randomized Intervention Trial of unstable Angina (RITA-3). *J Am Coll Cardiol*, 45, 221-8.
- KIM LG & THOMPSON S 2010. UNCERTAINTY AND VALIDATION OF HEALTH ECONOMIC DECISION MODELS. *HEALTH ECONOMICS*, 19.
- KLEINBAUM, D. G. & KLEIN, M. 2005. *Survival analysis : a self-learning text*, New York, NY, Springer.
- KUMARANAYAKE, L. 2000. The real and the nominal? Making inflationary adjustments to cost and other economic data. *Health Policy Plan*, 15, 230-4.
- LAGERQVIST, B., HUSTED, S., KONTNY, F., STAHLE, E., SWAHN, E. & WALLENTIN, L. 2006. 5-year outcomes in the FRISC-II randomised trial of an invasive versus a non-invasive strategy in non-ST-elevation acute coronary syndrome: a follow-up study. *Lancet*, 368, 998-1004.
- LATIMER, N. 2011. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available from <http://www.nicedsu.org.uk>.
- LATIMER, N. R. 2013. Survival Analysis for Economic Evaluations Alongside Clinical Trials-- Extrapolation with Patient-Level Data: Inconsistencies, Limitations, and a Practical Guide. *Med Decis Making*.
- LEAL, J., WORDSWORTH, S., LEGOOD, R. & BLAIR, E. 2007. Eliciting expert opinion for economic models: an applied example. *Value Health*, 10, 195-203.
- LEAMER, E. E. 1978. *Specification searches : ad hoc inference with nonexperimental data*, New York, Wiley.
- LEE, R. 2000. The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, 4, 80-91.
- LO´PEZ-BASTIDA, J. O., JUAN; ANTON~ANZAS, FERNANDO;. 2013. *Spanish Recommendations on Economic Evaluation of Health Technologies* [Online]. Available: http://www.ispor.org/PEguidelines/source/Spain-Guidelines-Abstract_English-Version-2010.pdf [Accessed 2013].
- LONGWORTH, L. & ROWEN, D. 2011. NICE DSU technical support document 10: the use of mapping methods to estimate health state utility values. *London: NICE*.
- M HENRIKSSON, D M EPSTEIN, S J PALMER, M J SCULPHER, T C CLAYTON, S J POCOCK, R A HENDERSON, M J BUXTON & K A A FOX 2007. The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. *Heart*, 94, 717-723.
- MAIN, C., PALMER, S., GRIFFIN, S., JONES, L., ORTON, V., SCULPHER, M., HENDERSON, R., SUDLOW, C., HAWKINS, N. & RIEMSMA, R. 2004. Clopidogrel used in combination with aspirin

- compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation. *Health Technol Assess*, 8, iii-iv, xv-xvi, 1-141.
- MAIN, C., PITT, M., MOXHAM, T. & STEIN, K. 2010. The clinical effectiveness and cost-effectiveness of rituximab for the first-line treatment of chronic lymphocytic leukaemia: an evidence review of the submission from Roche. *Health Technol Assess*, 14, 27-32.
- MAKRIDAKIS, S., ANDERSEN, A., CARBONE, R., FILDES, R., HIBON, M., LEWANDOWSKI, R., NEWTON, J., PARZEN, E. & WINKLER, R. 1982. The Accuracy of Extrapolation (Time-Series) Methods - Results of a Forecasting Competition. *Journal of forecasting*, 1, 111-153.
- MANCA, A., SARAMAGO, P., SCULPHER, M., ROYSTON, P., PARMAR, M. & COPAS, A. 2009. Fragile - handle with care! Issues in the long-term term extrapolation of individual patient level data for cost-effectiveness modelling. *Presented at the 2009 summer meeting of the Health Economists' Study Group. Work in progress.*
- MAUSKOPF, J., WALTER, J., BIRT, J., BOWMAN, L., COPLEY-MERRIMAN, C. & DRUMMOND, M. 2011. Differences among formulary submission guidelines: implications for health technology assessment. *Int J Technol Assess Health Care*, 27, 261-70.
- MCCABE, C., CLAXTON, K. & CULYER, A. J. 2008. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics*, 26, 733-44.
- MCCULLOUGH, P. A., O'NEILL, W. W., GRAHAM, M., STOMEL, R. J., ROGERS, F., DAVID, S., FARHAT, A., KAZLAUSKAITE, R., AL-ZAGOUM, M. & GRINES, C. L. 1998. A prospective randomized trial of triage angiography in acute coronary syndromes ineligible for thrombolytic therapy. Results of the medicine versus angiography in thrombolytic exclusion (MATE) trial. *J Am Coll Cardiol*, 32, 596-605.
- MCKENNA, C., MCDAID, C., SUEKARRAN, S., HAWKINS, N., CLAXTON, K., LIGHT, K., CHESTER, M., CLELAND, J., WOOLACOTT, N. & SCULPHER, M. 2009. Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis. *Health Technol Assess*, 13, iii-iv, ix-xi, 1-90.
- MCKENNA, C. C., K. 2011. Addressing Adoption and Research Design Decisions Simultaneously: The Role of Value of Information Analysis. *Medical Decision Making*, Forthcoming.
- MEHREZ, A. & GAFNI, A. 1989. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making*, 9, 142-9.
- MELTZER, D. 1997. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ*, 16, 33-64.
- MIHAYLOVA, B., BRIGGS, A., ARMITAGE, J., PARISH, S., GRAY, A. & COLLINS, R. 2006. Lifetime cost effectiveness of simvastatin in a range of risk groups and age groups derived from a randomised trial of 20,536 people. *BMJ*, 333, 1145.
- MORGAN, K., DIXON, S., MATHERS, N., THOMPSON, J. & TOMENY, M. 2004. Psychological treatment for insomnia in the regulation of long-term hypnotic drug use. *Health Technol Assess*, 8, iii-iv, 1-68.
- NICE 2006. Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer: Decision of the Panel.
- NICE 2013. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013.
- NORD, E. 1995. The person-trade-off approach to valuing health care programs. *Med Decis Making*, 15, 201-8.
- O'HAGAN, A. 1988. *Probability: Methods and Measurement*, Chapman and Hall London.
- O'HAGAN, A., BUCK, C., DANESHKHAH, A., EISER, J., GARTHWAITE, P. H., JENKINSON, D., OAKLEY, J. & RAKOW, T. 2006. *Uncertain Judgments: Eliciting Experts' Probabilities*, Wiley.
- OAKLEY, J. 2010. Eliciting univariate probability distributions. *Rethinking risk measurement and reporting*, 1.

- OFFICE FOR NATIONAL STATISTICS. 2012. Available: <http://www.ons.gov.uk/ons/index.html> [Accessed January 2012].
- OFFICE OF NATIONAL STATISTICS. 2011. *2010-based National Population Projections* [Online]. Office of National Statistics. Available: <http://www.ons.gov.uk/ons/datasets-and-tables/index.html?pageSize=50&sortBy=none&sortDirection=none&newquery=population+by+age&content-type=Reference+table&content-type=Dataset> [Accessed November 2012].
- PALMER, S., SCULPHER, M., PHILIPS, Z., ROBINSON, M., GINNELLY, L. & BAKHAI, A. 2002. *A cost-effectiveness model comparing alternative management strategies for the use of glycoprotein IIb/IIIa antagonists in non-ST-elevation acute coronary syndrome. Report to the National Institute for Clinical Excellence* [Online]. Available: <http://www.nice.org.uk/Docref.asp?d=32030> [Accessed February 2013].
- PALMER, S. & TORGERSON, D. J. 1999. Economic notes: definitions of efficiency. *BMJ*, 318, 1136.
- PAULDEN, M., RODGERS, M., GRIFFIN, S., SLACK, R., DUFFY, S., INGRAM, J. R., WOOLACOTT, N. & SCULPHER, M. 2010. Alitretinoin for the treatment of severe chronic hand eczema. *Health Technol Assess*, 14 Suppl 1, 39-46.
- PHARMACEUTICAL BENEFITS ADVISORY COMMITTEE. 2008. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3)* [Online]. Available: <http://www.ispor.org/PEguidelines/source/Australia-Guidelines-for-preparing-submissions-to-the-Pharmaceutical-Benefits-Advisory-Committee-2008.pdf> [2013].
- PHILIPS Z, CLAXTON K & PALMER S 2008. The Half-Life of Truth: What Are Appropriate Time Horizons for Research Decisions? *Medical Decision Making*, 28.
- PHILIPS, Z., BOJKE, L., SCULPHER, M., CLAXTON, K. & GOLDER, S. 2006. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*, 24, 355-71.
- PITMAN, R., FISMAN, D., ZARIC, G. S., POSTMA, M., KRETZSCHMAR, M., EDMUNDS, J. & BRISSON, M. 2012. Dynamic Transmission Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. *Value Health*, 15, 828-34.
- PLISKIN, J. 1980. Utility Functions for Life Years and Health Status. *Operations Research*, 28, 206-224.
- POLSKY, D. & BASU, A. 2006. Selection bias in observational data. In: JONES, A. (ed.) *The Elgar Companion to Health Economics*. Edward Elgar.
- PRICEWATERHOUSECOOPERS. 2014. *Economic projections: January 2014* [Online]. Available: <http://www.pwc.co.uk/economic-services/global-economy-watch/gew-projections.jhtml> [2014].
- RICE, J. A. 1995. *Mathematical statistics and data analysis*, Belmont, Calif., Duxbury Press.
- ROBINSON, M., PALMER, S., SCULPHER, M., PHILIPS, Z., GINNELLY, L., BOWENS, A., GOLDER, S., ALFAKIH, K., BAKHAI, A., PACKHAM, C., COOPER, N., ABRAMS, K., EASTWOOD, A., PEARMAN, A., FLATHER, M., GRAY, D. & HALL, A. 2005. Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling. *Health Technol Assess*, 9, iii-iv, ix-xi, 1-158.
- ROGOWSKI, W., BURCH, J., PALMER, S., CRAIGS, C., GOLDER, S. & WOOLACOTT, N. 2009. The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis. *Health Technol Assess*, 13, iii-iv, ix-xi, 1-77.
- ROYSTON, P. & PARMAR, M. K. 2002. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*, 21, 2175-97.
- RUSSELL, L. B. 2005. Comparing model structures in cost-effectiveness analysis. *Med Decis Making*, 25, 485-6.
- SCHOENFELD, D. 1982. Partial Residuals for The Proportional Hazards Regression Model. *Biometrika*, 69, 239-241.

- SCOTTISH MEDICINES CONSORTIUM. 2007. *Guidance to Manufacturers for Completion of New Product Assessment Form (NPAF) (Revised June 2007)* [Online]. Available: <http://www.ispor.org/PEguidelines/source/GuidanceinScotland-June2007.pdf> 2013].
- SCULPHER, M. 2012. Decision modelling for economic evaluation: Markov chains and beyond. *Advanced Modelling Course 2012*.
- SCULPHER, M. & CLAXTON, K. 2005. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty--when is there sufficient evidence? *Value in Health*, 8, 433-46.
- SCULPHER, M., CLAXTON, K., AKEHURST, R. 2005. It's just evaluation for decision making: recent developments in, and challenges for, cost-effectiveness research. In: SMITH, P. C. (ed.) *Health policy and economics. opportunities and challenges*. Maidenhead: Open University Press.
- SCULPHER, M., FENWICK, E. & CLAXTON, K. 2000. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics*, 17, 461-77.
- SCULPHER, M., SMITH, D., CLAYTON, T., HENDERSON, R., BUXTON, M., POCOOCK, S. & CHAMBERLAIN, D. 2002. Coronary angioplasty versus medical therapy for angina. Health service costs based on the second Randomized Intervention Treatment of Angina (RITA-2) trial. *Eur Heart J*, 23, 1291-1300.
- SCULPHER, M. J., CLAXTON, K., DRUMMOND, M. & MCCABE, C. 2006. Whither trial-based economic evaluation for health care decision making? *Health Economics*, 15, 677-87.
- SHYANGDAN, D., CUMMINS, E., ROYLE, P. & WAUGH, N. 2011. Liraglutide for the treatment of type 2 diabetes. *Health Technol Assess*, 15 Suppl 1, 77-86.
- SIMPSON, E., FITZGERALD, P., EVANS, P., TAPPENDEN, P., RECKLESS, J., BAKHAI, A. & KALITA, N. 2011. Bivalirudin for the treatment of ST-segment elevation myocardial infarction: A Single Technology Appraisal. *SCHARR*.
- SMITH, P. C. & STREET, A. D. 2012. On the Uses of Routine Patient-Reported Health Outcome Data. *Health Economics*.
- SOARES, M. O., BOJKE, L., DUMVILLE, J., IGLESIAS, C., CULLUM, N. & CLAXTON, K. 2011. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. *Stat Med*, 30, 2363-80.
- SPACEK, R., WIDIMSKY, P., STRAKA, Z., JIRESOVA, E., DVORAK, J., POLASEK, R., KAREL, I., JIRMAR, R., LISA, L., BUDESINSKY, T., MALEK, F. & STANKA, P. 2002. Value of first day angiography/angioplasty in evolving Non-ST segment elevation myocardial infarction: an open multicenter randomized trial. The VINO Study. *Eur Heart J*, 23, 230-8.
- STEVENSON, M., LLOYD-JONES, M. & PAPAIOANNOU, D. 2009. Vitamin K to prevent fractures in older women: systematic review and economic evaluation. *Health Technol Assess*, 13, iii-xi, 1-134.
- SULLIVAN, W. & PAYNE, K. 2011. The appropriate elicitation of expert opinion in economic models: making expert data fit for purpose. *Pharmacoeconomics*, 29, 455-9.
- SYMONS, M. J. & MOORE, D. T. 2002. Hazard rate ratio and prospective epidemiological studies. *J Clin Epidemiol*, 55, 893-9.
- TAN-TORRES EDEJER, T. & WORLD HEALTH ORGANIZATION 2003. *Making choices in health : WHO guide to cost-effectiveness analysis*, Geneva World Health Organization,.
- TAPPENDEN, P., CHILCOTT, J., WARD, S., EGGINGTON, S., HIND, D. & HUMMEL, S. 2006. Methodological issues in the economic analysis of cancer treatments. *Eur J Cancer*, 42, 2867-75.
- THE MEMBERS OF THE ITALIAN GROUP FOR PHARMACOECONOMIC STUDIES. 2001. *GUIDELINES FOR ECONOMIC EVALUATIONS IN ITALY: RECOMMENDATIONS FROM THE ITALIAN GROUP OF*

- PHARMACOECONOMIC STUDIES [Online]. Available: http://www.ispor.org/PEguidelines/source/Italy%20PE%20guidelines_Italy.pdf 2013].
- THE PHARMACEUTICAL BENEFITS BOARD. 2003. *General guidelines for economic evaluations from the Pharmaceutical Benefits Board* [Online]. Available: http://www.ispor.org/PEguidelines/source/Guidelines_in_Sweden.pdf 2013].
- TORGERSON, D. 2004. Randomised controlled trials: an introduction. *University of York Clinical Trials Unit*.
- TREASURY, H. M. S. 2003. *The Green Book: appraisal and evaluation in central government: Treasury guidance*, Stationery Office.
- TRUST, N. A. C. P. C. 2006. Appeal in respect of the Final Appraisal Determination: Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer.
- VAN BAAL, P. H., FEENSTRA, T. L., POLDER, J. J., HOOGENVEEN, R. T. & BROUWER, W. B. 2011. Economic evaluation and the postponement of health care costs. *Health Economics*, 20, 432-45.
- VAN DE WETERING, G., WOERTMAN, W. H. & ADANG, E. M. 2012. Time to incorporate time in cost-effectiveness analysis. *Eur J Health Econ*, 13, 223-6.
- VAN NOORTWIJK, J. M., DEKKER, A., COOKE, R. M. & MAZZUCHI, T. A. 1992. Expert judgment in maintenance optimization. *Reliability, IEEE Transactions on*, 41, 427-432.
- VAN WALRAVEN, C., DAVIS, D., FORSTER, A. J. & WELLS, G. A. 2004. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol*, 57, 672-82.
- VARDI, G. M., DAVIDSON, C. & WILLIAMS, E. 2005. Bifurcation lesion stent delivery using multiple guidewires. Google Patents.
- WALKER, D. G., HUTUBESSY, R. & BEUTELS, P. 2010. WHO Guide for standardisation of economic evaluations of immunization programmes. *Vaccine*, 28, 2356-9.
- WALKER, S., PALMER, S. & SCULPHER, M. 2007. The role of NICE technology appraisal in NHS rationing. *Br Med Bull*, 81-82, 51-64.
- WALKER, S., SCULPHER, M., CLAXTON, K. & PALMER, S. 2012. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value in Health*, 15, 570-579.
- WARD, S., PILGRIM, H. & HIND, D. 2009. Trastuzumab for the treatment of primary breast cancer in HER2-positive women: a single technology appraisal. *Health Technol Assess*, 13 Suppl 1, 1-6.
- WEINSTEIN, M. C., O'BRIEN, B., HORNBERGER, J., JACKSON, J., JOHANNESSEN, M., MCCABE, C. & LUCE, B. R. 2003. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value Health*, 6, 9-17.
- WILSON, J., CONNOCK, M., SONG, F., YAO, G., FRY-SMITH, A., RAFTERY, J. & PEAKE, D. 2005. Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation. *Health Technol Assess*, 9, 1-142.
- WINKLER, R. L. 1967. The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, 62, 776-800.
- WRIGHT, M., GRIEVE, R., ROBERTS, J., MAIN, J. & THOMAS, H. C. 2006. Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation. *Health Technol Assess*, 10, 1-113, iii.