The use of best practice guidelines and the effect of alternative model structures in results of cost-effectiveness: an analysis with emphasis in cardiovascular disease

by

MARIA CRISTINA PEÑALOZA RAMOS

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Health Economics Unit
Institute of Applied Health Research
College of Medical and Dental Sciences
The University of Birmingham
Birmingham, March 2017

Abstract

The growing use of decision analytic modelling (DAM) to aid decision making in healthcare has triggered the need for increased scrutiny of the methods used and the assessment of compliance with these methods. The assessment of structural uncertainty surrounding the choice of model structure and model external validity represent some of most frequent challenges faced by researchers.

This thesis used systematic reviews and two case studies focused on the self-management of hypertension in patients at high risk and thrombolysis in acute stroke to critically examine all available guidelines and statements of good practice and the adherence of current research to good practice guidelines. Two case studies were developed to assess structural uncertainty surrounding the choice of model structure and the impact of the exclusion of secondary events.

The results here indicate that DAM guidelines lack practicality due to the extensive amount of information available and their complexity; furthermore, researchers are failing to identify and correctly assess model structural uncertainty. This thesis makes an important contribution to current knowledge by developing and proposing the use of a practical five-dimension framework to improve the current standards of reporting results of DAM and by illustrating, through case studies, the assessment of structural uncertainty arising from the choice of model structure via scenario analysis and the use of extensive sensitivity analysis.

I would like to dedicate this thesis to my beautiful family

Para Wissam, Mamá, Papá (que en paz descanse), Ramón, Francia,

Carlos, Rosalba y Alejandro

Acknowledgements

This thesis could not have been completed without the support and help of many people and it is a pleasure to acknowledge and thank them. Firstly, I would like to thank my supervisors Dr Pelham Barton, Dr Sue Jowett and Dr Andrew Sutton for their insightful guidance and continuous support throughout this project. I am also thankful for the directions and advice I received from Professors Richard McManus and Jonathan Mant.

I would like to thank the CLAHRC Theme 7 Optimisation of the Management of Stroke & TIA project for funding my PhD and the Health Economics Unit, at the University of Birmingham for their great support and encouragement they provided throughout the development of my doctoral studies.

This thesis has also benefited from comments, suggestions and feedback from colleagues at the University of Birmingham and participants at conferences and seminars where parts of this work have been presented. I am particularly grateful to Lazaros, Hareth, Hugh, Tosin, Raymond, Ilias, Emma and Tracy who dedicated their time to either proofread individual chapters or provided valuable comments on the various pieces of work that were presented at internal seminars and annual reviews. Dr Laura Bojke and Professor Gary Ford graciously agreed to be my external examiners, and their input during the final stage of my thesis was invaluable.

I am indebted to all the individuals who volunteered to proofread individual chapters of my thesis, special thanks to Wissam Khalil, Dr. Robert Dunn, Kathleen Charest, Marianthi Dunn and Dr. Randolph Bruno.

Last but not least, I would like to thank my parents and in particular my late father, for their continuous encouragement and support along my life; my husband Wissam since he has been always there to support me unconditionally and somehow it feels like this doctoral research was a joint effort. They are the most important people in my life and I dedicate this thesis to them.

Table of Contents

List of Tables

List of Figures

Acronyms

BNF	British National Formulary
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CEAF	Cost-effectiveness acceptability frontier
CHD	Coronary heart disease
CI	Confidence interval
CRD	Centre for Reviews and Dissemination, University of York
CUA	Cost-utility analysis
CVD	Cardiovascular disease
DAM	Decision analytic modelling
DES	Discrete event simulation
DSA	Deterministic sensitivity analysis
EVPI	Expected value of perfect information
EVPPI	Expected value of partial perfect information
ICER	Incremental cost-effectiveness ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ICER	Incremental cost effectiveness ratio
ISM	Individual Sampling Model
mRS	Modified Rankin Scale
NHS	National Health Service
NHS EED	National Health Service Economic Evaluation Database
NICE	National Institute for Health and Care Excellence

NHB	Net health benefit
PRISMA	Systematic reviews and meta-analysis
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life year
RR	Risk ratio
RCT	Randomised controlled trial
SA	Sensitivity analysis
ScHARR	School of Health and Related Research, Sheffield
SMDM	Society for Medical Decision Making

Glossary

Cost-effectiveness analysis

CEA compares the relative costs and outcomes (consequences or effects) of different courses of action. The differences between the costs and consequences of the various alternatives are presented in the form of a ratio or the cost per unit of health effect. Cost-effectiveness results are usually presented as incremental costs, incremental effects and the incremental cost-effectiveness ratio (ICER)[1]. For example, given two options x and y, first their respective costs and effects are estimated; then the difference in costs and difference in effects are calculated; then the existence of dominance is checked (the intervention costs less and is at least as effective as the comparator); and, only if dominance is not found, the ICER is estimated as the difference in costs divided by the difference in effects:

$$ICER = \frac{Cost_x - Cost_y}{Effect_x - Effect_y} = \frac{\Delta \text{Cost}}{\Delta \text{Effect}}$$

Consequences or the effects of alternative interventions are estimated using different types of measurement units, for example, the effect of a blood pressure intervention can be measured by the cost per 1 mmHg reduction in systolic blood pressure. The choice of the measure of effect used in cost-effectiveness analysis is limitless, examples include cost per case detected, cost per case prevented, and life-years gained[1, 2].

Cost-utility analysis

A broader measure of the effect of a health care intervention is utility. Utility is viewed as a useful measure because it allows for health-related quality of life adjustments to a given set of treatment outcomes and at the same time provides a generic outcome measure for the comparison of costs and outcomes[2].

CUA allows the comparison of different health outcomes (prolongation of life, prevention of blindness or relief of suffering) by measuring them in terms of a single unit. This is most commonly the quality-adjusted life-year (QALY) although other generic outcome measures including the disability-adjusted life-year (DALY) and the healthy-years equivalent (HYE) have been proposed[2]. In general terms, any state of health or disability is assigned a utility on a scale ranging from 0 (equivalent to immediate death) to 1 (a state of perfect health). The outcome of any health intervention can then be calculated as the product of the increase in utility that it may cause and the time in years over which it may be enjoyed. When allocating scarce resources, those interventions that are expected to produce fewer QALYs for any given cost are given a lower priority and vice-versa.

The ability to compare directly costs of different health outcomes in monetary terms is attractive to the decision-maker. CUA has been at times termed as controversial because it is difficult to put a value on health status or on an improvement in health status as perceived by different individuals or societies. However, CUA has the advantage of considering individual preferences.

Cost-benefit analysis

"Cost-benefit analysis purports to be a way of deciding what society prefers. Where only one option can be chosen from a series of options, CBA should inform the decision maker as to which option is socially most preferred" (Dasgupta and Pearce[3], 1978. p. 46). Key to this definition is the identification of, measurement and valuation of benefits and costs arising from a change in the provision of a service or a good.

While economic evaluation is about assessing costs and consequences of competing uses of scarce resources, a new bridge, a new intervention, or a new drug, at root of applied CBA is the theory of welfare economics. Welfare economics allows the development of a social ordering where trade-offs are made between rankings of '*better than*', '*worse than*' or '*equally as good as*'.

The trade-offs that people make as they choose less of one good and substitute more of some other good reveal something about the value people place on these goods. Value measures based upon such substitutability are calculated in a number of ways, including the willingness to pay (WTP) and willingness to accept (WTA) measures.

CBA place monetary values on the gains and losses to those affected by a change in the level of provision of a good for which there is often no market, for example, health care. It allows the calculation of net gain or loss from a policy change, and determination of whether the change is potentially Pareto-improving[4] (meaning that the gainers from the change could hypothetically compensate the losers from the change).

Cost-effectiveness analyses and cost-utility analyses are techniques that aim to determine the best allocation of an existing budget[2]. However, CEA or CUA are not the appropriate instruments to give answer to questions, for example, what is the underlying value to society of gaining additional QALYs or how much in total is it worth spending on health care rather than on other social objectives (for example, education or defence).

CBA allows answering, for example, if a procedure reduces mortality, whether the monetary value of each death averted (benefits) was greater or less than the costs of obtaining these benefits. In addition, CBA, by having a common denominator expressed in monetary terms, permits comparisons not just within the health care budget but across different areas of spending[1].

Net monetary benefit

Net monetary benefit (NMB) of an intervention is calculated as the total health effects, for example, quality-adjusted life years (QALYs), multiplied by the willingness to pay (WTP) for a QALY minus the total costs of the intervention:

NMB = QALY*WTP - C

Expected net monetary benefit is defined as the mean of the net monetary benefits across all model iterations.

DECLARATION

The work presented in this thesis is the result of original research carried out by the author Maria Cristina Peñaloza-Ramos. During the course of the postgraduate study within the Health Economics Unit, University of Birmingham, the following articles and conference abstracts were accepted for publication and or presentation at conferences. Where listed, the secondary authors also advised on study design, data analysis and paper editing.

Publications

Chapter 2

Peñaloza Ramos MC, Barton P, Jowett S and Sutton AJ. A Systematic Review of Research Guidelines in Decision analytic Modeling. Value in Health. 2015; 18: 512-29

Chapter 3

Peñaloza-Ramos MC, Barton P, Jowett S and Sutton AJ. Do economic evaluations in primary prevention of cardiovascular disease conform to good practice guidelines? A systematic review. MDM Policy & Practice 2016; 1:1–15

Chapter 5

Peñaloza-Ramos MC, Jowett S, Sutton AJ, McManus R, Barton P. The importance of model structure in the cost-effectiveness analysis of primary care interventions for the management of hypertension. Value in Health. 2017; (In press)

Also, during the period of postgraduate study, the following papers were published and used as case studies in this thesis:

- Penaloza-Ramos MC, Jowett S, Mant J, et al. Cost-effectiveness of self-management of blood pressure in hypertensive patients over 70 years with suboptimal control and established cardiovascular disease or additional cardiovascular risk diseases (TASMIN-SR). Eur J Prev Cardiol. 2016 Jun;23(9):902-12. doi: 10.1177/2047487315618784. Epub 2015 Nov 24.
- Penaloza-Ramos MC, Sheppard JP, Jowett S, et al. Cost-Effectiveness of Optimizing Acute Stroke Care Services for Thrombolysis. *Stroke*, 45(2), pp. 553-562. ISSN (print) 0039-2499.

Conference presentations

HESG: Health Economics Study Group

- Paper accepted for oral presentation at the HESG, June 2013: Penaloza MC, Barton P, Jowett S, Sutton AJ. “Best practice in decision analytic modelling: a systematic review of good practice guidelines”

- Poster presentation at the HESG conference 25-27 June 2014, Glasgow: “The adherence to good practice in decision analytic modelling: a review of literature in primary and secondary prevention of cardiovascular disease”

CHAPTER 1. INTRODUCTION

Economic evaluation in health care has been increasingly used to allocate scarce health resources. The financial resources available for the provision of health care interventions are never sufficient to fund all health care needs, leaving those who plan, provide, or pay for health care services with the challenge of choosing how best to allocate scarce resources[1].

Decision Analytic Modelling (DAM) has been increasingly used in economic evaluation of health care to aid decision making, which has led to greater scrutiny of the methods used. The increasing use of DAM requires the use of sound analytic methods and consideration of the requirements of good practice.

Stroke is the fourth single largest cause of death in the UK and second in the world[5]. By 2020, cardiovascular disease is expected to become the leading cause of death and disability worldwide[6]. Hypertension is a leading risk factor for cardiovascular mortality and morbidity worldwide[7, 8].

This chapter introduces the concepts relevant to the development of this research work, specifies the aims and objectives, and outlines the structure of the thesis.

1.1. Economic evaluation

Economic evaluation[2, 9] offers a coherent and theoretically based approach to identifying, measuring and valuing resource use, costs and outcomes in a health care setting and handling uncertainty[1]. Furthermore, economic evaluation seeks to identify criteria useful in deciding between the competing uses of scarce resources[2]. In other words, it is not possible to establish the economic value of an intervention unless its costs and outcomes are compared with at least one other alternative option[2].

There are two main features that characterise an economic evaluation[2]: i) it considers both costs and consequences; and ii) choices: resource scarcity and the inability to produce all desired outputs necessitates that choices are made between competing alternatives in the allocation of resources. These two characteristics define economic evaluation as "the comparative analysis of alternative course of action in terms of both their costs and consequences"[2].

In economic evaluation the measurement of different types of costs is made in monetary units. Costs may be divided into direct medical costs (costs to the NHS), direct non-medical costs (family expenditure, social services) and indirect costs or productivity costs (changes associated with treatment such as time off work, earlier return to work).[2] The measurement of consequences in economic evaluation may differ considerably due to the nature of what is identified as an outcome[2]. Methods of economic evaluation currently in use include cost-effectiveness analysis (CEA), cost-benefit analysis (CBA) and cost-utility analysis (CUA). These are terms which should be understood, and it is important to realise that methods of economic evaluation are only an adjunct to decision making.

1.2. Decision analytic modelling in health care

Economic evaluations are often conducted alongside randomised controlled trials (RCTs), using patient level data describing costs and outcomes to estimate the cost-effectiveness of a specific health care technology or intervention. However, it has been argued that cost-effectiveness analysis based only on data from clinical trials does not provide a sufficient basis to inform regulatory or reimbursement decisions[10]. The design of clinical trials may not compare all the available options, or provide evidence on all relevant inputs or follow up patients over a period long enough to capture differences in economic outcomes. In addition, by relying on data from a single clinical trial, other sources of data, such as other trials or results of meta-analyses and data from observational studies may be ignored. Under these circumstances, DAM provides an alternative framework for economic evaluation[10].

Since the 1980s, DAM has been widely used in health care to synthesise clinical and economic evidence and to inform resource allocation decisions for the purpose of allowing scarce health care resources to be allocated more efficiently[9]. In the UK, the National Institute for Health and Care Excellence (NICE) recommends using the results of decision analysis as a basis for estimating cost-effectiveness and informing the allocation of health care resources[9, 11].

Decision analytic modelling has been defined as a systematic approach to the analysis of a decision problem under uncertainty[12]. In this thesis DAM is defined as a method that "uses

mathematical relationships to define a series of possible consequences that would flow from a set of alternative options being evaluated"[9] (p.6).

It has been emphasised that, when selecting the appropriate approach to DAM, care should be taken over identifying an appropriate model structure for the clinical question that is being considered[13, 14]. For example, cohort models examine the proportions of the population undergoing different events associated with costs and effects whilst patient or individual level models sample individuals with specific attributes and follow their progress over time[14]. In practice, model structure has usually been defined after considering the relationship between inputs and outputs required by the decision maker. The most common types of cohort models are decision trees and Markov models[9, 13, 14].

The decision tree is considered to have the simplest of the structures. Patient pathways are shown explicitly on a decision tree with associated probabilities and outcome measures. From Barton et al[13], "if the time frame is short and if the mortality of patients does not differ across strategies, a simple decision tree is usually appropriate" (p.111). Decision trees are particularly suited for acute care problems, once-only diseases and short-term diagnostic or screening decisions.

Markov models are recursive (repetitive) decision trees that are used for modelling conditions that have events that may occur repeatedly over time or for modelling predictable events that occur over time (screening for disease at fixed intervals). Markov models are suitable to handle the added complexity of modelling options such as the representation of progression of chronic disease, recurrent events and assessment of long term costs and effects. Some of the commonly identified limitations of Markov models

include the lack of memory with no account taken of history and the fact that they assume a uniform population and equal and constant risk. These limitations may be overcome by using a larger number of states or alternatively, by using other types of method (individual sampling model - ISM, discrete event simulation - DES)[14-16].

When it is important that a model captures interaction between individuals such as in the case of infectious diseases or models that focus on constraints on resources – discrete event simulation (DES) or system dynamic models are the recommended strategies[14, 15]. DES allows the representation of each individual's history and the interaction between specific individuals.

The ISM is used to represent individuals without interactions; the ISM tracks specific individuals with potentially heterogeneous characteristics that affect their progression through the model[14]. One of the key advantages of this type of model is that it allows modelling multiple co-morbidities which depend on multiple attributes or covariates[14, 15].

The growing use of modelling in economic evaluations, and increased need for the scrutiny of the methods used, has led to the emergence of best practice guidelines in the literature. Since 1985, guidelines for good practice in DAM have been developed; however, to date, there is no agreed standard on what constitutes a good model or how models should be formally assessed[17]. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Forces have provided leadership in producing and updating sets of guidelines[16, 18-25] while other authors have made great efforts to synthesise the available ones[17, 26].

To greater or less extent, DAM guidelines have all been concerned with some common aspects in the modelling process including the need to identify and synthesise evidence, consider all relevant comparators, adopt an appropriate time horizon and reflect uncertainty in evidence when presenting results. Each of these factors is considered in turn below.

1.2.1. Identifying and synthesising evidence

Model-based economic evaluations are expected to use all relevant and available sources of evidence including parameters relating to the effectiveness of interventions, resource use and utilities. Evidence cannot usually be found in a single source and instead needs to be drawn from a range of sources. Under these conditions, it has been suggested that a framework be adopted to ensure that all available evidence is properly used in characterising the decision problem[2, 9].

1.2.2. Inclusion of relevant comparators

An economic evaluation is expected to consider all relevant comparators related to different sequences of treatments or interventions. This requirement poses a challenge to researchers in terms of bringing together data from several clinical studies using appropriate statistical synthesis methods (in consideration of the fact that clinical trials will not always compare all the relevant alternatives)[9].

1.2.3. Time horizon

An economic evaluation should adopt a time horizon that is sufficiently long to reflect all the key differences in costs and outcomes between the alternative treatments being compared. For many interventions with a potential mortality effect and where survival curves are required to estimate life expectancy, extended time horizons (sometimes lifetime) are needed[9]. However, because the time horizon from trials usually does not reflect the full duration of the impact of interventions, decision modelling is needed as a framework to extend the results from a short-term trial over a longer time horizon[9].

1.2.4. Uncertainty

The results from an economic evaluation should indicate how uncertainty in the available evidence used to parameterise the model translates into decision uncertainty; this includes indicating the probability that a given decision is the correct one[9]. These results are most helpful when they are unbiased and the uncertainty about the estimated costs and outcomes is properly specified[27, 28]. The main sources of uncertainty in model predictions are related to the model input values and model structure[20, 28-30].

In order to quantify input uncertainty (which can lead to uncertainty in the model output), health economists use two main methods. Deterministic sensitivity analysis (DSA) which consists of varying manually individual input parameters to test the sensitivity of the model's results to specific parameters or sets of parameters[20], and probabilistic sensitivity analysis (PSA) where probability distributions are specified for the true values of the inputs

and then these distributions are propagated through the model using Monte Carlo sampling [20, 31-33]. In recent years, expected value of perfect information (EVPI) and expected value of perfect partial information (EVPPI) analysis have been increasingly used as an extension of PSA. These provide important information to policy makers on the consequences of adopting the wrong treatment strategy[34] - the decision to adopt and reimburse the strategy with the highest expected net monetary benefit is based on the information available at the time the decision is made with its accompanying uncertainty, and, as long as there is uncertainty, there will always be a chance the wrong decision is made[34].

Representing uncertainty in the model structure (or about the costs and health effects of the various decision options) is difficult since it requires judgements about the ability of a model to represent a complex real life decision problem faithfully[28]. Uncertainty related to a model structure is often difficult to handle since it requires judgements about the ability of a model to represent a complex real life decision problem faithfully[28].

1.3. Cardiovascular disease and decision analytic modelling

A disease area where DAM is commonly used to extend the results of clinical trials over a longer time horizon is cardiovascular disease. Cardiovascular disease is the main cause of death worldwide[35]. Self-management of hypertension can lead to significant reductions in blood pressure, thereby reducing the risk of cardiovascular disease[36, 37]. Cerebrovascular accident (commonly known as stroke) is the fourth single largest cause of death in the UK

and second in the world[5]. Treatment with thrombolytic therapy in patients presenting to hospital with a recent acute stroke increases the proportion of patients who are free of disability[38].

Due to the high morbidity and mortality burden of CVD and its equally important economic impact at the level of health systems and families, this is a clinical area where a large number of model-based economic evaluations covering both primary prevention of cardiovascular disease and acute care interventions for stroke have been undertaken[37, 39, 40].

1.4. Aim and objectives of this doctoral research

The increasing use of DAM in the economic evaluation of health care interventions requires the use of sound analytic methods and consideration of the requirements of good practice. Furthermore, there is a need to clarify what is understood by good practice in DAM, how existing DAM guidelines have been used, and most importantly, to what extent DAM guidance is lacking.

Which elements of the guidelines pose the greater challenges for modellers or correspond to deviances from guidelines in current practice are elements that need to be identified and investigated. To date few attempts have been pursued in this line of research.

It has been argued that alternative model structures can lead to variations in model predictions and that inappropriate model structures may lead to poorly informed policy decisions, resulting in the inefficient allocation of scarce resources. Therefore, assessing

the extent to which model predictions are influenced by choices made during the model development process is of utmost importance.

The analysis of uncertainty and structural uncertainty still seem to pose challenges, starting by the many issues that have been defined as structural uncertainty and the methods available to pursue such an assessment. No previous research has attempted to assess the extent to which structural uncertainty is considered as part of current practice in DAM.

Therefore, the overall aim of this thesis is to examine the contemporary understanding of 'good practice' in DAM, to assess the extent to which economic evaluations in primary prevention of cardiovascular disease and acute care interventions for stroke have adhered to good practice guidelines and to examine structural uncertainty arising from the choice of a model structure in applied studies.

The specific objectives are to:

- identify and critically assess good practice DAM guidelines, and highlight areas in which guidelines have failed to provide recommendations
- identify the extent to which recent model-based economic evaluations of interventions focused on lowering the blood pressure of patients with hypertension conform to published guidelines for DAM in health care
- assess the structural uncertainty surrounding the choice of model structure in previously published cost-effectiveness analyses in cardiovascular disease, and thereby gain insights into the optimum model structure in this setting

1.5. Thesis structure

Following this introductory chapter, Chapter 2 presents the results of a systematic review which aimed to identify and critically assess the best practice guidelines available to researchers undertaking DAM. Chapter 3 identifies and provides evidence on the adherence to best practice guidelines of contemporary economic evaluations using DAM in primary prevention of cardiovascular disease, with particular emphasis on the model structure and the assessment of structural uncertainty. Chapter 4 outlines the ways in which structural uncertainty has been described and understood within the decision analytic modelling process. Chapters 5 and 6 examine the impact of model structure on cost-effectiveness results through the use of case-studies:

The case study in Chapter 5 uses a model based economic evaluation of a primary prevention intervention aiming to reduce the risk of cardiovascular disease. The second case study in Chapter 6 explores whether the cost-effectiveness results for an acute care intervention after stroke would have been different if an alternative model structure had been used or if other elements pertaining to structural uncertainty had been explicitly considered.

Chapter 7 brings the study main results together, specifically it highlights the key findings of this study, interprets and discusses the results in light of methodological and theoretical strengths and weaknesses, and formulates recommendations for further research. Chapter 8 draws conclusions.

CHAPTER 2. A SYSTEMATIC REVIEW OF RESEARCH GUIDELINES IN DECISION ANALYTIC MODELLING

2.1. Introduction

Decision analytic modelling (DAM) in health care has been widely used to synthesise clinical and economic evidence and to inform resource allocation decisions for the purpose of allowing scarce health care resources to be allocated more efficiently[2]. In simple terms, in DAM, a model is structured to represent clinical pathways to examine whether an intervention, compared for example to current practice, is cost effective[10]. Building a model requires consideration of important elements including the complexity of the clinical area and the available evidence related to the disease, as well as other issues such as the scope or boundaries of the model, the appropriate time horizon, the perspective of the analysis, the availability of data and a formal synthesis of evidence within the model[10, 41]. The increasing use of DAM in the economic evaluation of health care interventions and health technology assessments (HTAs) requires the use of sound analytic methods and consideration of the requirements of good practice.

The aim of this chapter is twofold: to perform a review to identify and critically assess good practice guidelines, highlighting areas in which these have failed to provide recommendations, with emphasis being given to more recent developments, and to develop a practical framework to assess adherence to guidelines in DAM.

2.2. Methods

A systematic review of articles written in English was undertaken with the aim of identifying published guidelines on DAM in healthcare. The following types of studies were included: guidelines for DAM or HTA and other published articles on good practice in DAM. On the basis of an assessment of their title and abstract (if available), papers were deemed potentially relevant for inclusion if they: 1) provided general guidance in DAM for health care or HTA; or 2) provided general criteria against which to assess good practice in DAM (e.g., a checklist).

For the purpose of this review the following were excluded: 1) trials or economic evaluations alongside clinical trials; 2) other non-DAM studies including statistical or econometric models; and 3) conference abstracts or other non-DAM papers.

2.2.1. Search Strategy

An initial exploratory approach was utilized which employed search terms used in a previous systematic review[26] and this helped inform the final search terms identified for this review (see Appendix 1, Appendix 2). Further relevant literature was obtained by checking the references of the included articles.

The following bibliographic databases were searched: The Cochrane Library, Cochrane Methodology Register (CMR), Cochrane Health Technology Assessments, NHS Economic Evaluation Database, Embase, and MEDLINE via Ovid. To avoid duplication, the PROSPERO

database of prospectively registered systematic reviews in health and social care was searched for any existing or ongoing reviews that addressed similar topics, and none were identified. This review covered the period from January 1990 to March 2014, a period that has seen the development of guidelines for DAM in healthcare and the consolidation of good practice guidelines.

2.2.2. Data extraction

All studies were manually searched and data extracted from each paper using a data extraction form developed to retrieve and organise information from each paper on the basis of its main topic, model structure, model uncertainty, model transparency, and validation. The data extraction form was developed through a process in which the content of the articles informed the "areas" that the data were extracted under. This approach was used to ensure that the review did not miss any information related to the model-building process. Data were extracted as free text and in the form of a 'yes/no' response.

2.3. Results

Titles and abstracts (if available) were screened using the inclusion and exclusion criteria to identify potentially relevant articles. The database search yielded 4,150 studies and 11 additional studies were identified through other sources totalling 4,161 studies left for screening. A total of 3,976 records were excluded because they were economic evaluations alongside clinical trials, other non-DAM studies, conference abstracts or duplicates. One-

hundred and seventy eight full-text articles were assessed for eligibility, of which 145 were excluded on the basis of not providing criteria against which to assess good practice in DAM.

In total, thirty-tree studies, corresponding to general guidance or elements of good practice in DAM, were included in this review. A flow chart showing the study selection process is shown in Figure 2-1. The methodological quality of the articles included in this study was not comprehensively assessed using formal checklists because of the diversity of the literature included and the nature of the review.

Of the 33 articles included in this review, 15 studies provided general guidelines for good practice or criteria in the form of a checklist. Eighteen articles were focused on particular elements of good practice, for example, model structure or uncertainty, or model transparency and validation.

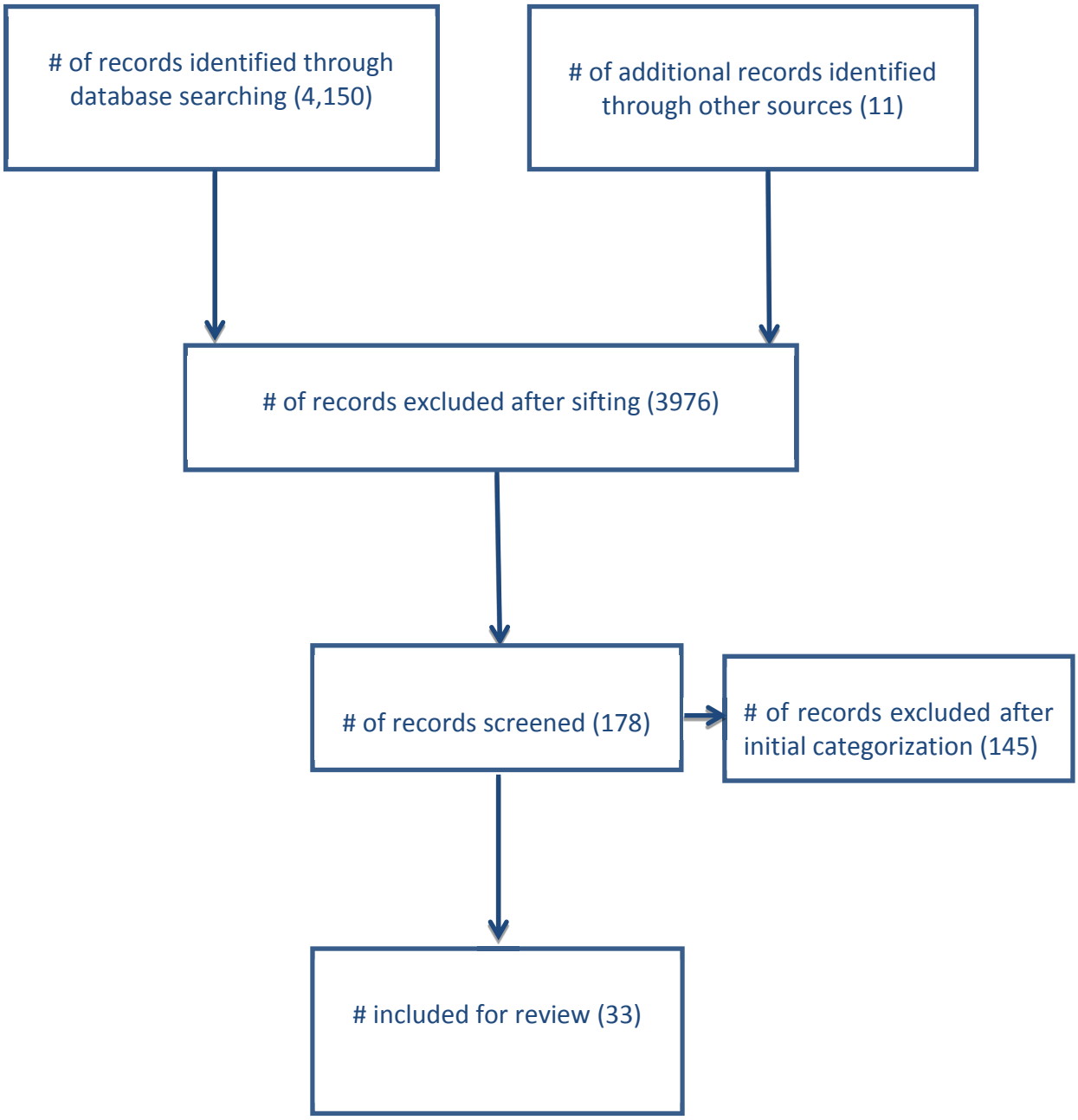


Figure 2-1 Flow Chart of DAM guidelines using the PRISMA statement

2.3.1. Elements of good practice for DAM

Fifteen studies provided general guidelines for good practice; 8 of the 15 guidelines were released before 2012[17, 18, 26, 42-47], with the remainder making up the ISPOR-SMDM[16, 20-25] set of guidelines. Table 2-1 presents a breakdown of the elements of good practice based on the main themes of the guidance, that is, model structure, identifying and synthesizing evidence, and model validity. These studies provided a source of complete information on the various stages that need to be covered in DAM. Some of the studies constituted a list of topics that need to be checked, or questions that modellers need to answer prior to constructing a model. Most commonly, guidelines have been presented as a series of good practice statements, starting with Weinstein et al[18], then Philips et al[17, 26] and more recently ISPOR-SMDM[16, 20-25]. DAM guidelines provide a set of principles that might lead, for example, to an appropriate model structure or else indicators of areas that require consideration in decision modelling[17].

Table 2-1 General guidelines

Paper ID		**Model Structure**								
Author(s) / year	**Topic**	**Decision problem/ objective**	**Scope / analytic perspective**	**Rationale for model structure**	**Comparators**	**Model type**	**Time horizon**	**Disease states/ time/events**	**Cycle length**	**Model Parsimony**
Sonnenberg et al[46]	Framework to judge adequacy	✓	✓	✓	✓	✓	✓	✓		✓
Sculpher et al[45]	Framework for validity and quality	✓		✓	✓		✓	✓	✓	
Soto[47]	Checklist for decision analytic modelling	✓	✓	✓	✓		✓			
Weinstein et al[18]	Good modelling practice		✓	✓			✓	✓	✓	✓
Philips et al[26]	General guidelines	✓	✓	✓	✓	✓	✓	✓	✓	✓
Philips et al[17]	Framework for quality assessment	✓	✓	✓	✓	✓	✓	✓	✓	

Paper ID		Model Structure								
Author(s) / year	**Topic**	**Decision problem/ objective**	**Scope / analytic perspective**	**Rationale for model structure**	**Comparators**	**Model type**	**Time horizon**	**Disease states/ time/events**	**Cycle length**	**Model Parsimony**
HTA, Canada [42]	General guidelines in Canada	✓	✓		✓	✓	✓			
Karnon et al [44]	Modelling issues						✓			✓
Earnshaw et al [43]	Guidelines for economic evaluation			✓			✓			
ISPOR-SMDM [16, 20-25]	Good research practice	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Ticks indicate the areas for which the different studies proposed statements of good practice or guidelines

Table 2-1 General guidelines (continuation)

Paper ID		Identifying and synthesizing evidence													
Author(s) / year	Topic	Baseline data	Bias in parameter estimates	Costs	Data identification	Data incorporation	Data modelling	Heterogeneity	Methodological	Parameter	Parameter estimates	Stochastic	Structural	Treatment effects	Utilities
Sonnenberg et al[46]	Framework				✓		✓	✓		✓	✓				✓
Sculpher et al[45]	Framework				✓	✓									
Soto[47]	Checklist			✓	✓	✓		✓	✓	✓			✓		✓
Weinstein et al[18]	Good modelling practice	✓				✓	✓	✓							
Philips et al[26]	General guidelines	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Philips et al[17]	Framework	✓		✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
HTA, Canada[42]	General guidelines			✓	✓			✓		✓	✓				✓
Karnon et al[44]	Modelling issues							✓							
Earnshaw et al[43]	Guidelines	✓	✓		✓					✓			✓	✓	
ISPOR-SMDM[16, 20-25]	Good research practices	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Ticks indicate the areas for which the different studies proposed statements of good practice or guidelines

Table 2-1 General guidelines (continuation)

Paper ID		Model validity			
Author(s) / year	Topic	Face/Internal/technical validity, verification or consistency	Cross validity	External validation	Predictive validation
Sonnenberg et al[46]	Framework to judge adequacy	✓			
Sculpher et al[45]	Framework for validity and quality	✓	✓		
Soto[47]	Checklist for decision analytic modelling			✓	
Weinstein et al [18]	Methodology regarded as good modelling practice	✓	✓	✓	✓
Philips et al[26]	General guidelines	✓		✓	✓
Philips et al[17]	Framework for quality assessment	✓	✓	✓	
HTA, Canada[42]	General guidelines in Canada				
Karnon et al[44]	Modelling issues				
Earnshaw et al[43]	Guidelines for economic evaluation	✓			
ISPOR-SMDM[16, 20-25]	Good research practices	✓	✓	✓	✓

Note: Ticks indicate the areas for which the different studies proposed statements of good practice or guidelines

To inform model construction and increase model credibility and validity, these guidelines provide a set of principles, checklists, or have stated the agreement of a common application[16, 21-23, 25, 44-47]. For example, guidelines have stated that model construction is likely to be influenced by the adoption of simplifying assumptions reflecting issues such as data availability, and that the design of a model should not be driven by the data at hand. Under these circumstances the identification of the explicit characteristics of the disease area that affect model selection, for example, the unit of representation, is considered important[16, 23, 25, 46, 47]. Other aspects in model construction that arise from the application of models to specific groups of patients or specific settings include the scope of the model, the model perspective, choice of model type, choice of utility structure (e.g. quality adjusted utility scale) and the interventions to be included in the model[16, 23, 25, 45-47]. These guidelines identify the characteristics of individuals as a key element aiding the process of model selection, that is, whether a model needs to represent individuals or groups or interactions between individuals[16]. Furthermore, guidelines recommend that 'the appropriate model type is determined by purpose, level of detail and complexity' (p.809)[16], and 'explicit processes' involving expert consultation, influence diagrams or similar should be used to convert the conceptualization of the problem into an appropriate model structure[16].

ISPOR-SMDM[16, 20-25] recognised the difficulty for all models in achieving all the recommended best practice for model validation, that is, face validity, internal validity, cross validity, external validity and predictive validity. Instead of establishing a minimum quality standard, guidelines recommend the adoption of optimal practices that all models

should aim for[22]. Among these, model transparency was identified as a key area of optimal practice that should be achieved by all models and is reflected by providing clear information on how the model was built, that is, describing its structure, parameter values, and assumptions[22].

ISPOR-SMDM[16, 20-25] reiterated statements of good practice emphasizing its appropriate conduct and furthermore establishing grounds for usage, for example, the use of time horizons sufficiently large to capture all health effects and costs relevant to the decision problem in cohort simulations[16, 20-25]; or insisting on the value of model simplicity as long as the model's face validity is not compromised[16].

2.3.2. Model structure

Good practice for selecting a model or the use of alternative model structures was discussed in ISPOR-SMDM[16, 21, 23] and in 4 of the 18 individual articles included in this review[13, 14, 48, 49]. Model structure should be considered in the initial stages in the process of model building (Table 2-2). Guidelines have suggested that before model building, researchers should identify the problem and objective of the project, the analytical perspective of the model, the scope, the rationale for selecting the particular structure, the target population, and the strategies and comparators and then give justification for choosing the model type, the time horizon and the disease states[13, 14, 16, 49]. These initial steps are important and will have important implications for the model structure, data requirements, and the reporting of the final results obtained from the model.

Table 2-2 Model structure

Author(s)	Area of guidelines	Criteria for selecting a modelling approach	Rationale for structure	Model-based evaluation	Parsimony	Key recommendations
Roberts[16]	Choice of model structure	Justified in line with policy context and aiming to inform resource allocation	Whether a model represents individuals or groups or interactions between individuals	Expert consultation and conceptualization in two stages: problem conceptualization and model conceptualization	Model simplicity however, preserving face validity	Early specification of the decision problem, modelling objectives and valuing outcomes will improve model efficiency (expert consultation, influence diagrams, concept mapping)
Siebert[25]	Structures and model complexity	Whether decision-problem requires time-dependent parameters or time to an event or repeated events	Markov models can handle memory by creating states that include history; but then model complexity	STM are comprehensive and powerful tools to guide decision in health care	Decision tree has limited ability to reflect time; then STM seems the simplest option	Markov model if decision problem has a manageable number of health states; if not, use an individual-level state-transition model (STM)

Author(s)	Area of guidelines	Criteria for selecting a modelling approach	Rationale for structure	Model-based evaluation	Parsimony	Key recommendations
Karnon[23]	Structures and model complexity	DES justified on model flexibility	Constraint resources; patient's interactions; time/dependencies	Value of DES to inform health care decisions; flexible and able to represent complex behaviour and interactions between individuals	Easy representation of complex systems	A good choice if individuals are subject to multiple or competing risks
Bentley et al[48]	Structures and model complexity	Subsequent event risk dependent on prior event history; simulation of event or disease risks over time; improving validity	Recurrent events and time dependency	The ability to incorporate past history is restricted to the number of model health states	Trade-off between model bias and model complexity	Failing to incorporate prior event history would overestimate the impact of an intervention; add dependency by using states that track event history; make event risks dependent on this history
Brennan et al[14]	Choice of model structure	Needs to be justified	Interactions between individuals; choice from taxonomy of model structures	Comparison of health technologies and synthesising evidence on costs and benefits	Simplest model that addresses objectives and structure of disease and treatment	Responsibility of developers to select the most appropriate modelling approach; taxonomy grid is a guidance

Author(s)	Area of guidelines	Criteria for selecting a modelling approach	Rationale for structure	Model-based evaluation	Parsimony	Key recommendations
Barton et al[13]	Choice of model structure	Needs to be justified	Interactions versus not interactions between individuals	Two distinct and independent aspects: mean estimate of cost-effectiveness and exploration of uncertainty in the model inputs	Simplicity (relates to the size of the model) is seen as an advantage	Check dependence or independence among individuals; model simplicity is an advantage; model validation; challenge the need for a complex model
Karnon[49]	Structures and model complexity	Assess relative advantages of alternatives according to areas of treatment	Model flexibility and analytic input (complexity of model building)	Choice depend on flexibility vs. time availability; there may be circumstances where DES provides a more accurate representation of the data	A simpler model was the optimal technique as compared to a complex DES model	Results of different models (Markov or DES) may produce likely results; model flexibility (DES) may be outweighed by greater time required to evaluate its results

Note: DES= Discrete Event Simulation; STM= State Transition Model

Guidelines for conceptualizing a model's structure have evolved from statements of general principles, for example by stating that the structure of a model should be consistent with the theory of the health condition and the available evidence[18], to more systematic processes describing how to select a model from competing alternatives[13, 14, 48, 49]. ISPOR-SMDM[16, 21, 23] described the development and construction of a model as a process that starts with *model conceptualization*[16] which consists of a two-step process: problem conceptualization and model conceptualization. Problem conceptualization in this context is transforming knowledge of the healthcare process into a representation of the decision problem. Model conceptualization is the representation of the components of the problem using a particular decision analytic method (Table 2-2). The nature of the problem and the project objectives are decisive in selecting the structure of a model. Furthermore, ISPOR-SMDM[16, 21, 23] has suggested that the early specification of the decision problem and project objectives will improve model-building and the structure of the model (data requirements, analytic strategy and reporting)[16].

The importance of the choice of model structure stems from the fact that alternative model structures can have an impact on model results and thereby affect decision making[13, 14, 16]. The appropriate model type should be determined according to its purpose, level of detail required, and complexity[16]. As previously demonstrated, guidelines aid the selection of an appropriate modelling approach by providing an overview of competing approaches and highlighting examples of where each alternative technique should be employed[13, 14, 16]. The most common issues affecting a model's selection are[16, 21, 23]: 1) the unit of representation, does a model represent individuals or groups? This affects the level of detail required for

the variables that predict outcomes[16]; 2) whether the decision problem requires the modelling of the effect of an intervention on disease spread or use of limited resources, in other words, if interactions among individuals need to be represented, then models designed for patient interactions are necessary[16]; and 3) the time horizon is dictated by the problem scope. For example, decision trees are considered appropriate for models with very short time horizons, while longer horizons require the use of models such as State-Transition (for example a Markov) or Discrete Event Simulation DES[16].

Among the most difficult stages in the conceptualization of a model is the selection of the appropriate level of model complexity, as very simple models may lose face validity if they do not incorporate all the aspects that experts feel are required; whereas complex models may be difficult to build, debug, analyse, understand and communicate[16]. Guidelines have generally supported the choice of simpler models as 'model simplicity is desirable for transparency, ease of analysis, validation and description'[16], while at the same time it is recognised that under certain circumstances, more complex models may be needed. Consensus-based guidelines, stating common grounds for the application of more complex model structures, have been developed, that is, state-transition models, discrete event simulation and dynamic transmission models[23-25].

2.3.3. Model uncertainty and synthesis of information

ISPOR-SMDM[20] and an additional eleven individual papers[27, 28, 31, 50-57] provided methodological guidelines for the analysis of model uncertainty (methodological, structural, parameter, heterogeneity and stochastic), and the use of sensitivity analysis.

Step by step guidelines and checklists have been developed (Table 2-3) to aid researchers in accounting for uncertainty or to identify how uncertainty is incorporated in a model or to address special model circumstances, for example where the evidence is insufficient to give a clear representation of the uncertainty through parameter distributions[20, 31]. The view presented by some of the studies included in this review is that many published models still fail to account correctly for the major sources of uncertainty, in particular structural uncertainty, indicating that a gap may still exist between techniques, guidelines, and what is done in practice[31, 54].

Table 2-3 Model uncertainty and synthesis of evidence

Author(s)	Area of guideline	General principles	Way of reporting	Methodological issues	Methods /recommendations
Briggs (ISPOR_12) [20]	Point estimate(s) & parameter uncertainty	Responsible reporting; use of terminology; justify its omission; decision maker's role; preferable parameterize uncertainty from structural assumptions if possible	Use tornado diagrams, threshold plots, or statements of threshold parameter to report DSA; describe assumption(s); report uncertainty around calibrated parameter(s); report EVPI if needed	Methodological, structural, patient heterogeneity; parameter uncertainty and stochastic uncertainty	For structural uncertainty, calibration approaches; for parameter uncertainty DSA or PSA; for point and interval estimates use CI or distributions; reflect absence of evidence
Bilcke et al[31]	Uncertainty: Step-by-step guide and checklist	Formulate decision-problem; specify sources of uncertainty; obtain information and evidence; report results; apportion uncertainty to sources	Report choices of normative approach(es); present sources of uncertainty; use distributions; assess the most influential sources of uncertainty; report results of PSA	Methodological, structural and parameter uncertainty	State if there are more than one approach that can be used; use distributions; assess the most influential sources of uncertainty; global sensitivity analysis; PSA

Author(s)	Area of guideline	General principles	Way of reporting	Methodological issues	Methods /recommendations
Jain et al[54]	Sensitivity Analysis	Report all sources of uncertainty; Strengths and limitations of SA should be acknowledge (interactions and correlations between parameters)	If long term analysis is needed, conduct CEA under various time horizons; use for instance, tornado diagrams, or threshold analysis to present results	Methodological, structural, parameter	Model averaging and parameterisation for structural uncertainty; methodological uncertainty can be addressed by providing results for a 'reference case'; DSA or PSA.
Koerkamp et al[55]	Uncertainty and patient heterogeneity	Consider range of assumptions for the natural course of a disease; provide model for every set of assumptions instead of using the single best model; trade-off between the realism of a model and time availability	Use tornado diagrams or threshold plots; describe assumption(s); report uncertainty; if the purpose of the PSA is the acquisition of information to reduce uncertainty, report EVPI	Parameter uncertainty, patient heterogeneity, stochastic uncertainty (first-order uncertainty)	PSA joint uncertainty; parameterisation model structure uncertainty; first-order Monte Carlo analysis for stochastic; DSA for parameter uncertainty; EVPI if needed
Jackson et al[27]	Structural uncertainty	Various sources: statistical models, evidence used, states or clinical events represented, or treatment strategies considered	Should be acknowledged, assessed and reported	Structural uncertainty	Reference case model; assign distributions; use PSA; for non-parameterised uncertainties use global

Author(s)	Area of guideline	General principles	Way of reporting	Methodological issues	Methods /recommendations
					model; if lack of data, elicit distributions
Strong et al[28]	How complex a model should be	Uncertainty in model structure is complex; it involves making judgements about model's ability to accurately represent a decision problem	Most commonly by PSA, however, it will only quantify uncertainty about the costs and consequences; problem when a model lacks accuracy	Uncertainty about the model input values and model structure	To properly represent uncertainty about the costs and outcomes, structural uncertainty must be presented; structural uncertainty measured with model averaging or the discrepancy approach
Bojke et al[51]	Structural uncertainty	Impossible to accurately predict mean costs and outcomes; sources are treatment effects and type of model	Importance of differentiating parameter and structural uncertainty: if uncertainty can be parameterised, then there is parameter uncertainty	Parameter, methodological and structural (little attention given to structural uncertainty)	Model selection (not plausible); model averaging (difficulty determining posterior distributions); parameterising (directly representing uncertainty by adding other 'uncertain' parameters)
Briggs et al[52]	Probabilistic probabilities	If there is a need to specify a distribution over	Has the Dirichlet distribution been specified over	Inconsistencies performing SA if a node has 2 or	Use Dirichlet distribution, a multivariate equivalent of the beta distribution

Author(s)	Area of guideline	General principles	Way of reporting	Methodological issues	Methods /recommendations
	over multiple branches	multiple branches at a chance node	multiple branches at a chance node?	more branches and the sum of the branch probability is different from 1	
Kuntz et al[56]	Patient heterogeneity	Cohorts are defined based on population characteristics; sometimes other characteristics may be overlooked (disease incidence or progression), causing heterogeneity	Heterogeneity bias may be evaluated as a function of 3 parameters: annual probability of developing the disease; RR of disease with vs. without the factor; iii) baseline prevalence of the factor	The assumption that each health state contains a homogenous population group does not always hold: for instance, in the presence of risk factors affecting the risk of developing disease	Adjust by introducing an heterogeneity factor; probability of transitioning to disease dependent on heterogeneity factor; transition probabilities averages to that of the model without adjustment
Briggs et al[53]	Uncertainty	Study designs included were modelling-type based approaches	The majority included some form of sensitivity analysis (one-way sensitivity analysis)	Mainly one-way SA; 5% attempted statistical analysis; 17% failed to provide any	Reference case (comparability of results); potential for ICER to vary; avoid selective comparison; uncertainty; interval estimates; SA; probabilistic nature of reported range; descriptive

Author(s)	Area of guideline	General principles	Way of reporting	Methodological issues	Methods /recommendations
				attempt to quantify uncertainty in their results	statistics; estimate CI; present CEAC
Andronis et al[58]	Sensitivity analysis	DSA requires variables and sources to be justified; for PSA distributions should be placed around all parameters (excluded parameters should be justified)	Repeated analysis should be run using different models and methods where uncertainties exist	Methodological and structural uncertainty	Univariate, multivariate, PSA and DSA; distributions in line with logical bounds; if correlation is expected, use joint distributions (do not assume independence)
Sendi et al[57]	Uncertainty & opportunity costs	Univariate and multivariate SA to assess robustness; however, SA does not inform joint uncertainty	Alternative approaches as a result of the intractability of the ICER: Net Health Benefit (NHB) and CEAC	ICER difficulty apparent if a distribution extends over more than one quadrant	NHB, however, a problem if lambda is unknown; CEAC, however, same problem with lambda; uncertainty can be accounted for using Bayesian methods

Note: DSA= Deterministic Sensitivity analysis; PSA= Probabilistic Sensitivity Analysis; EVPI= Expected Value of Perfect Information; CI= Confidence Intervals; SA= Sensitivity Analysis; CEA= Cost-Effectiveness Analysis; RR= Risk Ratio; ICER= Incremental Cost-Effectiveness Ratio; CEAC= Cost-Effectiveness Acceptability Curve; NHB= Net Health Benefit

Assumptions adopted in decision models determine their final structure and can consider the choice of relevant comparators and health states, or available clinical evidence that determines the type of adverse events, duration of treatment effects, time dependency of probabilities and prognostic implications of surrogate end points or the clinical events included[20]. Structural uncertainties arise when these *structural* assumptions are not formally quantified and it is uncertain whether they accurately reflect reality[20]. Current methods for addressing structural uncertainty include scenario analysis (presenting the results under different model structures); model averaging (presenting results of different models using different assumptions and an average across these models); parameterisation of structural uncertainty; and in the absence of data or presence of weak data, expert elicitation to translate expert beliefs into probability distributions[27]. Model structure plays an important role in defining the relationship between inputs and outputs to the point that it has been recognised that structural uncertainty may be at least as important, in terms of its impact on results, as parameter uncertainty[20]. ISPOR-SMDM[20] highlighted the emerging interest in calibration methods as an aid to ensure consistency of inputs and outputs in a model. Calibration is used when data are available to match model outputs rather than model inputs: it is then necessary to determine parameter values which give model results that match the data[20].

Many techniques that aim to capture the various sources of DAM uncertainty have been developed and have evolved. However, there still remains some areas in which more research is needed, such as: accounting for uncertainty surrounding quality of evidence for particular structural aspects; generalisability from one setting to another; and the way

multiple sources of evidence should be combined (heterogeneity of parameter values from different sources)[31]. ISPOR-SMDM[20] proposed the parameterisation of structural uncertainties into a model as an approach to deal with issues around the quality of evidence; however, this approach seems to become complex if a complete redesign/rebuilding of the model is required (nested structures)[20]. Under these circumstances, guidelines have stated that "where it is impossible to perform structural uncertainty analysis, it is important to be aware that this uncertainty may be at least as important as parameter uncertainty" and analysts are encouraged to be explicit about the structural assumptions that might impact their findings and suggest alternative assumptions for future modelling exercises[20].

2.3.4. Model transparency and validation

Four articles discussed methods to assess the consistency or validity of models and model transparency, (Table 2-4)[22, 59-61]. Model transparency reflects the extent to which a model's structure, equations, parameter values and assumptions can be reviewed, and a model is considered transparent if any interested reader with the necessary expertise who wants to evaluate the model is able to reproduce it[22]. Model validation has been recommended to enhance the credibility of models and as an indicator of reliability in practice guidelines[17, 22, 59-61]. Model transparency does not equal the accuracy of a model in making relevant predictions; a transparent model may yield the wrong answer, and vice versa, while a model may be correct and lack transparency. Thus, transparency and validation are both necessary for good practice in modelling[22].

Table 2-4 Model transparency and validation

Author(s	Area of guideline(s)	Methodology	Rationale for model transparency and validation	Best practice	Recommendations
Eddy (ISPOR_12) [22]	Transparency and validation of models	Recommendations on optimal practice	The model's non-technical and technical documentation is made available, written in sufficient detail to enable the reader to evaluate it	Face validity of a model's structure, evidence, problem formulation and results; transparency and validation	Models are instruments to help decision makers answer complex questions; model confidence and credibility is demonstrated by clarity in model structure, equations, parameters, and assumptions, and by subjecting models to tests of validity
Karnon[60]	Model validation	Empirical comparison	Identification of input parameter(s) that produce output that best predict observed data	Probabilistic calibration of models produced improvements in model's accuracy, and reduced uncertainty	Widespread of model calibration (probabilistic calibration); a process of validation against more theoretically grounded approaches is valuable (Bayesian updating approach)
Goldhaber-Fiebert[59]	External model validation	Literature review	Comparing model to independent data not used in the model	Heterogeneity in how results of model evaluation are reported	Evaluation via comparison(s) to independent studies; structured reporting format: empirical study description, baseline characteristics, study protocol, study outcomes, model outcomes and model consistency

Author(s	Area of guideline(s)	Methodology	Rationale for model transparency and validation	Best practice	Recommendations
Kim[62]	Model validation	Use of internal, prospective and external validation	Indication of reliability of assumptions adopted	A model should be generated which fits all available data	Model based on limited data may not be generalisable; uncertainty from model assumptions as important as parameter uncertainty; new model should be generated which fits all available data; model validation should assess: key events, rate of accrual of events and absolute and incremental costs and effects

Validation involves a set of methods for judging the accuracy of models when making predictions. More recent guidelines have used the terms '*model consistency'* or *'model validation'* to refer to five types of model validity:

- face validity (evaluation of model structure, data sources, assumptions and results)
- internal validity (the practical model should behave as the theoretical model predicts)
- cross validity (comparison of results with other models),
- external validity (comparing model results and real-world results)
- predictive validity (comparing model results with prospective observed events)[17, 22].

Principles and methods to enable researchers to assess model validity have been discussed and in some cases demonstrated[22, 60, 61]. However, Kim et al[61] established that health economic models based on limited follow-up data from one source may not be generalisable to longer follow-up periods or other contexts. Furthermore, in addition to the standard considerations of uncertainty about parameter estimates, it is important to assess the implications of model uncertainty on results, in other words, to undertake independent model validation[22].

Best practice recommends that face validity (due to its subjective nature) should be judged by people who have expertise in the problem area, but who are impartial and preferably blinded to the results[22]. Internal validity verifies that mathematical calculations are performed correctly and are consistent with the specification of the model. Methods to

assess internal validity will depend on the model's complexity, but two main stages of internal validity involve the verification of individual equations and their accurate implementation. It should be noted that internal validity does not evaluate the accuracy of a model's predictions[22]. Cross validity involves examining different models and comparing their results to then identify and analyse the causes of differences and similarities in these results. External validation compares the results of a model with actual data; however, the difficulty in identifying 'alternative data' has been noted[22]. Best practice to undertake external validation recommends following a formal process to compare a model's results to actual event data. Guidelines provide awareness of the important limitation that external validation can only address the parts covered by data sources[22]. Predictive validity remains a highly desirable type of independent model validation due to its potential ability to demonstrate the accuracy of the results obtained from the DAM. Its results, however, are potentially limited if there are changes in the design of the study or other factors outside the control of the study design change during the development of the study[22].

Even though the latest guidelines[22] have provided more detailed guidance on how best to ensure model transparency and undertake validity checks, which reflect the value of concise reporting of a model and advocate the quantification of uncertainties arising from differences in assumptions[22], some quandaries seem to prevail. For example, to examine external validity, modellers are advised to use actual event data. However, that same data in many instances will already have been used to parameterise the model – as guidelines suggest that the most representative data sources should be used in developing a model.

2.4. Development of a new decision-analytic modelling framework

This review demonstrates that although guidelines have been developed and are available to aid researchers to inform the results of their studies and, most importantly, to increase the credibility of their results, these guidelines lack practicality due to the extensive amount of information available and its complexity.

The data extraction instrument used in this Chapter was developed through a process in which the content of guidelines informed the areas that the data were extracted under. This is, the data extraction instrument is a reflection of the key areas of interest in the DAM process and consequently, each one of these areas should be considered as an integral element in the model-building process.

This review found that general guidelines concerned with the quality and adequacy of a model structure (see Table 2-1) indicated that the conceptualization of a model should at least comprise of two elements. Firstly, the conceptualization of the decision problem, including the knowledge of the healthcare process that is being represented (analytical perspective, target population, outcomes, comparators and time horizon) and secondly, the conceptualization of the model that matches the characteristics of a model type with the needs of the problem concept (model structure).

Other areas of guidelines identified through the data extraction instrument were synthesis of evidence and model uncertainty (see Table 2-3). In particular, the findings of this review indicate that, model uncertainty seems of utmost relevance as published models still fail to correctly account for the major sources of uncertainty, in particular structural

uncertainty. Finally, another area of relevance identified by this review was model transparency and validation which guidelines have agreed are both necessary and considered good practice in DAM.

Current standards of reporting could be improved if a single, comprehensive, user friendly and practical instrument is made available to direct researchers towards the key elements of good research practice in DAM, which should be assessed and reported to increase the credibility of their results. This single instrument should incorporate, as a minimum, the previously identified five areas of interest in DAM: problem concept, model concept, synthesis of evidence, model uncertainty and model transparency and validation. Furthermore, the availability of a single 'five-dimension' framework would allow modellers and researchers to assess adherence to guidelines in DAM.

2.5. Five-dimension framework

As previously discussed, the five-dimension framework proposed here incorporates and reflects much of the evidence from this systematic review, that is, it has synthesised all contemporary guidelines in a checklist instrument. To ensure its consistency, it adopted the most up to date and agreed guideline statement when components in each dimension were superseded or contradictory. For example, the attributes of good practice as stated by Philips et al[26] indicate that "the appropriate model type will be dictated by the stated decision problem and the choices made regarding the causal relationships within the model". This statement has been replaced by a more systematic process where the

appropriate model type is determined by its purpose, level of detail and complexity (see section 2.3.1). Therefore, the characteristics that affect model selection are the unit of representation, whether or not interactions between individuals are relevant, and the time horizon[13, 14, 16].

The five-dimension framework uses the following five-dimension checklist:

- Dimension 1: Problem concept;
- Dimension 2: Model concept;
- Dimension 3: Synthesis of evidence;
- Dimension 4: Analysis of uncertainty; and
- Dimension 5: Model transparency and validation

This framework does not attempt to replace the guidelines provided by ISPOR-SMDM 2012 or any other contemporary guidelines; instead it attempts to serve as a reference point and checklist for the thorough consultation of good practice guidelines (Table 2-5).

Table 2-5 Framework to assess adherence to good practice guidelines in decision analytic modelling (DAM)

DIMENSION 1: PROBLEM CONCEPT			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
Decision problem	Is there a written statement of the decision problem and scope of the study?		A clear statement of the decision problem and scope would determine the interventions and health outcomes to be measured
	Are the objective(s) of the study and model structure consistent with the stated decision problem and scope?		They are expected to be consistent
Analytical perspective	Has the perspective of the model been stated?		Most common perspectives are: patient, health system (insurer) and society
Target population	Has the target population been identified?		Target population should be defined in terms of features relevant to the decision (geography, patient characteristics, including co-morbid conditions, disease prevalence and stage)
Health outcomes	Are the outcomes of the model stated and consistent with the perspective, scope and overall objective(s) of the model?		Health outcomes may be events, cases of disease, deaths, life-years gained, quality-adjusted life-years, disability-adjusted life-years or other measures important to stakeholders and should be directly relevant to the question being asked

DIMENSION 1: PROBLEM CONCEPT			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
	Have any adverse effects of the intervention(s) been captured?		Interventions may cause negative health consequences that need to be modelled and discussed as part of the study's results. The impact of assumptions regarding adverse effects of interventions should be assessed as part of the structural uncertainty analysis
Comparators	Is there a clear definition of the alternative interventions under evaluation?		Usually the choice of comparators is governed by the scope of the model. Impact of assumptions adopted when deciding upon comparators should be assessed as part of the structural uncertainty analysis
	Is there a discussion around feasible options or justification for the exclusion of feasible options?		The choice of comparators affects results and should be determined by the decision problem, not by data availability. All feasible and practical strategies as determined by the scope of the model should be considered. Constraining the range of strategies should be justified
Time horizon	Is the time horizon of the model justified and sufficient to reflect all important differences between options?		Time horizon of the model should be long enough to capture relevant differences in outcomes across strategies (lifetime). Time horizon is dictated by the problem scope

DIMENSION 2: MODEL CONCEPT			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
Choice of model type	Has the unit of representation been given?		Usually stated in terms of groups or individuals. If groups are being modelled most frequently decision trees, Markov processes or infectious disease models are the correct choice; if individuals are being modelled then the choice is between DES, dynamic transmission models or agent-based models
	Is there a need to model the interaction between individuals in this model? Has this been discussed?		If interactions between individuals is required (when the disease or treatment includes interactions between individuals) then DES, dynamic-transmission, or agent-based models may be the correct choice
	Does the decision problem require a short time horizon?		For simple models or problems (short time horizon, few outcomes) a decision tree may be appropriate; time horizon should be large enough to capture all health effects and costs directed related to the decision problem
	Is it necessary to model time in discrete cycles?		Continuously for Individual STM or in discrete cycles for Markov STM; if the assumption that transition probabilities do not depend on history is not required, then individual STM are an alternative; If disease or treatment process need to be represented as health states, STM are appropriate (Markov type)
	Is there a need to model competition for resources or the development of		If the problem requires the ability of a model to incorporate interactions between individuals and other model parts for example to answer questions on resource allocation i.e., organ

DIMENSION 2: MODEL CONCEPT			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
	waiting lists or queues?		allocation for transplantation, distribution of antiretroviral medications in resource-poor environments, then a DES may be appropriate
	Has a type of model been chosen and discussed?		It is expected that studies report on the reasons for choosing a type of model
Model structure	Has the starting cohort been defined by demographic and clinical characteristics affecting the transition probabilities or state values?		If results may vary by subgroups (age, sex, risk factors) it is advisable to report results for different cohorts
	Has health states and transitions reflecting the biological/theoretical understanding of the disease or condition been modelled?		States should adequately capture the type of intervention (prevention, screening, diagnostics, and treatment) as well as the intervention's benefits and harms. States need to be homogeneous with respect to both observed and unobserved characteristics that affect transition probabilities

DIMENSION 3: SYNTHESIS OF EVIDENCE			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
Data sources	Has transition probabilities and intervention effects been derived from representative data sources for the decision problem?		Most common sources of data include population-based epidemiological studies, control arms of trials or literature
	Has (all) methods and assumptions used to derive transition probabilities and intervention effects been described/justified?		Attention should be given to the use of transition probabilities and rates; conversion of transition probabilities from one time unit to another should be done through rates and never presented as percentages
	Has parameters relating to the effectiveness of interventions derived from observational studies been controlled for confounding?		If results of meta-analyses are used consider how confounders are addressed and the likelihood of increased heterogeneity resulting from residual confounding and from other biases across studies. Efficacy derived from RCT may have to be adjusted for compliance to reflect real-world effectiveness. Effectiveness derived from observational studies must be adjusted for confounding. Adjustment for time-varying confounding (confounders that simultaneously act as intermediate steps in the pathway between intervention and outcome) require special methods such as marginal structural analysis or g-estimation. When results from observational studies are used in the model, causal graphs can be used to explicitly state causal assumptions

DIMENSION 3: SYNTHESIS OF EVIDENCE			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
	Has the quality of the data been assessed appropriately?		Sources of data and data limitations are expected to be discussed
	Has expert opinion been used, are the methods described and justified?		An expectation that strengths and limitations of assumptions adopted should be included
Utilities	Are the utilities incorporated into the model appropriate?		Methods used to obtain utility weights and methodology used to transform health estate estimates into quality of life scores
	Is the source for the utility weights referenced?		Sources of data and data limitations are expected to be discussed
Cycle length and half cycle correction	Has the choice of cycle length been justified?		It should be based on the clinical problem and remaining life expectancy
	Has the use of a half cycle correction been stated?		Any assumption adopted is expected to be disclosed
Resources/ costs	Are the costs incorporated into the model justified and sources described?		Sources of data and data limitations are expected to be discussed
	Has discount rates been reported and justified given the target decision-maker?		

DIMENSION 3: SYNTHESIS OF EVIDENCE			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
Patient heterogeneity	Has patient heterogeneity been considered?		For example, in a cohort model states need to be homogeneous to observed or unobserved characteristics affecting transition probabilities to observed or unobserved characteristics affecting transition probabilities
Parameter precision	Has mean values and distributions around the mean and the source and rationale for the supporting evidence been clearly described for *each parameter* included in the model?		Sources of data and data limitations are expected to be discussed

DIMENSION 4: ANALYSIS OF MODEL UNCERTAINTY			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
Uncertainty	Has analyses of uncertainty pertaining to the decision problem been included and reported? If not, has the reasons been explained for its omission?		Analysis of uncertainty is expected to be include as part of the DAM
Parameter estimation & uncertainty	Has one-way DSA or two-way sensitivity analysis been performed?		Tornado diagrams, threshold plots or simple statements of threshold parameter values, are all appropriate. Uncertainty of parameters may be represented by several discrete values, instead of a continuous range, called 'scenario analyses'. It is a good practice to include the specification of parameter's point estimate and a 95% CI range.
	Has a Probabilistic Sensitivity Analysis (PSA) been included?		The specific distribution (e.g. Beta, normal, lognormal) as well as its parameters should be disclosed. When PSA is performed without an accompanying EVPI, options for presenting results include CEAC and distributions of net monetary benefit or net health benefit. When more than two comparators are involved, curves for each comparator should be plotted on the same graph.
	Has correlation among parameters been assessed?		Lack of evidence on correlation among parameters should not lead to an assumption of independence among parameters

DIMENSION 4: ANALYSIS OF MODEL UNCERTAINTY			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
	If model calibration was used to derive parameters, has the uncertainty around calibrated values been tested using DSA or PSA?		Calibration is commonly used to estimate parameters or adjust estimated values such as overall and disease specific mortality and event incidence rates
Structural uncertainty	Has a discussion about the inclusion/exclusion of assumptions affecting the structure of the model been included?		For example: i) health states and the strategies adopted following the recurrence of events; ii) length of treatment effects; iii) types of adverse effects included; iv) duration of treatment effects; v) time dependency of probabilities (in a time dependent utility, the cost of delaying treatment as a function of the time a patient has remained in an untreated acute pathological state); vi) prognostic implications of surrogate end points or vii) clinical events. Although these structural assumptions are not typically quantified, it is uncertain whether they express reality accurately and for that reason they should be assessed as part of structural uncertainty analysis
Other reporting of uncertainty analyses	Has the EVPI being measured /discussed?		If the purpose of a PSA is to guide decisions about acquisition of information to reduce uncertainty in the results, EVPI should be presented in terms of expected value of information. EVPI is commonly reported in monetary terms using net monetary benefit or net health benefits; EVPI should be reported for specified ICER thresholds

DIMENSION 5: MODEL TRANSPARENCY AND VALIDATION			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
Transparency	Has a graphical description of the model been provided?		From observation / proof reading / peer review
	Has all sources of funding and their role been identified?		From observation / proof reading / peer review
	Have all methods the used been customised to specific application(s) and settings?		As per description in the methods section and peer review
	Has the report used nontechnical language and clear figures and tables to enhance the understanding of the model?		From observation / proof reading / peer review
	Has limitations and strengths been acknowledged/discussed?		From observation / proof reading / peer review
	Is there any reference as to whether technical documentation would be made available at request?		From observation / proof reading / peer review
Validation	Is there any evidence of model's face validity?		Can occur in several ways: the group that develop the model can appeal to members of the modelling group, people in the

DIMENSION 5: MODEL TRANSPARENCY AND VALIDATION			
Components of good practice	**Questions for review**	**Yes , No, or N/A**	**Attributes**
			same organisation who did not build the model, or external consultants. Any reader can perform his/her own evaluation. Peer review (previous to publication)
	Has internal validity been assessed?		Verification or technical validity; models should be subject to rigorous verification and the methods used should be described and results made available on request
	Has cross-validation been assessed?		It involves examining different models that address the same problem and comparing their results. Its meaningfulness depends on the degree to which methods and data are independent. Modellers should search for modelling analyses of the same or similar problems and discuss insights gained from similarities and differences in results
	Has external validity been assessed?		This compares the model's results with actual event data; a formal process needs to be developed including identifying suitable sources of data; results of external validation should be made available
	Has the model's predictive validity been assessed?		If feasible given the decision problem and future's sources availability

2.6. Discussion

The DAM guidelines identified in this chapter seem to have responded to the need to reflect on how good practice in the field has been defined; the need to keep pace with the rapid progress in the way that economic evaluation methodology has progressed since the 1980s; and as a means to ensure that guidelines for good practice remain current, effective, and helpful. More comprehensive guidelines, for example, Philips et al[26] or the set of the International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making (ISPOR-SMDM) guidelines, have been developed as part of bigger projects, that is, an HTA project involving experts from prestigious academic institutions or as part of a 'task force' respectively.

Recommendations and statements of good practice have been proposed following the application of different methods. For example, Philips et al[26] synthesised good practice guidance and accompanying checklist resulted after taking each theme and subtheme identified in a systematic review of guidelines followed by technical discussions among the research team of its relevance in relation to the development of general guidelines[26]. Guidelines produced by ISPOR-SMDM resulted from a 'task force' consisting of expert developers and experienced users of models from academia, industry, and government, with representation from many countries. A decision was made by the task force to divide the DAM topic into six components and working groups, respectively; three of these groups covered aspects relevant to all models such as the conceptualization of a model, the estimation of model parameters and handling of uncertainty, and the validation of models and issues of transparency. The other three components considered specific techniques:

state-transition modelling, discrete event simulation, and dynamic transmission models. The working groups produced draft reports for each section, and in contrast to Philips there was no systematic attempt to review the literature. The first draft of recommendations represented the opinions of the experts in the Task Force and these were posted on the ISPOR and SMDM Web sites for comment by the general membership of the societies. A second group of experts—again, with broad representation of modellers and users of models—was invited to formally review the articles. Their comments were addressed and after receiving additional comments and considering further revisions, the final version of each article was prepared and released to the public (see Section 2.3.1).

This review has critically compared statements of good practice in contemporary guidelines and identified areas in which further work may be needed.

This chapter demonstrates that: 1) good practice guidelines have been developed and agreed; adherence to these guidelines is considered as best practice in DAM; 2) guidelines should be seen as tools that if followed will lead to the results obtained being more credible; 3) there are common grounds in the application of guidelines; and 4) some aspects of the guidelines related to DAM require further development, for example, the choice of model structure, assessment of structural uncertainty and achieving predictive validity.

Common grounds have been identified for the application of guidelines in aspects such as the specification of a model's structure, the inclusion of incident cases over the time horizon of an evaluation, the use of time horizons, parsimonious model structure, and subgroup analysis in DAM.

Most decision problems can be conceptualized using one of the available model types, while the choice of model structure is unlimited. There is general acceptance of the special circumstances under which complex modelling needs to be taken into consideration, while at the same time, overly complex models should be avoided if a simpler model can accurately reflect all aspects of the decision problem. More research should be undertaken of case studies comparing the economic efficiency of simple versus complex models, the use of hybrid models which are considered to be very flexible and accurate with no restriction on how time is handled[21], and the trade-off between model complexities versus model transparency. This should be done in light of the advances in computing that make complex calculations feasible and economically efficient, opening the way for the more generalised use of individual-based simulations[21].

Whether model structure should be informed by data availability or not remains another conflicting aspect in DAM. Current guidelines have argued the case for building a model first and then looking for the data to populate it, as this strategy will result in more appropriate and relevant model structures[16]. An apparent drawback of this approach as argued by detractors is data availability. Alternatively, finding the data to populate the model might be possible perhaps by adopting more assumptions based on expert opinion[16, 21]. Independent of the assumptions adopted the model parameters should reflect the uncertainty due to the gaps in the available data, which in an ideal world would trigger the need for value of information analyses to show the value of this required data[20].

Structural uncertainty remains an area of controversy; an inappropriate structure can invalidate the conclusions drawn from cost-effectiveness analyses, while choices made

when structuring a model can significantly affect its results and the inferences from it. Until recently, even the definition of structural uncertainty was a matter of dispute[27, 51]; however, contemporary guidelines have clarified this concept by using an analogy with linear regression, and it is now recommended as good practice to factor in structural uncertainties into a model[20].

Another area in which issues have been raised has been with model validity. Guidelines have recognised that 'not all models will be able to achieve all these best practices', p. 844[22] while the 'inability to do so does not necessarily imply a model is not useful', p. 849[22]. Recent guidelines, however, seem to have provided a scope for analysts to use their own discretion to solve some issues, provided that the use of 'optimal practices', as described by methods and recommended practice is demonstrated[22]. Some aspects of model generalisability demand further research because it relies on the availability of follow-up data ideally from the same source, and follow-up data from other sources may not be generalisable to longer follow-up periods or to new contexts[61].

There are some areas in which there is a contradiction between the guidelines; however, it is suggested here these issues can be solved at the discretion of the analysts. A good example is when guidelines indicate the use of all feasible and practical comparators[16, 26]. The same guidelines indicate that the choice of comparators is governed by the scope of the model, which is a direct consequence of the research question. In other words, even though a broad range of feasible strategies may be available, the choice of comparators is expected to answer to the decision problem. However, the inclusion or exclusion of

potentially relevant comparators should be assessed as part of the structural uncertainty of the model[51].

Last but not least, this Chapter proposes the use of a practical five-dimension framework to improve current standards of reporting results of DAM. This framework consists of a single instrument that is comprehensive by synthesizing all contemporary guidelines in a single checklist instrument, is user friendly by reducing the extensive and complex amount of information in a single checklist tool and is practical by directing researchers towards the key elements of good practice in DAM.

2.6.1. Strengths and limitations

This chapter contains a comprehensive review of more than a decade of developments in DAM, including the most contemporaneous guidelines. Although this chapter has discussed all available general guidelines in a single document, the breadth of this field determined that this review focuses on aspects that are considered general to all models (model structure, model conceptualization, model parameters, model uncertainty, and model transparency and validation). The exclusion criteria adopted (abstracts, posters, conference papers and non-English language studies) may be considered as a limitation of this review; however, these were required to guarantee consistency in the analysis; furthermore, a negligible number of non-English language studies were identified pertaining to applied studies. Some databases such as HEED, Psychinfo and Cinhal were not included in this review mainly because the same references would be identified in Medline or else their focus was applied research. This review does not address the choice

of data and its processing to yield suitable inputs for the model; the view was taken that this is a topic that has been extensively developed in other fields such as epidemiology or statistics. Finally, as stated in the previous section, applied studies were excluded that are important for identifying which elements of guidelines pose greater challenges for analysts or correspond to deviances from guidelines in current practice. This undoubtedly triggers the need for research on the adherence of current practice to guidelines and its impact on results of decision-modelling emphasizing for example, on issues around the reporting of uncertainty analysis or the assessment of structural uncertainty or around areas of increasing interest such as the practical use and feasibility of generic models.

It is the purpose of Chapter 3 to demonstrate how this five-dimension framework might be used in practice to assess the adherence of published models to contemporary DAM guidelines. This type of exercise is more useful if undertaken for a single disease area or for one research question at a time, since this makes it possible to remove some of the variation between models which is not relevant for the purpose of assessing adherence to guidelines prior to comparing results across a disease area. Furthermore, an assessment of adherence, as the one proposed here, i.e. covering a disease area or focused on a particular research question, has the added value of contributing to improving current standards of reporting, identifying methodological challenges faced by modellers (for example, assessment of structural uncertainty or model validation), and contributing to identifying the methodological issues not covered in existing published guidelines.

2.7. Conclusions

The framework to judge the adequacy of decision analytic modelling has changed dramatically since it was first envisioned. Important attempts have been made to keep pace with the rapid progress in the way DAM guidelines have been developed. To date, ISPOR-SMDM 2012 constitutes the most contemporaneous, up-to date and agreed set of good practice guidelines. However, the results of this Chapter indicate that guidelines lack practicality due to the extensive amount of information available and its complexity. This Chapter proposes the use of a single instrument to aid researchers improving current standards of reporting results of DAM.

As previously stated, this review excluded applied studies that are important for identifying which elements of guidelines pose greater challenges for analysts or correspond to deviances from guidelines in current practice. This is taken forward in Chapter 3, where research is undertaken on the adherence of current practice to guidelines and its impact on results of decision analytic modelling.

CHAPTER 3. DO ECONOMIC EVALUATIONS IN PRIMARY CARE PREVENTION AND THE MANAGEMENT OF HYPERTENSION CONFORM TO GOOD PRACTICE GUIDELINES? A SYSTEMATIC REVIEW

3.1. Introduction

Cardiovascular disease, which incorporates coronary heart disease (CHD) and stroke, is the main cause of death worldwide[35] and in England and Wales[63]. Hypertension, defined as a persistent raised blood pressure of 140/90 mmHg[64], has been recognised as the most important modifiable risk factor for CVD[63, 64]. Poorly controlled high blood pressure can damage artery walls and increase the risk of developing a blood clot. Moreover, if it is not treated it can also damage organs such as the kidneys, heart and brain. DAM guidelines have recognised that RCTs are good sources of evidence to judge the effectiveness of treatments; however, because the time horizon for trials often does not reflect the full duration of the impact of interventions, DAM is typically used to extend the results of a short term trial over a longer time horizon[65, 66]. A primary outcome used in RCTs that are focused on hypertension is often change in blood pressure. However, this is only an intermediate outcome and DAM can be used to examine the impact of change in blood pressure on the risk of CVD events in the longer term.

Chapter Two and previously published research[67] identified the need for further investigation into the compliance of DAM to good practice and its impact on the conclusions drawn from

economic evaluations. The aim of this Chapter is to critically evaluate how DAM in primary prevention of CVD conforms to guidelines and, in doing so, demonstrate the usefulness of the five-dimension framework developed in Chapter Two to assess compliance to guidelines in a single setting. The focus here is on one particular clinical area since this makes it possible to remove some of the variation between models which is not relevant for the purpose of assessing compliance (for example, different outcomes, treatment options or sources of uncertainty). CVD prevention has been selected due to the wide number of recent and available model-based cost-effectiveness studies conducted in this topic area. This chapter focuses on interventions aimed at lowering blood pressure, as a modifiable risk factor for CVD, and seeks to answer the research question: 'to what extent do model-based economic evaluations of primary prevention interventions aimed at lowering blood pressure in patients with hypertension or at risk of developing hypertension conform to the published guidelines for DAM?'

3.2. Methods

Studies of interventions aimed at lowering blood pressure were reviewed and the challenges faced when applying DAM methods were identified and discussed. A systematic review was conducted, meeting the UK Centre for Review and Dissemination guidance and the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines for reporting[68].

The review followed a structured approach for framing research questions: patient population (P), intervention (I), the comparator group (C), outcome (O) and the study design (S), or PICOS[68]. Papers published from January 2000 to March 2015 and written in English were included in this review if they met all of the following inclusion criteria:

- the target population was individuals presenting with high blood pressure or at risk of developing hypertension;
- the intervention(s) aimed at lowering blood pressure;
- management of hypertension, as a modifiable risk factor for CVD, was part of a primary prevention strategy (when studies also included secondary prevention, we have concentrated on the results for primary prevention); and
- the study was a model-based economic evaluation.

This review excluded systematic reviews, guidelines, trials, protocols and conference abstracts. In addition, studies were also excluded where the interventions:

- were aimed at screening blood pressure;
- were part of a polypill strategy;
- measured non-adherence to treatment; or
- were part of a secondary prevention and treatment strategy.

Searches were undertaken using terms identified by expert clinical opinion and a list of synonyms identified for each term that helped inform the final search terms used in this review ("cost effectiveness", "mathematical model", "decision analysis", "Markov model",

"decision tree", "economic evaluation", "hypertension" and "lowering blood pressure"). The search was undertaken using truncations and wildcards and all synonyms were subsequently combined with appropriate medical subject heading terms (MeSH) or subject terms using Boolean operators (Appendix 3, Appendix 4).

The following databases were searched: EMBASE and Medline via the Ovid interface, and the Centre for Reviews and Dissemination's (CRD) NHS Economic Evaluation Database (NHS-EED). In addition, the reference lists of the studies included in this review were manually examined. All papers identified by database searching were exported into ENDNOTE-X7^{TM} and duplicate references were removed.

Titles identified by the searches were screened by reading the abstract. Articles that appeared to be relevant at this point were obtained and screened against the inclusion and exclusion criteria; several studies appeared relevant on reading the abstract but were subsequently excluded after reading the full paper.

All studies were manually searched and data were extracted; any point(s) requiring clarification were checked with at least one supervisor. The extraction tool consisted of the new framework proposed in Chapter 2 that synthesises contemporary DAM guidelines in a single checklist instrument. The tool aided the retrieval and organisation of information from each study across five dimensions (Table 2-5):

i) problem concept;

ii) model concept;

iii) synthesis of evidence;

iv) analysis of uncertainty; and

v) model transparency and validation.

This approach ensured that the review did not miss any information related to the model building process. Data were extracted as free text and in the form of a 'yes/no' response.

3.3. Results

The database search yielded 2,607 studies; after removing 27 duplicates, 2,580 studies were left for screening (

Figure 3-1). 2,549 studies were excluded because they did not consider a CVD related intervention, were not a model-based economic evaluation, or were focused on screening (

Figure 3-1). 31 full-text articles were assessed for eligibility, of which 18 were rejected as a secondary prevention strategy. 13 studies were included in this review, none of which were identified through other sources (

Figure 3-1).

Only two of the studies included were published prior to 2004. Thus it can be seen that the majority of studies (11/13) would have had access to DAM guidelines at the time of their publication such as, for example, Weinstein (2003[18]) or Philips (2004[26]).

Four studies evaluated programmes for the clinical prevention and treatment of hypertension[39, 69-71] and nine evaluated antihypertensive drug treatments to lower blood

pressure (Table 3-1)[72-80]. Ten studies were cost-utility analyses (CUA) or combined both CUA and a cost-effectiveness analysis (CEA)[39, 69, 70, 73-75, 77-80] while three studies were CEA[71, 72, 76] (Table 3-1). The intervention target (risk factor) examined was high blood pressure. The remainder of this section describes the main findings.

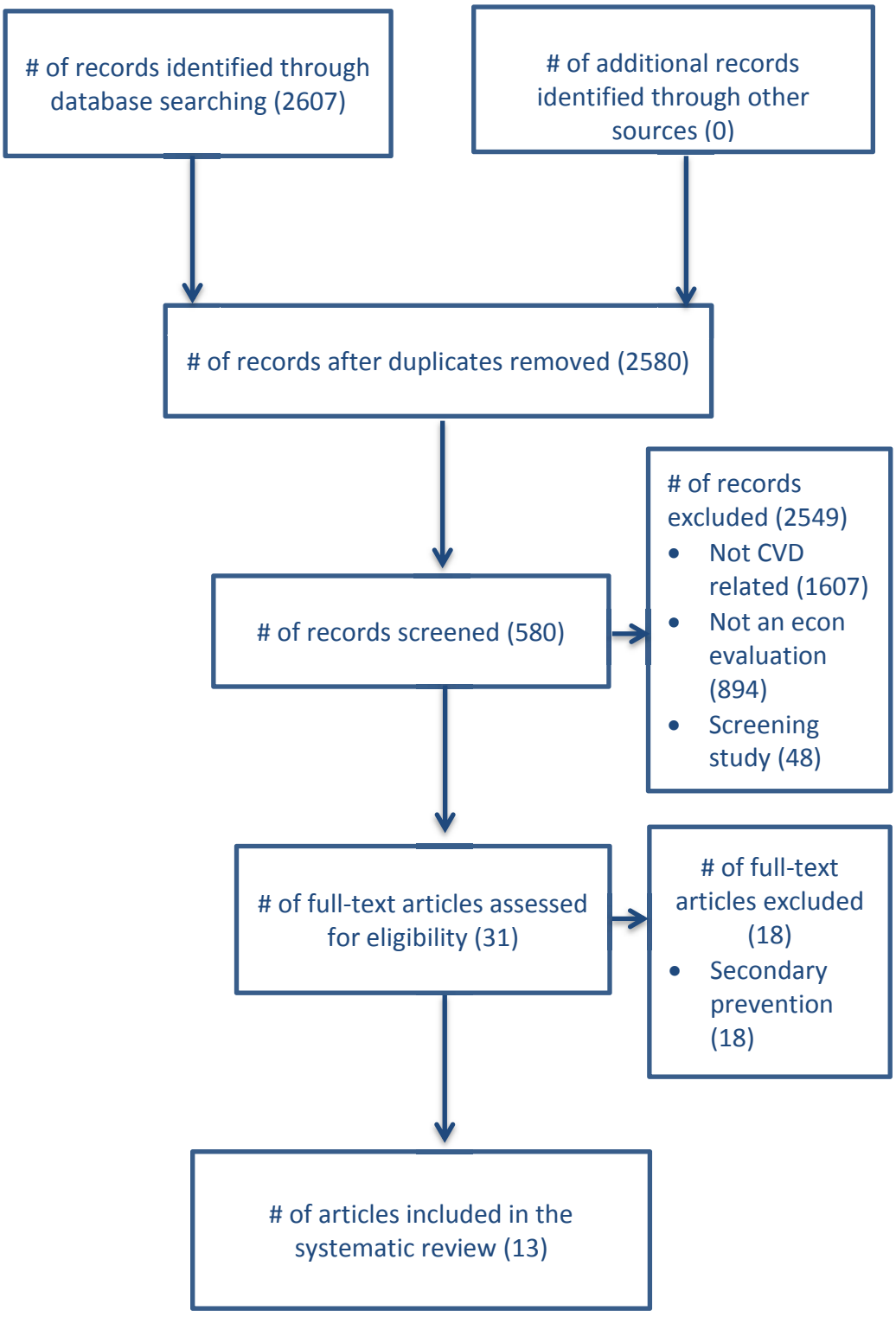


Figure 3-1 Flow chart for the identification of applied studies using the PRISMA statement

Table 3-1 Characteristics of studies, methods and model features

Author(s) / date	Research Question	Perspective/ Country	Comparators	Target Population/ Intervention	Time Horizon/ Discount Rate	Reported Baseline Results	Health States	Study Design and Outcomes
Kaambwa et al. (2014) [39]	LT CE of self-management of HPN	UK NHS	Self-management of HPN versus usual care	66-year old with HPN	Lifetime; 3.5% for both	Self-management of HPN was CE ICER £1,624/QALY	Well Stroke MI Angina HF Death	CUA QALYs
Stevanovic et al. (2014) [72]	CE of lowering blood pressure in patients with HPN and low CVD risk	Dutch HIS	Anti-HPN with HCTZ versus various combinations of HCTZ/Losartan (ACEIs) or HCTZ/ARBs versus no-treatment	Various age groups: 40, 50, 60 and 65, gender and various HPN groups	10 year and lifetime; 4 % for costs and 1.5 % for health	Systolic blood pressure reduction was found CE A 65-year old: -10 year lifetime: HCT €6,032/LYG man or €12,345/LYG woman; -Lifetime: HCT €3,076/LYG man or €3,074/LYG woman	Disease free-HPN Acute CVD (non-fatal) Stable CVD (non-fatal) Fatal CVD Non-CVD death	CEA LYG
Wu et al. (2013)[73]	CE of Amlodipine (CCB) versus ARB in the prevention of stroke and MI	Chinese Third party payer	Amlodipine (CCB) versus ARB	Average 65-year old cohort presenting HPN	5 years, 3% for both	Amlodipine was the dominant strategy	Disease free-HPN Stroke Post-stroke MI Post-MI Dead	CUA QALYs
Kourlaba et al. (2013) [74]	CE of a blood pressure lowering drug therapy in patients with mild-to-moderate HPN	Greek Third party payer	Telmisartan /HCTZ compared to Losartan/HCTZ and Valsartan/HCTZ	Average 57-year old cohort presenting HPN; analyses undertaken by gender	Lifetime, 3.5% for both	Telmisartan found to be CE Males: €3,002/QALY or €1,765/LYG Females: €10,856/QALY or €7,076/LYG	Disease free-HPN Non-fatal MI post- non-fatal MI Stroke Post-stroke death	CEA CUA LYG QALYs

Author(s) / date	Research Question	Perspective/ Country	Comparators	Target Population/ Intervention	Time Horizon/ Discount Rate	Reported Baseline Results	Health States	Study Design and Outcomes
Ekwunife et al. (2013) [75]	CE of drugs in the management of HPN	Nigerian third party payer perspective for costs	4 classes of antihypertensive medications: HCTZ, pranolol (Beta Blocker), lisinopril (ACE) and nifedipine (CCB)	Average 40-year olds with HPN	30 years; 3% for both	In the low CVRS ACEI had highest (15,000 $/QALY) NMB, however, in the medium and high risk CVRSs, CCB had highest WTP (15,000 and 12,500 $/QALY respectively)	Non-asympto-matic (disease free) Stroke Non-fatal Stroke CHD non-fatal CHD	CUA NMB EVPI US$/ QALYs
Wisloff et al. (2012) [76]	CE of various generic anti-HPN in the prevention of CVD	Norwegian HIS	CCB compared to no-treatment in various age groups and gender	HPN patients at different age groups (40, 50, 60 and 70)	Lifetime, 4% for both	CCB / male was CE: 40: -€456,838/LYG 50: -€445,018/LYG 60: -€410,510/LYG 70: -€352,875/LYG CCB /female was CE: 40: -€621,537/LYG 50: -€630,144/LYG 60: -€588,999/LYG 70: -€465,906/LYG	Disease free-HPN Stroke Stroke-Sequelae AMI Angina HF Post-CVD Dead	CEA LYG NHB
Baker et al. (2012[80]	CE of initiating hypertension treatment with valsartan and then switching to generic losartan in the prevention of CVD	US third party payer perspective	Two comparative analyses: 1.Continual Valsartan versus continual Losartan 2.Continual Valsartan versus Valsartan switch to generic Losartan	Moderate HPN patients – SBP 160-179- aged 18 and older	20-year time horizon and 3% discount for both	Treatment of moderate hypertension was considered CE with an ICER of $32,313/QALY or $27,268/LYG; Switching treatment resulted in an ICER of £30,170/QALY and $25,460/LYG	CVD event free with treated HPN Post-CVD with treated HPN Death	CVD event rates per arm CEA CUA LYG QALYs

Author(s) / date	Research Question	Perspective/ Country	Comparators	Target Population/ Intervention	Time Horizon/ Discount Rate	Reported Baseline Results	Health States	Study Design and Outcomes
Granstrom et al. (2012) [79]	Long-term CE of Candesartan versus Losartan in the primary prevention of HPN	Swedish HIS	Candesartan versus Losartan	Average 62-year old cohort presenting with HPN	Lifetime, 3% for both	Candesartan was the dominant strategy	Disease free-HPN HF Chronic IHD Post-MI PAD Post-stroke Arrhythmia Dead	CUA QALYs
Perman et al. (2011) [71]	CE multi-intervention programme versus pharmacological strategy	Argentinian Third party payer	HPN programme compared to usual care	Two target groups: 65-year old plus HPN; 65-year old, HPN and previous CVD	Lifetime, 5% for both	The HPN programme was cost-effective US$1,124/LYG	Acute myocardial event No event Death	CEA LYG
Ekman et al. (2008) [77]	CE of Irbesartan in combination with HCTZ in blood pressure reduction	Swedish third party payer	Four strategies in male and female population: Irbesartan Placebo Losartan Valsartan	55-year old male cohort presenting with HPN	Lifetime; 3% for both	Irbesartan was CE when compared to placebo in males and females; ICERs of €3,451/QALY and €7,704/QALY respectively Losartan & Valsartan were dominated by Irbesartan in males and females	Disease free-HPN Angina MI Post-MI CHF Stroke Post-Stroke Dead	CUA QALYs
Gandjour et al. (2007) [69]	CE of a national HPN programme for patients with essential HPN and without CVD	German HIS	National programme versus no programme (for low and high risk population)	Various age groups (40-49, 50-59, and 60-69); patients with essential HPN and without CVD	Lifetime; 3% for both	National programme is CE High risk male, 40: €800/QALY, 50: €880/QALY 60: €757/QALY High risk female, 40: -€17,347/QALY 50: -€26,987/QALY 60: -€1,263/ QALY	Disease free-HPN MI Stroke Renal- failure death	CUA QALYs

Author(s) / date	Research Question	Perspective/ Country	Comparators	Target Population/ Intervention	Time Horizon/ Discount Rate	Reported Baseline Results	Health States	Study Design and Outcomes
Montgomery et al. (2003) [70]	Effectiveness and CE of LT blood pressure lowering	UK Health service perspective	Anti HPN treatment versus non-treatment	Various pop cohorts: 30-39; 40-49; 50-59; 60-69; 70-79. Hypertensive population	Lifetime; 6% for costs and 1.5% for effects	Treatment found more CE than non-treatment. ICER was higher for low risk women compared to low risk men	Untreated Treated T_se U_cve_ua T_cve_ua T_se_cve_ua U_cve_af T_cve_af T_se_cve_af Death	CUA QALYs
Nordmann et al. (2003) [78]	CE of ACE as HPN first-line therapy versus conventional therapy	Canadian Third party payer	4 strategies: -Control or conventional therapy -ECG - EchoCar -ACE	40-year old male cohort presenting with HPN but without CVD	Lifetime; 5% for both	Unfavourable results of CE: ECG versus Control: US$ 0 /QALY/LYG; EchoCar vs Control: US$ 200,000/QALY/LYG ACE vs Control = US$700,000/QALY or US$525,000/LYG	Disease free- HPN (with or without LVH) CAD CVD CHF Dead	CEA CUA LYG QALYs

Note: ACE = Angiotensin-Converting-Enzyme Inhibitor; AMI = Acute Myocardial Infarction; ARB = Angiotensin-II-Receptor Blocker; BP = Blood pressure; CCB = Calcium-Channel Blocker; CE = cost-effectiveness or cost-effective; CEA = Cost-effectiveness analysis; CHF = Congestive Heart Failure; CUA = Cost-utility analysis; CVD = Cardiovascular disease; EchoCar = Echocardiography; EVPI = Expected Value of Perfect Information; HCTZ = Hydrochlorothiazide; HF = Heart Failure; HIS = Health Insurance System; HPN = Hypertension; IHD = Ischaemic Heart Disease; LYG = Life Year Gained; MI = Myocardial infarction; NHB = Net Health Benefit; NMB = Net Monetary Benefit; LT = Long term; PAD = Peripheral Artery Disease; T_se = Treated, side-effects (health state); U_cve_ua = Untreated, cardiovascular event, unaffected (health state); T_cve_ua = Treated, cardiovascular event, unaffected (health state); T_se_cve_ua = Treated, side-effects, cardiovascular event, unaffected (health state); U_cve_af = Untreated, cardiovascular event, affected (health state); T_cve_af = Treated, cardiovascular event, affected (health state); T_se_cve_af = Treated, side-effect, cardiovascular event, affected (health state)

3.3.1. Problem concept and model concept

The decision problem and study objective(s) were stated in all the studies, and all evaluated CE from a health care payer perspective. The target decision-maker audience was made explicit in 10/13 studies as that of the health care payer, i.e. including only the health effects experienced by patients receiving the intervention and costs for the medical services required to provide the intervention[16]. For the remaining studies[69, 72, 77] the perspective was left implicit. Ekman[77] commented that the analysis was "in a Swedish health-care setting", while Stevanovic[72] was interested "in the Dutch setting" and Gandjour[69] focused on those "insured by the German SHI", where SHI refers to the German Statutory Health Insurance.

For all studies, the target population was individuals with hypertension or at risk of developing hypertension (Table 3-1), frequently stratified by gender, presence of hypertension, age groups, and mean age. The target population was always modelled as closed thus reflecting patients entering only at the start of the analysis (Table 3-2).

Despite all the studies sharing a common aim, namely primary prevention of CVD via lowering blood pressure, these economic models compared a wide range of interventions and presented their results using outcome measures such as QALYs[39, 69, 70, 73-75, 77-80], life years gained (LYG)[71, 72, 74, 76, 78, 80], net health benefits (NHB)[76], net monetary benefits (NMB)[75] and expected value of perfect information (EVPI)[75] (Table 3-1).

Side effects were modelled in only one study[73]. Four studies[39, 69, 72, 80] acknowledged the lack of adverse events as a limitation of their results due to lack of data. Two studies argued that since 'previous clinical trials found that first-line hypertensive drugs do not have more side effects than placebo'[71] or they have 'mild side effects'[77] there was no need to model adverse

effects. Similarly another study argued that fatal side effects would have been already captured in the clinical trials via the measure of effectiveness[76].

All the studies commented on the reasons for the selection of their comparators, where their choice of comparators seems to have been governed by the scope of the study. Two studies acknowledged as a limitation the exclusion of relevant comparator(s) arguing that there may be more relevant comparators not included[77, 79]. Furthermore, the 'do nothing' option was considered in four of the studies[69, 72, 76, 77].

All the studies used Markov models and included a figure showing the model structure; in one study[71] the structure of the Markov model shown in the figure did not seem to reflect the structure of the model described in the text (Table 3-3). The model structures accounted for both acute and chronic health states. Five studies made explicit reference to how the structure of their models was defined either by using an existing generic model[76], being based on disease progression[39, 69] or consisting of health states designed to reflect the course and history of CVD events[80]. One study reported that 'health states in the Markov model are based on cardiovascular events measured in the previously reported registry study'[79]. For the remaining studies it was inferred that the model structure was based on disease progression.

Table 3-2 Dimension 1 - Problem concept

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Decision problem	Is there a clearly written decision problem?	Yes	Yes	Yes	Yes	Yes	Yes
	Are the study's objective(s) consistent with the decision problem and the study's scope?	Yes	Yes	Yes	Yes	Yes	Yes
Analytical perspective	Has the perspective being stated?	Yes	Yes	Yes	Yes	Yes	Yes
Target population	Has the target population being identified?	Yes	Yes	Yes	Yes	Yes	Yes
Health outcomes	Are the model's outcome(s) consistent with the perspective, scope and objective(s)?	Yes	Yes	Yes	Yes	Yes	Yes
	Have any adverse effect(s) been captured?	No	No	Yes	No	No	No
Interventions modelled	Are the options under evaluation clear?	Yes	Yes	Yes	Yes	Yes	Yes
	Were the inclusion/ exclusion of feasible options justified?	Yes	Yes	Yes	Yes	Yes	Yes

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Time horizon	Is the model time horizon sufficient to reflect all important differences between options?	Yes	Yes	No	Yes	Yes	Yes
	Have time horizon, duration of the treatment and the treatment effect(s) been described and justified?	Yes	Yes	No	Yes	Yes	Yes

Continuation

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Decision problem	Is there a clearly written decision problem?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Are the study's objective(s) consistent with the decision problem and the study's scope?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Analytical perspective	Has the perspective been stated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Target population	Has the target population been identified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health outcomes	Are model's outcome(s) consistent with the perspective, scope and model's objective(s)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Have any adverse effect(s) been captured?	No	No	No	No	No	No	No

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Interventions modelled	Are the options under evaluation clear?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were the inclusion/ exclusion of feasible options justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time horizon	Is the model time horizon sufficient to reflect all important differences between options?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Have time horizon, duration of the treatment and the treatment effect(s) described and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3-3 Dimension 2 - Model concept

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Choice of model type	Was the unit of representation given?	Yes	Yes	Yes	Yes	Yes	Yes
	Does interaction(s) among individuals need to be model? If yes, was this described?	No	No	No	No	No	No
	Does the decision problem require a short time horizon?	No	No	No	No	No	No
	Is it necessary to model time in discrete cycles?	Yes	Yes	Yes	Yes	Yes	Yes
	Was the type of model discussed and chosen?	Yes	Yes	Yes	Yes	Yes	Yes
Model structure	Was the starting cohort defined by demographic and clinical characteristics affecting transition probabilities or state values?	Yes	Yes	Yes	Yes	Yes	Yes
	Were health states and transitions reflecting the biological or theoretical understanding of the disease modelled?	Yes	Yes	Yes	Yes	Yes	Yes

Continuation

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Choice of model type	Was the unit of representation given?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Does interaction(s) among individuals need to be model? If yes, was this described?	No	No	No	No	No	No	No
	Does the decision problem require a short time horizon?	No	No	No	No	No	No	No
	Is it necessary to model time in discrete cycles?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Was a type of model discussed and chosen?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model structure	Was the starting cohort defined by demographic and clinical characteristics affecting transition probabilities or state values?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were health states and transitions reflecting the biological or theoretical understanding of the disease modelled?	Yes	Yes	Yes	Yes	Yes	Yes	Yes

A lifetime time horizon was adopted in all but two studies: of these, one used a five-year[73] time horizon for a population aged 65 years whilst the second used 20-years for a population aged 18 and over[80]. The five-year time horizon was justified as matching the five-year time span given to social security authorities in China for budget planning[73] whilst the 20-year time horizon was not discussed[80]. Cycle length, though rarely justified in the studies, was always 1 year. Only one study[69] justified their choice of time horizon as most of the data used in their model referred to a 1-year period.

3.3.2. Synthesis of evidence

Patient heterogeneity was considered in most of the studies; results were presented by age cohorts[69, 70, 72, 73, 76] and gender[39, 69, 70, 72-74, 76, 77, 79]. Some studies added further analyses based on the risk of CVD[69, 70, 75], scenarios of SBP reduction[72, 77] smoking[72] and patient adherence[72, 80]. The risks of secondary events were modelled in seven of the studies, e.g. the risk of a further stroke after a first stroke[70-72, 76-79]. In some instances, assumptions were acknowledged; for example, the study by Stevanovic[72] assumed the risk of secondary events to be equal to the risk of a first non-fatal CVD event. The authors acknowledged that this would lead to an under-estimation of the CVD risk, and so an increased risk of death in patients experiencing non-fatal CVD events was adopted[72]. In Wisloff[76], secondary non-fatal events were allowed, and a patient experiencing a secondary event was assumed to be in a health state which was worse than the state they were already in. For example, a patient with stroke sequelae that experiences a MI will have the risk and costs associated with the stroke sequelae

and not those related to MI). Perman[71] used expert opinion in the assessment of the risk of secondary events. Montgomery[70], due to a lack of data assumed that any second cardiovascular event was fatal and acknowledged this as a limitation. Some studies that did not use separate states to model secondary events[39, 69, 80] captured the increased mortality from secondary events through the mortality rate of patients surviving CVD events. Few of the studies acknowledged the lack of epidemiological data to model secondary events as a limitation (Table 3-4)[39, 80].

All studies applied discounting to their results: a discount rate of 3% was most common for costs and benefits[69, 73, 75, 77, 79, 80]; two studies used a different discount rate for costs and benefits (Stevanovic used 4% and 1.5%[72] while Montgomery used 6% and 1.5%[70]) (Table 3-1). Information on the parameters used as inputs were most frequently presented in tables showing mean values and the type of distribution(s) while some studies also included 95% confidence intervals or range intervals[39, 69, 78]. The methods used to report the sources of information varied from reporting a detailed list of sources per parameter in a table to mentioning the sources of data in the main text.

Table 3-4 Dimension 3 - Synthesis of evidence

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Patient heterogeneity	Was patient heterogeneity required/considered?	Yes	Yes	Yes	Yes	Yes	Yes
Data sources	Were transition probabilities and intervention effects derived from representative data sources?	Yes	Yes	Yes	Yes	Yes	Yes
	Were (all) methods and assumptions used to derive the model's inputs described?	Yes	Yes	Yes	Yes	Yes	Yes
	Were parameters derived from observational studies controlled for confounding?	NA	NA	NA	NA	NA	NA
	Was data quality discussed?	Yes	Yes	No	Yes	Yes	Yes
	If expert opinion was used, was its methods described and justified?	No	NA	NA	NA	NA	Yes
Utilities (HSUV-weights & benefits)	Are the utilities incorporated into the model appropriate?	Yes	NA	Yes	Yes	Yes	NA
	Was the source for the utility weights referenced?	Yes	NA	Yes	Yes	Yes	NA

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Half cycle correction	Was the use of a half cycle correction stated?	Yes	No	No	No	No	No
Resources including costs	Were the costs used in the model justified and its sources described?	Yes	Yes	Yes	Yes	Yes	Yes
	Were discount rates reported and justified?	Yes	Yes	Yes	Yes	Yes	Yes
Communicating results	Did the report present results using non-technical language aided by figures or tables?	Yes	Yes	Yes	Yes	Yes	Yes
Parameter precision	Were mean value(s), distribution(s), source(s) of data and rationale for the supporting evidence described?	Yes	Yes	No	Yes	Yes	Yes

Continuation

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Patient heterogeneity	Was patient heterogeneity required/considered?	No	Yes	No	Yes	Yes	Yes	Yes
Data sources	Were transition probabilities and intervention effects derived from representative data sources?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were (all) methods and assumptions used to derive the model's inputs described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were parameters derived from observational studies controlled for confounding?	NA	Yes	NA	NA	NA	Yes	NA
	Was data quality discussed?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	If expert opinion was used, were its methods described and justified?	NA	NA	NA	NA	No	NA	Yes

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Utilities (HSUV-weights & benefits)	Were the utilities incorporated into the model appropriate?	Yes	Yes	NA	Yes	Yes	Yes	Yes
	Was the source for the utility weights referenced?	Yes	Yes	NA	Yes	Yes	Yes	Yes
Half cycle correction	Was the use of a half cycle correction stated?	No	Yes	No	No	Yes	No	Yes
Resources including costs	Were the costs used in the model justified and its sources described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were discount rates reported and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Communicating results	Did the report present results using non-technical language aided by figures or tables?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parameter precision	Were mean value(s), distribution(s), source(s) of data and rationale for the supporting evidence described?	Yes	Yes	Yes	Yes	Yes	No	No

3.3.3. Analysis of uncertainty

The studies examined and reported uncertainty surrounding their identified outcomes through sensitivity analysis. Uncertainty in parameter estimates was most commonly handled through deterministic sensitivity analysis (DSA) and probabilistic sensitivity analysis (PSA). Five studies used only one-way DSA[70, 73, 77, 78, 80], whilst another four[39, 74-76] only used PSA. Only one study measured EVPI[75] (Table 3-5).

Elements pertaining to structural uncertainty were acknowledged as such in six studies[39, 69, 71, 72, 77, 78]. Most commonly structural uncertainty was assessed through sensitivity analysis by varying the time horizon[39], the duration of the effectiveness of the treatment[39, 72], the discount rate[71, 77] or by using alternative measures of outcomes[77]. One study examined the impact of assumptions related to secondary events[76]. Lack of clinical evidence for key parameters such as the treatment effect of drugs[39, 69, 72] was identified as a source of structural uncertainty. Two studies acknowledged that they could have included more relevant comparators had they had more information[77, 79], and another two acknowledged that they had excluded a potentially relevant state due to lack of epidemiological data[69] or insufficient evidence regarding its relevance[78].

The decision about which events and health states were included was partially discussed. Some studies acknowledged that they subdivided a health state[39] (CHD into MI, HF and angina), or excluded a potentially relevant health state[73] (combined stroke and MI event). All studies included chronic health states (post events); however, few discussed having modelled the progression of disease[39, 69, 80]. Most frequently, the studies acknowledged the adoption of

assumptions, i.e. assuming the duration of treatment effects to be lifetime or as long as the time horizon in the model[39, 69, 72], or five years[77] or varied[79].

Table 3-5 Dimension 4 - Analysis of uncertainty

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Analysis of uncertainty	Was analysis of uncertainty pertaining to the decision problem included and reported?	Yes	Yes	Yes	Yes	Yes	Yes
Parameter estimation & uncertainty	Were one-way or two-way DSA sensitivity analysis performed?	No	Yes	Yes	No	No	No
	Was probabilistic sensitivity analysis (PSA) included?	Yes	Yes	No	Yes	Yes	Yes
Multivariate estimation and correlation	Was correlation among parameters considered?	NA	NA	NA	NA	NA	NA
Structural uncertainty	Was there any discussion /evidence of uncertainty in structural assumptions?	Yes	Yes	No	No	No	Yes
Other reporting of	Was EVPI measured/ discussed?	No	No	No	No	Yes	No

uncertainty analyses	If model calibration was used to estimate parameters, was uncertainty tested?	NA	NA	NA	NA	NA	NA

Continuation

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Analysis of uncertainty	Was analysis of uncertainty pertaining to the decision problem included and reported?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parameter estimation & uncertainty	Were one-way or two-way DSA sensitivity analysis performed?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Was probabilistic sensitivity analysis (PSA) included?	No	Yes	Yes	No	Yes	No	No
Multivariate estimation and correlation	Was correlation among parameters considered?	NA	NA	NA	NA	NA	NA	NA
Structural uncertainty	Was there any discussion /evidence of uncertainty in structural assumptions?	No	No	Yes	Yes	Yes	No	Yes

Other reporting of uncertainty analyses	Was EVPI measured/ discussed?	No	No	No	No	No	No	No
	If model calibration was used to estimate parameters, was uncertainty tested?	NA	NA	NA	NA	NA	NA	NA

3.3.4. Model transparency and validation

All the studies included a graphical description of the Markov model they used (Table 3-6). Sources of funding were identified in 11 studies: five were funded by the pharmaceutical industry[72-74, 77, 79, 80], one benefited from joint funds from government and pharmaceutical sources[76], three were exclusively government-funded[39, 70, 78], and one was privately funded[75]. None of the studies stated any means for accessing more detailed information about the model. All the studies had a clear policy context with an explicit statement by the funder and developer.

Validation according to guidelines[22] is a set of methods for judging the accuracy of a model in making relevant predictions, in other words, validation helps readers understand what a model does and how it does it. This review checked for five main types of validation. All the studies were subjected to face validity checks (having been peer reviewed and published in a journal) and they were subjected to verification (internal validity checking). The methods used were justified to a greater or lesser extent in each study. All studies undertook sensitivity analysis of parameters as a way to double check that the direction and magnitude of their outputs were as expected.

In terms of cross validation, results were mixed. Eight studies[39, 69-72, 75, 77, 78] examined different models that addressed the same problem and compared their results; however, the level of detail provided varied. Five studies presented limited or no evidence of cross-validation[73, 74, 76, 79, 80]; only Wisloff[76] undertook an exercise of external validation by comparing their estimated lifetimes to those reported by Statistics Norway and in doing so they found that the

input into their model needed to be adjusted to fit Norwegian mortality data. An assessment of predictive validity was not included in any of the studies considered.

Table 3-6 Dimension 5 - Model transparency and validation

Information	Review question	Kaambwa et al. (2014) [39]	Stevanovic et al. (2014) [72]	Wu et al. (2013)[73]	Kourlaba et al. (2013) [74]	Ekwunife et al. (2013) [75]	Wisloff et al. (2012) [76]
Transparency	Were the purpose, type and graphical description of the model provided?	Yes	Yes	Yes	Yes	Yes	Yes
	Was the source(s) of funding and its role identified?	Yes	Yes	Yes	Yes	Yes	Yes
	Were data sources identified/ described?	Yes	Yes	Yes	Yes	Yes	Yes
	Were methods customised to specific application(s) and settings?	Yes	Yes	Yes	Yes	Yes	Yes
	Were the effects of uncertainty measured?	Yes	Yes	Yes	Yes	Yes	Yes
	Were limitations acknowledged/ discussed?	Yes	Yes	Yes	Yes	Yes	Yes
	Was any reference made to the availability of the model's documentation at request or the terms and conditions to access it?	No	No	No	No	No	No
Validation	Was there any evidence of the model's face validity?	Yes	Yes	Yes	Yes	Yes	Yes
	Was internal validity (verification or technical validity) assessed?	Yes	Yes	Yes	Yes	Yes	Yes
	Was cross-validation (external consistency) assessed?	Yes	Yes	No	Yes	Yes	No
	Was external validity assessed?	No	Yes	No	No	No	No
	Was the model's predictive validity assessed?	NA	NA	NA	NA	NA	NA

Continuation

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
Transparency	Were the purpose, type and graphical description of the model provided?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Was the source(s) of funding and its role identified?	Yes	Yes	No	Yes	No	Yes	Yes
	Were data sources identified/ described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were methods customised to specific application(s) and settings?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were the effects of uncertainty measured?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Were limitations acknowledged/ discussed?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Was there any reference to the availability of the model's documentation on request, or	No	No	No	No	No	No	No

Information	Review question	Baker et al. (2012)[80]	Granstrom et al. (2012) [79]	Perman et al. (2011) [71]	Ekman et al. (2008) [77]	Gandjour et al. (2007) [69]	Montgomery et al. (2003)[70]	Nordmann et al. (2003) [78]
	the terms and conditions to access it?							
Validation	Was there any evidence of the model's face validity?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Was internal validity (verification or technical validity) assessed?	Yes	Yes	Yes	Yes	Yes	Yes	No
	Was cross-validation (external consistency) assessed?	No	No	Yes	Yes	Yes	Yes	Yes
	Was external validity assessed?	No	No	No	No	No	No	No
	Was the model's predictive validity assessed?	NA	NA	NA	NA	NA	NA	NA

3.4. Discussion

Using the practical framework described in Chapter 2, 13 published economic evaluations were evaluated to judge whether they conformed to contemporaneous good practice guidelines. It was found that published economic evaluations of interventions aimed at lowering blood pressure in patients with hypertension, as part of a primary prevention strategy of CVD, demonstrate limited compliance to DAM guidelines which has usually been justified by lack of data or imperfect data. This was particularly apparent in the assessment of structural uncertainty (or lack of) and model external validation.

This review identified common grounds in terms of the adherence to, and use of, guidelines. The conceptual modelling in all the studies included in this review was based on a disease process where the focus was on the definition of the health states (conditions) as opposed to treatment (pathways) received and where the decision problem posed required the evaluation of the reduction in the risk of developing hypertension, thus explaining the use of Markov models.

It has been argued that alternative model structures can lead to variations in model predictions[27], most importantly, in the context of a primary prevention strategy, an inappropriate model structure may lead to poorly informed policy decisions, resulting in inefficient allocation of scarce resources[81]. Models are by nature sensitive to choices made at every single stage during the model development process (i.e., model concept, model structure). There will almost always be more than one set of choices, for this reason, guidelines have suggested assessing the extent to which model predictions are influenced by

the choices made during the model development process, and have suggested methods to do so, such as scenario analyses[20, 82].

Lifetime time horizons should be adopted (or be justified when constrained by the cohort's lifetime) or at the very least, time horizons should be 'long enough' to capture relevant differences in outcomes across strategies[16]. Lack of data or imperfect data still poses important challenges for researchers - for example, when modelling the risk of secondary events and disease progression or to attempt the assessment of model validity. Even though elements pertaining to structural uncertainty were identified by various authors, the assessment of structural uncertainty cannot be considered common practice in this particular clinical area and additional guidelines are still needed to aid researchers identifying and quantifying structural uncertainty.

External validity still poses a challenge to researchers and more importantly, to future guidelines due to the apparent unavailability of actual extra data (from RCT or patient level data) to undertake the exercise. It has been suggested that instead of using all the data available to create a model, some data be set aside to use during the validation process (for example, one-third of the data)[30]. This may or may not always be possible, and will depend on how much data a researcher has to build a model.

Studies included in this review shared similar research questions and yet there was a great diversity in the structures of the Markov models used. Some of these were simple and some more complex, and were generally developed with limited justification[81]. These indicate, as suggested by Squires et al[83], that the methods for the development of the model structure are

still underdeveloped. This can lead to errors including poor validity, credibility, and no basis for model verification and the analysis of structural uncertainty.

Caro and Möller[30] described the above as the 'disposable approach' to modelling: models are built for a single use, focused on a particular product for a relatively short time. This explains to some extent the reduced motivation for undertaking model validation[30]. Future research should examine whether the development of 'generic models', or, as proposed by Caro and Möller, the development of multi-use models over time, can capture sufficient detail to be realistic and to avoid these aspects for which there are no data. This would allow the economic evaluation of interventions targeting CVD in any setting, and bridge the knowledge gap and, most importantly, allow ease of comparison between the results obtained from different studies.

3.4.1. Strengths and limitations

This is the first study that has critically reviewed compliance to DAM guidelines using a previously developed practical framework. It has covered more than a decade of published DAM studies of interventions aimed at lowering blood pressure in patients with hypertension. The inclusion of recent studies from European, American and Asian countries has helped to reflect current practice worldwide.

The exclusion criteria adopted may be considered as limitation; however, these were required to guarantee consistency in the analysis. Furthermore, a negligible number of non-English-language studies were identified pertaining to applied studies. The fact that none of the

studies included was published after the release of the 'five-dimension framework' and the selection of one particular clinical area (and any impact on generalisability this may have) may also be considered a limitation.

These findings seem in line with the recent debate around the methodological challenges being faced by DAM where model validation and structural uncertainty have been identified as fundamental problems due to the lack of motivation, time and data to validate models and, in the case of structural uncertainty, a lack of methods[30].

3.5. Conclusions

This Chapter, focusing on one particular clinical area, found limited compliance to DAM of economic evaluations, which was most commonly explained by lack of data or imperfect data. This result was particularly apparent in the assessment of structural uncertainty or lack of assessment of structural uncertainty.

Model predictions are influenced by choices made during the process of model development, in particular the choice of model structure. The results of this review indicate that model structures used by studies that shared similar research questions varied from simple to more complex and most importantly, that model structures were generally developed with limited justification. Chapters 4 to 6 take the findings of this chapter forward and explore the effect(s) of the choice of model structure on cost-effectiveness results; starting with Chapter 4, which outlines the issue of structural uncertainty in the decision analytic modelling process.

CHAPTER 4. OVERVIEW OF STRUCTURAL UNCERTAINTY IN DECISION ANALYTIC MODELLING

Chapters 2 and 3 identified good practice guidelines and critically assessed compliance of published model-based economic evaluations to DAM guidelines for primary care prevention and the management of hypertension. Main findings in Chapter 3 indicated that even though DAM guidelines are available, compliance to guidelines in one particular clinical area was limited, particularly when identifying structural uncertainty arising from the choice of model structure.

4.1. Introduction

The UK National Institute for Health and Care Excellence (NICE) has recommended that formal decision analytic processes rely on results of decision analytic modelling (DAM) as the standard framework for health technology assessment[11]. Mathematical models applied in cost-effectiveness or cost-utility analysis aim to examine whether an intervention is cost-effective when compared to one or more alternatives; to do so, models estimate the population mean costs and health effects. Results of such forms of analysis are most helpful when they are unbiased and uncertainty about estimated costs and consequences is properly specified[27, 28]. The main sources of uncertainty in model predictions are related to the model input values and model structure[20, 28-30] (see Section 1.2.4).

As demonstrated in Chapter 2 there are statistical literature and other sources of information available on structural uncertainty, including ISPOR-SMDM[20] and other individual articles[27, 28, 31, 51] which propose a set of guidelines for the analysis of structural uncertainty. However, when current DAM practice was assessed against compliance to good practice DAM guidelines, the findings in Chapter 3 have indicated that although elements pertaining to structural uncertainty are at times identified, the characterisation of structural uncertainty is not 'common practice' and in fact, it was shown that current DAM practice tends to omit testing for structural uncertainty. These findings are in line with those from Bojke et al[51], Afzali et al[29], and Caro and Moller[30].

The aim of this chapter is to outline the ways in which structural uncertainty has been described, characterised and assessed within the DAM process.

4.2. Methods

A review of literature using the best evidence was undertaken to identify the various ways in which structural uncertainty has been described and understood within DAM. Following the snowball technique of reference identification, relevant studies were identified that examined the following:

- Definition and/or characterisation of structural uncertainty
- Sources of structural uncertainty
- Quantification or assessment of structural uncertainty

Initial references were identified from personal and supervision team knowledge; references from serendipitous discovery, such as identifying a relevant paper when looking for something else, were included as well.

4.3. Definition and characterisation of structural uncertainty

Structural uncertainty has been referred to in a number of ways in the DAM literature. Manning[84] used the term 'model uncertainty' to refer to structural uncertainty and defined it as situations when there is uncertainty about the mathematical forms by which the parameters are combined. For example, is the response to a treatment linear with regards to increasing dose levels, or does it exhibit decreasing effects as dosage increases?

Bojke et al[51] referred to structural uncertainty as uncertainty that is not easily described as parameter or methodological uncertainty and indicated that the sources of structural uncertainty are 'the different types of simplifications and scientific judgements that have to be made when constructing and interpreting a model of any sort' (p. 739). Briggs et al[20] used an analogy with a simple regression model to explain structural uncertainty: 'just as a linear regression imposes a structural relationship between independent and dependent variables, a decision analytic model is characterized by assumptions reflected in its structure but not formally expressed numerically (types of adverse events, duration of treatment effects or time dependency of probabilities)' (page 837). For Strong et al[85] structural uncertainty relates to whether or not all relevant processes are represented in a model and asks: "does the model adequately reflect reality?" Health economists build deterministic models that can be

represented by $y = \eta(X)$, where the model is a function $\eta(.)$ that takes a vector of input parameters *X* and generates an output *y*. Whilst uncertainty in input parameters *X* is usually considered and acknowledged through sensitivity analysis, uncertainty about output *y* due to uncertainty about the model's function, $\eta(.)$, is not frequently considered. When the answer to the question "does the value of $y = \eta(X)$ represent the 'true' target value *Ÿ*?" is negative, there is model structural uncertainty[85]. NICE defines structural uncertainty as the uncertainty relating to the range of assumptions and judgements necessary in constructing a model. This can include designed features of the model (for example, the assumed standard pathway of care) as well as judgements about the relevance of evidence, assumptions about appropriate distributions for parameters, and alternative methods of estimation[11]. Structural uncertainty has been more generally used to describe all structural aspects of a model (conceptual framework, implementation platform) or the mathematical representation of a decision problem[29, 33].

Independently of its definition, structural uncertainty will almost always arise once a choice of model structure or choice of relationships between inputs and outputs is defined within the model development process. In other words, in a health economics context it is not safe to assume that one of a plausible set of model alternatives is necessarily correct. Furthermore, there should be an acceptance that, in words of Strong et al[85] 'we will almost never have adequate data against which to calibrate the output of a health economic decision model' p. 9. Therefore it is essential to assess the extent to which model predictions are influenced by making alternative choices within the model development process[29].

4.4. Sources of structural uncertainty

The results of a review of 241 HTA reports commissioned by the NHS-HTA programme from 1997 to 2005 to examine how structural uncertainty had been described and resolved in the literature found that only 15% of the reports discussed issues pertaining to structural uncertainty[51]. Among the reports discussing structural uncertainty, the most common sources of structural uncertainty identified were:

i) selection of comparators including the do nothing alternative;

ii) decision about which events, stages or health states to include;

iii) uncertainty about the correct statistical method to use to estimate parameters, and

iv) lack of clinical evidence from randomised control trials triggering the use of expert opinion[51].

For Afzali & Karnon[29] (p. 436), sources of structural uncertainty can include:

i) "health states and events represented in a model;

ii) transitions between health states;

iii) the choice of modelling technique;

iv) the relationship between a transition probability and time (time dependency);

v) the relationship between a transition probability and the clinical profile or baseline characteristics of a patient; and

vi) the choice of a particular functional form used to estimate model inputs (for example, the choice of alternative survival models to estimate time-to-event parameters)"[29].

Structural uncertainty arises from the different types of simplifications and judgements that researchers make when building a model. For example, hazard functions may be estimated as either multiplicative or additive functions of risk factors. The choice between multiplicative or additive functions, in absence of evidence, may be based on mathematical convenience. However, the choice of functional form affects the resulting hazard estimate, which in turn affects the cost-effectiveness results[84]. Clinical trials provide estimates of relative risk or risk reduction during the follow-up period. However, a trial restricted to a particular demographic or clinical group may give little or no indication on how to estimate the survival curve for individuals beyond the end of the trial, or of different age and/or sex, or for individuals with comorbidities. In this case, the researcher would need to adopt assumptions regarding the appropriate basis for extrapolating beyond the period of observation and to the population with different survival curves. A straightforward assumption could be that the age and sex-specific risk of death is modified by the disease in question, the intervention being evaluated and any comorbidity relative to the general population. A key choice to make is whether these three effects are to be considered as additive or multiplicative. The choice made, in absence of evidence, affects the health outcomes which in turn affects the results of cost-effectiveness[86].

Other issues leading to structural uncertainty include the effect of the intervention versus current practice on, for example:

i) mortality in the immediate year after the end of trial data;

ii) the effect of an intervention versus current practice on the long-term cardiovascular mortality;

iii) the types of adverse effect(s) included or ignored;

iv) assumptions regarding the duration of treatment effects;

v) time dependency of probabilities (in a time dependent utility, the cost of delaying treatment as a function of the time a patient has remained in an untreated acute pathological state); and

vi) prognostic implications of surrogate end points or clinical events.

Jackson et al[27] demonstrated that alternative plausible model structures, combining hypothetical scenarios and expert opinion, can lead to wide variation in model predictions with potential impact on funding decisions. The authors suggested that appropriate characterisation of structural uncertainty should consider the parameterisation of structural choices[27]. Results of a systematic review of structural model properties for cost-effectiveness analysis indicated that the associated impact of differences in model structure and parameterisation had a substantial effect on outcomes[87]. An important challenge in addressing structural uncertainty seems to be posed by the many elements that are grouped under structural uncertainty[30].

4.5. Assessment of structural uncertainty in decision analytic modelling

In an ideal world, an approach to DAM would consider identifying a set of plausible models as providing useful simplifications of the process being investigated. Parameters in each of the identified models be estimated and the fit of each model investigated using relevant statistical criteria. Ideally, these different statistics will consistently identify the same model as having the best fit, and then this model would be shown as the best option.

In DAM a set of alternative model structures would always be plausible because of inherent uncertainty about the structure of the model (lack of complete knowledge of the system), or because decisions have been adopted to simplify other rather complex models. Strong et al[85] defined the following formulation for describing a set of possible models as M-Closed or M-Open[85].

Assume that the complete set of plausible models is denoted as $(M_i,\ i = 1, \dots I)$. The set of models $(M_i, i = 1, \dots I)$ is described as M-Closed if there is a belief that one of the models in (M_i) is the 'true' model, but it is not known which one. In this framework, a Bayesian would use probabilities $p(M_i)$ to represent prior beliefs about the 'truth' of model (M_i). In contrast, a set of models is described as M-Open if there is a belief that none of the models in (Mi) is the correct one[85].

Methods for quantifying structural uncertainty are less well described compared to methods for characterising parametric or methodological uncertainty[27, 29, 51, 85]. Whilst standard PSA will typically be undertaken to quantify parametric and methodological uncertainty, PSA does not take into account any structural uncertainty that may be present, potentially leading to

spuriously precise estimates of the model results[85]. This section discusses strategies that have been proposed to aid researchers in resolving the problem of model structural uncertainty.

4.5.1. Scenario analysis

Scenario analysis (sometimes considered a form of deterministic sensitivity analysis) is perhaps among the most widespread methods in the assessment of structural uncertainty. Scenario analysis requires the modeller to identify competing and credible model structures and to compute results for each alternative model specification, representing alternative sets of judgements or assumptions that are plausible. Briggs[82] has suggested running repeated analysis using different models with specified prior probabilities of different models across this model space. If different credible scenarios suggest conflicting decisions, then structural uncertainty is important. But, even if this is not the case, structural uncertainty will affect decision uncertainty and value of information in ways that are difficult to assess[51].

A weakness of this approach is that there are no established methods to formally assess the plausibility of alternative models and it is not clear which type of, or how many, scenarios should be considered. Furthermore, it is not possible to establish the value of conducting further research to resolve the source of these structural uncertainties[51]. Presenting multiple model structures to the decision maker may lead to inappropriate decisions by implicit averaging of alternative results[27].

4.5.2. Parameterisation

Some approaches to structural uncertainty have sought to parameterise the source of structural uncertainties into a model by adding uncertain parameters to represent the choice between multiple model scenarios. This approach, sometimes referred to as an expanded model[20], is analogous to model averaging, where weights are given to the parameterised structural assumptions depending on what prior knowledge is known about the likelihood of the scenarios representing the 'true' scenario. However, unlike model averaging where the objective is to synthesise all evidence on the structure of a decision model to assess if a treatment is cost-effective, by parameterising the uncertainty directly in the model, estimates of the value of further research on uncertain parameters can also be made. Parameterisation of structural uncertainty may be trivial for a nested structure, where a model is a restricted version of another, but challenging for non-nested model structures that require a re-building of the model[20].

For notational convenience, consider a model with four states. Given the transition probabilities from the multinomial likelihood for a homogeneous Markov model (where the transition parameters are independent of time and dependent only on an observed cycle length), assume homogeneity of the transition parameters across subjects (this is because the data are assumed to be in aggregate form) and let $p_{kij}(t)$ denote the probability of making a transition from state i to state j, where:

$i, j \in [1,2,3,4]$ where numbers 1 to 4 refer to the different health states in a model

The data are usually in aggregate form y_{kij}, which represents the total number of observed transitions from state *i* to state *j* in any given period, for each treatment *k*. The data are observed over the cycle period providing direct information on $p_{kij}(1)$. The likelihood for the data is given by:

$$(y_{ki1}, y_{ki2}, y_{ki3}, y_{ki4}) \sim Multinomial\ (p_{ki1}(1), p_{ki2}(1), p_{ki3}(1), p_{ki4}(1);\ n_{ki})$$

Subject to:

$$\sum_{j=1}^{4} p_{kij}(1) = 1, \quad 0 \leq p_{kij}(1) \leq 1, \quad \sum_{j=1}^{4} y_{kij} = n_{ki}$$

It is possible to parameterise structural uncertainty using transition rates and convert these to transition probabilities using Kolmogorov's forward equations, to conform to the multinomial likelihood[88] of the data:

$$\frac{d}{dt}\mathrm{p_k}(t) = \mathrm{p_k}(t).\mathrm{G_k}$$

Here $\mathrm{p_k}(t)$ is the transition probability matrix and G_k is the transition rate matrix. Transition rates have the advantage of being specified as hazard ratios, which are transportable across trials, rather than relative risks that depend on the Markov cycle length[88].

Parameterisation of structural uncertainties seems a preferred option if it is feasible to internalise structural uncertainty by adding parameters into the model. However, researchers have been advised to be explicit about the structural assumptions that might have an impact

on the findings and to suggest alternative assumptions that future modelling exercises might employ[20, 51].

4.5.3. Model averaging

The model averaging approach takes an M-Closed view and considers that $(\mathrm{M}_i, \mathrm{i} \in \mathrm{I})$ represents a set of plausible models, and the best approximation of the true value of the target parameter Ÿ corresponds to a weighted mean value of the individual model output.

Weights can be assigned with differing criteria, for example by choosing the model from the set that is believed to be the 'correct' while discarding the rest (effectively placing all the weight on a simple model), or alternatively assigning weights to the competing structures based on beliefs about how likely the different models are to be the 'correct', and weight the outputs by these probabilities. Weights could be based too on some measure of model adequacy, for example, AIC or the Bayesian framework[85].

Measures of model adequacy require data availability, D, to weight the model outputs by some function of the adequacy measure. Within a Bayesian framework prior model probabilities, $p(M_i)$, could be specified to calculate posterior probabilities via:

$$p(M_i|\mathrm{D}) = \frac{p(D|M_i)p(M_i)}{\sum_{i \in I} p(D|M_i)p(M_i)}$$

leading to a weighted mean output[85]

$$p(\ddot{Y}|\mathrm{D}) = \sum_{i \in I} p(\ddot{Y}|\, M_i, D) p(M_i|D)$$

It is important to note that model averaging methods can be used to assess structural uncertainty if a complete set of plausible competing models can be built and weighted according to some measure of model adequacy as previously described[28].

In practice, model averaging approach has some limitations. The first difficulty is that of the M-Closed assumption. In reality researchers believe that none of the models is correct, and they have instead an M-open set in which they do not believe Y=$\eta_i(X)$ for any i.

Another limitation is the form of the data available, D. Model averaging involves evaluating a likelihood function for each model, however, the data will not typically be in the form where a likelihood p(D|fi (X)) could be identified and this would have been the reason for building the model in the first place.

4.5.4. Model selection

An alternative to model averaging is to choose the 'best' model on the basis of some measure of predictable performance. The model selection criteria can take either M-Closed or M-Open view and consists of selecting a set of credible and plausible models which are then ranked according to measures of goodness-of-fit (representing the fit of the model predictions to the observed data). There exist a number of measures of goodness-of-fit to discriminate and select among alternative fitted models. For nested models measures like an F-test or

likelihood ratio test can be used. Nested and non-nested models can be ranked according to the Akaike information criterion (AIC), the Deviance information criterion (DIC), among others[89].

The AIC is calculated as:

$$AIC = -2\log(L) + 2P,$$

where *L* is the maximum likelihood estimate of parameters by the model fitted to the observed data, and *P* is the number of parameters in the model. Models are ranked comparing the value of the AIC for each model, the model with the best approximation (smallest value of AIC) is selected[29]. The maximised likelihood is a measure of fit to the sample data. The penalty size *2P* has an information-theoretic justification and ensures that over-complex models, which cannot be generalised outside the observed data are not chosen[27].

However, in an M-Open scenario, candidate models are chosen to approximate a true model of unknown form. The true model may be extremely complex, and more complex models would give better predictions in large samples but complex models would be over penalised by, for example, the AIC.

In the view of Jackson et al[32] an optimal model choice criterion could be derived by approximating the true model. In the context of health economics, models approximate the highly complex processes of progression of disease and response to treatment. An assumption is made that the true process underlying the data is too complex to be completely identified even in the presence of large sample data. Instead of using, for example, AIC, the

authors suggest taking a predictive approach by judging model adequacy by the expected utility of predicting a replicate data set *y* based on model *Mk* fitted to data *x*.

A Bayesian model assessment measure of this type is the DIC[32], a Bayesian extension to the AIC that can be used to compare the fit of the different models either nested or non-nested. The DIC is an estimate of an expected predictive utility $\mathrm{E}[\mathrm{U.y} \mid x, Mk/]$ defined as:

$D[yE.\theta\ /, x, Mk] = -2log[f(y|E.\theta\ /, x, Mk)]$, the predictive deviance based on 'plugging in' the expected parameter values of model *Mk*. This method assumes that the models under consideration are reasonable approximations to the true process, and that the posterior distribution of θ is approximately normal. A parameterisation should be chosen so that θ has an approximately normal posterior.

It has been seen so far that the best model, among a set of alternative models, is selected on the basis of how well a model's outputs match observed data for one or more model outputs. After putting each model through a validation process, the best-validated model is chosen. However, to some critics, even if models can be hypothetically ranked according to some criterion, it may not be advantageous to determine the "best" model. By choosing the best model, useful evidence may be discarded violating the requirements of the DAM process; in addition, uncertainty relating to the choice of the best model would also be ignored[51]. Last but not least, the need for additional resources to build alternative models has been identified as a limitation of this approach[29].

4.5.5. Discrepancy approach

This method assesses the difference between the model run as its 'best' or 'true' input, and the value of the output quantity. This method assumes that models will always be an imperfect description of reality, and makes judgements about the *discrepancy* between the model output and the true costs and consequences of interest defined via the following function:

$$Z = f(X) + \delta z$$

where δz corresponds to the discrepancy term. The beliefs about the discrepancy term are quantified via a weight based on a probability $p(\delta z)$. The aim is to assess how large an error might be due to the structure of the model at hand (as opposed to making assessments about a 'true' model structure)[85].

While PSA quantifies input uncertainty, the discrepancy approach quantifies uncertainty about the costs and health effects of the various decision options (uncertainty in the model structure). Rather than focusing on generating weights for models within some set, this method makes inferences about model 'discrepancy': the difference between the model run at its best or true input and the true value of the output quantity (structural error).

This method recognises that a given model may be deficient, not in the sense of mistakes, 'slips', 'lapses' or other errors of implementation, rather these errors refer to deficiencies arising as a result of the gap between any model of reality and reality itself[28].

For a decision maker to base their decision on the model output, the model must have credibility. The model must be judged to be sufficiently good to support the decision being made. The primary goal of this analysis is therefore to provide a means for quantifying judgements about structural error and specifically to determine the relative importance of structural compared with input uncertainty in addressing a decision problem. If it can be demonstrated that the uncertainty about the structural error is small in comparison with that due to input uncertainty, then a claim can be made that a model is credible[28].

Making meaningful judgements about the model discrepancy is difficult (and would require some form of elicitation of beliefs). For that reason this method proposes instead, to make judgements about this discrepancy indirectly, or at the sub-function level by decomposing the model into a series of sub-functions that link the model inputs to the model output. The idea is that, because each sub-function represents a much simpler process than the full model *f(X)*, making judgements about discrepancy in f_i will be easier than making judgements about discrepancy in *f( )*. It should be noted that not all sub-functions are required to have a structural error and similarly there is not a unique decomposition of the model f into a series of sub-functions. Once discrepancy terms have been introduced within a model, judgements about the size of the discrepancy should be made via the specification of the joint probability distribution $p(X, \delta)$ where it is assumed that discrepancies are independent of inputs henceforth the joint density $p(X, \delta) = p(X)p(\delta)$[28].

A main advantage of this approach is that only one model needs to be developed and this allows analysts to determine the relative importance of the alternative structural

assumptions and estimating EVPI surrounding decision uncertainty. However, the discrepancy approach requires a subjective estimation of the magnitude and variance of the discrepancy between model predictions from the sub-function and reality and the elicitation of expert judgements to quantify structural errors may introduce additional uncertainties[29].

4.6. Discussion

Structural uncertainty arises because of the nature of models being mere simplifications of reality. Many assumptions need to be adopted during the process of building a model such that the distance between a model and reality will always be unknown, or whether other plausible models would be a better alternative. This can potentially lead to a wide variation in model predictions with potential impact on funding decisions[27].

Even though the impacts of structural uncertainty on cost-effectiveness results have been thoroughly documented, the availability of guidelines providing in-depth guidance on how to characterise, address, and report this type of uncertainty seems less apparent and may explain the findings in Chapter 3 that the assessment of structural uncertainty is not common practice.

Various alternative statistical methods have been proposed to address the impact of structural uncertainty on the results of cost-effectiveness[20, 27-29, 31, 32, 51, 53, 82, 85, 88, 90, 91] whilst some other authors have provided examples on how to implement some of these methods in different clinical areas[87, 92, 93]. However, it has been recognised that methods for quantifying

structural uncertainty are less well described if compared to methods for characterising parametric or methodological uncertainty[27, 29, 51, 85].

A main challenge in addressing structural uncertainty is posed by the many issues that have been identified as 'structural uncertainty' making it a complex task (which may not even be cost-effective) to address properly[30]. Structural uncertainty will almost always arise once a choice of model structure or choice of relationships between inputs and outputs is defined within the model development process[85]. Therefore, it is essential to assess the extent to which model predictions are influenced by such choices within the model development process[29]. Consequently, in this thesis structural uncertainty is defined as the uncertainty associated with all aspects of model structure, i.e., health states and the relationships between health states. This is in contrast to parameter uncertainty, which is very much focused on the parameters used in a model and their uncertainty. Differences in model structure are dependent on the importance given to various aspects of the process being modelled, allowing in some instances for model simplifications. In some cases, these originate when data are not available, although their inclusion could potentially still be relevant for the analysis.

The challenges posed by the assessment of structural uncertainty might be overcome if additional research is undertaken on an experimental basis. Case studies aimed at measuring the impact of changing or adapting chosen model structures on previous results of cost-effectiveness could provide insightful evidence of how much results would be altered when alternative model structures are implemented. This would also provide evidence of what

other elements, besides the model structure, may be critical in affecting results of cost-effectiveness.

This thesis aims to contribute towards that end by presenting two case studies in Chapters 5 and 6. The aim is to assess the extent to which model predictions are influenced by the choice of model structure within the model development process. These chapters aim to provide practical illustrations of the impact of changing or adapting model structures on the results of cost-effectiveness.

The focus of the first case study is on the prevention of cardiovascular disease whilst the second case study focuses on the optimisation of treatment after a stroke. These case studies are focused on one particular clinical area, cardiovascular disease, since this makes it possible to remove some of the variation between models which is not relevant for the purpose of assessing structural uncertainty (for example, sources of uncertainty).

4.7. Conclusions

The aim of this chapter was to present an overview of structural uncertainty in DAM. This chapter has reviewed what has been understood by structural uncertainty. It has discussed the sources of structural uncertainty and methods that have been employed to deal with issues pertaining to structural uncertainty. Given the inherent uncertainty in estimates produced in DAM, the assessment of structural uncertainty is an essential part of the DAM process and should be considered in any model-building process as part of good practice guidelines.

Many issues has been identified as pertaining to structural uncertainty, additional guidelines are needed in order to aid researchers identifying what elements are of greater importance and how best to identify, address and report results of structural uncertainty.

Statistical methods to assess structural uncertainty have been proposed and developed, however, all of them have been subjected to criticism rendering uncertainty as to whether it is good value for money to attempt to address it.

The results of this overview will contribute towards the design and development of case studies aiming to explore the effect of choice of model structure in results of cost-effectiveness in the following two chapters. Chapters 5 and 6 of this thesis constitute research on the effects of structural uncertainty arising from the choice of model structure and are aimed at providing practical illustration, through two case studies, of the impact on results of cost-effectiveness of changing or adapting model structures. These case studies will also provide evidence on whether other elements besides the model structure affect the results of cost-effectiveness.

CHAPTER 5. THE IMPORTANCE OF MODEL STRUCTURE ON RESULTS OF COST-EFFECTIVENESS: A CASE STUDY OF PRIMARY CARE INTERVENTIONS FOR THE MANAGEMENT OF HYPERTENSION

Chapter 4 outlined the ways in which structural uncertainty has been described and understood within the decision analytic modelling process. This chapter aims to provide a practical illustration of the impact on the results of cost-effectiveness of changing or adapting model structures in a model-based economic evaluation. Structural uncertainty arising from the model structure is assessed for a Markov model first used to measure the cost-effectiveness of self-management of blood pressure as a strategy in the prevention of cardiovascular disease via management of hypertension. Evidence of whether other elements besides the structure of a model may affect the results of cost-effectiveness are also examined.

5.1. Introduction

High blood pressure (hypertension, defined as blood pressure persistently ≥140/90mmHg) is one of the most important but preventable causes of premature morbidity and mortality in the UK and worldwide[7, 8, 64]. Hypertension is a major risk factor for ischaemic and

haemorrhagic stroke, myocardial infarction (MI), heart failure (HF), chronic kidney disease (CKD), cognitive decline and premature death. It has been estimated for England that a 2 mmHg reduction in average systolic blood pressure for 40-69 year olds could save 1,500-2,000 lives per year[94]. One of the most common interventions in primary care is the management of hypertension. Self-management of hypertension, in which individuals monitor their own blood pressure and adjust their own medication, has been shown to lead to significantly lower blood pressure in hypertension, including individuals with higher cardiovascular risk[37, 39, 40].

Economic evaluations can be undertaken alongside randomised controlled trials (RCTs) where costs and health outcomes are measured. The primary outcome of RCTs in hypertension is often a change in blood pressure. However, a change in blood pressure corresponds to an intermediate outcome, and the final outcome of interest, in this case, is the risk of CVD. As RCTs rarely follow patients over the long-term or lifetime, Decision analytic modelling (DAM) provides a vehicle to extrapolate the impact of a change in blood pressure on the risk of CVD events in the long-term. Modelling the course of CVD can be challenging, requiring CVD risk factors (smoking, cholesterol, and diabetes), interactions among the risk factors, adverse events and the resulting health states (e.g. stroke sequelae and angina) to be considered.

Structural uncertainty is defined in this chapter as uncertainty associated with all aspects of model structure, i.e., health states and relationships between health states. This is in contrast to parameter uncertainty, which is very much focused on the parameters used in a model and their uncertainty. Structural uncertainty reflects the extent to which a given model differs from the real system it is intended to reflect[29, 33]. Differences in model structure are

dependent on the importance given to various aspects of the process being modelled, allowing in some instances for model simplifications. In some cases, these originate when data are not available, although their inclusion could potentially still be relevant for the analysis.

The fact that a model is a simplification of reality means that many assumptions need to be adopted during the model building process[84, 85, 95]. This can potentially lead to a wide variation in model predictions with potential impact on funding decisions [27].

Various alternative statistical methods have been proposed to address the impact of structural uncertainty on the results of cost-effectiveness (Chapter 4) whilst some other authors have provided examples on how to implement some of these methods in different clinical areas[87, 92, 93]. However, it has been recognised that methods for quantifying structural uncertainty are less well described when compared to methods for characterising parametric or methodological uncertainty[27, 29, 51, 85]. The main challenge in addressing structural uncertainty is posed by the many issues that have been identified as 'structural uncertainty' making it a complex task (which may not even be cost-effective) to address properly[30].

Previous studies[51, 67, 96, 97] indicate that even though elements pertaining to structural uncertainty are occasionally considered, the assessment of structural uncertainty is not common practice and most modelling tends to omit testing for structural uncertainty. However, it is essential to assess the extent to which model predictions are influenced by such choices made within the model development process[96].

Case studies aimed at measuring the impact of changing or adapting chosen model structures on previous results of cost-effectiveness could provide insightful evidence of how much results would be altered when alternative model structures are implemented. This would also provide evidence of what other elements besides the model structure might be critical in affecting the results of cost-effectiveness.

Structural uncertainty will almost always arise once a choice of model structure or choice of relationships between inputs and outputs is defined within the model development process[85]. Therefore it is essential to assess the extent to which model predictions are influenced by such choices within the model development process[29].

The TASMIN-SR[37] trial aimed to determine the effect of self-monitoring with self-titration (self-management) of antihypertensive medication on systolic blood pressure among hypertensive patients with suboptimal blood pressure control and pre-existing CVD, diabetes mellitus and CKD compared to usual care. The trial involved 552 patients at 59 UK primary care practices and it was conducted between March 2011 and January 2013.

A model-based economic evaluation was undertaken to assess the cost-effectiveness of the self-management intervention compared with usual care[40]. The main results indicated that self-management of blood pressure in high risk patients with poorly controlled hypertension not only reduced blood pressure compared to usual care, but also represented a cost-effective use of healthcare resources.

The case study developed in this chapter is based on the results of the TASMIN-SR study[40] and aims to provide a practical illustration of the impact, on results of cost-effectiveness, of changing or adapting model structures in a model-based economic evaluation on the primary prevention of CVD.

5.2. Description of the TASMIN-SR model

The economic evaluation in the Tasmin-SR study consisted of a model-based cost-utility analysis[40]. The aim of the model was to assess the long-term cost-effectiveness of the self-management intervention in a 'high risk' patient population compared with usual care. A Markov model was built to extrapolate the results of the TASMIN-SR trial[37], given in terms of blood pressure, to the long-term risk of cardiovascular endpoints. The study considered a cohort of 70 year old patients (39% female) with sub-optimal hypertension (BP>= 130/80 mmHg at baseline), combined with a history of stroke, diabetes, CHD, and CKD. The model was run over a lifetime time horizon using a six-month time cycle, with results presented from a UK National Health Service (NHS) and Personal Social Services (PSS) perspective.

The structure of the TASMIN-SR model is shown in Figure 5-1. Patients start in an initial 'HR' or high risk health state representing individuals with hypertension and a history of stroke, CHD, diabetes and CKD. The model simulates the lifetime of these patients until any of three possible events occur (stroke, myocardial infarction (MI) and unstable angina (UA)) or the patient dies from other causes. Individuals that survive an acute phase in any of the health

states progress into a post event or chronic phase for that condition until death, with no recurrences of cardiovascular events being possible. A lower quality of life was permanently applied until death in all chronic health states.

The CVD history of patients entering the model was informed by the TASMIN-SR [37] trial data. Transition probabilities of suffering a stroke, MI, or UA were obtained from the literature for each of the high risk conditions. Age-related risk reductions from treatment for MI, UA, and stroke were estimated using trial based systolic blood pressure reductions at 6 and 12 months (Appendix 5). Resource use and costs were obtained from trial data and published studies (Appendix 6).

5.3. Methods

Taking the TASMIN-SR as the case study, the research methods of this chapter are outlined as follows:

- systematic review to identify plausible alternative model structures to TASMIN-SR
- definition and implementation of changes to the structure of the TASMIN-SR model
- inclusion of secondary events in the TASMIN-SR model
- identification of alternative model inputs

5.3.1. Alternative model structures

Structural uncertainty was addressed here by assessing issues such as the adequacy of the type of model used (Markov), the structure of the model (health states and transition

probabilities) that translates into plausible alternative model structures, and data availability to inform input parameters, for example the risk of secondary events.

The results of the systematic review developed in Chapter 3 were used to inform plausible alternative model structures in this chapter (see Section 3.2). Thirteen model-based economic evaluations identified from the literature in Chapter 3 were used to inform the changes implemented to the TASMIN-SR model. A data extraction instrument was developed to extract information from each paper. This included the inclusion or exclusion of potentially relevant comparators, type of model used, health states included, recurrence of events, choice of covariate effects used in the transition probabilities, and the inclusion or exclusion of any other assumption(s) pertaining to structural uncertainty (see Table 5-1).

The review indicated that all thirteen included studies used Markov models but only two justified their use. Kourlaba[74] justified the use of a Markov model in his study by saying that it is 'a conventional model that describes restricted transition probabilities between important health states' (p.87). Kaambwa[39] indicated that 'the Markov model overcame limitations associated with within-trial analyses' (p.1527). In the TASMIN-SR[40] study, it was stated that 'arguably, a more complex model such as individual patient level simulation could be more appropriate' (p.9) by incorporating patients' histories more efficiently. The use of Markov models can overcome limitations associated with within-trial analyses, specifically by allowing the modelling of effects and costs of long-term events and the assessment of the long term cost-effectiveness beyond the trial period[9]. Even though individual patient level simulation models have long been praised for their flexibility and ability to record patient

attributes[41], because cardiovascular diseases are chronic with recurring events and often result in health states with persistently reduced quality of life, the use of a Markov model is often preferred as a more parsimonious approach[13, 49].

Table 5-1 Data extraction instrument for the assessment of structural uncertainty

Author	Inclusion/exclusion of potentially relevant comparators	Health states included/excluded, recurrence of events; type of model	Inclusion/exclusion of other assumptions affecting the structure of the model
Kaambwa et al[39]	Yes	Yes	Yes
	Authors argued that self-monitoring of hypertension (as a means to lower blood pressure) has been largely evaluated; previous CE results found to be inconsistent plus not been extrapolated to the longer term. Their study examined the long-term cost-effectiveness of self-monitoring combined with self-titration (i.e., self-management) of blood pressure	Four acute health states (Stroke, MI, Angina, and HF) and death were considered. It was not mentioned how health states were identified; authors acknowledged to have made an assumption that CHD consisted of MI, HF and angina (this was reflected in the structure of their Markov model); the risk of secondary events, including progression of disease, was not modelled and was acknowledged as a weakness	Adverse effects such as anxiety or drug side effects were not modelled due to lack of data, however, trial data found minimal differences; effectiveness of the intervention after the year of the trial was unknown however the effect of various potential reductions in efficacy was tested in SA. Lifetime time horizon was tested in SA
Stevanovic et al[72]	Yes	Yes	Yes
	Authors argued that health and economic consequences of newer anti-hypertensive agents such as ACEIs and ARBs were not available at the time of the study in Netherlands. As a result, authors compared HCT 25 mg (diuretics) versus HCT/ACEIs versus HCT/ARBs versus no treatment	One acute health state (Acute CVD), a chronic health state (Stable CVD) and death were considered. The inclusion of states in the Markov model was not justified. Risk of secondary events was assumed to be equal to the risk of a first non-fatal CVD event. This assumption was acknowledged to lead to an under-estimation of the CVD risk and compensated with the adoption of an increased risk of death in patients experiencing non-fatal CVD events	Adverse effect(s) from antihypertensive treatment was not considered; large uncertainty ranges around the expected values of the SCORE input parameters (model for ten year risk of fatal cardiovascular disease) used in the model for both 10-year and lifetime horizons, as tested through PSA and ANCOVA analyses

Author	Inclusion/exclusion of potentially relevant comparators	Health states included/excluded, recurrence of events; type of model	Inclusion/exclusion of other assumptions affecting the structure of the model
Wu et al[73]	Yes	No	No
	Comparators resulted from the results of a meta-analysis study indicating that Norvasc (Amlodipine) was superior to ARBs in the prevention of stroke and MI in hypertensive patients.	Two acute (Stroke and MI) and its corresponding chronic health states were considered. No justification was given for the inclusion of states in their Markov model; authors did not discuss the possibility of recurrent events, however they acknowledged as a weakness in the model not including the risk of patients having both stroke and MI due to lack of data	Even though an assumption was adopted that the risk of stroke or MI and the mortality risk during the lifetime of the model (5-years) will remain fixed, this assumption was not tested in SA
Kourlaba et al[74]	Yes	No	No
	Comparators resulted from answering the research question in light of recent guidelines in Greece for the use of combined therapy to treat hypertension	Two acute (MI and Stroke) health states and its corresponding chronic health states were modelled. No movement from MI to stroke was assumed; it was acknowledged as a limitation; risk of secondary events in their Markov model was not considered; same risk of CVD death was assumed (independently of whether a patient has experienced a previous CVD). None of these assumptions was tested in SA	No evidence or discussion presented on this respect
Ekwunife et al[75]	Yes	Yes	No
	Comparators were identified from hypertension guidelines in Nigeria	Two acute health states (Stroke and CHD) were modelled and two chronic post event health states. The model reflected the pathway of patients with hypertension starting in an asymptomatic health state, and then moving to a cardiovascular state (CHD or stroke) and death. The authors did not consider secondary events	

Author	Inclusion/exclusion of potentially relevant comparators	Health states included/excluded, recurrence of events; type of model	Inclusion/exclusion of other assumptions affecting the structure of the model
		and this was not discussed. The authors used a Markov model	
Wisloff et al[76]	Yes	Yes	No
	Alternatives were aimed at contribute towards the discussion around intervention thresholds and the choice of first-line drug and 'add on' drugs	Four acute health states (Stroke, AMI, Angina and HF) and two post event health states (Post-Stroke and Post-CVD) were considered. Health events in the Markov model reflected the asymptomatic stages, cardiovascular life and death of patients; the model allowed for secondary events after which the model assumed patients will move to the worst health state; some assumptions regarding risk of secondary events were based on expert opinion. These assumptions were not tested in SA	The authors used observed incidence rates to reflect risk factors using registry data; this was acknowledged as a limitation however was not tested in SA
Granstrom et al[79]	Yes	Yes	No
	Justified on the grounds that no head to head randomised comparative studies were previously performed comparing Candesartan and Losartan; authors acknowledged as a limitation that there may be ARB comparators more relevant to Candesartan than Losartan in other health care setting	The authors considered health states (HF, PAD and Arrhythmia), post event health states (post-MI and post Stroke) and a chronic health state (IHD). Health states in the Markov model, including post MI and post stroke were based on CVD events measured in a registry study (authors commented and acknowledged a potential risk of confounding); after a CVD event an increased risk of mortality was applied (no SA to test for these assumptions)	
Baker et al[80]	Yes	Yes	Yes

Author	Inclusion/exclusion of potentially relevant comparators	Health states included/excluded, recurrence of events; type of model	Inclusion/exclusion of other assumptions affecting the structure of the model
	Comparators were justified in view of the concerns surrounding non-medical ARB switching after Simvastatin became a generic product leading to a number of patients being switched from branded atorvastatin to generic Simvastatin for economic rather than medical reasons	Health states were designed to reflect the course and history of CVD events in a typical patient with hypertension (CVD event free, post CVD, and death). Although secondary CVD events were not explicitly considered in the model, patients in the post-event state were subject to an increased risk of death reflecting their disease state	The model assumed that Valsartan remained on patent for the first 2.75 years of the model time horizon and Losartan for only 4 months after which generic formulations would become available. No side effects were modelled, which was acknowledged as a weakness
Perman et al[71]	Yes	Yes	Yes
	Justified on the basis that previous evidence was favourable for hypertension programmes as compared to drug treatments	The health states were: an acute MI event, no event and death. The inclusion/exclusion of health states in the Markov model was not discussed but rather just introduced; an interesting assumption in the model was that patients presenting an acute CVD event could have or not have hospital attention? Risk of secondary events was considered. This was not tested in SA	The discount rate was considered as a structural variable and thus analyses were performed with different discount rates ranging from 0%-12%
Ekman et al[77]	Yes	No	Yes

Author	Inclusion/exclusion of potentially relevant comparators	Health states included/excluded, recurrence of events; type of model	Inclusion/exclusion of other assumptions affecting the structure of the model
	The comparators corresponded to those found to have mild side effects as per previous clinical trials; it was acknowledged as a weakness not to include other comparators such as diuretics	Four acute health states (Angina, MI, CHF and stroke), a post MI, and post stroke health state were modelled. No particular explanation was given for the inclusion/exclusion of health states. The Markov model assumed that patients may undergo revascularization procedures while in the MI or angina health states; recurrence of events was modelled however acknowledged that data was limited to reflect how risks of recurrent strokes or MIs vary depending on various disease histories.	Treatment effects were supposed to last five years; SA tested the sensitivity of the model to changes in the duration of the antihypertensive treatment and variation of discount rate (between 0% to 8%) and measure of LYG instead of QALYs
Gandjour et al[69]	Yes	No	No
	Comparators resulted from the research question which is whether the health service can reduce the underuse of hypertensive medication among the German population through a national programme	Three acute health states (MI, stroke and renal failure) were modelled. No particular reason argued for the inclusion/exclusion of health states in the Markov model; treatment and its effect were assumed to last a lifetime; secondary events were not modelled but captured through the mortality rates of patients after a CVD event. Model assumptions were not tested in SA	
Montgomery et al[70]	Yes	No	No
	Comparators resulted from the research question which is whether incorporating patients' preferences into the decision-making process may have an important influence on treatment recommendations for individual patients	A single acute CVD health state was considered with variations to account for the impact of side effects and treatment or lack of it. No explanation was given for the inclusion/exclusion of health states in the Markov model; secondary events were not modelled and the assumption that any second cardiovascular event was	

Author	Inclusion/exclusion of potentially relevant comparators	Health states included/excluded, recurrence of events; type of model	Inclusion/exclusion of other assumptions affecting the structure of the model
		fatal was adopted. Model assumptions were not tested in SA	
Nordmann et al[78]	Yes	No	No
	Comparators were chosen in line with hypertension guidelines for first-line antihypertensive therapy from both, the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (JNC-VI) and the World Health Organization (WHO)	Three acute health states (CAD, CVD and CHF) were considered. Authors argued to have included the most common CVD outcomes as health states in the Markov model; the model allowed one opportunity to switch from conventional therapy to ACE inhibitors in response to adverse effects or lack of efficacy; recurrence of events was modelled	

The complexity of the different model structures considered in this analysis was varied, this being due to the different approaches to the inclusion of the acute or post-event health states modelled. Model structures were most frequently a reflection of the course and history of CVD events or disease progression. The most common initial state was disease-free and the most common acute states modelled were stroke and MI followed by angina, heart failure and CVD. Few studies modelled only a single health state to describe an acute cardiovascular event[70-72]. Some studies modelled additional states such as congestive HF[77], coronary artery disease[78], renal failure[69] or peripheral artery disease[79]. Absorbing states consisted of death and non-CVD death. Some studies acknowledged they had excluded states[69, 78] or combinations of health states (HF and stroke)[76] due to data limitations. Compared to the TASMIN-SR model, the review identified a variety of model structures ranging from a simplistic (single CVD morbidity)[71, 72, 80] to more complex approaches (four states including stroke, MI, HF, angina)[76, 77, 79].

The risk of secondary events was modelled in seven[70-72, 76-79] of the studies reviewed. It could be argued that some of these studies adopted assumptions which would add extra uncertainty to the results. These included assuming that the risk of secondary events was equal to the risk of a first non-fatal event[72], assuming that the patient with a second event will be in a health state worse than the state prior to the event[76], or using expert opinion to inform risks of secondary events[71]. Lack of epidemiological data was acknowledged as the main reason for the exclusion of secondary events by some authors[39, 80]. The TASMIN-SR model assumed no recurrence of CV events due to the lack of data describing secondary events for a high risk population. The choice of modelling approach should be considered as

part of the investigation into the impact of structural uncertainty. However in this Chapter, only the Markov model structure is considered as informed by the results of the review.

5.3.2. Implementation of model structures

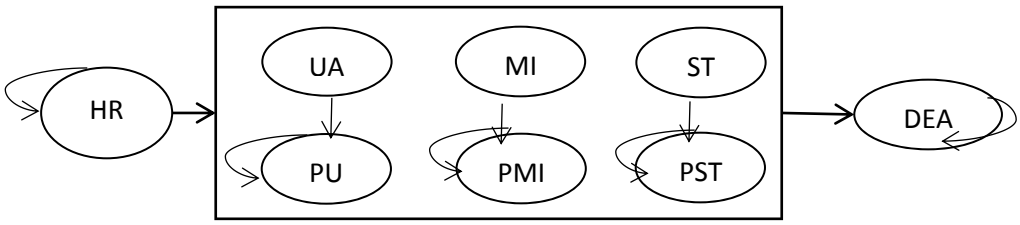
Alternative Markov model structures were primarily identified based on the findings of the systematic review. Validation of the adequacy of this type of model over competing structures such as decision trees, DES or individual sampling model was checked using a framework to select the appropriate model type[13]. The validation check indicated that Markov was just the right type of model, considering that estimating interactions between individuals was not necessary whilst modelling health states was important (patient pathways would not be adequately represented by decision trees) and an excessive number of health states was not required (excluding the option of individual sampling model).

These alternative model structures, were labelled Model 1 through Model 3, were developed by varying the number of health states used from a simplistic structure to one of increased complexity (Figure 5-1). Model 1 uses a simplified approximation of the TASMIN-SR model structure. It was informed by Stevanovic[72] and consists of a single CVD state with progression to a chronic state (Figure 5-1). Following NICE Statin guidelines[98] it was assumed that CVD is a combined state consisting of CHD (MI and UA) and stroke. Assumptions were adopted to estimate transition probabilities, utilities and costs due to lack of data in the literature to inform a single CVD health state (Table 5-2). Parameters for the CVD state correspond to a

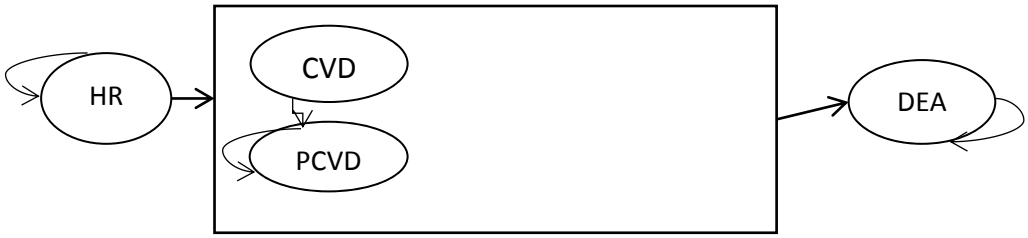
weighted average of input parameters used in the case study model for the states stroke, MI and UA (Table 5-2).

Model 2 applied the assumption that if the costs and utilities for two health states are the same, then it may not be necessary to distinguish between those two states to estimate lifetime costs and effectiveness (Figure 5-1)[32].

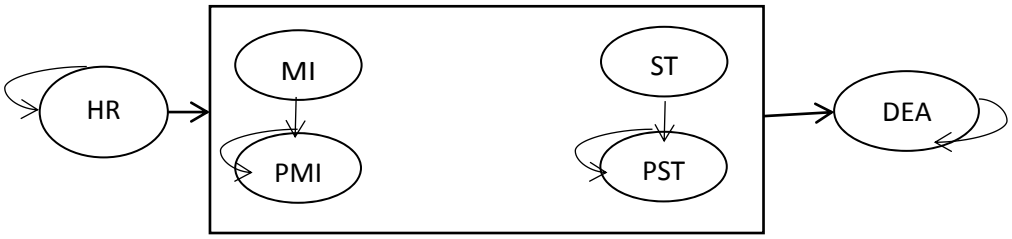
TASMIN-SR model



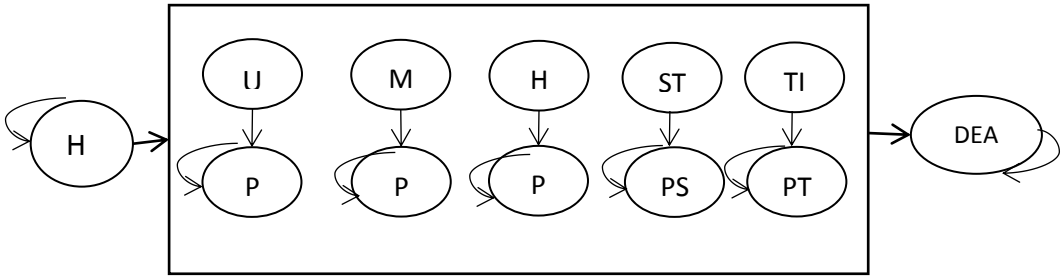
Model 1 Single state structure



Model 2 Two health state structure



Model 3 Expanded structure



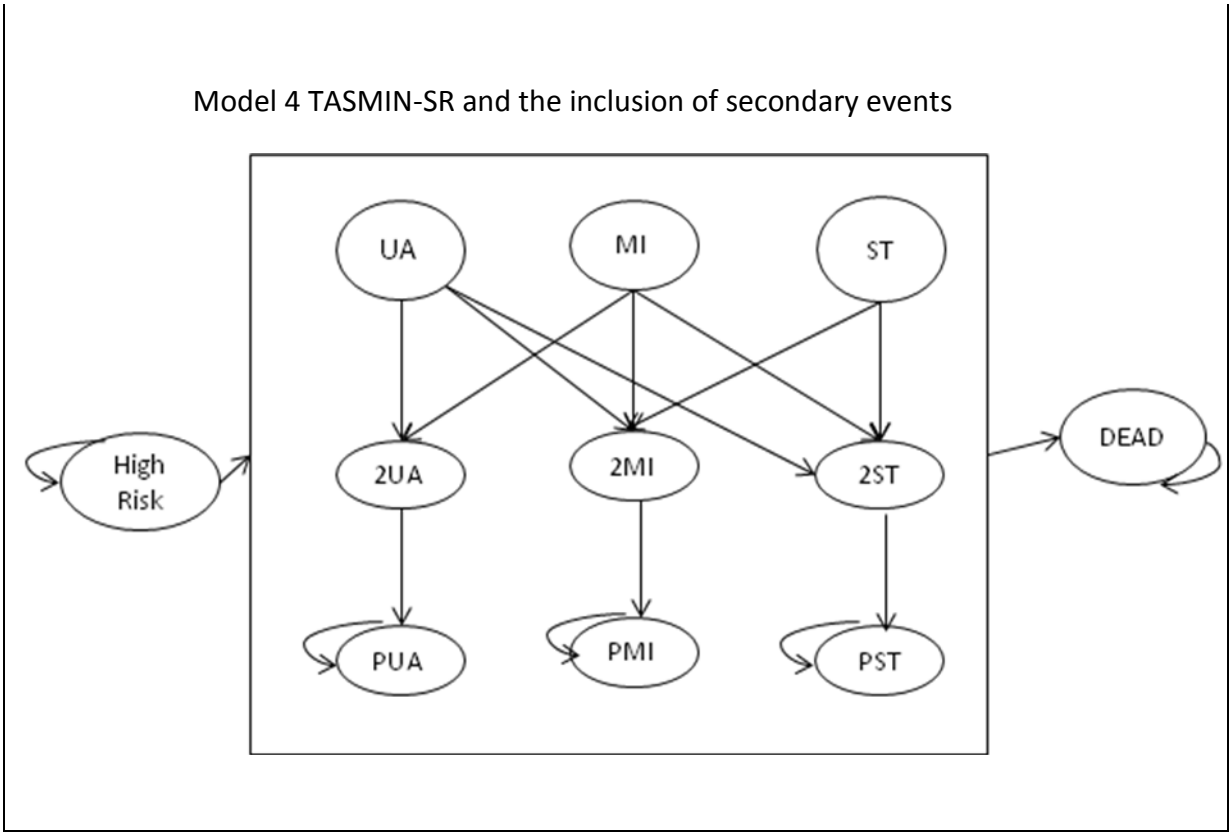


Figure 5-1 Model structures for the TASMIN-SR and models 1 - 4

HR = High Risk, UA= Unstable Angina; MI= Myocardial Infarction; HF = Heart Failure; ST = Stroke; TIA = Transient Ischemic Attack; CVD = Cardiovascular Disease.

Names preceded by a 'P', for example, PMI refers to a post event (chronic) health state for a patient surviving an event (MI)

Names preceded by '2', for example, 2UA refers to the occurrence of a second event consisting of a UA

Patients can move to the 'Dead' state from any of the health states in the models

In the TASMIN-SR model, treatment effects and the long-term costs and utilities for states MI and UA were assumed to be the same due to lack of data on UA (Table 5-2). Under these circumstances, it may not be necessary to include a state UA. Model 2 reflects a restricted version of the TASMIN-SR model consisting of two health states Stroke and MI. The review identified studies using a model structure consisting of two states, named stroke and a MI[73, 74] or stroke and CHD[75]. Model 2 was implemented consisting of health states stroke and MI with progression to a chronic phase for individuals who survive (Figure 5-1 and Table 5-2).

Model 3 adopted an expanded structure that was informed by the structure of the most complex models[76, 77, 79], using an increased number of health states (Table 5-1). In Model 3, high risk patients can move to one of a number of primary CVD events, MI, stroke, HF, UA and transient ischemic attack (TIA) or dead from CVD or other causes. Individuals that survive an acute CVD phase naturally progress to a chronic phase where quality of life is lower and where they remain until death.

5.3.3. Inclusion of secondary events

TASMIN-SR did not consider recurrence of cardiovascular events due to the lack of suitable epidemiological data to reflect the transition of elderly and high risk patients after a primary cardiovascular event. After carefully reviewing sources of data and literature, including relevant NICE guidelines, no additional suitable data were identified. Therefore in this chapter, assumptions based on expert clinical advice were adopted. In Model 4, individuals that survive a primary acute event can either move into a chronic post event phase

(asymptomatic) or may experience a recurrent cardiovascular event one year after the first event. In Model 4 it was assumed that patients will experience only one cardiovascular event per year and following a primary event, patients may experience a second event one year after the first event with the same probability as for the first event. Transitions from a more severe health state (e.g. stroke) to a less severe state (e.g. unstable angina) were omitted from the model because such transitions would imply lower costs and improvements in quality of life that may not reflect clinical reality (Figure 5-1)[99].

5.3.4. Model input parameters and analysis

Information from the literature was sought to populate all input parameters for models 1 to 4 for a UK setting (Table 5-2). When information on transition probabilities or age-related relative risks was not readily available, figures were estimated using a weighted average based on the distribution of patients to primary CVD events[40]. Costs were derived from a combination of standard unit costs[100, 101] and previously published literature and models[101, 102], and were adjusted using the Hospital and Community Health Service index to the price year of 2014/15[100]. The acute and chronic costs of CVD were estimated using a weighted average based on the distribution of patients to primary CVD events[102]. The probabilities of death due to cardiovascular events within a year of the event are reported in Table 1 and were applied to the first year after an event (first two cycles in the model). Life tables were used to determine the overall mortality for each model dependent on age and gender[103]. Risks

of death following a second event and utility values following a second event used in Model 4 were taken from the literature (Table 5-2).

A cost-utility analysis was undertaken for all models to calculate the cost per quality-adjusted life year (QALY) gained. Results from each alternative model specification are presented as scenario analysis. Deterministic and probabilistic sensitivity analysis was conducted to assess parameter uncertainty. The PSA was run with 10,000 Monte Carlo simulations allowing cost-effectiveness planes (CEP) and cost-effectiveness acceptability curves to be constructed to estimate the probability of self-management being cost effective at different willingness-to-pay thresholds.

Sensitivity analyses were conducted to assess uncertainty in the results of each model (TASMIN-SR and Model 1-4). Deterministic sensitivity analysis was undertaken around key parameters and assumptions. The time horizon for each model was varied from 30 years (lifetime) to 10, 5, 3, 2 and 1 year to determine whether the intervention was cost-effective in the shorter and the long-term. All cost variables were increased by 40% and 200% or decreased by 40% and 50%. Additional sensitivity analyses for Model 4 was undertaken to examine the impact of doubling or halving the probabilities of having a second cardiovascular event.

The impact of structural uncertainty was presented in terms of the impact on the cost-effectiveness results of a model and expected value of perfect information (EVPI). Including different parameters in the model can be expected to alter the extent of uncertainty captured

in the EVPI calculation. Because the models have different parameter sets, comparisons of expected value of partial perfect information would not be helpful.

Table 5-2 Input parameters used in the case study and each one of the alternative model structures

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
Annual CVD events for patients with DM						
Stroke						
60-69 years old	0.0196		0.0196	0.0196	0.0196	NICE, Diabetes guidelines[104]
70-79 years old	0.0262		0.0262	0.0262	0.0262	
80-89 years old	0.0298		0.0298	0.0298	0.0298	
MI						
60-69 years old	0.0089		0.0089	0.0089	0.0089	NICE, Diabetes guidelines[104]
70-79 years old	0.0100		0.0100	0.0100	0.0100	
80-89 years old	0.0111		0.0111	0.0111	0.0111	
UA						
60-69 years old	0.0041			0.0041	0.0041	NICE, Diabetes guidelines[104]
70-79 years old	0.0047			0.0047	0.0047	
80-89 years old	0.0052			0.0052	0.0052	
TIA						
60-69 years old				0.0053		NICE, Diabetes guidelines[104]
70-79 years old				0.0059		
80-89 years old				0.0066		
HF						
60-69 years old				0.0197		NICE, Hypertension Guidelines[64]
70-79 years old				0.0236		

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
80-89 years old				0.0264		
CVD						
60-69 years old		0.0323				Added risks*
70-79 years old		0.0405				
80-89 years old		0.0456				
Annual CVD events for patients with CKD						
Stroke						
60-69 years old	0.0072		0.0072	0.0072	0.0072	Kerr et al[105]
70-79 years old	0.0147		0.0147	0.0147	0.0147	
80-89 years old	0.0189		0.0189	0.0189	0.0189	
MI						
60-69 years old	0.0051		0.0051	0.0051	0.0051	Kerr et al[105]
70-79 years old	0.0113		0.0113	0.0113	0.0113	
80-89 years old	0.0171		0.0171	0.0171	0.0171	
UA						
60-69 years old	0.0024			0.0024	0.0024	Kerr et al[105]
70-79 years old	0.0054			0.0054	0.0054	
80-89 years old	0.0081			0.0081	0.0081	
TIA						
60-69 years old				0.0600		Koren-Morag et al[106]
70-79 years old				0.1303		
80-89 years old				0.1867		

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
HF						
60-69 years old				0.0269		Shiba et al[107]
70-79 years old				0.0585		
80-89 years old				0.0838		
CVD						
60-69 years old		0.0146				Added risks*
70-79 years old		0.0311				
80-89 years old		0.0435				
Annual CVD events for patients with a previous stroke						
Stroke						
60-69 years old	0.0348		0.0348	0.0348	0.0348	PROGRESS (1999) & NICE, Lipid modification guidelines[108, 109]
70-79 years old	0.0590		0.0590	0.0590	0.0590	
80-89 years old	0.0715		0.0715	0.0715	0.0715	
MI						
60-69 years old	0.0139		0.0139	0.0139	0.0139	PROGRESS (1999) & NICE, Lipid modification guidelines[108, 109]
70-79 years old	0.0232		0.0232	0.0232	0.0232	
80-89 years old	0.0232		0.0232	0.0232	0.0232	
UA						
60-69 years old	0.0139			0.0139	0.0139	PROGRESS (1999) & NICE, Lipid
70-79 years old	0.0232			0.0232	0.0232	
80-89 years old	0.0232			0.0232	0.0232	

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
						modification guidelines[108, 109]
TIA						
60-69 years old				0.5000		Hankey GL (2003)[110]
70-79 years old				0.0848		
80-89 years old				0.1027		
HF						
60-69 years old				0.0115		NICE, Hypertension guidelines[64]
70-79 years old				0.0193		
80-89 years old				0.0207		
CVD						
60-69 years old		0.0615				Added risks*
70-79 years old		0.1022				
80-89 years old		0.1141				
Annual CVD events for patients with CHD						
Stroke						
60-69 years old	0.0348		0.0348	0.0348	0.0348	NICE, Lipid modification and Hypertension guidelines[64, 108]
70-79 years old	0.0590		0.0590	0.0590	0.0590	
80-89 years old	0.0715		0.0715	0.0715	0.0715	
MI						
60-69 years old	0.0666		0.0666	0.0666	0.0666	

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
70-79 years old	0.1112		0.1112	0.1112	0.1112	NICE, Lipid modification and Hypertension guidelines[64, 108]
80-89 years old	0.1112		0.1112	0.1112	0.1112	
UA						
60-69 years old	0.0528			0.0528	0.0528	NICE, Lipid modification and Hypertension guidelines[64, 108]
70-79 years old	0.0882			0.0882	0.0882	
80-89 years old	0.0882			0.0882	0.0882	
TIA						
60-69 years old				0.0499		NICE, Lipid modification guidelines[108]
70-79 years old				0.0820		
80-89 years old				0.1046		
HF						
60-69 years old				0.0304		NICE, Lipid modification guidelines[108]
70-79 years old				0.0512		
80-89 years old				0.0653		
CVD						
60-69 years old		0.1467				Added risks*
70-79 years old		0.2373				
80-89 years old		0.2475				
Probability of death for those who have suffered an event						

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
Fatal Stroke	0.23		0.23	0.23	0.23	Bamford et al[111]
Fatal MI						
65-74 years old	0.23		0.23	0.23	0.23	ONS, Deaths registry & Kerr et al[103, 105]
75-84 years old	0.39		0.39	0.39	0.39	
85 and over	0.52		0.52	0.52	0.52	
Fatal TIA				0.11		Mant et al & Gattellary et al[112, 113]
Fatal HF						
Male				0.17		NorCAD model (2008)[99]
Female				0.16		
Fatal CVD						
65-74 years old		0.20				Weighted average†
75-84 years old		0.25				
85 and over		0.29				
Probability of death from a second cardiovascular event, one year after the first event						
Stroke after a first stroke					0.34	NICE, Statins guidelines[102]
UA after first UA					0.02	
MI after first MI					Same as first year event	
Age-related relative risks at 12 months						
MI, UA and HF – self-management						

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
60-69 years old	0.63		0.63	0.63	0.63	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.69		0.69	0.69	0.69	
80-89 years old	0.75		0.75	0.75	0.75	
Stroke and TIA – self-management						
60-69 years old	0.54		0.54	0.54	0.54	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.59		0.59	0.59	0.59	
80-89 years old	0.75		0.75	0.75	0.75	
CVD – self-management						
60-69 years old		0.60				Weighted average†
70-79 years old		0.65				
80-89 years old		0.75				
MI, UA and HF - usual care						
60-69 years old	0.82		0.82	0.82	0.82	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.85		0.85	0.85	0.85	
80-89 years old	0.88		0.88	0.88	0.88	
Stroke and TIA - usual care						
60-69 years old	0.76		0.76	0.76	0.76	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.81		0.81	0.81	0.81	
80-89 years old	0.88		0.88	0.88	0.88	
CVD - usual care						
60-69 years old		0.80				Weighted average†
70-79 years old		0.83				

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
80-89 years old		0.88				
Age-related relative risks at 6 months						
MI, UA and HF – self-management						
60-69 years old	0.71		0.71	0.71	0.71	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.75		0.75	0.75	0.75	
80-89 years old	0.80		0.80	0.80	0.80	
Stroke and TIA – self-management						
60-69 years old	0.62		0.62	0.62	0.62	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.68		0.68	0.68	0.68	
80-89 years old	0.80		0.80	0.80	0.80	
CVD – self-management						
60-69 years old		0.68				Weighted average†
70-79 years old		0.72				
80-89 years old		0.80				
MI, UA and HF - usual care						
60-69 years old	0.83		0.83	0.83	0.83	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.85		0.85	0.85	0.85	
80-89 years old	0.89		0.89	0.89	0.89	
Stroke and TIA - usual care						
60-69 years old	0.77		0.77	0.77	0.77	TASMIN-SR trial & Law et al[37, 114]
70-79 years old	0.81		0.81	0.81	0.81	

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
80-89 years old	0.89		0.89	0.89	0.89	
CVD - usual care						
60-69 years old		0.80				Weighted average†
70-79 years old		0.84				
80-89 years old		0.89				
Costs of acute disease one-off cost (UK 2014/15 £)						
Stroke	11,433		11,433	11,433	11,433	Youman et al[115]
MI	5,693		5,693	5,693	5,693	Palmer et al[116]
UA	3,416			3,416	3,416	Assumed 60% of MI
TIA				1,715		NHS Reference costs 2013-14[101]
HF				2,797		NHS Reference costs 2013-14[101]
CVD		7,235				Weighted average‡
Costs for long-term (chronic) disease per year (UK 2014/15 £)						
Stroke	2,823		2,823	2,823	2,823	Youman et al[115]
MI	593		593	593	593	Cooper et al[117]
UA	593			593	593	Cooper et al[117]
TIA				333		NICE, Statins guidelines[102]
HF				1,274		Stewart et al[118]
CVD		1,432				Weighted average†

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
Utilities for initial health states						
Self management & Usual care						
65-74 years old	0.81	0.81	0.81	0.81	0.81	TASMIN-SR trial[37]
75-84 years old	0.74	0.74	0.74	0.74	0.74	
85 and over	0.71	0.71	0.71	0.71	0.71	
Utilities for acute events						
UA	0.77			0.77	0.77	NICE, Lipid modification, Hypertension and Statins guidelines; TASMIN-SR trial[37, 64, 102, 108]
MI	0.76		0.76	0.76	0.76	
Stroke	0.63		0.63	0.63	0.63	
TIA				0.90		
HF				0.68		
CVD		0.76				
Stroke after stroke					0.479	Ara, et al[119]
UA after UA					0.615	
MI after MI					0.700	
MI and Stroke					0.479	
Angina and Stroke					0.596	
Angina and MI					0.541	
Utilities for long term (chronic) disease						
UA	0.88			0.88	0.88	NICE, Lipid modification and Statins guidelines,
MI	0.88		0.88	0.88	0.88	
Stroke	0.63		0.63	0.63	0.63	

Parameter	TASMIN-SR	Model 1	Model 2	Model 3	Model 4	Sources
TIA				0.90		TASMIN-SR trial[37, 102, 108]
HF				0.68		
CVD		0.78				NICE, hypertension guidelines[64]
Dead	0	0	0	0	0	By definition
Annual discount rate for costs and utility	0.035	0.035	0.035	0.035	0.035	Gray et al[1]

*The probability of CVD was estimated as the added risks of the individual risk probabilities for stroke, MI and UA

†Weighted averages were estimated based on the distribution of patients to primary event health states in the ScHARR economic model

‡Weighted average using TASMIN-SR trial data

5.4. Results

The main cost-effectiveness results obtained in the TASMIN-SR study were found to be robust to changes in model structure (Table 5-3) and to the inclusion of secondary events. Self-management of blood pressure remained dominant (more effective and cheaper than usual care) for all models. Expected costs for each intervention were noticeable higher for Model 4 while the expected QALYs for both interventions were noticeably lower for Model 3.

The highest QALY outcomes for both interventions were found by implementing Model 2 (restricted version). Higher incremental QALYs were found for Models 3 and 4 between self-management and usual care. Differences found between incremental QALYs for TASMIN-SR and Models 2 and 3 were marginal (0.0001 and 0.0002 respectively) (Table 5-3).

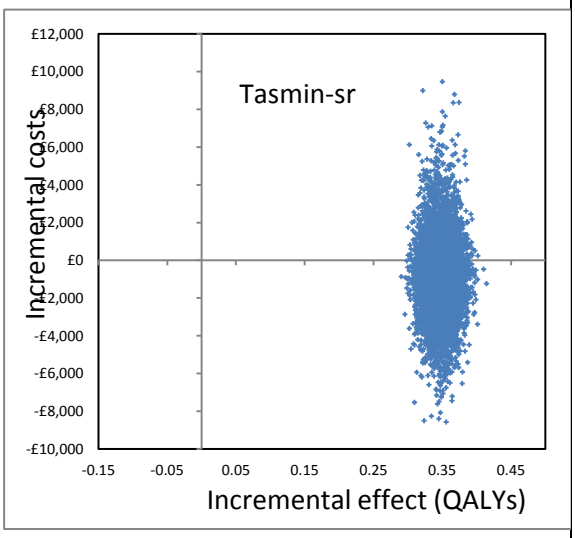
The cost effectiveness plane (CEP, Figure 5-2) shows the results from the Monte Carlo simulation for 10,000 replications. All the results were in the north-east and south-east quadrants indicating that self-management was always more effective but may be more or less costly. The cost-effectiveness acceptability curves (CEAC) shown in Figure 5-3 were derived from the joint density of incremental costs and incremental QALYs for the self-management of blood pressure. Each CEAC presents the probability that the self-management intervention is cost-effective for the different model structures. For a willingness to pay of £20,000 per QALY, the proportion of model replications that were cost-effective was higher than 99% for all model structures (Figure 5-3).

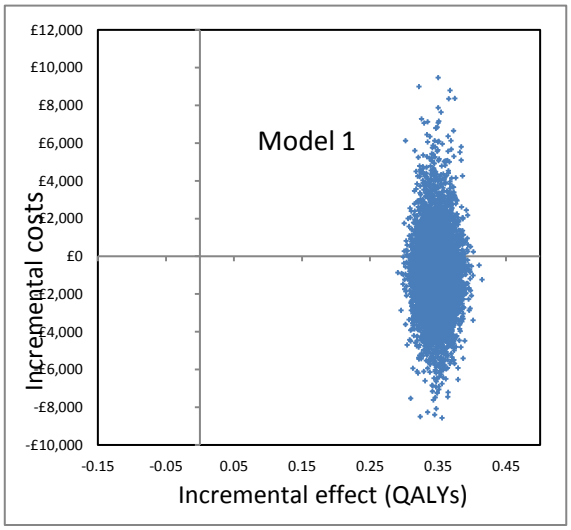
All sensitivity analyses undertaken appear to indicate that individual results for the various models remained aligned after increasing or decreasing all costs (Table 5-4), varying the

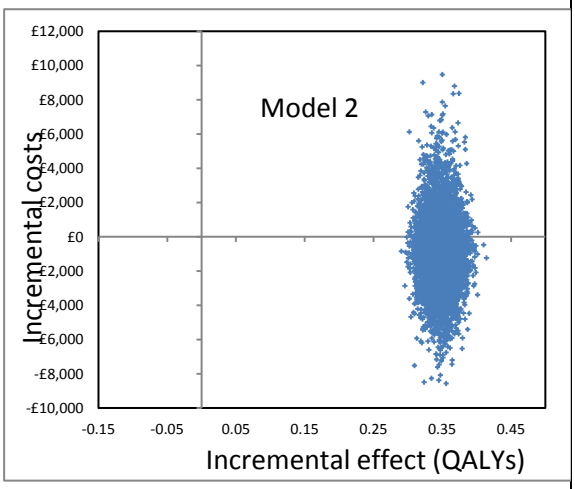
length of time horizons (Table 5-5), and varying transition probabilities to secondary events (Table 5-6), in other words, self-management of hypertension always remained dominant. Self-management in Models 1-4 was found to be dominant if the time horizon was two years or more (Table 5-5). Per-patient lifetime EVPI for alternative model structures compared to TASMIN-SR was reduced substantially for Model 1 at all willingness to pay thresholds. For all other model structures, there was a smaller decrease, again observed at all thresholds (see Figure 5-4).

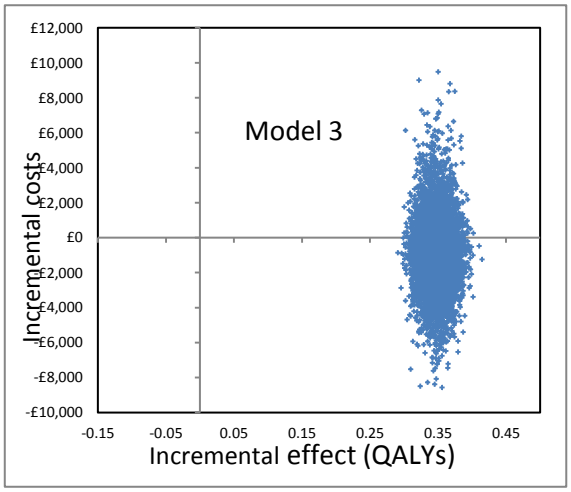
Table 5-3 Cost-Effectiveness results for the case study and each one of the alternative model structures

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
TASMIN-SR model					
Usual care	9,860	7.0946			
Self-management	8,997	7.4390	-864	0.3444	Dominant
Model 1					
Usual care	9,452	6.9102			
Self-management	8,813	7.2311	-639	0.3210	Dominant
Model 2					
Usual care	9,854	7.1612			
Self-management	8,858	7.5057	-996	0.3445	Dominant
Model 3					
Usual care	9,696	5.9274			
Self-management	9,156	6.2721	-539	0.3446	Dominant
Model 4					
Usual care	11,651	7.0704			
Self-management	10,378	7.4207	-1,273	0.3503	Dominant


Tasmin-sr
Incremental costs
£12,000
£10,000
£8,000
£6,000
£4,000
£2,000
£0
-£2,000
-£4,000
-£6,000
-£8,000
-£10,000
-0.15
-0.05
0.05
0.15
0.25
0.35
0.45
Incremental effect (QALYs)


Model 1
Incremental costs
£12,000
£10,000
£8,000
£6,000
£4,000
£2,000
£0
-£2,000
-£4,000
-£6,000
-£8,000
-£10,000
-0.15
-0.05
0.05
0.15
0.25
0.35
0.45
Incremental effect (QALYs)


Model 2
Incremental costs
£12,000
£10,000
£8,000
£6,000
£4,000
£2,000
£0
-£2,000
-£4,000
-£6,000
-£8,000
-£10,000
-0.15
-0.05
0.05
0.15
0.25
0.35
0.45
Incremental effect (QALYs)


Model 3
Incremental costs
£12,000
£10,000
£8,000
£6,000
£4,000
£2,000
£0
-£2,000
-£4,000
-£6,000
-£8,000
-£10,000
-0.15
-0.05
0.05
0.15
0.25
0.35
0.45
Incremental effect (QALYs)

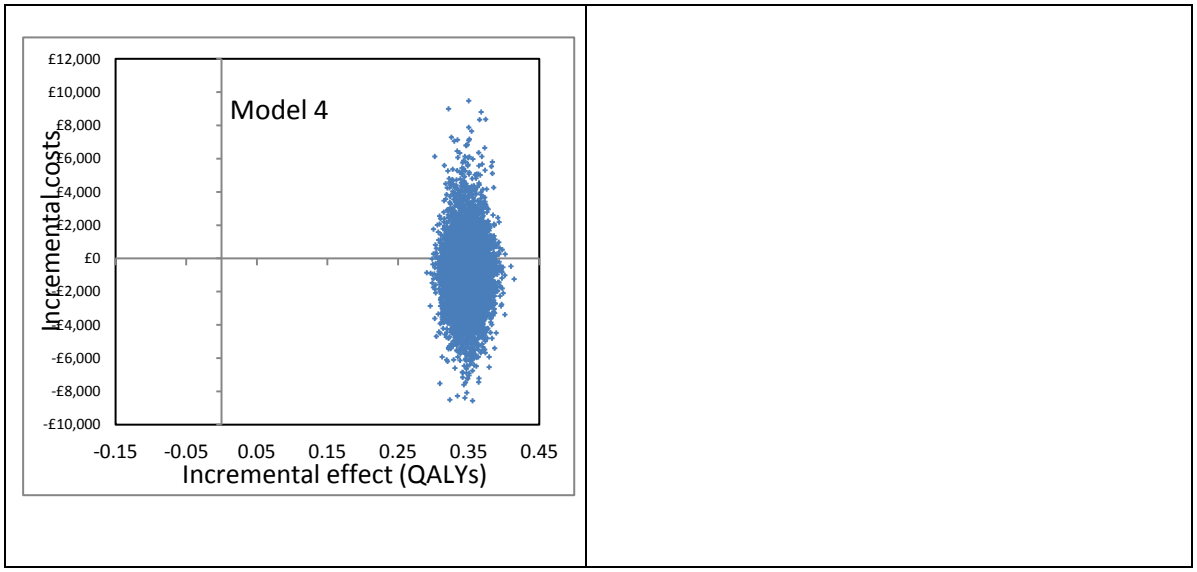


Figure 5-2 Cost-effectiveness plane (CEP) for the case study and each one of the alternative model structures

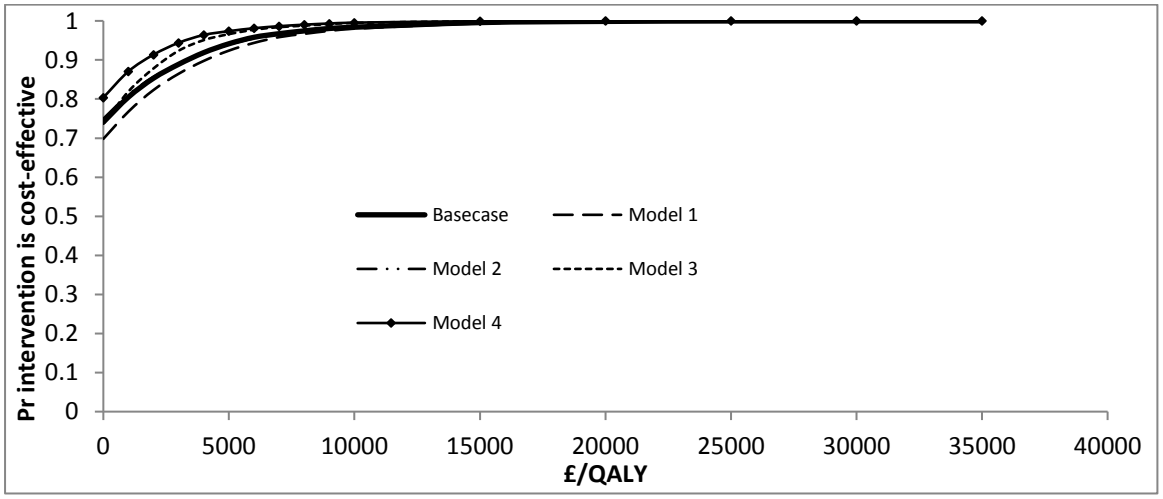


Figure 5-3 Cost-effectiveness acceptability curve (CEAC) for the probability that self-management of blood pressure is cost-effective compared to usual care – TASMIN-SR and each one of the alternative model structures

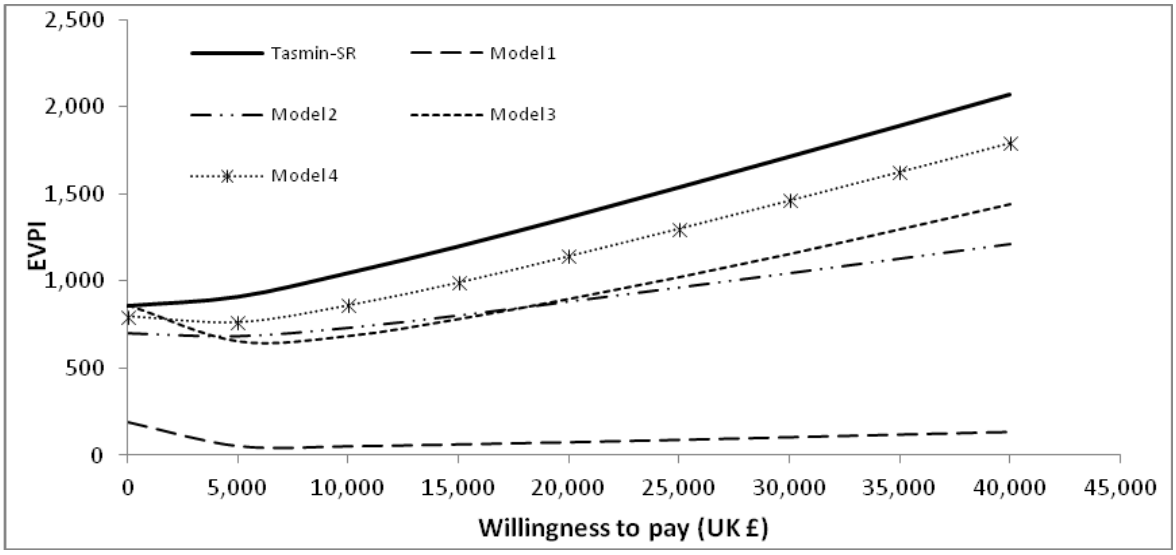


Figure 5-4 Per-patient expected value of perfect information (EVPI) across varying willingness to pay values for the TASMIN-SR and each one of the alternative model structures

Table 5-4 Results of cost-effectiveness after increasing and decreasing total costs

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
TASMIN-SR					
Increasing costs 200%					
Usual care	19,728	7.0946			
Self-management	17,899	7.4390	-1,829	0.3444	Dominant
Increasing costs 40%					
Usual care	13,805	7.0946			
Self-management	12,596	7.4390	-1,209	0.3444	Dominant
Decreasing costs 40%					
Usual care	5,917	7.0946			
Self-management	5,398	7.4390	-518	0.3444	Dominant
Decreasing costs 50%					
Usual care	4,931	7.0946			
Self-management	4,550	7.4390	-382	0.3444	Dominant
Model 1					
Increasing costs 200%					
Usual care	18,905	6.9102			
Self-management	17,528	7.2311	-1,376	0.3210	Dominant
Increasing costs 40%					
Usual care	13,234	6.9102			
Self-management	12,339	7.2311	-895	0.3210	Dominant
Decreasing costs 40%					
Usual care	5,671	6.9102			
Self-management	5,287	7.2311	-384	0.3210	Dominant
Decreasing costs 50%					
Usual care	4,726	6.9102			
Self-management	4,455	7.2311	-271	0.3210	Dominant

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Model 2					
Increasing costs 200%					
Usual care	19,717	7.1612			
Self-management	17,626	7.5057	-2,091	0.3445	Dominant
Increasing costs 40%					
Usual care	13,797	7.1612			
Self-management	12,402	7.5057	-1,394	0.3445	Dominant
Decreasing costs 40%					
Usual care	5,913	7.1612			
Self-management	5,315	7.5057	-598	0.3445	Dominant
Decreasing costs 50%					
Usual care	4,928	7.1612			
Self-management	4,479	7.5057	-449	0.3445	Dominant
Model 3					
Increasing costs 200%					
Usual care	19,396	5.9274			
Self-management	18,218	6.2721	-1,179	0.3446	Dominant
Increasing costs 40%					
Usual care					
Self-management	13,574	5.9274			Dominant
Decreasing costs 40%	12,819	6.2721	-755	0.3446	
Usual care					
Self-management	7,069	5.9274			Dominant
Decreasing costs 50%	6,607	6.2721	-462	0.3446	
Usual care	4,848	5.9274			
Self-management	4,628	6.2721	-220	0.3446	Dominant
Model 4					
Increasing costs 200%					
Usual care	31,071	7.0489			
Self-management	27,771	7.4085	-3,301	0.3596	Dominant
Increasing costs 40%					
Usual care	16,097	7.0489			
Self-management	14,464	7.4085	-1,633	0.3596	Dominant

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Decreasing costs 40%					
Usual care	6,899	7.0489			
Self-management	6,199	7.4085	-700	0.3596	Dominant
Decreasing costs 50%					
Usual care	7,766	7.0489			
Self-management	7,011	7.4085	-755	0.3596	Dominant

Table 5-5 Results of cost-effectiveness after altering the time horizon

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
TASMIN-SR					
10 years					
Usual care	5,860	5.1741			
Self-management	5,237	5.3506	-623	0.1765	Dominant
5 years					
Usual care	3,109	3.2475			
Self-management	2,753	3.3079	-356	0.0605	Dominant
3 years					
Usual care	1,792	2.1372			
Self-management	1,626	2.1564	-166	0.0192	Dominant
2 years					
Usual care	1,173	1.4889			
Self-management	1,110	1.4957	-63	0.0068	Dominant
1 year					
Usual care	629	0.7791			
Self-management	652	0.7797	23	0.0006	35,391
Model 1					
10 years					
Usual care	5,729	5.1310			
Self-management	5,231	5.3029	-498	0.1719	Dominant
5 years					
Usual care	3,066	3.2371			
Self-management	2,762	3.2955	-304	0.0584	Dominant
3 years					
Usual care	1,772	2.1331			
Self-management	1,630	2.1516	-141	0.0185	Dominant
2 years					
Usual care	1,162	1.4871			
Self-management	1,111	1.4937	-51	0.0066	Dominant
1 year					
Usual care	624	0.7788			
Self-management	650	0.7795	25	0.0006	40,799

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Model 2					
10 years					
Usual care	5,738	5.2904			
Self-management	5,045	5.4608	-692	0.1704	Dominant
5 years					
Usual care	2,923	3.3047			
Self-management	2,565	3.3580	-357	0.0533	Dominant
3 years					
Usual care	1,644	2.1624			
Self-management	1,492	2.1786	-152	0.0161	Dominant
2 years					
Usual care	1,066	1.5018			
Self-management	1,014	1.5075	-52	0.0056	Dominant
1 year					
Usual care	566	0.7830			
Self-management	593	0.7835	27	0.0005	50,960
Model 3					
10 years					
Usual care	6,114	4.4454			
Self-management	5,641	4.6811	-473	0.2357	Dominant
5 years					
Usual care	3,481	2.9515			
Self-management	3,132	3.0489	-348	0.0974	Dominant
3 years					
Usual care	2,100	2.0112			
Self-management	1,903	2.0448	-197	0.0336	Dominant
2 years					
Usual care	1,402	1.4255			
Self-management	1,314	1.4378	-89	0.0123	Dominant
1 year					
Usual care	771	0.7607			
Self-management	785	0.7619	14	0.0012	11,701

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Model 4					
10 years					
Usual care	6,880	5.2107			
Self-management	5,975	5.3803	-905	0.1696	Dominant
5 years					
Usual care	3,470	3.2711			
Self-management	2,985	3.3237	-485	0.0527	Dominant
3 years					
Usual care	1,889	2.1446			
Self-management	1,685	2.1611	-203	0.0165	Dominant
2 years					
Usual care	1,194	1.4907			
Self-management	1,123	1.4969	-71	0.0062	Dominant
1 year					
Usual care	629	0.7791			
Self-management	652	0.7797	23	0.0007	35,334

Table 5-6 Results of cost-effectiveness after increasing and decreasing the probability of having a second event

Model structure	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Model 4					
Doubling the probability of having a second event					
Usual care	15,528	6.0034			
Self-management	13,925	6.3859	-1,603	0.3825	Dominant
Halving the probability of having a second event					
Usual care	8,213	7.9957			
Self-management	7,351	8.2767	-862	0.2810	Dominant

5.5. Discussion

Decision-analytic modelling represents an organised way to synthesise the evidence currently available on the outcomes and costs of alternative health care interventions[12, 14]. The results obtained from a DAM will depend on how the model structure has been defined and the data used to populate the model. The analysis of uncertainty in DAM has mainly focused on parameter uncertainty, taking account of any uncertainties in the data inputs[27, 29, 51, 85, 97]. Such analyses are usually based on the premise that the model has been correctly specified. However, an inappropriate model structure can potentially invalidate estimates of cost-effectiveness and, therefore, is also of little value to a decision maker[27, 29, 51]. Although limitations in model structure are usually acknowledged, there is a lack of clarity about the methods used to evaluate structural uncertainty[27, 29, 51].

This chapter identified and implemented alternative model structures in the assessment of the cost-effectiveness of primary care interventions for the management of hypertension in patients at risk of or with established CVD. The results of each alternative model structure, including the results of EVPI, were presented and compared.

The main cost-effectiveness results obtained in the TASMIN-SR study did not change when alternative model structures (Model 1 to 3) were implemented or after adjusting the TASMIN-SR model for the effect of secondary events (Model 4) suggesting that structural uncertainty was not important in this model. This case study gave similar results for EVPI across the range of model structures, except for Model 1, where the restricted parameter set meant that a large part of the decision uncertainty was not apparent in the model.

The illustration of various scenarios representing structural uncertainty offers the decision maker the opportunity to decide on which model structure or assumption(s) he/she believes and make policy decisions on that basis. However, it does not provide any explicit framework for quantifying the uncertainty or offer any guidance to decision makers that have no clear preferences over alternative model assumptions.

The assessment of structural uncertainty shown in published studies in the area of primary prevention of CVD has mainly focused on assessing parameter uncertainty and there have been relatively few studies that have attempted to examine structural uncertainty in the extent that this study has done. Studies that considered the assessment of structural uncertainty varied in scope[39, 71, 72, 77] however none attempted to show the effect of different model structures on the cost-effectiveness of anti-hypertension treatments.

Model 1 (single CVD state) produced lower QALYs and lower costs compared to TASMIN-SR and this can be explained by the increased overall risk of CVD due to the added individual risks of stroke, MI and UA used to estimate the risk of CVD and the lower weighted average acute and chronic costs of CVD. The findings in terms of the highest QALY outcomes and reduced costs found by implementing Model 2 may be explained by the fact that when compared with TASMIN-SR, the population entering Model 2 was exposed to an overall reduced risk of CVD and reduced acute and chronic costs due to the exclusion of the angina state, thus leading to increased QALYs and reduced overall costs.

The lowest QALY gained and higher self-management costs from Model 3 (more complex model) can be explained by the additional burden of mortality for patients presenting with HF and TIA. The results of Model 4 show self-management to be even more cost-effective than usual care when compared with results from the case study and alternative Models 1-3. This can be explained by the increased overall risk of CVD due to the occurrence of additional events, and therefore more scope for preventing these events. The fact that expected costs for each intervention were noticeable higher for Model 4 can be explained by the acute costs of having experienced a second CVD event while lower expected QALYs for both interventions in Model 3 are explained by the additional burden of mortality for patients presenting HF and TIA.

The main conclusions drawn from the cost-effectiveness analyses were not altered when alternative model structures were implemented or in the presence of secondary events, and were driven by greater uncertainty around costs, reflected in the CEP where the results from

the Monte Carlo simulations were in the north-east and south-east quadrants indicating that self-management was always more effective but could be more or less expensive.

These results may well lead to the conclusion that the use of a simple model will suffice when examining the potential impact of anti-hypertensive strategies on the primary care prevention of CVD.

This case study reflects the level of the complexities typically faced in current practice when undertaking an assessment of structural uncertainty. Currently, guidance regarding the assessment of structural uncertainty in DAM by bodies such as ISPOR and NICE (in the UK) goes as far as recommending that modellers should parameterize uncertainties[20] and, if this is not possible, use sensitivity analysis and scenario analysis[11, 20].

The wide variation in the model structures that were identified by the systematic review supports the need for improved guidance to handle the implications of potential sources of structural uncertainty. Most importantly, this may be an indication that disease-specific or generic models to examine the cost-effectiveness of self-management of hypertension in patients with established CVD may need to be considered.

Challenges across different disease areas are so varied that it may well be the case that only studies such as this can shed any light on the importance of model uncertainty in different settings. In fact, there is evidence from UK HTA appraisals of metastatic end-stage cancer of biases associated with the use of models such as state transition versus partitioned survival analyses (PartSA), leading NICE to recommending[120] that further research is warranted to

understand the conditions under which the PartSA and state transition modelling approaches perform well.

Current practice seems bound by data availability whilst methods proposed to assess structural uncertainty, borrowed from other disciplines, seem oblivious to the needs in a health care setting where patient level data is most of the time not readily available.

5.5.1. Limitations

A limitation of the approach adopted in this Chapter to assess structural uncertainty is that there are no established methods to formally assess the plausibility of alternative models and it is not clear which type of, or how many scenarios should be considered.

The choice of model type in this case study was limited to a cohort Markov model. Some may argue that a microsimulation or Discrete Event Simulation (DES) may offer some advantages such as flexibility in incorporating individual heterogeneity and tracking individual event history. However, the review indicated that all economic evaluations in this disease area had used Markov models, presumably based on the trade-off between model flexibility and analytical input[49]. Furthermore, chronic and recurring diseases are often reflected by using Markov models in which individuals move between clinical states of interest in discrete time periods, and each state is associated with a cost and utility[9]. In addition, giving the information available for a UK setting to populate the model, developing a more complex model structure would have required the adoption of additional assumptions and therefore

would have meant adding more uncertainty to the model. Due to a lack of epidemiological data, Models 1 to 3 did not capture structural uncertainty arising from the exclusion of secondary events of CVD for high risk patients. However, using assumptions based on expert opinion, the risk of secondary events in Model 4 were assessed. The exclusion of secondary events in Models 1 to 3 was a conservative assumption, as a reduction in blood pressure was expected to reduce the risk of these events in addition to the primary events already considered, making self-management even more cost-effective as demonstrated in Model 4.

More sophisticated methods could not be implemented, for example, model selection, model averaging, or discrepancy approach to select the best model on the basis of how well the model's output match observed data (commonly judged by the likelihood-based information criteria). This was because only single point estimates were available for key parameters (transition probabilities) taken from the literature. These do not allow the estimation of the maximum likelihood of parameters: for that, actual patient level data is required. Furthermore, results of previous research seem to indicate that the standard likelihood-based approaches may be unsuitable when the underlying datasets are different[93]. Renal failure and peripheral artery disease were not considered in this case study. These additional health states are part of the broader set of diseases that may indirectly lead to CVD and data to populate input parameters for these states was not available.

The results of cost-effectiveness for self-management of blood pressure in this case study were of dominance for all competing model structures. For this reason it is difficult to draw conclusions regarding the importance of correctly exploring structural uncertainty in this

setting. It may be that if the results were near the £20,000 threshold, changes in model structure could have led different results of cost-effectiveness and possible EVPI.

The assessment of structural uncertainty shown in published studies in the area of primary prevention of CVD has mainly focused on assessing parameter uncertainty and there have been relatively few studies that have attempted to examine structural uncertainty in the extent that this study has done, showing the effect of different model structures on the cost-effectiveness of anti-hypertension treatments and implementing extensive sensitivity analyses and EVPI.

5.6. Conclusions

The results of this chapter indicate that the main conclusions from the TASMIN-SR cost-effectiveness model are robust to changes in model structure. The cost-effectiveness results and the EVPI were not sensitive to model structure specification.

Even though the results from Model 1 were not similar to those of TASMIN-SR, the fact that the main conclusions were the same raises the question, whether in this particular case study a more parsimonious model would have sufficed. Currently there are no available guidelines indicating how structural uncertainty, in particular structural uncertainty arising from the structure of a model, should be identified, assessed, and reported. Therefore, further research should focus on the strengthening of generally agreed guidelines on how to address issues pertaining to structural uncertainty and, more specifically, how to deal with challenges

across different disease areas, perhaps incentivising the development of analyses such as that presented in this chapter, focusing on disease specific areas.

Based on the findings of this chapter, the following recommendations are put forward:

1. The assessment of structural uncertainty should not be ignored as it is an integral part of good practice in DAM
2. The reasons why an assessment of structural uncertainty is not possible or not needed should be always stated as a limitation of the research
3. Data limitations to undertake an assessment of structural uncertainty should be clearly stated and discussed
4. If there is a reason to believe that structural uncertainty is an issue that may have affected the results of CE, then an assessment of structural uncertainty should be included
5. Ideally, sound statistical methods should be used in the assessment of structural uncertainty e.g. the discrepancy approach, model averaging, parameterization, model selection, scenario analysis, etc. but if none of these are possible due to data limitations, then at the very least, appropriate sensitivity analysis should be routinely conducted, as per current ISPOR-SMDM guidelines

Chapter 6 develops a similar analysis to assess the structural uncertainty arising from the model structure of a decision tree.

CHAPTER 6. IMPORTANCE OF MODEL STRUCTURAL UNCERTAINTY FOR THE RESULTS OF COST-EFFECTIVENESS: A CASE STUDY OF OPTIMISING ACUTE STROKE CARE SERVICES FOR THROMBOLYSIS

This chapter builds on the results of Chapter 5 by considering the issue of model structural uncertainty for another case study. This chapter provides another practical illustration of the impact on results of cost-effectiveness, of changing or adapting model structures in a model-based economic evaluation. Structural uncertainty arising from model structure is assessed for a decision tree used to measure the cost-effectiveness of alternative strategies aimed at increasing thrombolysis rates among patients who have suffered a stroke. This case study also examines whether there is evidence of other elements, besides the structure of a model, that affect the results of cost-effectiveness.

6.1. Introduction

The most important public health measurements of stroke impact are mortality and its sequelae. Resulting from the sequelae are a loss of productivity, reduced quality of life, and productivity. It has been estimated that 87% of strokes are of ischaemic origin, with the remainder being haemorrhagic strokes or transient ischaemic attacks (TIA)[5]. Mortality during the first 30 days after a ischaemic stroke is approximately 10%, and is mainly associated with

neurological sequelae, and then 40% in the subsequent 12 months[111]. As the world population ages, the burden of disease from stroke is expected to increase over the coming decades[35].

A previous study aimed at quantifying the annual cost of illness of stroke to the UK in 2009 found that the treatment of and productivity loss arising from stroke resulted in total societal costs of £8.9 billion a year, with treatment costs accounting for approximately 5% of total UK NHS costs. Direct care accounts for approximately 50% of the total, informal care costs 26% and indirect costs 24%[121].

Thrombolytic therapy or thrombolysis is a treatment to dissolve clots in blood vessels, improve blood flow, and prevent damage to tissues and organs. Thrombolysis may involve the injection of clot-busting drugs through an intravenous (IV) line or through a long catheter that delivers drugs directly to the site of the blockage. It may also involve the use of a long catheter with a mechanical device attached to the tip that either removes the clot or physically breaks it up. In patients presenting to hospital with a recent acute ischaemic stroke, thrombolysis (for example, alteplase) can dissolve the clot blocking the arteries and cause a stroke. Randomised clinical trials have previously shown that if alteplase is given within a window timeframe of no more than 4.5 hours of the onset of symptoms, it increases the proportion of patients who are free of disability by 3 months, with larger benefits seen when alteplase is able to restore the brain's blood supply earlier[38].

However, the use of thrombolysis in a small proportion of patients comes with the risk of a brain bleed (cerebral haemorrhage), which in some cases results in serious disability or death[38].

Health economic evaluations in a number of countries have shown that thrombolysis is a cost-effective treatment for ischaemic stroke. The introduction of thrombolytic therapy with recombinant tissue plasminogen activator (rt-PA) for the treatment of acute ischaemic stroke has resulted in improved patient outcomes for those eligible to receive thrombolytic treatment[122-124]. The treatment is safe and effective if administered within a short window of time (4.5 hours and 6 hours post-onset of symptoms depending on a patient's age). Therefore, the early recognition of symptoms, prompt arrival at hospital, and timely computed tomography (CT) scanning have been recognised as determinants for successful treatment.

A previously published economic evaluation of the cost-effectiveness of optimising acute stroke care services for thrombolysis[123] (referred to as the 'Stroke study' in the remainder of this chapter) estimated the cost-effectiveness and potential implementation costs of a series of interventions aimed at increasing thrombolysis rates through optimisation of the care pathway for acute stroke[123]. A decision tree was used to represent the acute care pathway for patients with stroke and lifetime costs and outcomes per 100,000 population. This chapter aims to explore whether the results of cost-effectiveness in the Stroke study would have been different if the analysis had considered: i) an alternative model structure (for example a Markov model or a combined decision tree and a Markov model); or ii) other elements pertaining to structural uncertainty.

6.2. Description of the Stroke study

Thrombolysis in the context of the Stroke study refers to an intravenous injection of clot-busting drugs. Thrombolysis in acute stroke is effective up to 4.5 hours after the onset of

symptoms but relies on early recognition, prompt arrival in hospital, and timely brain scanning. The economic evaluation in the Stroke study aimed to establish the cost-effectiveness and potential implementation costs of increasing thrombolytic rates through a series of hypothetical interventions designed to optimize the acute care pathway for stroke (see Appendix 7 for details).

The Stroke study used a decision tree built in TreeAge PRO Suite 2011 software, to reflect the acute care pathway for stroke patients (see

Figure 6-1 below). The decision tree describes the care pathway of individuals presenting to hospital with a suspected stroke[123]. Routes to hospital and the demographic characteristics of patients (age, sex, and ethnicity) were based on patient level data collected from participating hospitals. All patients who entered the model were assumed to have been admitted via the hospital emergency department and could have arrived via the Emergency Medical Services (EMS), been referred by a general practitioner, travelled via private transport, transferred from another hospital or had a stroke as an existing inpatient (i.e., patient had a stroke as a inpatient during an admission for an unrelated complaint), see Figure 6-1 below. The study included 488 stroke events in the base-case analysis including 133 patients presenting with nonvascular disease (stroke mimics). It was assumed that once in hospital, all patients with suspected stroke would received a CT scan and the results of the CT scan were used to dichotomise patients as having had a stroke (whether ischaemic or haemorrhagic) or not having had a stroke (stroke mimics), see Figure 6-1. Those patients found to have had a stroke were further dichotomised as having had an ischaemic or

haemorrhagic stroke. Those with ischaemic stroke were considered potentially eligible for thrombolysis, see Figure 6-1.

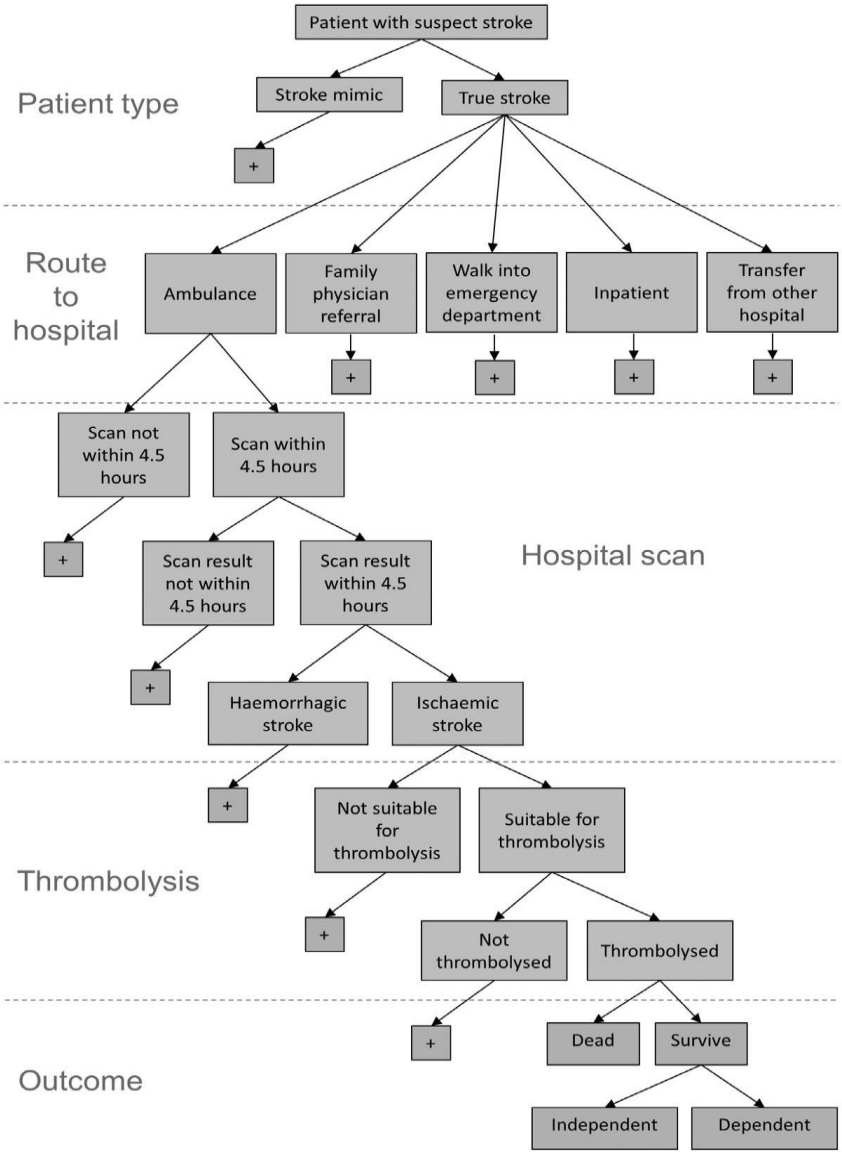


Figure 6-1 Decision Tree model structure

The Decision-tree model is displayed in vertical format rather than the traditional horizontal format. The model is identical at every node ending (grey box with +) with final outcomes (shown in grey boxes at the bottom of the figure) considered to be death, dependency, or independency.

The cost utility analysis in the Stroke study took a National Health Service (NHS) and Personal Social Services (PSS) perspective. Short and long-term costs of initial assessment and treatment in hospital along with increased dependency because of stroke-related disability were considered and adjusted to a price year of 2010/11. The Stroke study justified the use of a decision tree model, as recommended for acute diseases where any intervention affecting prognosis is settled in a short time frame[1].

Model pathways were identified by following the care pathway of individuals presenting to hospital with a suspected stroke[123]. Patient level data were collected from patients admitted to three participating hospitals with stroke or stroke-like symptoms over a 12-month period. Patients in the model were assumed to have been admitted via the hospital emergency department and could have arrived via the Emergency Medical Services (EMS), referred by a general practitioner (GP), travelled via private transport, transferred from another hospital, or have had a stroke as an existing inpatient (Appendix 8). Once in hospital, all patients received a CT scan; patients were dichotomised as having had a stroke (either ischaemic or haemorrhagic) or a mimic stroke as per the results of the CT scan[123]. Only patients with ischaemic stroke were considered eligible for thrombolysis. Life expectancy and the time horizon in the model varied depending on whether an individual had a stroke or not and after a stroke whether they were dependent or independent[123].

Several assumptions were made with respect to current practice and the data for the base-case. The time window for the administration of thrombolysis treatment was set at 4.5 hours as this represented current practice at the participating centres at the time of data collection.

Patients aged up to 85 years old were considered eligible for treatment; the impact of thrombolysis on reducing dependency was modelled using results from a Cochrane Review[125].

The primary outcome measured was changes in QALYs with utilities being identified from the literature. The modified Rankin Scale (mRS) – a commonly used scale for measuring the degree of disability or dependence in the daily activities of people who have suffered a stroke was used to assess dependency. This scale runs from 0 (least dependent) to 5 (most dependent) and 6 (dead). Scores of 0-2 were classified as independent and 3-5 as dependent. QALYs were calculated by multiplying life expectancy by the utility associated with a given outcome. Costs and outcomes were discounted at the standard annual rate of 3.5% [1]. In the case of events that imitated a stroke, only the costs incurred by transport and initial assessment in hospital were modelled. The model evaluated a series of hypothetical interventions designed to reduce delay to thrombolysis treatment; these interventions were identified from the literature (Appendix 7). The results of the Stroke study indicate that all intervention strategies that increased thrombolysis rates in acute stroke are cost-effective because of a reduction in dependency after stroke and the subsequent reduction in long-term care costs.

6.3. Methods

In order to explore how the results from the Stroke study may have differed if the original model was different; this chapter examines the adequacy of the type of model used, the

structure of the model as well as changes to other key parameters pertaining to structural uncertainty. The resulting cost-effectiveness results are then compared with those from the Stroke study. The structure of the model used in the Stroke study constitutes the base-case model (

Figure 6-1). The analysis of structural uncertainty was performed in three steps:

- Identification of plausible alternative model structures
- Definition and implementation of alternative types of model
- Identification of model parameters

6.3.1. Identification of plausible alternative model structures

As described in Chapter 5, model structural uncertainty is addressed by assessing issues such as the adequacy of the type of model used, the structure of the model (model pathways or health states and transition probabilities) and data availability to inform input parameters, for example the risk of secondary events.

In order to identify plausible alternative types of model and model structures, a literature review was undertaken.

6.3.1.1. Methods of the literature review

The literature review identified studies of interventions aimed at increasing or optimising thrombolysis in acute stroke care services. The review followed a structured approach to identify papers: patient population (P), intervention (I), the comparator group (C), outcome (O) and the study design (S), or PICOS[68]. Studies published from Jan 2000 to June 2016 and written in English were included if they met all of the following conditions:

- Target population was individuals presenting with stroke symptoms
- Intervention was thrombolysis
- Study was a model-based economic evaluation

The review excluded systematic reviews, conference papers, commentaries or letters and non-English articles. Studies were also excluded if the interventions were part of an intermittent pneumatic compression or stent retrievers or CT perfusion of screening or mechanical thrombectomy or other non-thrombolytic therapies.

Searches were undertaken using reasonable variations of the following terms 'cost-effectiveness' or 'decision-analysis' or 'economic evaluation' or 'cost utility' combined with the term 'stroke' and the terms 'thrombolysis' or 'rt-PA' or 'alteplase'. The search was undertaken using truncations and wildcards, and all synonyms were subsequently combined with appropriate medical subject heading terms (MeSH) or subject terms using Boolean operators (Appendix 9)

The following databases were searched: the Centre for Reviews and Dissemination's (CRD) NHS Economic Evaluation Database (NHS-EED) and EMBASE and Medline via the Ovid

interface. In addition, the reference lists of the studies included in this review were manually examined. All papers identified by database searching were exported into ENDNOTE-X7^{TM} and duplicated references were removed.

Titles identified by the searches were screened by reading the abstract. Articles that appeared to be relevant at this point were obtained and screened against the inclusion and exclusion criteria; several studies appeared relevant on reading the abstract but were subsequently excluded after reading the full paper.

6.3.1.2. Results of the literature review

The database yielded 328 studies; 302 studies were excluded as duplicates, systematic reviews, conference papers, commentaries, letters or non-English language papers (Figure 6-2). 26 full-text articles were assessed for eligibility, of which 8 were rejected as interventions that were part of an intermittent pneumatic compression or stent retrievers of CT perfusion or screening or mechanical thrombectomy in acute ischaemic stroke or other. 5 further studies were excluded as non-economic evaluations. 13 studies (including the Stroke study) were reviewed[15, 122-124, 126-134].

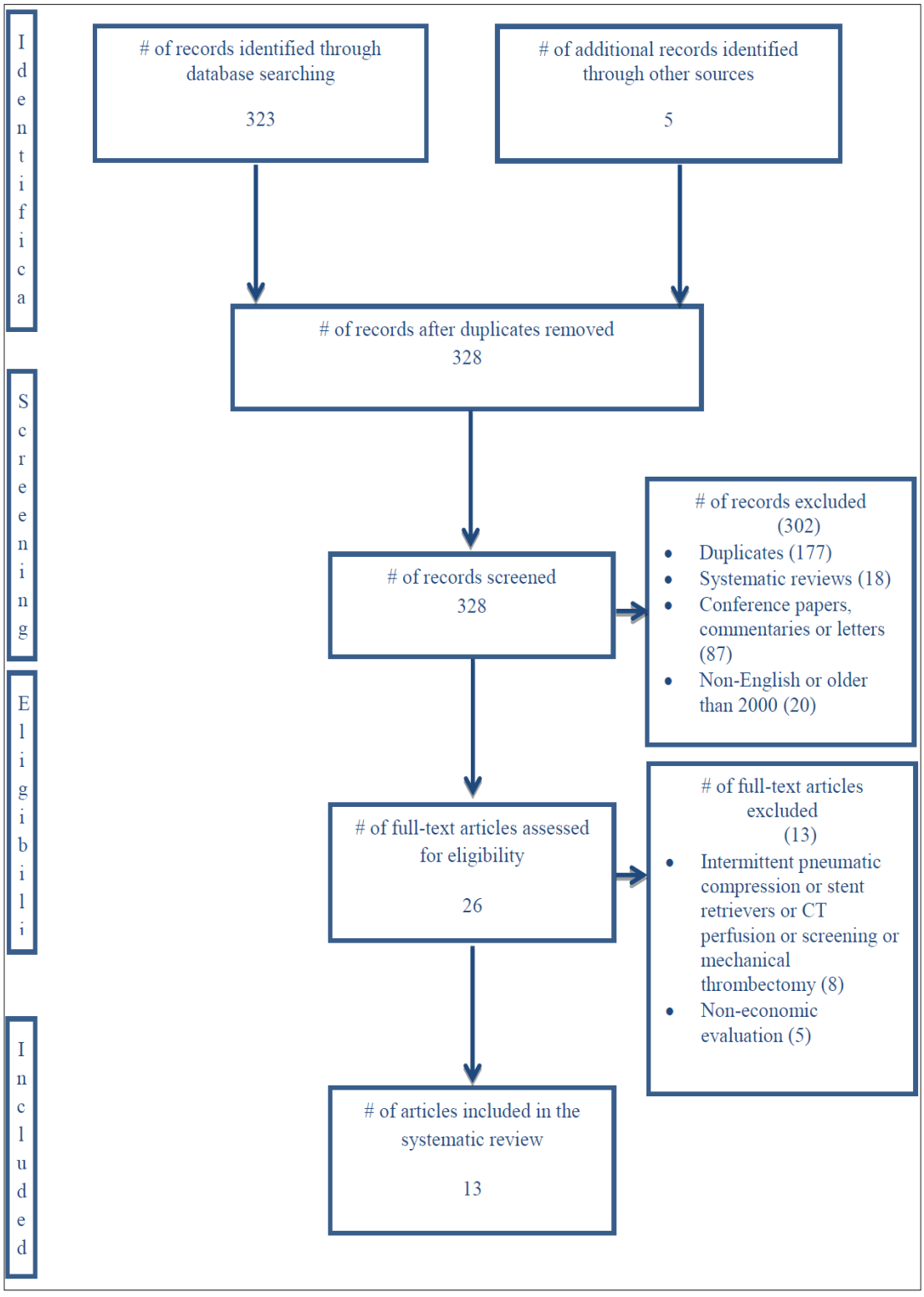


Figure 6-2 Flow chart using the PRISMA statement

All studies were manually searched and data were extracted; any doubtful point(s) were checked with at least one of the supervisors. Data were retrieved, and organised across the following items: author/year, model comparators, type of model/description and structural assumptions. This approach ensured that the review did not miss any information related to the model type, model structure, model assumptions, model justification and the assessment of structural uncertainty. Table 6-1 summarises the results of the review.

The included studies aimed to assess the use and cost-effectiveness of thrombolysis for acute stroke. The comparators considered were thrombolysis compared with placebo or no thrombolysis. One study[127] assessed the proportion of patients thrombolysed based on a breakthrough implementation programme compared to the do nothing (laissez-faire) alternative.

Table 6-1 Overview of published models on thrombolysis treatment for patients presenting with acute ischaemic stroke

Author, year	**Model comparators**	**Type of model / description**	**Structural assumptions**
Araujo et al[122]	Thrombolysis (up to 3hrs after symptom onset) compared to placebo from the Brazilian Public Health System perspective	Markov model consisting of four health states and a death state simulating a patient's transit from an acute phase of coronary vascular accident (CVA) to an intracranial bleeding event or no intracranial bleeding and to different levels of severity of post-stroke sequelae (Rankin 0 to Rankin 5) or death; cycle length varied from 3 months in the first year to 1-year from year 2 onwards. Used a lifetime time horizon of up to 30 years. Outcomes using QALYs; discount rate 5% for costs and outcomes	Side effects of thrombolysis modelled were intracerebral bleeding (6.4% with rt-PA and 0.6% with the conservative treatment)
Barton et al[15]	Two initial groups: Thrombolysed patients versus not thrombolysed patients from the perspective of the Northern Ireland integrated health and community social services	Discrete Event Simulation (DES) representing 3 health states, death, survival in an independent state or survival in a dependent state. Patients were distributed among six subgroups reflecting their final destination (home, rehabilitation centre or no rehabilitation centre, or institutional care). The DES model tracked six groups of patients, in which each group represented a patient pathway within the simulation and their LOS in hospital. Even though SIMUL8 was used, ideal for DES models, authors did not take advantage of all its features (queuing /patient competition for services). Outcomes were measured as QALYs (discounted at 3.5%). No discount rate was provided for costs.	Side effects of thrombolysis were increased risk of haemorrhage and death

Author, year	Model comparators	Type of model / description	Structural assumptions
Boudreau et al[126]	rtPA administered (in the 3-to 4.5-hour therapeutic window) versus rtPA not administered. Analysis was conducted from the payer perspective.	A short-term disease-based decision tree model to simulate 90-day outcomes for a hypothetical cohort of patients presenting to hospital with acute ischaemic stroke by treatment with or without rtPA. Short term outcomes were disabled, nondisabled or death. Patients surviving to 90 days in the short-term model entered a Markov model to simulate long-term clinical outcomes such as disability, recurrent stroke and death during the patient's remaining years of life. 3% discount rate for costs and outcomes was applied	Side effects, a disutility of -0.38 for 2 weeks was applied if symptomatic intracranial haemorrhage ocurred Secondary events: the risk of a recurrent stroke was 1.52-fold for disabled patients (2-5 mRS) and 1.10 for nondisabled (0-1 mRS) taken from literature Model structure: simplified model structure was justified after evaluating the validity of grouping mRS rankings compared with using separate values by comparing the weighted values for costs, quality of life and mortality. They found that grouping mRS lead to similar results for rtPA versus no rtPA. Authors verified that the weighted utilities for the corresponding health states were equivalent in both arms

Author, year	Model comparators	Type of model / description	Structural assumptions
Dirks et al[127]	Breakthrough Series-based implementation programme to increase the proportion of patients thrombolysed compared to a laissez-faire implementation of thrombolysis (both within 4hrs from onset) from a healthcare perspective	Type of model used not specified, however, it seems a modified Markov model or ISM was used. Even though the use of this type of model was not justified, presumably it was linked to the outcomes measured: thrombolysis rates and timing of response to assess health benefits. Patients from both arms entered the model at hospital admission. They may have a recurrent stroke and be readmitted, become more disabled or die. Model used half year time steps until death. 3% annual discount rate used for costs and health effects.	Secondary events: the risk of a recurrent stroke was considered using information from individual patient data
Ehlers et al[128]	Thrombolysis within 3hrs of symptom onset by MRI imaging selection compared to conservative treatment, Danish National Health perspective	Decision tree with Markov modelling The decision tree followed the care pathway of individuals with acute ischaemic stroke; the model assumed that patients could receive either thrombolysis or conservative treatment and according to treatment received, patients were at risk of haemorrhage The Markov model reflects the transition of patients between seven post-stroke disability states (R0 to R5 or death) according to the functional outcome after 3 months based on a mRS. After hospitalization patients are assumed to be discharged home, to rehabilitation or to a nursing home Time horizons of 1,2,3 and 30 years; 5% discount rate for costs and outcomes; outcomes were measured using QALYs	Side effects: risk of intracranial haemorrhage of 5.9% on thrombolysis and 1.1% on conservative treatment Secondary events: the risk of recurrence of stroke was assumed to be 5.2% per year (for survival after first year and recurrence an equal rate for all patients was assumed). Information was taken from the NINDS tPA Stroke Study

Author, year	Model comparators	Type of model / description	Structural assumptions
Ehlers et al[129]	National use of thrombolysis with alteplase for acute ischaemic stroke via telemedicine compared to conservative treatment from a Danish National Health perspective	This study used the same Decision tree with Markov model as per Ehlers et al (2007); in this case the model was expanded to include thrombolysis at satellite clinics linked to a larger thrombolysis centre	Side effects: risk of intracranial haemorrhage of 5.9% on thrombolysis and 1.1% on conservative treatment Secondary events: the risk of recurrence of stroke was assumed to be 5.2% per year (for survival after first year and recurrence an equal rate for all patients was assumed). Information was taken from the NINDS tPA Stroke Study
Mar et al[130]	Thrombolysis versus thrombolysis and intravenous administration of tPA (up to 3hrs from symptom onset) in Spain from the perspective of the health care system	Markov model, the initial health state reflects patients presenting with stroke, patients can transition after 1-year cycle to an autonomous, disabled or death state. The model considers the recurrence of stroke. 3% discount rate was applied and a lifetime time horizon. Outcomes were measured using QALYs	Side effects the risk of haemorrhage associated with both alternatives refers to the number of events that appeared in the NINDS study Secondary events (recurrent stroke was modelled using a 0.051 probability within a year). Figures were taken from literature

Author, year	Model comparators	Type of model / description	Structural assumptions
Pan et al[131]	tPA treatment versus non-tPA treatment (within 4.5hrs from symptom onset) from the perspective of the healthcare payers, including the government, medical insurance and patients in China	Decision tree with a Markov model The decision tree followed the care pathway of individuals presenting with acute ischaemic stroke that would or would not receive intravenous tPA. Patients could or could not have been affected by symptomatic intracerebral haemorrhage (sICH). The Markov model estimated the long-term costs and outcomes per disability levels modelling recurrence of stroke. Functional states were identified from the mRS. Outcomes were measured as QALYs by multiplying years of life by utility scores derived from literature. Outcomes and costs were modelled over the short-term (2 years) and the long-term (30 years)	Secondary events: recurrence of stroke after the first 90 days and assumed an increase in stroke recurrence rates by 1.01—fold per life year according to the relative risk estimated from patients of ischemic stroke in the China National Stroke Registry (CNSR), a nationwide registry for patients with acute cerebrovascular events
Sandercock et al[124]	Thrombolytic treatment for acute ischaemic stroke compared with standard care (within 3hrs from symptom onset), from a broad health care and personal social services perspective in the UK	Decision tree with Markov model A decision analysis model reflected the pathways that acute stroke patients follow after being admitted to hospital. A Markov model was used to predict the health and economic outcomes of rt-PA after the first year. The model used age-specific mortality, risk of recurrent stroke and stroke-specific case-fatality to estimate the probabilities of being dead, dependent and independent at the beginning of each year. One year cycles were assumed. Outcomes were measured as QALYs modelled over short-term (1 year) and longer-term (lifetime) time horizons. Annual costs and health benefits were discounted at 6%.	Side effects The risk of haemorrhage modelled as part of the decision tree based on data from the Lothian Stroke Register Secondary events: Recurrence of stroke modelled as an annual risk stroke recurrence of 0.05 after 1 year. Information was taken from the Lothian Stroke Register (LSR)

Author, year	Model comparators	Type of model / description	Structural assumptions
Sinclair et al[132]	t-PA therapy within 3hrs of stroke onset versus no t-PA from the perspective of the Canadian healthcare system	Markov model to describe the short and long term outcomes associated with treatment versus usual stroke management. Patients with ischaemic stroke, presenting to hospital within 3 hours of the onset of symptoms enter the model. During hospitalisation patients could experience a symptomatic intracranial haemorrhage. At discharge, individuals within the cohort were classified according to the their mRS. Outcomes considered were the costs of stroke, management of stroke sequelae and 30-year post stroke, quality-adjusted survival (measured by QALYs). Costs discounted at an annual rate of 5%	Side effects: Major bleeding complications were assumed as t-PA 6.4% versus no t-PA 0.6% Secondary events: The risk of recurrent stroke was taken from literature (5.2% per annum)
Tan et al[133]	Intravenous tPA treatment (within 4.5 hrs of symptom onset) compared to no-tPA in Australia	A decision tree modelled the pathway of patients presenting with acute ischaemic stroke that would or would not be treated with intravenous tPA. Patients who survived the treatment would move to one of six disability levels as per mRS (mRS 0 to mRS 5). The time horizon was 1 year after tPA treatment, with the assumption that patients' clinical and functional status would remain stable between 90 days and 1 year. Health benefits were measured using QALYs	Side effects patient level data indicated that a substantial proportion of patients (7.1% experienced haemorrhage after tPA – symptomatic intracranial haemorrhage or parenchyma hematoma type 2

Author, year	Model comparators	Type of model / description	Structural assumptions
Tung et al[134]	Intravenous tPA administered in the 3-to-4.5hr time window compared to non-thrombolytic therapy	Decision analytic tree and Markov model, a patient enters the model when he/she is admitted to hospital within 3 to 4.5hrs after the onset of stroke symptoms. Patients who survived the treatment would move to one of six disability levels as per mRS. After the end of each annual cycle, patients may remain in the same health state, have a recurrent stroke and transition to a lower health state, or die. Costs and health benefits were discounted at a rate of 3.0%. The analysis was conducted from the societal perspective, however, indirect economic costs such as lost work productivity were not considered in the model.	Side effects: The risk of symptomatic intracerebral haemorrhage was modelled Secondary events: the risk of recurrent stroke was assumed to be 0.051 per time step and the risk of death after a second stroke was assumed to be 0.19

Six studies[124, 126, 128, 129, 131, 134] used decision trees combined with Markov models: a decision tree to reflect the pathways that acute stroke patients follow after being admitted in hospital and a Markov model to predict the health and economic outcomes in the long-term (or after the end of the acute stage). Three studies[122, 130, 132] used Markov models to reflect the transition from an acute phase to different levels of severity of post-stroke sequelae and/or recurrence of stroke. One study[133] used a decision tree to model the pathway of patients treated with tPA after a stroke and until one year after treatment. Another study[15] used a discrete event simulation (DES) model to track the final destination (home, rehabilitation centre, no rehabilitation centre or institutional care) of patients after stroke. This type of model does not seem relevant in this context and this is apparent by the fact that even though the authors used SIMUL8 software, ideal for modelling DES models, they did not take advantage of all its features, for example, tracking individuals competing for resources. One study[127] did not specify the type of model used. However, it seems an ISM was used to measure thrombolysis rates and the timing of response after the implementation of a programme (Table 6-1).

In terms of model structure, only Boudreau et al[126] evaluated the validity of grouping mRS values compared with using separate scores by comparing the weighted values for costs, quality of life, and mortality. They found that the distribution of mRS within the nondisabled and disabled categories was similar on average between rt-PA and no rt-PA, thus justifying their simplified modelling approach.

Side effect(s) of thrombolysis, for example the risk of intracranial haemorrhage or intracerebral bleeding was considered in most of the studies reviewed (Table 6-1). For two studies[127, 131] it was not possible to establish if the side effect(s) of thrombolysis had been considered. One study[15] found excess risk of death after thrombolysis via the increased risk of haemorrhage; however, the authors did not find statistical support for an increased risk of dependency[15]. 9 studies[124, 126-132, 134] modelled the risk of a recurrent stroke (another stroke after a first stroke) using data from population registers[124, 131], patient level data[127], clinical trials[128, 129] and secondary sources[126, 130, 132, 134] (Table 6-1).

6.3.2. Methods - model structure

Based on the findings of the review, alternative model structures were identified and assessed in light of the primary research aim of the Stroke study i.e. estimate the cost-effectiveness of alternative strategies aimed at increasing thrombolysis rates through the optimisation of the care pathway for acute stroke. As previously described, the review identified three main types of model structure that have previously been used in the literature: i) combined decision tree and Markov model[124, 126, 128, 129, 131, 134]; ii) Markov model alone[122, 130, 132]; or iii) decision tree alone[133].

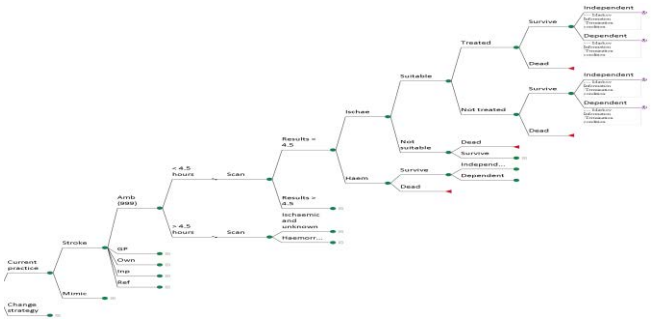
A Markov model alone, such as the structure used by Araujo et al[122] who used 1-year cycles or Mar et al[130] who used 3-month cycles during the first year and 1-year cycles afterwards was not considered as a plausible alternative structure to implement in this case study due to the short time frame nature of the intervention of interest (thrombolysis). Furthermore, in

the Stroke model, critical pathways during the thrombolysis treatment (recognition of symptoms, arrival to hospital, computed tomography - CT) occur over an instantaneous discrete period of time. In other words, there is no need to measure an explicit time for events to occur (for example, time from onset of symptoms to arrival to A&E, to CT scan, etc.). The simplified version of a decision tree as per Tan et al[133] who focused on final outcomes after thrombolysis treatment in terms of disability levels does not seem a plausible alternative either because it would not allow for the modelling of critical pathways for the treatment of acute stroke such as early recognition of symptoms, prompt arrival in hospital, and timely CT. Other types of models identified in the review, for example a DES or ISM were rejected on the basis of model parsimony. An essential feature of a DES model is that patients compete for limited resources. In the Stroke study patients with stroke symptoms, arriving at A&E within the window timeframe and being suitable for thrombolysis would all be thrombolysed. ISM on the other hand would be ideal in cases where the aim is mainly to assess the timing of the response to an event. However, the aim of the Stroke study was to increase thrombolysis rates by expediting the acute care pathway through a series of intermediate hypothetical strategies.

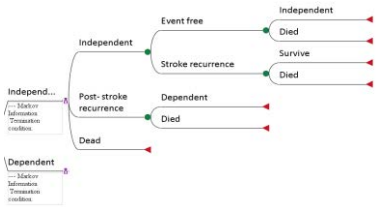
Based on the above considerations this case study used as an alternative model structure a combined decision tree and Markov model (Figure 6-3, Part A and Part B). The short-term decision tree describes the care pathway up until 12 months after stroke, whilst the Markov structure reflects lifetime costs and outcomes (beyond 12 months). Validation of the adequacy of this type of model over other competing structures such as DES or individual sampling model was checked using a framework to select an appropriate model type[13]. The

validation check indicated that either a decision-tree or a decision-tree followed by a Markov model was a correct type of model, considering that estimating interactions between individuals was not necessary whilst an excessive number of health states was not required.

The Stroke study did not model the recurrence of stroke after a primary stroke. This case study adopts assumptions based on the literature to reflect the risk of stroke recurrence[119]. Individuals that survived a primary acute event at the end of the 12 months (short term decision tree) move into the Markov component of the model as independent or dependent survivors (Figure 6-3, Part B). As in the Stroke model, the mRS was used to assess dependency where scores of 0 to 2 were classified as independent and 3-5 as dependent. Patients could experience a recurrent stroke event one year after the first event. Following a recurrent event, patients were assumed to become dependent with a lower quality of life or die.



Part A Decision Tree followed by a Markov model - complete model structure



Part B Expansion of the Markov component in the Decision Tree followed by a Markov model

Figure 6-3 Decision Tree followed by a Markov model

6.3.3. Model parameters and analyses

The model was populated with the dataset used in the Stroke study, that is, patient level data collected from 3 participating English NHS hospital Trusts. Information from the literature for a UK setting was sought to populate all the input parameters in the Markov component of the model. Transition probabilities for patients with a previous stroke, the risk of death following a second event, and utility values following a second event were all taken from secondary sources (Table 6-2 below).

One-off costs of acute disease after a second stroke were derived from the literature and adjusted using the Hospital and Community Health Service Index (Table 6-2). ONS Life tables were used to determine overall mortality rates[103].

Table 6-2 Input parameters and distributions

Variables	Value (range)	Distribution	Parameters	Parameter 1	Parameter 2	Source
Patients with symptoms who have suffered a "true" stroke						
Percentage who contact ambulance services (999)	67%	Beta	n , r	355	237	Stroke study, Patient level data[123]
Percentage who contact their GP surgery first	16%	Beta	n , r	355	57	
Percentage who make their own way to the hospital	11%	Beta	n , r	355	40	
Percentage of patients who are already inpatients	5%	Beta	n , r	355	18	
Percentage of patients who are referred/transferred	1%	Beta	n , r	355	3	
Patients with symptoms who have suffered a stroke and contacted 999 services						
Percentage who get to A&E within 4.5h (hours) of stroke onset	53.6%	Beta	n , r	237	127	Stroke study, Patient level data[123]
Percentage who get a CT scan and results within 4.5h of stroke onset	39.8%	Beta	n , r	237	94	
Percentage with an ischaemic stroke and suitable for thrombolysis	30.1%	Beta	n , r	237	71	

Variables	Value (range)	Distribution	Parameters	Parameter 1	Parameter 2	Source
Percentage who have an ischaemic stroke and receive thrombolysis	12.7%	Beta	n , r	237	30	
All stroke patients						
Percentage of ischaemic stroke in all stroke patients	89.9%	Beta	n , r	355	319	Stroke study, Patient level data[123]
Percentage within 4.5h of stroke onset who have an ischaemic stroke	42.0%	Beta	n , r	355	149	
Percentage of death* after stroke due to an haemorrhagic infarct	19.4%	Beta	n , r	36	7	
Percentage of death* after stroke due to ischaemic stroke	19.1%	Beta	n , r	319	61	
Percentage of death* after an ischaemic stroke and thrombolysis treatment	15.2%	Beta	n , r	33	5	
Percentage of being dependent/disabled after an ischaemic stroke	35.0%	Beta	n , r	545	191	Bamford et al[111]
Patients with "true" stroke symptoms						
Percentage with no previous contraindications for thrombolysis	95.5%	Beta	n , r	88	84	Stroke study, Patient level data[123]
Effectiveness of thrombolysis						
Decrease in dependency after stroke due to thrombolysis treatment	OR 0.67 (95% CI	Log normal	mean, se	-0.4005	0.0527	Wardlaw et al[125]

Variables	Value (range)	Distribution	Parameters	Parameter 1	Parameter 2	Source
	0.61 to 0.75)					
Life expectancy (number of additional years)						
Life expectancy for a 72 year old female with no history of stroke**	10.8					ONS, Life tables (stroke Study)[123]
Life expectancy for a 68 year old male with no history of stroke**	11.2					
Life expectancy for a 72 year old female with history of stroke**	10.0					
Life expectancy for a 68 year old male with history of stroke**	10.7					
Utility score						
Utility score for a dependant stroke patient	38%	Beta	α , β	41.80	68.20	Sandercock et al[124]
Utility score for an independent stroke patient	74%	Beta	α , β	216.48	76.06	
Resource use						
Cost of a GP home visit	£82					Curtis et al[135]
Cost of a GP phone call	£15					
Cost of a GP surgery visit	£25					
Cost of transportation by ambulance including paramedics	£251					Reference costs 2009-10

Variables	Value (range)	Distribution	Parameters	Parameter 1	Parameter 2	Source
Cost of attending A&E leading to admission	£237					
Cost of a CT scan	£95					
Cost of Thrombolysis (drug only)	£720					BNF 2011[136]
Stroke costs						
Cost of an independent acute stroke	£4,100	Gamma	α , λ	1	0.0001	
Cost of a dependent acute stroke	£14,935	Gamma	α , λ	1	0.0002	
Cost of long-term independent acute stroke	£1,341	Gamma	α , λ	1	0.0001	
Cost of long-term dependent acute stroke	£17,651	Gamma	α , λ	1	0.0001	
Cost of a fatal stroke	£9,664	Gamma	α , λ	1	0.0007	Sandercock et al[137]
Stroke recurrence one year after the first stroke						
70-79 years old	5.9%	Beta	n , r	75	4	PROGRESS (1999) & NICE, Lipid Modification Guidelines[108, 109]
80-89 years old	7.1%	Beta	n , r	110	8	
Probability of death from a second stroke, one year after the first stroke	34%					NICE, Statins guidelines[102]
Utility for acute stroke after stroke	0.479					Ara et al[119]
Dead	0					By definition
Annual discount rate for costs and utility	3.5%					Gray et al[1]

*Death within 1 year of stroke
**Office for National Statistics (ONS); adjusted by mortality rates in post-stroke years

This case study implemented four alternative hypothetical interventions, based on the strategies implemented in the Stroke study to reduce delay on the current thrombolysis pathway. To assess the impact of change to the model structure on the results of cost-effectiveness, the same alternative hypothetical interventions (scenarios 1 to 4) were implemented for each of the two competing model structures (see Table 6-3 below) for details:

- Scenario 1 assumed full (100%) implementation of a timely GP referral consisting of diverting GP calls to the ambulance service after stroke recognition
- Scenario 2 assumed partial (64%) implementation of a timely GP referral consisting of diverting GP calls to the ambulance service after stroke recognition
- Scenario 3 assumed full (100%) implementation of a timely CT scan to ensure that all who presents with stroke receive an immediate CT scan
- Scenario 4 assumed partial (25%) implementation of a timely CT scan to ensure that all who presents with stroke receive an immediate CT scan

The main results of the cost-effectiveness analyses for the hypothetical interventions were compared using the two competing model structures (Table 6-3).

Table 6-3: Description of hypothetical interventions and scenarios for implementation in the case study

Area for improvement based on baseline data (n=355)	**Hypothetical interventions with reference to literature (Estimates noted)**	**Maximum # of patients benefiting from removal of block in pathway n(%) – or theoretical hypothetical interventions (n=355)**	**Predicted benefit based on that previously recorded in the literature n(%) – or achievable hypothetical intervention**
Timely referral: 16% of patients contacted General Practitioner following onset of stroke and arrival in hospital was delayed compared to those who called the emergency services.	**Divert GP calls to ambulance service** General practice staff are trained to better recognise stroke, resulting in patients who initially contact their General Practitioner being referred immediately to hospital. Up to 64% immediately referred to EMS	**Scenario 1** 57 patients (100%)	**Scenario 2** 36 patients (64%)
Timely CT scan 14% of patients who arrived in hospital within 4.5 hours of symptom onset did not receive an immediate CT scan within the time window for thrombolysis.	**Ensure all who present with stroke receive immediate CT scan** In-hospital stroke services are reorganised to ensure patients with stroke receive a timely CT scan (e.g. CT scanner moved closer to the emergency department ward). Reduce time to CT scan by an hour	**Scenario 3** 51 patients (100%)	**Scenario 4** 13 patients (25%)

Note: Scenarios 1 to 4 were run using a decision tree model whilst scenarios 5 to 8 were run using a decision tree followed by a Markov model

A cost-utility analysis was undertaken for all models to calculate the cost per quality-adjusted life year gained. Results from each alternative model specification are presented as scenario analyses. One-way sensitivity analyses were undertaken to reflect the uncertainty and imprecision surrounding specific costs: CT scan costs were doubled; the cost of an individual dose of alteplase was increased by 50%; and long-term care costs (for dependent and independent patients) were increased by 20%. The method of recruitment used in the Stroke model meant that it was possible for patients with more severe stroke to have been excluded from the sample population (because patients who were seriously ill or died in hospital could not be approached for consent)[123]. The impact of this potential bias was examined by increasing the proportion of dependent patients with stroke with an mRS score of >=3 by 5% and 10%). Probabilistic sensitivity analysis was used to explore the implications of parameter uncertainty, where parameter values were sampled from distributions describing each variable in the model (Table 6-2). Using 10,000 Monte Carlo simulations, uncertainty was propagated through the model by randomly selecting values from distributions describing each model parameter (shown in Table 6.2). Cost-effectiveness planes and cost-effectiveness acceptability curves were constructed to estimate the probability of the hypothetical interventions to be cost-effective at different willingness-to-pay thresholds.

The impact of structural uncertainty in terms of the impact on the cost-effectiveness results of a model and EVPI are presented. Including different parameters in the model can be expected to alter the extent of uncertainty captured in the EVPI calculation. As the models

have different parameter sets, comparisons of expected value of partial perfect information would not be helpful.

6.4. Results

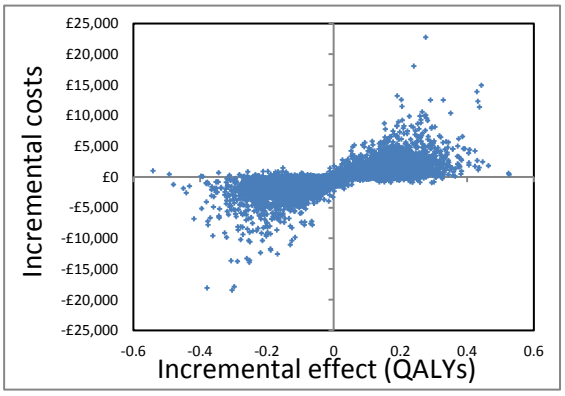
The main cost-effectiveness results obtained in the Stroke study were found to be robust when a new model structure (decision tree followed by a Markov model) was implemented, and when secondary events were included. In other words, any intervention that increased thrombolysis rates in acute stroke remained dominant, i.e. it was more effective and cheaper. All implemented hypothetical interventions (either full implementation or partial implementation) reduced costs and increased QALYs (Table 6-4). Both model structures, a single decision tree and a decision tree followed by a Markov model, indicated that the hypothetical intervention with the largest cost saving was a timely CT scan (Table 6-4).

The cost effectiveness plane shows the results from the Monte Carlo simulation for 10,000 replications. The generated estimates appearing as points on the plane for all scenarios 1 to 4 corresponding to the decision tree and to the decision tree followed by a Markov model are scattered across all four quadrants and clustered around the origin. Substantial portions of the joint density ($\Delta C/\Delta E$) (Figure 6-4) are contained within the NE quadrant (between 45 to 48 per cent of the points in all scenarios considered) and SW quadrant (between 43 to 45 per cent of the points in all scenarios considered). This seems to indicate a strong linear correlation between costs and QALYs: to gain an additional extra QALY is increasingly more expensive in this setting and vice versa.

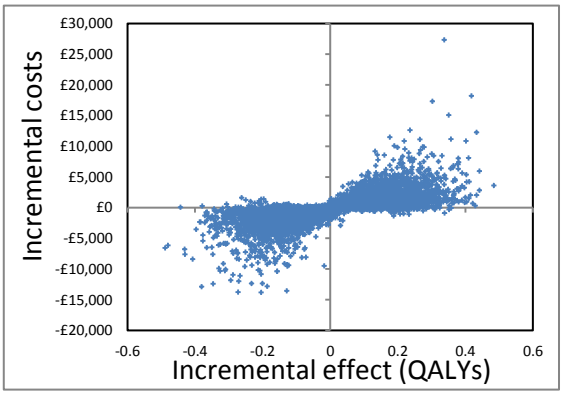
Table 6-4: Results of cost-effectiveness per patient

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Decision tree					
Current Practice	43,954	6.9996			
1. Timely GP referrals (100% implementation)	43,861	7.0105	-93	0.0109	Dominant
2. Timely GP referrals (65% implementation)	43,888	7.0074	-66	0.0078	Dominant
3. Timely CT scan (100% Implementation)	43,804	7.0185	-150	0.0189	Dominant
4. Timely CT scan (25% Implementation)	43,916	7.0042	-38	0.0047	Dominant
Decision tree plus Markov model					
Current Practice	60,424	6.4595			
1. Timely GP referrals (100% implementation)	60,401	6.4679	-23	0.0085	Dominant
2. Timely GP referrals (65% implementation)	60,408	6.4656	-16	0.0061	Dominant
3. Timely CT scan (100% Implementation)	60,390	6.4714	-33	0.0120	Dominant
4. Timely CT scan (25% Implementation)	60,414	6.4631	-10	0.0036	Dominant

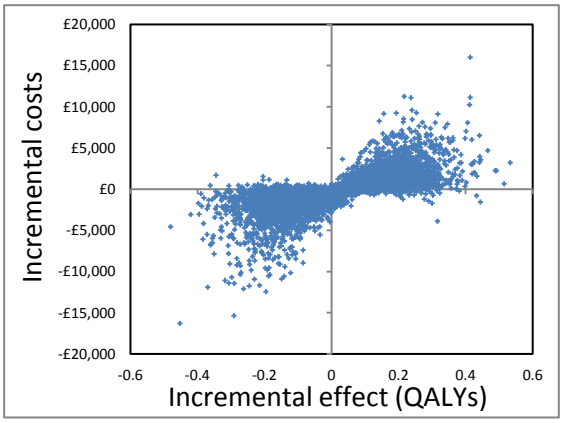
Note: Results of incremental costs and incremental effectiveness are against current practice



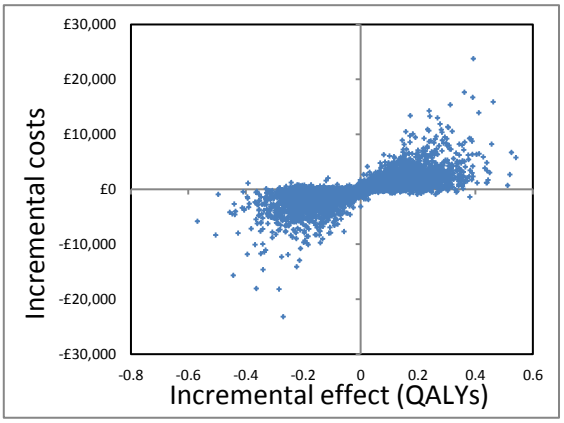
Decision-tree (scenario 1)



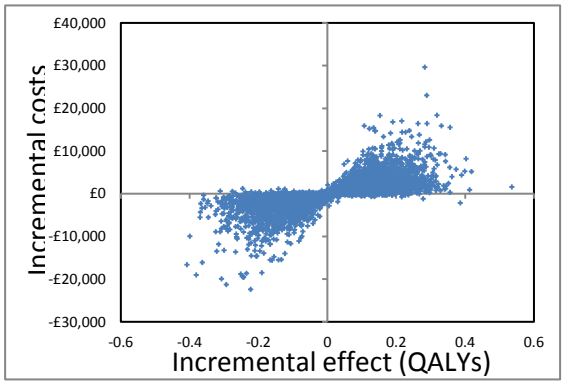
Decision-tree (scenario 2)



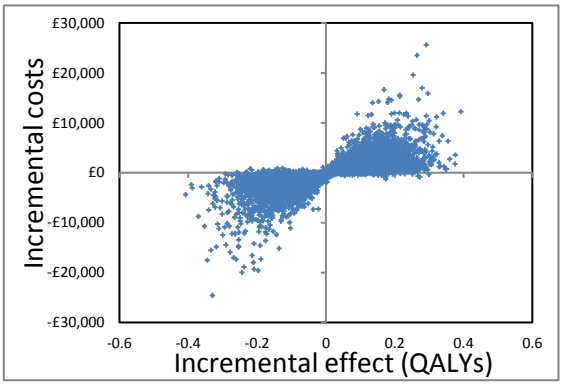
Decision-tree (scenario 3)



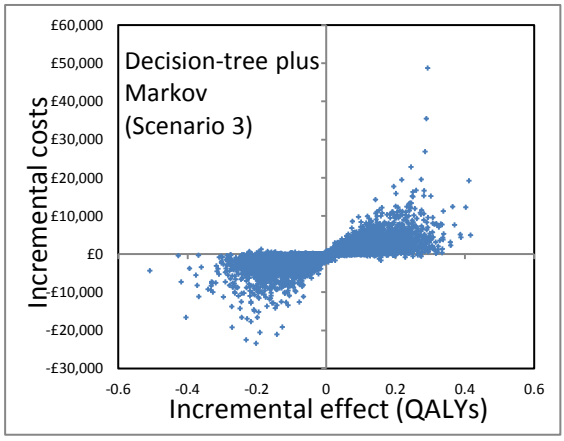
Decision-tree (scenario 4)



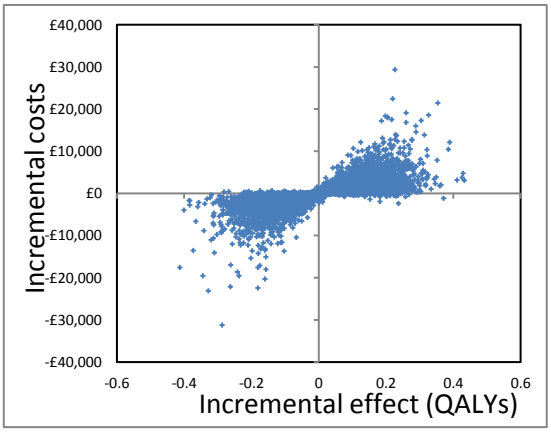
Decision-tree plus Markov model (scenario 1)



Decision-tree plus Markov model (scenario 2)



Decision-tree plus Markov model (scenario 3)



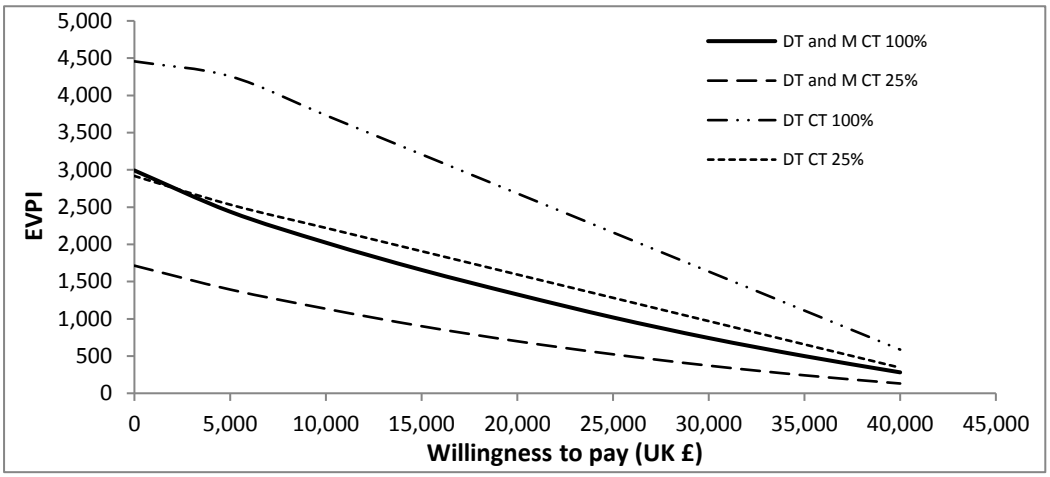
Decision-tree plus Markov model (scenario 4)

Figure 6-4 Cost effectiveness plane (CEP) for scenarios 1 to 4 – Decision Tree and Decision Tree followed by a Markov model

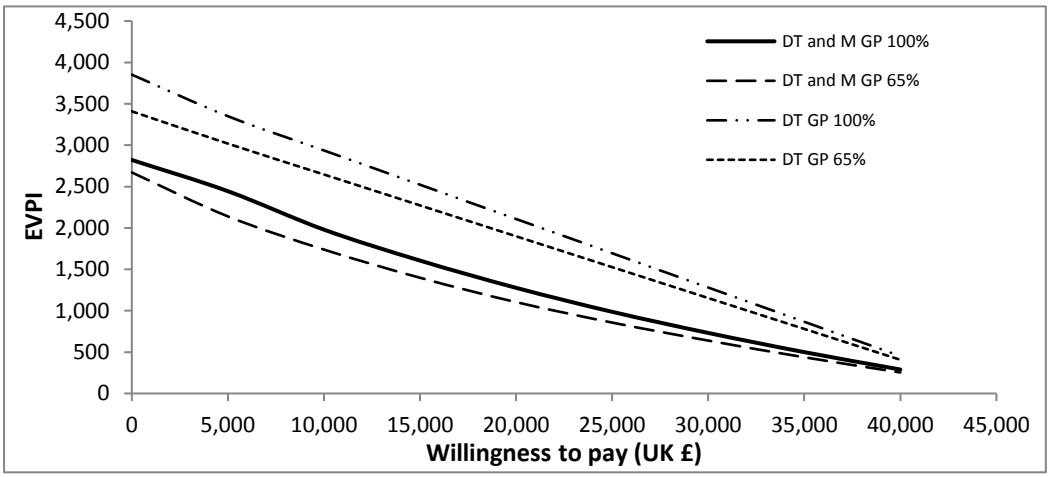
Lifetime EVPI for the decision-tree followed by a Markov model compared to the base-case decision-tree was reduced substantially at all willingness to pay thresholds and under the various scenarios (Figure 6-5).

The cost-effectiveness acceptability curves shown in

Figure 6-6 were derived from the joint density of incremental costs and incremental QALYs for the hypothetical interventions. Each CEAC presents the probability that a hypothetical intervention is cost-effective for the different scenarios and under different model structures. For a willingness to pay of £20,000 per QALY gained, scenarios 3 and 4 had the highest probability of being cost-effective (66%) when the decision tree model was used. This CEAC increased and after a threshold of £20,000 per QALY started decreasing. Scenario 3 had the highest probability of being cost-effective (62%) if the decision tree followed by a Markov model was used. The particular shape of this CEAC is explained by the fact that the confidence ellipse lies around the origin with most of the values contained within the NE and SE quadrant of the CE plane. In other words, around half of the starting points of the curve are at 50% (since half of the circle is below the x-axis) and the limit of the curve is around 50% (since half of the circle is left of the y-axis). Since for increasing ceiling ratios around half of the circle is below the threshold line (the acceptability curves show little variation in the probability that the alternative is cost-effective when the threshold changes quite substantially), the acceptability curve will resemble a straight line.

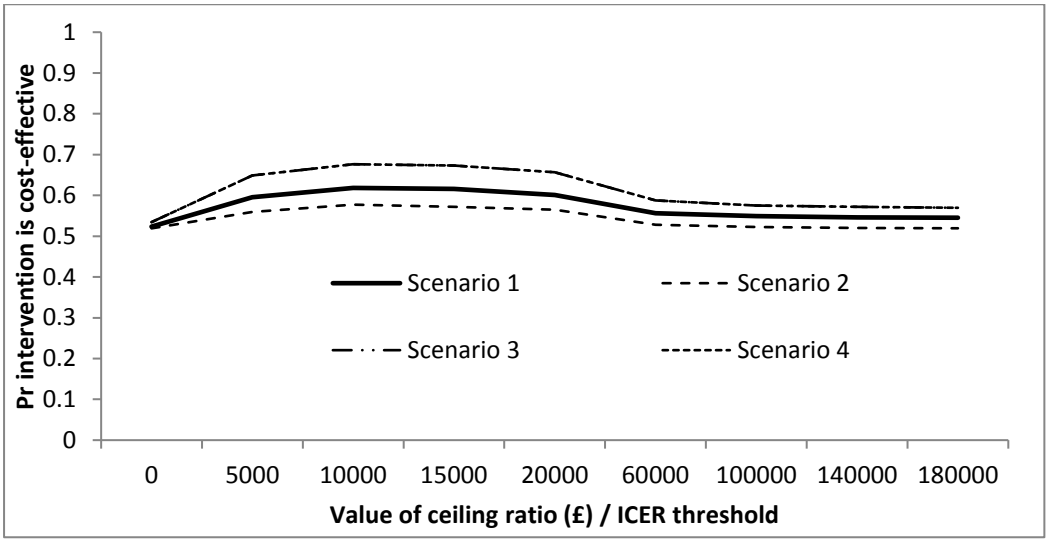


Part A Scenarios 1 and 2

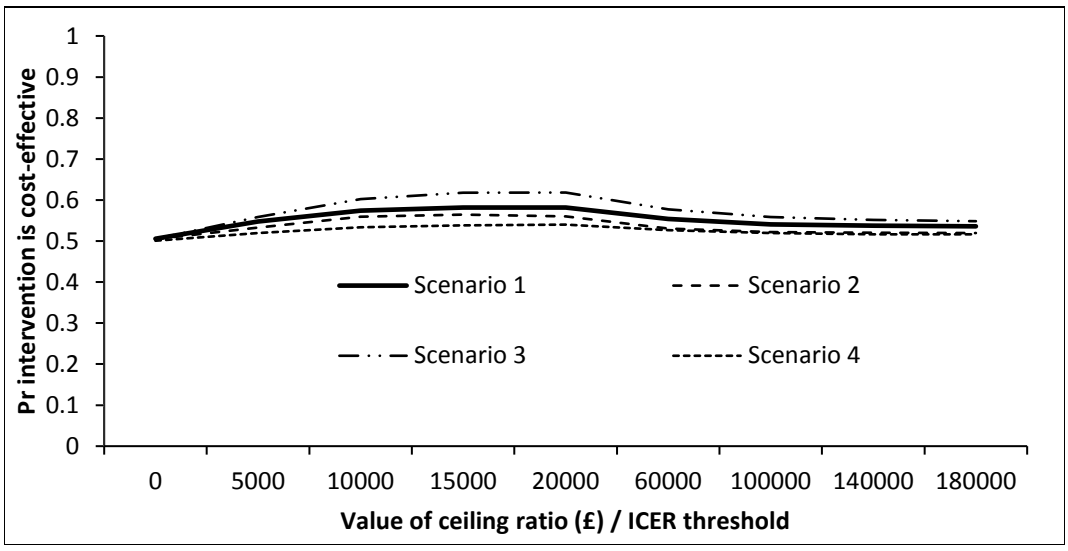


Part B Scenarios 3 and 4

Figure 6-5 Per-patient expected value of perfect information (EVPI) across varying willingness to pay values for the Decision-Tree model and the Decision-Tree Followed by a Markov model - Part A scenarios 1 and 2; Part B scenarios 3 and 4



Part A Decision Tree – scenarios 1 to 4



Part B Decision tree followed by a Markov model – scenarios 1 to 4

Figure 6-6 The cost effectiveness acceptability curves (CEAC) for the probability that hypothetical interventions are cost-effective: Part A – Decision Tee model; Part B – Decision Tree followed by a Markov model

The various sensitivity analyses undertaken seem to indicate that individual results from the various scenarios for the different model structures remained aligned after doubling the cost of a CT scan (Table 6-5), increasing the cost of alteplase by 50% (Table 6-6) and after increasing the cost of long-term care by 20% (Table 6-7). Increasing the proportion of patients being dependent after stroke by 5% did not change the results of cost-effectiveness (Table 6-8). When the proportion of patients being dependent after stroke was increased by 10%, under the new model structure (decision tree followed by Markov model), hypothetical intervention were not dominant (Table 6-9).

Table 6-5: Lifetime costs and outcomes per patient after doubling costs of CT scan

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Stroke Study (decision tree)					
Current Practice	44,049	6.9996			
1. Timely GP referrals (100% implementation)	43,956	7.0105	-93	0.0109	Dominant
2. Timely GP referrals (65% implementation)	43,983	7.0074	-66	0.0078	Dominant
3. Timely CT scan (100% Implementation)	43,899	7.0185	-150	0.0189	Dominant
4. Timely CT scan (25% Implementation)	44,011	7.0042	-38	0.0047	Dominant
Stroke Study (Decision tree plus Markov model)					
Current Practice	60,519	6.4595			
1. Timely GP referrals (100% implementation)	60,496	6.4679	-23	0.0085	Dominant
2. Timely GP referrals (65% implementation)	60,503	6.4656	-16	0.0061	Dominant
3. Timely CT scan (100% Implementation)	60,485	6.4714	-33	0.0120	Dominant
4. Timely CT scan (25% Implementation)	60,509	6.4631	-10	0.0036	Dominant

Note: Results of incremental costs and incremental effectiveness are against current practice

Table 6-6: Lifetime costs and outcomes per patient after increasing costs of Alteplase by 50%

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Stroke Study (decision tree)					
Current Practice	43,978	6.9996			
1. Timely GP referrals (100% implementation)	43,890	7.0105	-88	0.0109	Dominant
2. Timely GP referrals (65% implementation)	43,916	7.0074	-62	0.0078	Dominant
3. Timely CT scan (100% Implementation)	43,837	7.0185	-141	0.0189	Dominant
4. Timely CT scan (25% Implementation)	43,942	7.0042	-36	0.0047	Dominant
Stroke Study (Decision tree plus Markov model)					
Current Practice	60,515	6.4595			
1. Timely GP referrals (100% implementation)	60,496	6.4679	-19	0.0085	Dominant
2. Timely GP referrals (65% implementation)	60,502	6.4656	-13	0.0061	Dominant
3. Timely CT scan (100% Implementation)	60,487	6.4714	-28	0.0120	Dominant
4. Timely CT scan (25% Implementation)	60,507	6.4631	-8	0.0036	Dominant

Note: Results of incremental costs and incremental effectiveness are against current practice

Table 6-7: Lifetime costs and outcomes per patient after increasing costs of long-term care by 20%

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Stroke Study (decision tree)					
Current Practice	51,461	6.9996			
1. Timely GP referrals (100% implementation)	51,351	7.0105	-110	0.0109	Dominant
2. Timely GP referrals (65% implementation)	51,384	7.0074	-78	0.0078	Dominant
3. Timely CT scan (100% Implementation)	51,283	7.0185	-178	0.0189	Dominant
4. Timely CT scan (25% Implementation)	51,416	7.0042	-45	0.0047	Dominant
Stroke Study (Decision tree plus Markov model)					
Current Practice	70,615	6.4595			
1. Timely GP referrals (100% implementation)	70,588	6.4679	-27	0.0085	Dominant
2. Timely GP referrals (65% implementation)	70,597	6.4656	-18	0.0061	Dominant
3. Timely CT scan (100% Implementation)	70,575	6.4714	-40	0.0120	Dominant
4. Timely CT scan (25% Implementation)	70,604	6.4631	-11	0.0036	Dominant

Note: Results of incremental costs and incremental effectiveness are against current practice

Table 6-8: Lifetime costs and outcomes per patient by increasing the proportion of dependent patients by 5%

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Stroke Study (decision tree)					
Current Practice	45,698	6.9569			
1. Timely GP referrals (100% implementation)	45,631	6.9672	-67	0.0103	Dominant
2. Timely GP referrals (65% implementation)	45,651	6.9643	-47	0.0074	Dominant
3. Timely CT scan (100% Implementation)	45,593	6.9747	-104	0.0178	Dominant
4. Timely CT scan (25% Implementation)	45,671	6.9613	-27	0.0044	Dominant
Stroke Study (Decision tree plus Markov model)					
Current Practice	61,775	6.4317			
1. Timely GP referrals (100% implementation)	61,773	6.4398	-3	0.0080	Dominant
2. Timely GP referrals (65% implementation)	61,774	6.4375	-1	0.0058	Dominant
3. Timely CT scan (100% Implementation)	61,771	6.4431	-5	0.0114	Dominant
4. Timely CT scan (25% Implementation)	61,774	6.4352	-1	0.0034	Dominant

Note: Results of incremental costs and incremental effectiveness are against current practice

Table 6-9: Lifetime costs and outcomes per patients by increasing the proportion of dependent patients by 10%

	Costs	QALYs	Incremental cost	Incremental QALYs	ICER
Stroke Study (decision tree)					
Current Practice	47,458	6.9138			
1. Timely GP referrals (100% implementation)	47,421	6.9233	-37	0.0095	Dominant
2. Timely GP referrals (65% implementation)	47,433	6.9206	-25	0.0069	Dominant
3. Timely CT scan (100% Implementation)	47,405	6.9303	-53	0.0165	Dominant
4. Timely CT scan (25% Implementation)	47,444	6.9178	-14	0.0041	Dominant
Stroke Study (Decision tree plus Markov model)					
Current Practice	63,139	6.4037			
1. Timely GP referrals (100% implementation)	63,160	6.4113	20	0.0076	2,688
2. Timely GP referrals (65% implementation)	63,154	6.4092	15	0.0054	2,803
3. Timely CT scan (100% Implementation)	63,167	6.4144	28	0.0107	2,626
4. Timely CT scan (25% Implementation)	63,148	6.4070	9	0.0032	2,760

Note: Results of incremental costs and incremental effectiveness are against current practice

6.5. Discussion

This case study was developed in response to findings in Chapters 3 and 4 indicating that: 1) the assessment of structural uncertainty is not common practice; and 2) additional research on the impact of changing model structures on results of cost-effectiveness is needed. Specifically, this chapter has explored whether the cost-effectiveness results from the Stroke study would have been different had the structure and type of model been different or had other elements pertaining to structural uncertainty been considered.

This case study identified and implemented a plausible alternative type of model structure based on results of a systematic review of economic evaluations focused on a similar research question. The selected model type consisted of a decision-tree followed by a Markov model. The decision tree reflected the short term care pathway and the Markov structure reflected the lifetime costs and outcomes. Other types of models such as a DES or ISM were rejected on the basis of model parsimony. The results of cost-effectiveness for each alternative model and scenarios specified, including the EVPI, were presented and compared.

The main cost-effectiveness results in the stroke model did not change when an alternative decision-tree followed by a Markov model was implemented, including an adjustment to allow for the recurrence of stroke. This may suggest that structural uncertainty was not important in the stroke model. Results for EVPI in the decision-tree followed by a Markov model decreased for all scenarios, indicating that a large part of the decision uncertainty was not apparent in the model.

The results here corroborate previous findings indicating that all intervention strategies that increase thrombolysis rates in acute stroke are cost-effective because of a reduction in dependency after stroke and the subsequent reduction in long-term care costs[123].

The results presented in this chapter for this particular clinical area and research question, lead to the conclusion that the simplest model (decision tree) would suffice in projecting the potential impact of alternative interventions aiming at increasing thrombolysis rates through optimisation of the care pathway for acute stroke.

However, an important variation in model structures was identified by the systematic review for similar research questions, and this corroborates findings in Chapter 5 indicating that improved guidance to handle the implications of potential sources of structural uncertainty are needed.

The question remains as to whether this results would have been different, had the research question be different, or the clinical area, or the results of cost-effectiveness. Further research is required to identify the forms of structural uncertainty analyses that might be relevant to best inform decision making according to the characteristics of the disease being modelled.

6.6. Strengths and limitations

As previously indicated in Chapter 5, a weakness in the approach followed to assess structural uncertainty in this chapter is the lack of established methods to formally assess the plausibility of alternative models or to start understanding how many scenarios should be considered.

More sophisticated methods to assess structural uncertainty such as the discrepancy approach were not implemented due to data restrictions to estimate the maximum likelihood of parameters.

This chapter used as a case study a model taken from work previously published on the cost-effectiveness of optimising acute stroke care services for thrombolysis[123]. The model structure of the decision-tree was affected by some of the nuances associated with the pathway (service evaluation) type of models These nuances include the healthcare interventions received based upon what was known or believed by healthcare practitioners at the time, and the extent of the model complexities which may not or could not be appropriately captured by the model.[138] However, on building the model, relevant NICE clinical guidelines and expert opinion were incorporated.

The choice of model type for the case study was limited to a decision tree followed by a Markov model. Some may argue that a microsimulation or DES may offer some advantages in the particular case of the Stroke study. An essential feature of a DES is that patients compete for limited resources however, in the Stroke study patients were not competing for resources as any patient presenting in A&E within 4.5 hours of the onset of symptoms and was found eligible, received thrombolysis. An ISM on the other hand is more useful when the aim is to assess the timing of the response to an event. However, the aim of the Stroke study was to assess the effectiveness of key hypothetical interventions in increasing the rates of thrombolysis among patients who had suffered an ischaemic stroke rather than changes to the timing of the intervention.

The results of cost-effectiveness analyses indicated that increasing thrombolytic rates via the implementation of interventions designed to optimize the acute care pathway for stroke patients is clinically superior and cost saving. This case study may have limited ability to be generalisable to other disease areas since thrombolysis for acute stroke is associated with substantial QALY gains, typically not seen with many other therapies. It may be that if the results were near the £20,000 per QALY threshold, changes in model structure could have led to different conclusions regarding cost-effectiveness and possible EVPI.

Other limitations may relate to the model perspective adopted. A National Health Service (NHS) and Personal Social Services (PSS) perspective was utilized here meaning that other benefits associated with service evaluation, for example, the potential improvement in people's perceptions of urgent care, were not measured.

The assessment of structural uncertainty shown in published studies in the area of secondary prevention of Stroke/TIA has mainly focused on assessing parameter uncertainty. This chapter has extensively examined structural uncertainty showing the effect of a plausible alternative model structure and scenarios on the results of cost-effectiveness through implementing extensive scenario and EVPI analyses.

6.7. Conclusion

The results in this chapter indicate that the main conclusions of cost-effectiveness analysis from the Stroke model are robust to plausible changes to the model structure. The cost-

effectiveness results and the EVPI were not sensitive to model structure specification. These results were also robust to the various sensitivity analyses undertaken which lead to the conclusion that, in this particular case study, a parsimonious model would suffice. As is normal in the economic evaluation of healthcare technology using decision-analytic modelling, the mathematical model represents a simplification of reality and the results presented here need to be interpreted in relation to the assumptions used and evidence available.

The results in this chapter corroborate the findings in Chapter 5 and have demonstrated the need to strengthen the current guidelines to indicate more clearly how structural uncertainty should be identified, assessed, and reported. Most importantly, in which particular circumstances or for which disease areas structural uncertainty arising from the structure of a model is more likely to affect the results of cost-effectiveness.

Based on the findings of this chapter, the following recommendations are proposed:

1. The assessment of structural uncertainty should not be ignored as it is an integral part of good practice DAM
2. The reasons why an assessment of structural uncertainty is not possible or not needed should be always stated as a research limitation
3. Data limitations that impact on the assessment of structural uncertainty should be stated and discussed
4. If there is a reason to believe that structural uncertainty is an issue that may have affect the results of CE, then an assessment of structural uncertainty should be included

5. If sophisticated methods to assess structural uncertainty are not implemented due to data limitations (e.g. discrepancy approach, model averaging, parameterization, model selection), then at least scenario analysis or appropriate sensitivity analysis should be routinely conducted, as per current ISPOR-SMDM guidelines

CHAPTER 7. DISCUSSION

The overall aim of this thesis was to examine the understanding of good practice in DAM, to assess the extent to which economic evaluations in primary prevention of cardiovascular disease and acute care interventions for stroke adhered to good practice guidelines, and to assess how structural uncertainty arising from the choice of model structure impacts on cost-effectiveness results in two alternative settings.

In particular, this thesis has combined the use of methodological research and quantitative analysis to:

1. Examine all available guidelines and statements of good practice in decision-analytic modelling between 1990 to 2012 to identify currently available good practice guidelines
2. Develop a five-dimension framework to assess the compliance of current practice to guidelines in decision-analytic modelling
3. Assess to what extent model-based economic evaluations of primary prevention interventions aimed at lowering blood pressure in patients with hypertension or at risk of developing hypertension have complied with published decision-analytic modelling guidelines using the framework developed in Chapter 2

4. Develop two case studies illustrating the assessment of structural uncertainty arising from the choice of model structure using scenario analysis and extensive sensitivity analysis

This chapter discusses the key findings, draws conclusions, and make recommendations for further research.

7.1. Main Findings

There is wide acceptance that economic methods such as DAM contribute to the decision making process in the health care setting by offering a coherent and theoretically based approach to identifying, measuring and valuing resource use, costs and outcomes, and by handling uncertainty[1, 2, 9]. Since the 1990s, DAM has been widely used to synthesise clinical and economic evidence and to inform resource allocation decisions for the purpose of allowing scarce health care to be allocated more efficiently[9].

The growing use of modelling in economic evaluations has led to increased scrutiny of the methods used, including clear requirements for researchers in terms of good practice in DAM. Good practice considers factors such as the need to incorporate appropriate evidence, to consider all relevant comparators, and reflect uncertainty in the evidence when presenting results.

This thesis has demonstrated that even though contemporaneous DAM guidelines have kept pace with recent progress in the way economic evaluation methodology has progressed, there are aspects of DAM that require further development and/or strengthening.

7.1.1. Contemporary understanding of 'good practice' in decision analytic modelling

A comprehensive review of DAM guidelines was undertaken that included 33 studies, and was the first to critically assess all available guidelines and statements of good practice since 2006. The results of the systematic review have demonstrated that contemporaneous DAM guidelines have kept pace with recent progress in the way economic evaluation methodology has progressed and have ensured guidelines remain current, effective and helpful. Furthermore, DAM guidelines should be seen as tools that if followed will lead to the results obtained in economic evaluations being more credible, and adherence to these guidelines is considered as best practice in DAM.

The review identified some aspects of DAM guidelines that require further development. The choice of model structure: whether model structure should be informed by data availability or not remains an ongoing issue. Structural uncertainty remains an area of controversy. An inappropriate structure can invalidate the conclusions drawn from cost-effectiveness analyses, while choices made when structuring a model can significantly affect its results and the inferences from it. Model validity, in particular some aspects of model generalisability that require examining whether model's predictions are reliable, demands further research. It is argued that model external validation be based on new data that is external to the data used

for model development. However, most available data will typically be used in model development and there will often be relatively little or even no data left available for undertaking external validation.

This thesis found that new sets of guidelines have been developed since 1990, kept up to date and are generally available. However, these may at times lack practicality due to the extensive amount of reports and information available, making them complex for researchers to follow and implement. From the findings of this research it is apparent that standards of reporting cost-effectiveness results could be improved if additional effort were made to produce a single, comprehensive, user-friendly, up-to-date and practical instrument to direct researchers towards the key elements of good practice in DAM.

Following on from this finding, as part of this thesis, a new DAM framework incorporating and reflecting on all the evidence found in the review was developed. The framework synthesised current and agreed guidelines in a five-dimensional checklist instrument: problem concept, model concept, synthesis of evidence, analysis of uncertainty, and model transparency and validation. The framework constitutes the first of its type since Philips et al[17]. It is recommended here *that ISPOR, as part of their mission to improve decision making for health globally,* undertakes a periodic review and update of this framework, *whenever new statements of good practice are made available or otherwise, at least every two years,* as this would allow sufficient time for the application of any new guidelines to be adopted into practice.

7.1.2. Compliance to good practice guidelines

The results of the review on DAM guidelines indicated that compliance to guidelines is considered best practice. However, few studies in the past have been concerned with assessing the extent to which current practice has adhered to DAM guidelines.

In assessing adherence to DAM guidelines, care was taken to remove all possible variation between models which was not relevant for the purpose of assessing compliance (for example, different outcomes, treatment options or sources of uncertainty). This was achieved by focusing on one particular clinical area. The clinical area of cardiovascular disease prevention was selected for the assessment of compliance to DAM guidelines due to the large number of recent model-based cost-effectiveness analyses conducted in this area, and the fact that CVD is a commonly modelled clinical area where intermediate outcomes such as blood pressure reduction can be tracked. The focus here was on interventions aimed at lowering blood pressure in patients at risk of developing cardiovascular disease.

Limited compliance was more commonly justified by a lack of data or imperfect data, and this was particularly apparent in the assessment of structural uncertainty and external model validation. This appeared to pose challenges to researchers for example, when modelling the risk of secondary events and disease progression or when undertaking model external validity.

Even though elements pertaining to structural uncertainty were acknowledged in some of the studies reviewed, the assessment of structural uncertainty was not found to be common practice. Guidelines have recommended assessing structural uncertainty as part of good practice. Furthermore, this thesis found literature illustrating the assessment of structural

uncertainty using various sophisticated statistical methods (Chapter 4). However, it was found that these methods are not being implemented in current practice to assess structural uncertainty, indicating a possible lack of practicality, excess complexity, or a lack of adequate data necessary for them to be implemented. This findings seem to indicate that additional guidelines or the strengthening of current guidelines are still needed to aid researchers in identifying and quantifying the effect of structural uncertainty on cost-effectiveness results.

Similarly, model external validity was not found to be common practice and model external validation poses important challenges to researchers in terms of actual data or the lack of availability of big data, randomised controlled trial, or patient-level data to allow the exercise.

7.1.3. Structural uncertainty surrounding the choice of model structure

This thesis found that many issues related to structural uncertainty have been identified, making it a complex task to address properly. As previously mentioned, statistical methods have been developed to address and to report the results of structural uncertainty, however most of them have been the object of criticism (Chapter 4). Specific guidelines are needed in order to aid researchers in identifying what elements are of greater importance for different disease areas, and how best to identify, address and report the results of structural uncertainty. Findings in Chapters 5 and 6 have demonstrated that for this particular clinical area, disease conditions, and research questions, parsimonious models sufficed and model structural uncertainty did not seem to affect the results of cost-effectiveness.

The case studies developed in this thesis to assess the impact of choice of model structures on cost-effectiveness results modelled the course of prevention and treatment of cardiovascular disease. For each case study, alternative model structures were identified from the literature, implemented, and then the results of cost-effectiveness analyses compared. An important consideration in the assessments of these results was the extent to which cost-effectiveness results obtained in the case studies were robust to the assumptions adopted (for example, the effect of secondary events) and to sensitivity analyses (deterministic and probabilistic).

After implementing competing model structures that varied between a simplified structure and a complex one, it was found that the main cost-effectiveness results obtained in the TASMIN-SR model and in the Stroke model were robust to changes in model structure and to the inclusion of secondary events. In other words, dominance was still demonstrated in the results, with higher QALYs and lower costs for the intervention strategy in the case of TASMIN-SR and for any strategy leading to the increased of thrombolysis rates in the case of the Stroke model. For this reason, it is difficult to draw conclusions regarding the importance of correctly exploring model structural uncertainty based on the results of the case studies. Therefore, these results are not generalisable to other clinical areas or disease conditions for which the results may have been different and for which structural uncertainty surrounding model structure may still remain an area of controversy, for example, metastatic end-stage cancer modelling where results of cost-effectiveness vary depending on the choice of model type and model structure.

Inappropriate model structures can invalidate the conclusions drawn from cost-effectiveness analyses, while choices made when structuring a model can significantly affect the results obtained and the inferences from it.

It has long been argued[14-16] that the art of building models rests on the principle of parsimony, according to which model selection should value both accuracy and simplicity. Finding the right balance between simplicity of modelling and avoidance of over simplification is difficult. However, it is getting this balance right that makes the difference in terms of model transparency and model credibility: excessive model details and complexity reduces transparency and can lead to distrust in models[21]. Simplicity in models which relates to the size of the model and not to the modelling technique used, has been recognised as an advantage because simpler models are usually easier to understand and to validate[13].

Challenges posed by the assessment of structural uncertainty might be overcome if additional research were undertaken on an experimental basis, for example by measuring the impact of changing model structures on cost-effectiveness results. Case studies aimed at measuring the impact of changing or adapting model structures could provide insightful evidence as to how much cost-effectiveness results would be altered when alternative model structures are implemented. This may or may not be always feasible, depending on how complicated models might be or whether patient level data are available to support the development of this type of exercises.

Current practice seems bound by data availability whilst methods proposed to assess structural uncertainty have been borrowed from other disciplines oblivious to the needs in a health care setting where patient level data is not always readily available.

The main recommendations arising from the case studies are:

1. The assessment of structural uncertainty should not be ignored as it is an integral part of good practice DAM
2. The reasons why an assessment of structural uncertainty is not possible or not needed should always be stated as a research limitation
3. Data limitations to undertake an assessment of structural uncertainty should be clearly stated and discussed
4. If there is a reason to believe that structural uncertainty is an issue that may have affected the results of CE, then an assessment of structural uncertainty should be included
5. Ideally, sound statistical methods should be used in the assessment of structural uncertainty (e.g. discrepancy approach, model averaging, parameterization, model selection) but if none are possible due to data limitations, then at least scenario analysis and appropriate sensitivity analysis should be conducted, as per the case studies implemented in this thesis
6. ISPOR-SMDM guidelines should be followed at the very least by implementing sensitivity analysis in the assessment of structural uncertainty

7.2. Strengths and limitations

This thesis has a number of strengths in terms of presenting novel research in three main areas. First, it developed, for the first time since Philips et al[26] an updated framework to assess the quality of decision analytic models. This framework, named the five-dimension framework, proposes that the criteria for assessing the quality of model-based economic evaluation fall into five areas: problem concept, model concept, synthesis of evidence, model uncertainty, and model transparency and validation. Second, it undertook an in-depth assessment of the quality of decision analytic models and the value of the five-dimension framework in assessing compliance to DAM guidelines. The assessment of compliance to DAM guidelines was conducted on economic evaluations of primary prevention interventions aimed at lowering blood pressure in patients with hypertension or at risk of developing hypertension. Compliance to DAM guidelines was assessed using the five-dimension framework. The value of the five-dimension framework was assessed from the point of view of its usefulness in reviewing results of DAM. Third, this thesis developed an assessment of the impact of model structural uncertainty in the area of prevention and treatment of cardiovascular disease. This thesis has demonstrated how structural uncertainty arising from model choice can be assessed using scenario analysis, extensive sensitivity analyses and in terms of expected value of perfect information. The main results of the cost-effectiveness here have indicated that the main conclusions were robust to changes in model structure, and thus raise the question, whether in this particular clinical setting, more parsimonious model structures suffice.

However, there are several limitations of the work presented in this thesis. The exclusion criteria adopted for inclusion of studies in the systematic reviews may be considered a limitation; however, these were required to guarantee consistency in the analysis. For example, the focus on one particular clinical area (CVD prevention) made it possible to remove some of the variation between models that was not relevant for the purpose of assessing the quality of model based economic evaluations (e.g., different outcomes, treatment options, or sources of uncertainty), but meant that these results cannot not be more generalisable to other settings. Similarly, the fact that this review did not look for studies published in a non-English-language may also be considered a limitation.

The choice of model type used in the case studies was limited to decision tree and Markov models and this may also be seen as a limitation. However, as previously defined, model complexity is not only related to the type of model used but as well to how complex it is. Some may argue that a microsimulation or discrete event simulation may have offered some advantages in the case of the TASMIN-SR model, such as the flexibility in incorporating individual heterogeneity and tracking individual event history. However, as per DAM guidelines, model structure is identified by considering the natural history of the disease and the answer to questions: is interaction between patients important? No; do you need to model recursive events? Yes; do you require the model to represent a lot of health states? No. All of which indicated that a Markov model was the right approach[49 13].

In the case studies considered in this thesis it was not possible to implement a model selection method to select the best model on the basis of how well the model's output matched

observed data (judged by the likelihood-based information criteria). This was because only single point estimates for key parameters were available, which did not allow the estimation of the maximum likelihood of parameters.

The cost-effectiveness results in both case studies were of dominance. It may be that if the results were near the £20,000 per QALY threshold, changes in model structure could have led to different conclusions being drawn from the cost-effectiveness analysis.

7.3. Comparison with similar research studies

There is agreement that the development of good practice guidelines is an ongoing process[21], and, as such, efforts should be made to review and update DAM guidelines in accordance with the methods being used in current practice[17, 21, 30]. Whilst it is recognised that a single framework cannot determine the quality of a model, it can serve as general guidance regarding what modelling issues are important and the details of how a model is reported[17]. The use of a framework has the additional advantage of providing a systematic approach to the assessment of quality in a model review, allowing the reviewer to focus on the key characteristics of a model and thereby identify strengths and weaknesses[17, 26]. Furthermore, the aim of a framework is to synthesise DAM guidelines that are currently available, providing pointers towards the particular aspects that require consideration in developing and appraising models[17, 26].

Model validation and the assessment of structural uncertainty represent areas posing key methodological challenges in DAM due to the lack of motivation, time and data to validate models and in the case of structural uncertainty, the lack of methods[21, 29, 30, 96]. Model structures are commonly described by linking them to the natural history disease, however they are less justified from the point of view of the adequacy of the type of model being used over competing model structures when reporting cost-effectiveness results. In the other hand, concerns have been expressed by the diversity of alternative model structures submitted to national reimbursement bodies, such as NICE, for the same disease and to answer similar research questions[96]. This ultimately reduces the comparability of evaluations of competing health technologies for the same condition, potentially leading to inconsistent public funding decisions[96]. However, decision problems across different disease areas are so varied that it may well be the case that only analysis such as the analysis undertaken in the case studies developed in this thesis can shed any light on the importance of model uncertainty in different clinical settings.

7.4. What these findings mean and the need for future research

DAM in health economics has traditionally been used to synthesise best available data or to extrapolate beyond the clinical trial to endpoints. Assessing the quality of DAM is paramount in light of the unresolved long term concerns regarding DAM. These include less well defined methodology, the incorporation of many assumptions and the use of analyst discretion[139], the increasing use of DAM in the economic evaluation of health care interventions and health

technology assessments (HTA), and newer methods being used including more sophisticated modelling techniques.

Good practice guidelines in DAM should aim to reflect the current consensus regarding the attributes that characterise a good model in terms of structure, data, and validation. To serve as such, DAM guidelines should be subjected to periodic evaluation and update. It is only through this means that DAM guidelines would remain, in words of Caro et al[21], "current, helpful and effective" p. 668.

Evaluation and the update of DAM guidelines require that reviews of their relevance as in this thesis be undertaken periodically. Chapter 2 has shown that since Philips et al[17, 26], no further attempts have been made to synthesise DAM guidelines in a single framework. To remain effective, frameworks such as the five-dimension framework developed in Chapter 2, need to be updated and adjusted in response to new developments in current practice.

Future research should aim to develop a systematic approach to evaluate and update good practice frameworks. The options to do so could consider focusing on the following questions: firstly, do current results of economic evaluation comply with good practice DAM guidelines? The answer to this question would help identify to what extent good practice frameworks are considered by researchers prior to submitting their results for publication. However this does imply that a view has been established about what is considered good practice in DAM. Secondly, do DAM guidelines improve the transparency of published results of economic evaluations? The answer to this question would allow the assessment of the transparency of reporting results of DAM and would allow users of models to evaluate the quality of models

according to criteria of good practice. However some may argue that this would not necessary answer the question of whether DAM guidelines have improved the quality of published results of model based economic evaluations[139].

Overall, limited compliance in terms of the development of model validation was identified in the area of cardiovascular disease, but this has been justified by the lack of sufficient event data to undertake the exercise. DAM guidelines have suggested that models should be based on the best evidence available at the time they are built, but at the same time, external model validity should use actual event data. This poses the difficulty of splitting available data in a way that would serve both purposes. Some experts have suggested setting aside one-third of the data for the purpose of external validation[30]. Other issues affecting model validation besides data availability are a lack of motivation, time and a short-term vision used during the process of building models[30]. It has been suggested that the development of multi-use models[30] progressively validated over time would contribute towards solving the issue of model validation.

The wide variations in the model structures identified in the literature to answer similar research questions supports the need for improved guidance to handle the implications of potential sources of model structural uncertainty. If there is a choice, a model should aim for simplicity and how simple a model should be depend upon the sensitivity of the policy implications to added complexity[139]. The best model for a given decision is the simplest model that is adequate to represent the decision problem, where adequacy may be defined as

allowing the decision maker to make the same decision(s) as would be made with a perfect model.

Modelling in health economics poses the challenge that models are developed for a particular decision making context making it difficult for other decision makers to use the published results from the modelling. This requires the results to be presented in a way that makes clear how much they are dependent on that particular context.

Model uncertainty is largely related to the choice of appropriate model structure. This thesis has demonstrated how it is possible, on a case-by-case basis, to identify and assess model adequacy from among both simplistic and more comprehensive model representations of the disease area and its impact on cost-effectiveness results. This exercise could be used to serve as a basis for the future identification and development of an appropriate common model structure for specific diseases. These types of models have been referred to as reference[96, 140] or global models[27]. The development of reference models would not eliminate structural uncertainty particularly if there is insufficient or conflicting evidence to support an appropriate model structure[96].

The assessment of structural uncertainty is not common practice and in fact, structural uncertainty is rarely addressed when presenting results of economic evaluation. Some of the challenges addressing structural uncertainty have been posed by the many elements that have been defined as pertaining to structural uncertainty including assumptions about the disease process (events and states). The case studies developed in this thesis indicated that many of the sophisticated methods proposed to assess structural uncertainty are difficult to

implement due to data restrictions. Further research is needed to develop statistical methods suitable for a health economics arena where patient level data is insufficient or sometimes unavailable in many clinical settings. Similarly, decision-analytic modelling guidelines need to be strengthened in order to aid researchers to address structural uncertainty.

Currently there are no available guidelines indicating how structural uncertainty arising from the structure of a model should be assessed and reported. Therefore, further research should focus on the development of general guidelines on how to address these issues and more specifically, how to deal with challenges across different disease areas, perhaps incentivising the development of more case studies focusing on specific areas such as those developed in Chapters 5 and 6.

CHAPTER 8. CONCLUSIONS

Good practice guidelines in decision analytic modelling should aim to reflect current consensus regarding the attributes that characterise a good model in terms of structure, data and validation. To serve as such, DAM guidelines should be subjected to periodic evaluation and update. Evaluation and update of DAM guidelines require that exercises such as these reflected in this doctoral research be undertaken periodically. It is recommended here *that ISPOR, as part of their mission to improve decision making for health globally,* undertakes a periodic review and update of this framework, *whenever new statements of good practice are made available or otherwise, at least every two years,* as this would allow sufficient time for the application of any new guidelines to be adopted into practice.

The wide variations in the model structures identified in the literature for similar research questions supports the need for improved guidance to handle the implications of potential sources of model structural uncertainty. If there is a choice, a simple model is the preferred option and how simple the model should depend upon the sensitivity of the policy implications to added complexity[139]. The best model for a given decision is the simplest model adequate to represent the decision problem, where adequacy may be defined as allowing the decision maker to make the same decision(s) as would be made with a perfect model.

This thesis has: 1) responded to the need to update and reflect on current DAM practice by developing an up-to-date five-dimensional DAM framework and has demonstrated its

usefulness in assessing compliance to DAM guidelines; and 2) answered to the need for more case studies comparing the efficiency of simple versus complex model structures in the results of cost-effectiveness analyses and by doing so, has set the foundation for the development of reference or global models.

APPENDICES

Appendix 1. Search results for Decision Analytic Modelling guidelines

EMBASE AND MEDLINE(R)

Steps	Search criteria	Number of hits
1	(checklist$ or check list$ or standards or standardi?ation or peer review or rules or critiquin or criteria or good or bad or correct$ or bias or fundamentals recommend$ or best or strength$ or weakness$ or quality or qualities or validity or guideline$ or validation or checkpoint$ or critically appraise or problems or limitations or rating scale$ or framework$ or protocol$ or audit or principles or methodolog$ or validate or validation or evaluating or properties or guidance or integrity or evaluation or pros or cons).m_titl.	2261664
2	limit 1 to abstracts	1470011
3	limit 2 to English language	1266438
4	limit 3 to yr="1990 -Current"	1052800
5	limit 4 to humans	684348
6	(economic model$ or Markov model$ or mathematical model$ or cost model$ or decision model$ or pharmacoeconomic$ model$ or decision tree$ or decision data or decision analytic$ or decision analysis or economic evaluation? or economic analysis).m_titl.	30075
7	limit 6 to abstracts	24245
8	limit 7 to English language	21989
9	limit 8 to yr="1990 -Current"	19501
10	limit 9 to humans	12259
11	5 and 10	3930
12	**remove duplicates from 11**	**2486**

Appendix 2. Search results Decision Analytic Modelling Guidelines from COCHRANE library

Steps	Search criteria	Number of hits
1	economic model* or economic analysis* or economic evaluation* or decision analytic* or decision analysis* or economic study* or economic submission* from 1990 to 2013, in Methods Studies, Technology Assessments and Economic Evaluations (Word variations have been searched)	14677
2	guideline*	13983
3	#1 and #2	**1664**

Appendix 3. Search strategy for the applied economic evaluations

Cochrane databases (searched 20 March 2015 for the period 2000 to 2015) and NHS EED (economic evaluations)

ID	Searches - CRD (NHS-EED)
#1	MeSH blood pressure EXPLODE PERMUTE
#2	MeSH hypertension EXPLODE PERMUTE
#3	cost utility analys*
#4	mathematical model
#5	decision analys*
#6	Markov chain* or Markov process* or decision tree
#7	Economics
#8	cost effective* or cost effective* analys*
#9	#3 OR #4 OR #5 OR #6 OR #7 OR #8
#10	#1 OR #2
#11	#9 AND #10
#12	MeSH primary prevention EXPLODE PERMUTE
#13	#11 AND #12

NHS EED National Health Service Economic Evaluation Database, */$* wildcard characters

Appendix 4. Search results for case study 1

EMBASE AND MEDLINE databases

ID	Searches (via OVID)
#1	(lowering blood pressure or lowering-blood-pressure or blood pressure lowering).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui]
#2	(hypertensi$ or antihypertensi$ or anti-hypertensi$).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui]
#3	1 OR 2
#4	(cost effective$ OR cost-effective$ OR mathematical model OR decision-analys$s OR decision analys$s OR Markov OR decision tree OR economic evaluation OR cost utility).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui]
#5	3 AND 4
#6	limit 5 to English language
#7	limit 6 to yr="2000 -Current"
#8	limit 7 to humans
#9	Exclude conference abstracts, methodological papers, commentaries, editorials, notes
#10	remove duplicates from 9

Appendix 5. Model input parameters from the TASMIN-SR model

Parameter	Input	Sources
CVD risk in patients with DM		
Stroke		
60-69 years old	0.0196	NICE, Diabetes[104]
70-79 years old	0.0262	
80-89 years old	0.0298	
MI		
60-69 years old	0.0089	NICE, Diabetes[104]
70-79 years old	0.0100	
80-89 years old	0.0111	
UA		
60-69 years old	0.0041	NICE, Diabetes[104]
70-79 years old	0.0047	
80-89 years old	0.0052	
CVD risk in patients with CKD		
Stroke		
60-69 years old	0.0072	Kerr et al [105]
70-79 years old	0.0147	
80-89 years old	0.0189	
MI		
60-69 years old	0.0051	Kerr et al [105]
70-79 years old	0.0113	
80-89 years old	0.0171	
UA		
60-69 years old	0.0024	Kerr et al [105]
70-79 years old	0.0054	
80-89 years old	0.0081	
CVD risk in patients with stroke		
Stroke		
60-69 years old	0.0348	PROGRESS (1999) & NICE, Lipid modification guidelines[108, 109]
70-79 years old	0.0590	
80-89 years old	0.0715	
MI		
60-69 years old	0.0139	

Parameter	Input	Sources
70-79 years old	0.0232	PROGRESS (1999) & NICE, Lipid modification guidelines[108, 109]
80-89 years old	0.0232	
UA		
60-69 years old	0.0139	PROGRESS (1999) & NICE, Lipid modification guidelines[108, 109]
70-79 years old	0.0232	
80-89 years old	0.0232	
CVD risk in patients with CHD		
Stroke		
60-69 years old	0.0348	NICE, Lipid modification and Hypertension guidelines[64, 108]
70-79 years old	0.0590	
80-89 years old	0.0715	
MI		
60-69 years old	0.0666	NICE, Lipid modification and Hypertension guidelines[64, 108]
70-79 years old	0.1112	
80-89 years old	0.1112	
UA		
60-69 years old	0.0528	NICE, Lipid modification and Hypertension guidelines[64, 108]
70-79 years old	0.0882	
80-89 years old	0.0882	
Age-related relative risks		
MI, UA – self-management		
60-69 years old	0.63	TASMIN-SR trial & Law at al (2009)[37, 114]
70-79 years old	0.69	
80-89 years old	0.75	
Stroke – self-management		
60-69 years old	0.54	TASMIN-SR trial & Law at al (2009)[37, 114]
70-79 years old	0.59	
80-89 years old	0.75	
MI, UA - usual care		
60-69 years old	0.82	TASMIN-SR trial & Law at al (2009)[37, 114]
70-79 years old	0.85	
80-89 years old	0.88	
Stroke - usual care		

Parameter	Input	Sources
60-69 years old	0.76	TASMIN-SR trial & Law at al (2009)[37, 114]
70-79 years old	0.81	
80-89 years old	0.88	
Cost of death	0	By definition
Utilities		
Utilities for initial health states		
Self-management and usual care		
65-74 years old	0.81	TASMIN-SR trial[37]
75-84 years old	0.74	
85 and over	0.71	
Utilities for acute events		
UA	0.77	NICE, Lipid modification guidelines[108]
MI	0.76	
Stroke	0.63	
Utilities for long-term (chronic) disease		
UA	0.88	NICE, Lipid modification guidelines[108]
MI	0.88	
Stroke	0.63	
Annual discount rate for costs	0.035	Gray et al [1]
Annual discount rate for utility	0.035	Gray et al [1]
Death utility	0	By definition
Average age of cohort at time of intervention (years)	70	TASMIN-SR trial[37]

Appendix 6. Input parameters and their distributions from the TASMIN-SR model

Description	Input	Distribution	a	b
Probability of death from Stroke	0.23	Beta	125	420
Probability of death from MI	0.23	Beta	155	520
			alpha	lambda
Cost of well state self-monitoring	74	Gamma	1	0.0136
Cost of well state for Usual care arm	62	Gamma	1	0.0161
Cost acute angina	3292	Gamma	1	0.0003
Cost acute MI	5487	Gamma	1	0.0002
Cost acute Stroke	11020	Gamma	1	0.0001
Cost chronic angina	286	Gamma	1	0.0035
Cost chronic MI	286	Gamma	1	0.0035
Cost chronic Stroke	1361	Gamma	1	0.0007
Cost of intervention	35	Gamma	1	0.0286
			mean	s.d.
Multiplier used to adjust for initial health states by age	1	Normal	1	0.0125

Appendix 7. Hypothetical interventions in the optimisation of acute stroke care for thrombolysis

	Area for improvement based on baseline data (n=355)	Hypothetical intervention with reference to literature (Estimates noted)	Max theoretical number of patients benefiting from removal of block in pathway n(%) – or theoretical hypothetical intervention (n=355)	Predicted max benefit based on that recorded in the literature or estimated n(%) – or achievable hypothetical intervention
1.	**Timely referral**: 16% of patients contacted a General Practitioner following onset of stroke and arrival in hospital was delayed compared to those who called the emergency services.	**Divert GP calls to ambulance service** General practice staff is trained to better recognise stroke, resulting in patients who initially contact their General Practitioner being referred immediately to hospital. Up to 64% immediately referred to EMS	57 patients (100%)	36 patients (64%)
2.	**Wake-up strokes**: According to previous research, approximately 16% (range 8-27%) of all stroke patients suffer symptom onset upon waking. Currently, these patients are ineligible for thrombolysis due to unknown onset time.	**Use imaging to estimate onset time for wake up strokes** New imaging techniques are introduced which make it possible to estimate onset time in wake-up stroke, thus increasing the number of patients potentially eligible for thrombolysis. Estimate that up to 80% feasible with new imaging.	57 patients (100%)	45 patients (80%)
3.	**Unknown onset time in stroke**: Symptom onset time was unknown in 37% of stroke patients presenting in hospital. Approximately 22% (57% of those with unknown onset) do **not** have wake-up stroke. These patients are ineligible for thrombolysis.	**Improve recognition and recording of onset time** Media campaigns improve public awareness of the importance of noting onset time in stroke. Additional training of healthcare professionals facilitates better recording of onset time in patient medical notes. 92% known onset time once wake up stroke excluded	77 patients (100%)	49 patients (64%)

	Area for improvement based on baseline data (n=355)	Hypothetical intervention with reference to literature (Estimates noted)	Max theoretical number of patients benefiting from removal of block in pathway n(%) – or theoretical hypothetical intervention (n=355)	Predicted max benefit based on that recorded in the literature or estimated n(%) – or achievable hypothetical intervention
4.	**Patient delay from onset of symptoms to contact with emergency services:** Approximately 15% of stroke patients had a known onset time but arrived in hospital greater than 4.5 hours after symptom onset. This was often caused by a delay in contacting medical services due to lack of awareness of stroke as an emergency or failure to identify stroke symptoms.	**Reduce time to call emergency services** Better public awareness of stroke through a series of educational interventions result in an increase in the proportion of patients contacting emergency services immediately after the onset of stroke symptoms. Estimate additional 25% of those arriving outside 4.5 hours now arrive within 4.5 hours	54 patients (100%)	14 patients (25%)
5.	**Timely CT scan** 14% of patients who arrived in hospital within 4.5 hours of symptom onset did not receive an immediate CT scan within the time window for thrombolysis.	**Ensure all who present with stroke receive immediate CT scan** In-hospital stroke services are reorganised to ensure patients with stroke receive a timely CT scan (e.g. CT scanner moved closer to the emergency department ward). Reduce time to CT scan by an hour	51 patients (100%)	13 patients (25%)
6.	**FAST Acute care pathway:** 22% of patients travelling via ambulance and who received a FAST check (or 14% of total patients travelling via ambulance) were Face Arm Speech Test negative and did not have urgent CT scan.	**Better stroke recognition tools** New stroke recognition tools are introduced which ensure FAST negative stroke patients are recognised by emergency services as having stroke. Use ROSIER score with sensitivity 93% vs. 79% for FAST in the present study.	51 patients (100%)	35 patients (68%)

	Area for improvement based on baseline data (n=355)	Hypothetical intervention with reference to literature (Estimates noted)	Max theoretical number of patients benefiting from removal of block in pathway n(%) – or theoretical hypothetical intervention (n=355)	Predicted max benefit based on that recorded in the literature or estimated n(%) – or achievable hypothetical intervention
7.	**Thrombolysis for >85 years old**:† 6% of stroke patients older than 85 years old were scanned within 4.5 hours. †Current licensing makes these patients ineligible for thrombolysis.	**Extend thrombolysis eligibility to over 85s** New evidence from the International Stroke Trial [IST-3] suggests that patients over the age of 80 could benefit from thrombolysis up to 4.5 hours after symptom onset. Assume same proportion of those over 85 who get scan within 4.5 hours receive thrombolysis.	19 patients (100%)	19 patients (100%)

†In the present CLAHRC cohort, patients up to 85 years were thrombolysed. Note, current guidance indicates that thrombolysis should be administered with caution in over 80s from 3 to 4.5 hours from symptoms onset.

Appendix 8. Parameters used in the Stroke study

Parameter	Value	Distribution	Source
Percentage of patients with symptoms who have suffered a "true" stroke and…			
contact ambulance services (999)	67%	Beta	Patient level data
contact their GP surgery first	16%	Beta	Patient level data
make their own way to the hospital	11%	Beta	Patient level data
are already inpatients	5%	Beta	Patient level data
are referred/transferred	1%	Beta	Patient level data
Percentage of patients with symptoms who have suffered a stroke, contact 999 services and			
get to A&E within 4.5h (hours) of stroke onset	53.4%	Beta	Patient level data
get a CT scan and results within 4.5h of stroke onset	39.8%	Beta	Patient level data
are suitable for thrombolysis (ischaemic stroke and younger than 85 years)	30.1%	Beta	Patient level data
have an ischaemic stroke and receive thrombolysis	12.7%	Beta	Patient level data
For all stroke patients, percentage of…			
ischaemic stroke in all stroke patients	89%	Beta	Patient level data
ischaemic stroke patients arriving within 4.5h of stroke onset	42%	Beta	Patient level data
death* after stroke due to a haemorrhagic stroke	19%	Beta	Patient level data
death after stroke due to ischaemic stroke	19%	Beta	Patient level data
death after an ischaemic stroke and thrombolysis treatment	15%	Beta	Patient level data
likelihood of being dependent/disabled after an ischaemic stroke	35%	Beta	Bamford (1990)

Parameter	Value	Distribution	Source
Patients with true stroke symptoms			
Percentage with no prior contraindications for thrombolysis	95%	Beta	Patient level data
Effectiveness of thrombolysis			
Decrease in dependency after stroke due to thrombolysis treatment	0.67 †	Log normal	Cochrane Review (2009)
Life expectancy (in years)			
74 year old female with no history of stroke	9.9	Beta	Adjusted Life tables§ ONS,
69 year old male with no history of stroke	10.8	Beta	Adjusted Life tables§ ONS,
74 year old independent female with history of stroke	9.6	Beta	Adjusted Life tables§ ONS,
69 year old independent male with history of stroke	10.5	Beta	Adjusted Life tables§ ONS,
74 year old dependent female with history of stroke	8.7	Beta	Adjusted Life tables§ ONS,
69 year old dependent male with history of stroke	10.0	Beta	Adjusted Life tables§ ONS,
Utility score			
Dependent stroke patient	0.38	Beta	Sandercock (2004)
Independent stroke patient	0.74	Beta	Sandercock (2004)
Unit costs			
GP home visit	$120		PSSRU 2011
GP phone call	$22		PSSRU 2011
GP surgery visit	$37		PSSRU 2011

Parameter	Value	Distribution	Source
Transportation by ambulance including paramedics	$369		NHS Reference Costs 2010/11
Attending A&E leading to admission	$348		NHS Reference Costs 2010/11
CT scan	$140		NHS Reference Costs 2010/11
Thrombolysis (drug only)	$1,057		BNF September 2011**
Stroke costs (per year) ¶			
Independent acute stroke	$6,021	Gamma	Sandercock (2002) HTA Report
Dependent acute stroke	$21,931	Gamma	Sandercock (2002) HTA Report
Long-term independent stroke	$1,969	Gamma	Sandercock (2002) HTA Report
Long-term dependent stroke	$25,919	Gamma	Sandercock (2002) HTA Report
Fatal stroke	$14,191	Gamma	Sandercock (2002) HTA Report

* Death within one year of stroke

† Odds ratio (95% Confidence intervals 0.61 to 0.75)

§ Office for National Statistics (ONS); adjusted by mortality rates in post stroke years for each mRS level using published data[28]

**BNF, March 2011. Cost refers to intravenous administration over 60 minutes, 900 micrograms/kg (max 90 mg); initial 10% of dose by intravenous injection (with diluents), remainder by intravenous infusion (with diluents, transfer device and infusion bag)

¶ All figures have been adjusted using the Hospital and Community Health Service (HCHS) pay and price Index 2010/11

NOTE: all figures were converted to US Dollars using the purchasing power parity (PPP) exchange rate 2012 = 0.681 per USD.[25]

Appendix 9. Search results for case study 2

Search results via Cochrane Library 27/06/2016

No.	Searches	Results
1	MeSH descriptor: [thrombolytic therapy] explode all trees	1764
2	MeSH descriptor: [Stroke] explode all trees	6543
3	1 AND 2	246

Search results via Ovid MEDLINE 27/06/2016

No.	Searches	Results
1	Stroke	69958
2	Thrombolytic therapy/ or Tissue Plasminogen Activator/	31876
3	Rt-PA.mp	2010
4	2 or 3	32119
5	1 and 4	5294
6	Cost-effectiveness.mp or cost-benefit analysis/	80899
7	Cost-utility.mp	2768
8	Economic evaluation.mp or cost-benefit analysis/	68707
9	6 or 7 or 8	83115
10	9 and 5	82

LIST OF REFERENCES

1. Gray AM, Clarke PM, Wolstenholme JL and Wordsworth S. *Applied Methods of Cost-Effectiveness Analysis in Health Care*. Oxford: Oxford University Press, 2011.
2. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ and Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, 2005.
3. Dasgupta AK and Pearce DW. *Cost-Benefit Analysis: Theory and Practice*. Macmillan Education UK, 1972.
4. McIntosh E, Clarke P, Frew E and Louviere J. *Applied Methods of Cost-Benefit Analysis in Health Care*. Oxford University Press, 2010.
5. Stroke Statistics. State of the Nation. Stroke Association, 2016.
6. WHO. *The World Health Report: Reducing Risks, Promoting Healthy Life*. Geneva2002.
7. Lewington S, Clarke R, Qizilbash N, Peto R and Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002; 360: 1903-13.
8. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, AlMazroa MA, Alvarado M, Anderson HR, Anderson LM, Andrews KG, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2010; 380: 2095-128.
9. Briggs AH, Claxton K and Sculpher MJ. *Decision Modelling for Health Economic Evaluation*. Oxford, University Press, 2006.
10. Petrou S and Alastair G. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ*. 2011; 342.
11. NICE. Guide to the methods of technology appraisal 2013. *Process and methods guides*. London: National Institue for Health and Care Excellence, 2013.
12. Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. 1968. *MD computing : computers in medical practice*. 1993; 10: 312-28.
13. Barton P, Bryan S and Robinson S. Modelling in the economic evaluation of health care: Selecting the appropriate approach. *Journal of Health Services Research and Policy*. 2004; 9: April.
14. Brennan A, Chick SE and Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics*. 2006; 15: December.
15. Barton M, McClean S, Gillespie J, Garg L, Wilson D and Fullerton K. Is it beneficial to increase the provision of thrombolysis?-- a discrete-event simulation model. *Qjm*. 2012; 105: 665-73.
16. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P and Krahn M. Conceptualizing a Model: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Value in Health*. 2012; 15: 804-11.

17. Philips Z, Bojke L, Sculpher M, Claxton K and Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: A review and consolidation of quality assessment. *PharmacoEconomics*. 2006; 24: 2006.
18. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C and Luce BR. 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 in Health*. 2003; 6: January/February.
19. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics*. 2006; 24: 1043-53.
20. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ and Paltiel AD. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health*. 2012; 15: 835-42.
21. Caro JJ, Briggs AH, Siebert U and Kuntz KM. Modeling Good Research Practices - Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health*. 2012; 15: 796-803.
22. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM and Wong JB. Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value in Health*. 2012; 15: 843-50.
23. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J and Moller J. Modeling using Discrete Event Simulation: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value in Health*. 2012; 15: 821-7.
24. Pitman R, Fisman D, Zaric GS, Postma M, Bretzschmar M, Edmunds J and Brisson M. Dynamic Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. *Value in Health*. 2012; 15: 828-34.
25. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ and Kuntz KM. State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value in Health*. 2012; 15: 812-20.
26. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, Woolacoot N and Glanville J. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health technology assessment (Winchester, England)*. 2004; 8: Sep.
27. Jackson CH, Bojke L, Thompson SG, Claxton K and Sharples LD. A framework for addressing structural uncertainty in decision models. *Medical Decision Making*. 2011; 31: 662-74.
28. Strong M, Oakley JE and Chilcott J. Managing structural uncertainty in health economic decision models: a discrepancy approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2012; 61: 25-45.
29. Afzali HH and Karnon J. Exploring structural uncertainty in model-based economic evaluations. *Pharmacoeconomics*. 2015; 33: 435-43.
30. Caro JJ and Moller J. Decision-analytic models: current methodological challenges. *Pharmacoeconomics*. 2014; 32: 943-50.
31. Bilcke J, Beutels P, Brisson M and Jit M. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. *Medical Decision Making*. 2011; 31: 675-92.

32. Jackson CH, Sharples LD and Thompson SG. Structural and parameter uncertainty in Bayesian cost-effectiveness models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2010; 59: 233-53.
33. Stevenson M, Tappenden P and Squires H. Methods for handling uncertainty within pharmaceutical funding decisions. *International Journal of Systems Science*. 2014; 45: 60-8.
34. Oostenbrink JB, Al MJ, Oppe M and Rutten-van Molken MP. Expected value of perfect information: an empirical example of reducing decision uncertainty by conducting additional research. *Value Health*. 2008; 11: 1070-80.
35. WHO. *GLOBAL STATUS REPORT on noncommunicable diseases 2014*. Geneva2014.
36. McManus RJ, Mant J, Bray EP, Holder R, Jones MI, Greenfield S, Kaambwa B, Banting M, Bryan S, Little P, Williams B and Hobbs FDR. Telemonitoring and self-management in the control of hypertension (TASMINH2): a randomised controlled trial. *Lancet*. 2010; 376: 163-72.
37. McManus RJ, Mant J, Haque MS, Bray EP, Bryan S, Greenfield SM, Jones MI, Jowett S, Little P, Penaloza C, Schwartz C, Shackleford H, Shovelton C, Varghese J, Williams B, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. *Jama*. 2014; 312: 799-808.
38. Whiteley WN, Emberson J, Lees KR, Blackwell L, Albers G, Bluhmki E, Brott T, Cohen G, Davis S, Donnan G, Grotta J, Howard G, Kaste M, Koga M, von Kummer R, et al. Risk of intracerebral haemorrhage with alteplase after acute ischaemic stroke: a secondary analysis of an individual patient data meta-analysis. *The Lancet Neurology*. 15: 925-33.
39. Kaambwa B, Bryan S, Jowett S, Mant J, Bray EP, Hobbs FDR, Holder R, Jones MI, Little P, Williams B and McManus RJ. Telemonitoring and self-management in the control of hypertension (TASMINH2): A cost-effectiveness analysis. *European Journal of Preventive Cardiology*. 2014; 21: 1517-30.
40. Penaloza-Ramos MC, Jowett S, Mant J, Schwartz C, Bray EP, Sayeed Haque M, Richard Hobbs FD, Little P, Bryan S, Williams B and McManus RJ. Cost-effectiveness of self-management of blood pressure in hypertensive patients over 70 years with suboptimal control and established cardiovascular disease or additional cardiovascular risk diseases (TASMIN-SR). *Eur J Prev Cardiol*. 2015.
41. Jowett S. Using Decision Analytical Modelling Techniques in Health Economics: An Application to Screening for and Treatment of Atrial Fibrillation. PhD Thesis. University of Birmingham, 2007.
42. Canadian Coordinating Office for Health Technology A. *Guidelines for the economic evaluation of health technologies: Canada (3rd Edition) (Special)*. Canadian Coordinating Office for Health Technology Assessment (CCOHTA), 2006.
43. Earnshaw J and Lewis G. NICE guide to the methods of technology appraisal: Pharmaceutical industry perspective. *PharmacoEconomics*. 2008; 26: 2008.
44. Karnon J, Brennan A and Akehurst R. A critique and impact analysis of decision modeling assumptions. *Medical Decision Making*. 2007; 27: July/August.
45. Sculpher M, Fenwick E and Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. [Review] [44 refs]. *PharmacoEconomics*. 2000; 17: 461-77.

46. Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M and Kent DL. Toward a peer review process for medical decision analysis models. *Medical care*. 1994; 32: Jul.
47. Soto J. Health economic evaluations using decision analytic modeling: Principles and practices - Utilization of a checklist to their development and appraisal. *International Journal of Technology Assessment in Health Care*. 2002; 18: 2002.
48. Bentley TG, Kuntz KM and Ringel JS. Bias associated with failing to incorporate dependence on event history in Markov models. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2010; 30: 2010-2Dec.
49. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics*. 2003; 12: 01.
50. Andronis L, Barton P and Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health technology assessment (Winchester, England)*. 2009; 13: Jun.
51. Bojke L, Claxton K, Sculpher M and Palmer S. Characterizing structural uncertainty in decision analytic models: A review and application of methods. *Value in Health*. 2009; 12: July/August.
52. Briggs AH, Ades AE and Price MJ. Probabilistic sensitivity analysis for decision trees with multiple branches: Use of the Dirichlet distribution in a Bayesian framework. *Medical Decision Making*. 2003; 23: July/August.
53. Briggs AH and Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technology Assessment*. 1999; 3: 1999.
54. Jain R, Grabner M and Onukwugha E. Sensitivity analysis in cost-effectiveness studies: From guidelines to practice. *PharmacoEconomics*. 2011; 29: 2011.
55. Koerkamp BG, StijnenT, Weinstein MC and Hunink MG. The combined analysis of uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making*. 2011; 31: 650-61.
56. Kuntz KM and Goldie SJ. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2002; 22: 2002-2Jun.
57. Sendi P, Gafni A and Birch S. Opportunity costs and uncertainty in the economic evaluation of health care interventions. *Health Economics*. 2002; 11: 2002.
58. Andronis L, Barton P and Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *NIHR Health Technology Assessment programme: Executive Summaries*. Southampton (UK): NIHR Journals Library, 2009.
59. Goldhaber-Fiebert JD, Stout NK and Goldie SJ. Empirically evaluating decision-analytic models. *Value in Health*. 2010; 13: July-August.
60. Karnon J and Vanni T. Calibrating models in economic evaluation: A comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. *PharmacoEconomics*. 2011; 29: 2011.
61. Kim LG and Thompson SG. Uncertainty and validation of health economic decision models. *Health Economics*. 2010; 19: January.

62. Kim SY, Goldie SJ and Salomon JA. Exploring model uncertainty in economic evaluation of health interventions: the example of rotavirus vaccination in Vietnam. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2010; 30: E1-e28.
63. NICE. Lipid modification: cardiovascular risk assessment and the primary and secondary prevention of cardiovascular disease. NICE Guidelines CG67. London: National Institute for Health and Clinical Excellence, 2008.
64. NICE. *Hypertension: clinical management of primary hypertension in adults. NICE clinical guideline 127*. London: National Institute for Health and Clinical Excellence, 2011.
65. Gray AM. Cost-effectiveness analyses alongside randomised clinical trials. *Clin Trials*. 2006; 3: 538-42.
66. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitti D and Reed S. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health*. 2005; 8: 521-33.
67. Peñaloza Ramos MC, Barton P, Jowett S and Sutton AJ. A Systematic Review of Research Guidelines in Decision-Analytic Modeling. *Value in Health*. 2015; 18: 512-29.
68. Moher D, Liberati A, Tetzlaff J and Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA StatementThe PRISMA Statement. *Annals of Internal Medicine*. 2009; 151: 264-9.
69. Gandjour A and Stock S. A national hypertension treatment program in Germany and its estimated impact on costs, life expectancy, and cost-effectiveness. *Health Policy*. 2007; 83: 257-67.
70. Montgomery AA, Fahey T, Ben-Shlomo Y and Harding J. The influence of absolute cardiovascular risk, patient utilities, and costs on the decision to treat hypertension: A Markov decision analysis. *Journal of Hypertension*. 2003; 21: 1753-9.
71. Perman G, Rossi E, Waisman GD, Aguero C, Gonzalez CD, Pallordet CL, Figar S, Gonzlez Bernaldo de Quiros F, Canning J and Soriano ER. Cost-effectiveness of a hypertension management programme in an elderly population: A Markov model. *Cost Effectiveness and Resource Allocation*. 2011; 9.
72. Stevanovic J, O'Prinsen AC, Verheggen BG, Schuiling-Veninga N, Postma MJ and Pechlivanoglou P. Economic evaluation of primary prevention of cardiovascular diseases in mild hypertension: A scenario analysis for the Netherlands. *Clinical Therapeutics*. 2014; 36: 368-84.e5.
73. Wu Y, Zhou Q, Xuan J, Li M, Zelt S, Huang Y, Yin H and Huang M. A cost-effectiveness analysis between amlodipine and angiotensin ii receptor blockers in stroke and myocardial infarction prevention among hypertension patients in China. *Value in Health Regional Issues*. 2013; 2: 75-80.
74. Kourlaba G, Fragoulakis V, Theodoratou D and Maniadakis N. Economic evaluation of telmisartan, valsartan and losartan in combination with hydrochlorothiazide for treatment of mild-to-moderate hypertension in Greece: A cost-utility analysis. *Journal of Pharmaceutical Health Services Research*. 2013; 4: 81-8.
75. Ekwunife OI, Okafor CE, Ezenduka CC and Udeogaranya PO. Cost-utility analysis of antihypertensive medications in Nigeria: A decision analysis. *Cost Effectiveness and Resource Allocation*. 2013; 11.

76. Wisloff T, Selmer RM, Halvorsen S, Fretheim A, Norheim OF and Kristiansen IS. Choice of generic antihypertensive drugs for the primary prevention of cardiovascular disease - A cost-effectiveness analysis. *BMC Cardiovascular Disorders*. 2012; 12.
77. Ekman M, Bienfait-Beuzon C and Jackson J. Cost-effectiveness of irbesartan/hydrochlorothiazide in patients with hypertension: An economic evaluation for Sweden. *Journal of Human Hypertension*. 2008; 22: 845-55.
78. Nordmann AJ, Krahn M, Logan AG, Naglie G and Detsky AS. The cost effectiveness of ACE inhibitors as first-line antihypertensive therapy. *PharmacoEconomics*. 2003; 21: 573-85.
79. Granstrom O, Levin L and Henriksson M. Cost-effectiveness of candesartan versus losartan in the primary preventive treatment of hypertension. *ClinicoEconomics and Outcomes Research*. 2012; 4: 313-22.
80. Baker TM, Goh J, Johnston A, Falvey H, Brede Y and Brown RE. Cost-effectiveness analysis of valsartan versus losartan and the effect of switching. *Journal of medical economics*. 2012; 15: 253-60.
81. Squires H, Chilcott J, Akehurst R, Burr J and Kelly MP. A systematic literature review of the key challenges for developing the structure of public health economic models. *International journal of public health*. 2016; 61: 289-98.
82. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*. 2000; 17: 479-500.
83. Squires H, Chilcott J, Akehurst R, Burr J and Kelly MP. A Framework for Developing the Structure of Public Health Economic Models. *Value in Health*.
84. Manning WG, Fryback DG and Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR SJ, Russell LB, et al, (ed.). *Cost-effectiveness in health and medicine*. New York (NY): Oxford University Press, 1996, p. 247-75.
85. Strong M, Pilgrim H, Oakley J and Chilcott J. Structural uncertainty in health economic decision models. *ScHARR Occasional Paper*. Sheffield: University of Sheffield, 2009.
86. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, members of the Panel on Cost-Effectiveness in H and Medicine. Assessing the Effectiveness of Health Interventions for Cost-Effectiveness Analysis. *Journal of General Internal Medicine*. 1997; 12: 551-8.
87. Frederix GW, van Hasselt JG, Schellens JH, Hovels AM, Raaijmakers JA, Huitema AD and Severens JL. The impact of structural uncertainty on cost-effectiveness models for adjuvant endocrine breast cancer treatments: the need for disease-specific model standardization and improved guidance. *Pharmacoeconomics*. 2014; 32: 47-61.
88. Price MJ, Welton NJ, Briggs AH and Ades AE. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. *Value Health*. 2011; 14: 205-18.
89. Schaber J, Liebermeister W and Klipp E. Nested uncertainties in biochemical models. *IET systems biology*. 2009; 3: 1-9.
90. Bilcke J, Chapman R, Atchison C, Cromer D, Johnson H, Willem L, Cox M, Edmunds WJ and Jit M. Quantifying Parameter and Structural Uncertainty of Dynamic Disease Transmission Models Using MCMC: An Application to Rotavirus Vaccination in England and Wales. *Medical Decision Making*. 35: 633-47.

91. Jackson CH, Thompson SG and Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *Journal of the Royal Statistical Society Series A, (Statistics in Society)*. 2009; 172: 383-404.
92. Brisson M and Edmunds WJ. Impact of model, methodological, and parameter uncertainty in the economic analysis of vaccination programs. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2006; 26: 434-46.
93. Thom HH, Jackson CH, Commenges D and Sharples LD. State selection in Markov models for panel data with application to psoriatic arthritis. *Stat Med*. 2015; 34: 2456-75.
94. NHS England. Factsheet: Improved detection and management of hypertension. 2014.
95. Mentz RJ, Bakris GL, Waeber B, McMurray JJV, Gheorghiade M, Ruilope LM, Maggioni AP, Swedberg K, Pina IL, Fiuzat M, O'Connor CM, Zannad F and Pitt B. The past, present and future of renin-angiotensin aldosterone system inhibition. *International Journal of Cardiology*. 2013; 167: 1677-87.
96. Afzali HH, Karnon J and Merlin T. Improving the accuracy and comparability of model-based economic evaluations of health technologies for reimbursement decisions: a methodological framework for the development of reference models. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2013; 33: 325-32.
97. Peñaloza-Ramos MC, Barton P, Jowett S and Sutton AJ. Do economic evaluations in primary prevention of cardiovascular disease conform to good practice guidelines? A systematic review. (In Press) *MDM Policy & Practice*. 2016.
98. Ward S, Jones ML, Pandor A, Holmes M, Ara R, Ryan A, Yeo W and Payne N. A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technology Assessment*. 2007; 11: iii-160.
99. Wisløff T. *Norwegian Cardiovascular Disease Model (NorCaD): a simulation model for estimating health benefits and cost consequences of cardiovascular interventions*. Oslo: Norwegian Knowledge Centre for the Health Services, 2008, p.69 s. digital, PDF-fil.
100. Curtis L and Burns A. *Unit Costs of Health & Social Care*. Kent: Personal Social Services Research Unit, 2015.
101. Department of Health. NHS Reference Costs Schedule 2013-14. In: Health Do, (ed.). London2013.
102. NICE. Statins for the prevention of cardiovascular events. NICE guideline TA94. London: National Institute for Health and Clinical Excellence, 2006.
103. Office for National Statistics. Interim Life Tables for England.
104. NICE. National guidelines for the management of blood glucose levels in people with type 2 diabetes. NICE guidelines CG87. London: National Institue for Health and Clinical Excellence, 2002.
105. Kerr M, Bray B, Medcalf J, O'Donoghue DJ and Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. *Nephrology Dialysis Transplantation*. 2012; 27.
106. Koren-Morag N, Goldbourt U and Tanne D. Renal dysfunction and risk of ischemic stroke or TIA in patients with cardiovascular disease. *Neurology*. 2006; 67: 224-8.

107. Shiba N and Shimokawa H. Chronic kidney disease and heart failure—Bidirectional close link and common therapeutic goal. *Journal of Cardiology*. 2011; 57: 8-17.
108. NICE. Lipid modification: Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. NICE guidelines CG67. London: National Institute for Health anc Clinical Excellence, 2008.
109. PROGRESS Collaborative Group. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *The Lancet*. 2001; 358: 1033-41.
110. Hankey GJ. Long-term outcome after ischaemic stroke/transient ischaemic attack. *Cerebrovascular diseases (Basel, Switzerland)*. 2003; 16 Suppl 1: 14-9.
111. Bamford J, Sandercock P, Dennis M, Burn J and Warlow C. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project--1981-86. 2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral and subarachnoid haemorrhage. *Journal of neurology, neurosurgery, and psychiatry*. 1990; 53: 16-22.
112. Gattellari M, Goumas C, Garden F and Worthington JM. Relative survival after transient ischaemic attack: results from the Program of Research Informing Stroke Management (PRISM) study. *Stroke*. 2012; 43: 79-85.
113. Mant J, Barton P, Stevens A, Ford G and Rothwell PM. What is the optimum model of service delivery for Transient Ischaemic Attack? Report for the National co-ordinating Centre for NHS Service Delivery and Organisation R&D (NCCSDO). 2008.
114. Law MR, Morris JK and Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemilogical studies. *BMJ*. 2009; 338: b1665.
115. Youman P, Wilson K, Harraf F and Kalra L. The economic burden of stroke in the United Kingdom. *Pharmacoeconomics*. 2003; 21: 43-50.
116. Palmer AJ, Annemans L, Roze S, Lamotte M, Lapuerta P, Chen R, Gabriel S, Carita P, Rodby RA, De Zeeuw D and Parving HH. Cost-effectiveness of early irbesartan treatment versus control (standard antihypertensive medications excluding ACE inhibitors, other angiotensin-2 receptor antagonists, and dihydropyridine calcium channel blockers) or late irbesartan treatment in patients with type 2 diabetes, hypertension, and renal disease. *Diabetes Care*. 2004; 27: 1897-903.
117. Cooper A and O'Flynn N. Risk assessment and lipid modification for primary and secondary prevention of cardiovascular disease: summary of NICE guidance. *BMJ*. 2008; 336: 1246-8.
118. Stewart S, Jenkins A, Buchan S, McGuire A, Capewell S and McMurray JJ. The current cost of heart failure to the National Health Service in the UK. *European journal of heart failure*. 2002; 4: 361-71.
119. Ara R and Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition specific data are not available. In: 10/11 HDP and ScHARR UoS, (eds.). 2011.
120. Woods DS, E.; Palmer, S.; Latimer, N.; Soares, M.;. NICE DSU Technical Support Document 19: Partitioned Survival Analysis for decision modelling in health care: a critical review. In: Decision Support Unit S, University of Sheffield, (ed.). 2017.

121. Saka Ö, McGuire A and Wolfe C. Cost of stroke in the United Kingdom. *Age and Ageing*. 2009; 38: 27-32.
122. Araujo DV, Teich V, Passos RB and Martins SC. Analysis of the cost-effectiveness of thrombolysis with alteplase in stroke. *Arq Bras Cardiol*. 2010; 95: 12-20.
123. Penaloza-Ramos MC, Sheppard JP, Jowett S, Barton P, Mant J, Quinn T, Mellor RM, Sims D, Sandler D and McManus RJ. Cost-Effectiveness of Optimizing Acute Stroke Care Services for Thrombolysis. *Stroke*. 2014.
124. Sandercock P, Berge E, Dennis M, Forbes J, Hand P, Kwan J, Lewis S, Lindley R, Neilson A and Wardlaw J. Cost-Effectiveness of Thrombolysis With Recombinant Tissue Plasminogen Activator for Acute Ischemic Stroke Assessed by a Model Based on UK NHS Costs. *Stroke*. 2004; 35: 1490-7.
125. Wardlaw JM, Murray V, Berge E and Del Zoppo GJ. Thrombolysis for acute ischaemic stroke. *The Cochrane database of systematic reviews*. 2009: Cd000213.
126. Boudreau DM, Guzauskas G, Villa KF, Fagan SC and Veenstra DL. A model of cost-effectiveness of tissue plasminogen activator in patient subgroups 3 to 4.5 hours after onset of acute ischemic stroke. *Annals of emergency medicine*. 2013; 61: 46-55.
127. Dirks M, Baeten SA, Dippel DW, van Exel NJ, van Wijngaarden JD, Huijsman R, Koudstaal PJ and Niessen LW. Real-life costs and effects of an implementation program to increase thrombolysis in stroke. *Neurology*. 2012; 79: 508-14.
128. Ehlers L, Andersen G, Clausen LB, Bech M and Kjolby M. Cost-effectiveness of intravenous thrombolysis with alteplase within a 3-hour window after acute ischemic stroke. *Stroke*. 2007; 38: 85-9.
129. Ehlers L, Muskens WM, Jensen LG, Kjolby M and Andersen G. National use of thrombolysis with alteplase for acute ischaemic stroke via telemedicine in Denmark: a model of budgetary impact and cost effectiveness. *CNS drugs*. 2008; 22: 73-81.
130. Mar J, Begiristain JM and Arrazola A. Cost-effectiveness analysis of thrombolytic treatment for stroke. *Cerebrovascular diseases (Basel, Switzerland)*. 2005; 20: 193-200.
131. Pan Y, Chen Q, Zhao X, Liao X, Wang C, Du W, Liu G, Liu L, Wang C, Wang Y and Wang Y. Cost-effectiveness of thrombolysis within 4.5 hours of acute ischemic stroke in China. *PLoS One*. 2014; 9: e110525.
132. Sinclair SE, Frighetto L, Loewen PS, Sunderji R, Teal P, Fagan SC and Marra CA. Cost-Utility analysis of tissue plasminogen activator therapy for acute ischaemic stroke: a Canadian healthcare perspective. *Pharmacoeconomics*. 2001; 19: 927-36.
133. Tan Tanny SP, Busija L, Liew D, Teo S, Davis SM and Yan B. Cost-effectiveness of thrombolysis within 4.5 hours of acute ischemic stroke: experience from Australian stroke center. *Stroke*. 2013; 44: 2269-74.
134. Tung CE, Win SS and Lansberg MG. Cost-effectiveness of tissue-type plasminogen activator in the 3- to 4.5-hour time window for acute ischemic stroke. *Stroke*. 2011; 42: 2257-62.
135. Curtis L. *Unit Costs of Health and Social Care*. Kent: Personal Social Services Research Unit, University of Kent, 2012.
136. *British National Formulary*. London: BMJ Publishing group and RPS Publishing, 2011.
137. Sandercock P, Berge E, Dennis M, Forbes J, Hand P, Kwan J, Lewis S, Lindley R, Neilson A, Thomas B and Wardlaw J. A systematic review of the effectiveness, cost-effectiveness and

barriers to implementation of thrombolytic and neuroprotective therapy for acute ischaemic stroke in the NHS. *Health Technol Assess*. 2002; 6: 1-112.
138. Tappenden P. Conceptual modelling for health economic model development. *HEDS Discussion*. 2012; Paper 12/05.
139. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T and Vray M. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ*. 1997; 6: 217-27.
140. Ransohoff DF, Pignone M and Russell LB. Using models to make policy: an inflection point? *Medical decision making : an international journal of the Society for Medical Decision Making*. 2011; 31: 527-9.