KUOPION YLIOPISTON JULKAISUJA E. YHTEISKUNTATIETEET 152
KUOPIO UNIVERSITY PUBLICATIONS E. SOCIAL SCIENCES 152

JANNE MARTIKAINEN

Application of Decision-Analytic Modelling in Health Economic Evaluations

Doctoral dissertation

To be presented by permission of the Faculty of Social Sciences
of the University of Kuopio for public examination in Auditorium ML1,
Medistudia building, University of Kuopio,
on Saturday 16th February 2008, at 12 noon

Department of Health Policy and Management
Faculty of Social Sciences
University of Kuopio

KUOPION YLIOPISTO

KUOPIO 2008

Distributor: Kuopio University Library
P.O. Box 1627
FI-70211 KUOPIO
FINLAND
Tel. +358 17 163 430
Fax +358 17 163 410
http://www.uku.fi/kirjasto/julkaisutoiminta/julkmyyn.html

Series Editors: Jari Kylmä, Ph.D.
Department of Nursing Science

Markku Oksanen, D.Soc.Sc. Philosophy
Department of Social Policy and Social Psychology

Author's address: Department of Social Pharmacy
University of Kuopio
P.O. Box 1627
FI-70211 KUOPIO
FINLAND
Tel. +358 17 162 600
Fax +358 17 163 464

Supervisors: Professor Hannu Valtonen
Department of Health Policy and Management
University of Kuopio

Professor Olli-Pekka Ryynänen
School of Public Health and Clinical Nutrition
University of Kuopio

Docent Miika Linna, D.Sc. (Tech.)
Centre for Health Economics - CHESS
STAKES

Reviewers: Professor Alan Lyles
Division of Government and Public Administration
University of Baltimore, USA

Doctor Elisabeth Fenwick, Ph.D.
Public Health and Health Policy
University of Glasgow, UK

Opponent: Professor Pekka Rissanen
Tampere School of Public Health
University of Tampere

ISBN 978-951-27-0811-6
ISBN 978-951-27-1062-1 (PDF)
ISSN 1235-0494

Kopijyvä
Kuopio 2008
Finland

Martikainen, Janne. Application of decision-analytic modelling in health economic evaluations. Kuopio University Publications E. Social Sciences 152. 2008. 147 p.
ISBN 978-951-27-0811-6
ISBN 978-951-27-1062-1 (PDF)
ISSN 1235-0494

ABSTRACT

Background: Western Societies are facing increasing challenges in funding their health care systems. Therefore, health care decision-makers have started increasingly to demand evidence that a health technology is cost-effective before it is financed or recommended to be adopted for wide use. In medicine, scientific evidence produced by randomised controlled trials (RCTs) is typically ranked highly in the hierarchy of evidence. Unfortunately, RCTs are restricted in their ability to produce evidence in the form that is needed for health care decisions. The applications of decision-analytic models have been suggested to offer a potential vehicle to produce valid evidence, which is relevant to the health care decision-makers.

Aim of the study: To develop the applications of decision-analytic models and to explore the applicability of a set of Bayesian methods for evidence synthesis and decision-analytic modelling in health economic evaluations.

Material and methods: The study is based on four separate case studies, where the decision-analytic models were developed and evidence synthesis methods were applied to aid in real world decision-making processes. Probabilistic modelling techniques were applied in all four case studies. Furthermore, a set of Bayesian methods was applied to synthesise the available evidence and to handle uncertainties in the decision-analytic models.

Results: The conducted case studies showed that the decision-analytic models can be applied, when there is a need to assess all relevant evidence, to link intermediate outcomes to final outcomes, to make results applicable to the decision-making context due to a gap between clinical trial evidence and the requirements for a decision, or to estimate cost-effectiveness for specific subgroups. The case studies also showed that the identification, selection, and critical appraisal of evidence are the most time consuming parts of the model development process. Evidence synthesis proved to be challenging due to the timing of cost-effectiveness evaluations, since they tend to focus on a time period at or around the implementation of health technology, when experience and evidence about its clinical and economic consequences may still be relatively limited. Furthermore, the case studies proved that the probabilistic modelling approach does offer an efficient way to reflect decision uncertainty but data used in the probabilistic models requires very often some preparation, which in turn, increases the number of the additional sources of methodological and process uncertainties.

Conclusions and suggestions for further research: The study proved that the applications of decision-analytic models offer a clear and coherent mathematical structure to combine all relevant evidence and to assess in advance the consequences, such as expected costs and health outcomes, of different decisions. However, further developments to standardise the modelling processes and to reflect the quality of evidence used in the decision-analytic models are needed. In addition, further developments to improve the handling of model uncertainty and to increase transparency are to be welcomed.

National Library of Medicine Classification: W 74

Medical Subject Headings: Decision Making; Economics; Models, Economic; Cost and Cost Analysis; Bayes Theorem

Martikainen, Janne. Päätösanalyyttisten mallien soveltaminen terveystaloudellisessa arviointitutkimuksessa. Kuopion yliopiston julkaisuja E. Yhteiskuntatieteet 152. 2008. 147 s.
ISBN 978-951-27-0811-6
ISBN 978-951-27-1062-1 (PDF)
ISSN 1235-0494

TIIVISTELMÄ

Tutkimuksen tausta: Länsimaiset terveydenhuoltojärjestelmät ovat viime vuosina kohdanneet enenevässä määrin haasteita palvelujärjestelmiensä rahoittamisessa. Tämän vuoksi tarve kustannusvaikuttavuustutkimusten käyttöön terveydenhuollossa tehtävien rahoitus- ja käyttöönottopäätösten tukena on lisääntynyt huomattavasti. Lääketieteessä on perinteisesti arvostettu tutkimusnäyttöä, joka on tuotettu satunnaistetussa ja kontrolloidussa tutkimusasetelmassa. Kyseenomaiset tutkimusasetelmat eivät kuitenkaan tuota kaikkea tarvittavaa tutkimusnäyttöä, jota tarvitaan tehtäessä terveydenhuollon taloudellisia päätöksiä. Viimeaikaisessa tutkimuskirjallisuudessa on esitetty päätösanalyyttisten mallien laaja-alaisempaa käyttöönottoa validin ja päätöksentekoa tukevan tutkimusnäytön tuottamiseksi.

Tutkimuksen tarkoitus: Suunnitella päätösanalyyttisten mallien käytännönsovelluksia terveysteknologioiden kustannusvaikuttavuuden arviointia varten ja tutkia bayesilaisten menetelmien hyödynnettävyyttä tutkimusnäytön synteesissä sekä päätösanalyyttisessä mallinnuksessa.

Aineisto ja menetelmät: Tutkimus perustuu neljään itsenäiseen case-tutkimukseen, joissa päätösanalyyttisia malleja ja näytönsynteesimenetelmiä käytettiin kustannus-vaikuttavuusinformaation tuottamisessa. Tutkimuksen kaikissa case-tutkimuksissa sovellettiin bayesilaista todennäköisyysjakaumien käyttöön perustuvaa lähestymistapaa.

Tulokset: Päätösanalyyttisia malleja voidaan käyttää syntetisoimaan olemassa olevaa tutkimusnäyttöä, yhdistämään surrogaattitason muutokset todellisiin lopputilamuutoksiin, muuntamaan kliinisen tutkimuksen tulokset taloudellista päätöksentekoa tukevaan muotoon ja arvioimaan terveysteknologian kustannusvaikuttavuutta potilasalaryhmissä. Case-tutkimuksissa tutkimusnäytön identifiointi, valinta ja arviointi muodostivat mallinnuksen yksittäisistä työvaiheista eniten aikaa vievän osuuden. Lisähaasteen mallien suunnitteluun ja tutkimusnäytön synteesiin aiheutti kustannusvaikuttavuusarviointien ajoittuminen terveysteknologioiden käyttöönottovaiheeseen, jolloin käyttökokemusta ja tutkimusnäyttöä terveysteknologioiden hyödyistä ja kustannuksista on saatavissa vielä rajallisesti. Lisäksi case-tutkimukset osoittivat, että bayesilainen lähestymistapa tarjoaa päätösanalyyttiseen mallinnukseen tehokkaan tavan havainnollistaa päätösepävarmuutta, joka johtuu tutkimusnäytön epätarkkuudesta. Bayesilaisen lähestymistavan soveltaminen vaatii kuitenkin perinteiseen mallinnukseen nähden enemmän välivaiheita, jotka voivat tuoda mukanaan uusia metodologiaan ja mallinnusprosessiin liittyviä epävarmuustekijöitä.

Johtopäätökset ja jatkotutkimusehdotukset: Tutkimus osoitti, että päätösanalyyttisten mallit tarjoavat selkeän ja johdonmukaisen matemaattisen rakenteen, jonka avulla voidaan yhdistää käytettävissä oleva tutkimusnäyttö ja arvioida erilaisten terveysteknologioiden kustannusvaikuttavuutta ennen niiden laajamittaista käyttöönottoa. Kehitystyötä tarvitaan kuitenkin vielä mallinnusprosessien standardoinnissa, tutkimusnäytön laadun sisällyttämisessä malleihin sekä menetelmissä, joilla voidaan parantaa malliepävarmuuden käsittelyä ja mallien ”läpinäkyvyyttä” päätöksentekijöille.

Yleinen suomalainen asiasanasto: päätöksenteko; mallit; vaikuttavuus; kustannukset; terveystaloustiede; bayesilainen menetelmä

“Essentially, all models are wrong … some are useful.”

- George Box (1987)

ACKNOWLEDGEMENTS

The present study was carried out in the Department of Health Policy and Management and the Department of Social Pharmacy, University of Kuopio, during the years 2002-2008.

Firstly, I wish to express my deepest gratitude to my supervisor Professor Hannu Valtonen, who introduced me to the field of economic evaluation and gave me opportunity to write this thesis. I would also like to thank my co-supervisors, Professor Olli-Pekka Ryynänen and Docent Miika Linna, for the stimulating discussions related to economic evaluation and health care as a whole over the years.

I am very grateful to the pre-examiners of the study Professor Alan Lyles, University of Baltimore, and Dr Elisabeth Fenwick, University of Glasgow, who read the manuscript thoroughly and whose constructive comments helped improve the final version of this thesis. I also warmly thank Professor Pekka Rissanen for accepting the invitation to be the opponent in the public examination of this thesis.

The Department of Social Pharmacy has provided a stimulating working environment and an excellent opportunity to engage in academic studies; therefore I wish to thank Professor Hannes Enlund and Professor Riitta Ahonen for granting me this opportunity.

I express my gratitude to my coauthors Professor Tuula Pirttilä, Akseli Kivioja, MSc, Taru Hallinen, MSc, Piia Vihinen, PhD, Anne-Mari Ottelin, MSc, Vesa Kiviniemi, Lic.Phil, Professor Helena Gylling, Timo Purmonen, MSc, Erkki Soini, BSc, Docent Vesa Kataja, Ms Riikka-Liisa Vuorinen, and Professor Pirkko-Liisa Kellokumpu-Lehtinen. My special thanks go to my research fellows Erkki Soini and Taru Hallinen for many constructive discussions and for providing me with valuable comments during the finalisation phase of this thesis. Also Piia Peura, MSc, Juha Turunen, PhD, and other colleagues in the Department of Social Pharmacy are recognised for their help and convivial company at work and beyond. In addition, the efforts of Ewen MacDonald, PhD, for revising the language, as well as, Eija Hiltunen, BSc, for revising the lay-out of the thesis, are greatly appreciated.

I owe my sincere thanks to my friends outside the University. Especially, I would like to thank the members of "*Savon Mafia*", since the membership of this particular entourage has given me a great opportunity to relax and enjoy splendid company during the frequent get-together parties arranged over these years. I am a very lucky guy, that I have such true friends! I would also like to thank the Huotari brothers, who have invited me to participate in hunting for game birds and hares every autumn in Raate. Our strolls in the autumnal woods with the dogs and other related activities have been unforgettable experiences!

I also want to express my deepest gratitude to my parents Leena and Kyösti, who have always supported and encouraged me to study, and to my siblings Riika-Mari and Antti for their continuous support and interest when my thesis will be ready - it is ready now, finally!

Nelli, our dear daughter, I would like to thank you for being a wonderful little child, whose smile and laugh make me happy every day!

Finally, I owe my deepest gratitude to my dear Laura for her love and continuous encouragement. I hope that I can support you in the finalisation of your own thesis project at least partly as well as you have supported and encouraged me over these past years.

This study was supported in part by research grants from the Finnish Cultural Foundation of Northern Savo (Emeritus Professor Sirkka Sinkkonen's Fund), National Technology Agency of Finland (TEKES), the Yrjö Jahnsson Foundation, Schering-Plough Oy, Raisio Benecol Oy, and Pfizer Finland Oy those are warmly acknowledged.

Kuopio, January 2008

Janne Martikainen

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ORIGINAL PUBLICATIONS

This doctoral thesis is based on the following original publications:

I Martikainen J, Valtonen H & Pirttilä T. Potential Cost-effectiveness of a Family-Based Programme in Mild Alzheimer's Disease Patients. *European Journal of Health Economics* 2004;5:136-142.

II Martikainen J, Kivioja A, Hallinen T & Vihinen P. Economic Evaluation of Temozolomide in the Treatment of Recurrent Glioblastoma Multiforme. *Pharmacoeconomics* 2005;23:803-815.

III Martikainen J, Ottelin A-M, Kiviniemi V & Gylling H. Plant Stanol Esters are Potentially Cost-Effective in the Prevention of Coronary Heart Disease in Men: Bayesian Modelling Approach. *European Journal of Cardiovascular Prevention & Rehabilitation* 2007;14:265-272

IV Purmonen T*, Martikainen J*, Soini E, Kataja V, Vuorinen R-L & Kellokumpu-Lehtinen P. Economic Evaluation of Sunitinib Malate in the Treatment of Cytokine-Refractory Metastatic Renal Cell Carcinoma (mRCC). *Clinical Therapeutics*, *in press*

***** Authors share equal contribution

ABBREVIATIONS

CEAC	Cost-Effectiveness Acceptability Curve
ΔC	Incremental (mean) Cost
ΔE	Incremental (mean) Effect
ΔNB(λ)	Incremental Net Benefit
ICER	Incremental Cost-Effectiveness Ratio
MCMC	Markov Chain Monte Carlo
NHB	Net Health Benefit
NMB	Net Monetary Benefit
PSA	Probabilistic Sensitivity Analysis
RCT	Randomised Controlled Trial
QALY	Quality Adjusted Life Year

TERMINOLOGY

Bayes' Theorem
Bayes' Theorem is a result that allows the use of new information to update the conditional probability of an event.

Comprehensive decision modelling
The synthesis of all available sources of evidence into a single coherent model that can be used to evaluate cost-effectiveness of alternative treatments.

Decision-analytic model
A mathematical model that reflects the course of a disease in the presence of a treatment and provides estimates of (long-term) costs and effects of the compared intervention.

Evidence
Evidence is an observation or organised body of information, offered to support or justify inferences or beliefs in the demonstration of some proposition or matter at issue.

Health economic evaluation
Economic evaluation can be thought of as a method of assessing the most efficient use of available resources in health care, defined in terms of costs and health outcomes.

Health technology
Health technology covers a wide range of methods of intervening to promote health, including the prevention, diagnosing or treatment of disease, the rehabilitation or long-term care of patients, as well as drugs, devices, clinical procedures and healthcare settings.

Meta-analysis
Meta-analysis is a systematic review or overview which uses quantitative methods to summarise the results.

Probability
Frequentist definition - probability is the proportion of times an event will occur in an infinitely long series of repeated identical situations.

Bayesian definition - probability measures the degree of belief about any unknown but potentially observable quantity, whether or not it is one of a number of repeatable experiments.

Systematic Review
Systematic Review is a literature review focused on a single question which tries to identify, appraise, select and synthesise all high quality research information relevant to that question.

1 INTRODUCTION

1.1 General background

Western Societies are facing increasing challenges in funding their health care systems. Due to these increased pressures, many countries have made explicit use of economic evaluation to make decisions about which new health technologies should be funded from collective resources (Hjelmgren et al. 2001). Scientific evidence from cost-effectiveness analyses allows decision-makers to improve efficiency by spending the limited health care budget on those health technologies that generate the greatest health outcomes per euros spent (Drummond et al. 2005b).

General requirements for cost-effectiveness analysis to inform the allocation decisions in health care have been defined as follows (Scuplher et al. 2005, Drummond et al. 2005b):

- *Defining the decision question.* The need for a clear statement of the decision question. A study setting should be consistent with the stated decision problem.
- *The appropriate time horizon*. From a normative perspective, the time horizon of an analysis should be sufficient to indicate when cost and effect differences between health technologies are stable. For example, for any health technologies that may have a plausible effect on mortality, a lifetime horizon is required.
- *Evidence synthesis*. The study setting should provide an analytic framework within which all evidence relevant to the study question can be brought to bear.
- *Evaluation*. The analysis needs to identify the optimal decision according to the defined decision rules for cost-effectiveness analysis.
- *Uncertainty*. The analysis needs to quantify the uncertainty associated with the decision. In addition, the study setting should facilitate an assessment of the various types of uncertainty relating to the analysis.
- *Additional evidence*. The results of analysis should provide a basis for prioritising future research, which can generate further evidence to re-assess the study question in the future.

In the field of health care, scientific evidence produced by randomised controlled trials (RCTs) is typically ranked highly in the hierarchy of evidence (Sackett et al. 1996). The features of RCTs such as strict inclusion criteria, randomisation, and blinding ensure that the evidence is endowed with high internal validity. However, RCTs pose several threats to “real world” relevance, such as inadequate follow-up times and the use of placebo as a comparator treatment that should be taken into account in cost-effectiveness evidence generation, since the decision-makers need evidence with high internal and external validity (Fayers & Hand 1997, Drummond 1998, Baltussen et al. 1999, Revicki & Frank 1999, Backhouse et al. 2002, Claxton et al. 2001).

Due to the limitations of RCTs, the RCT-based analyses are recognised as having limitations in their ability to produce both internally and externally valid evidence. Buxton (1997) has determined this trade-off problem between internally and externally valid evidence as follows: "…*we seek both scientific rigour and policy relevance. There is no point in having a very precise answer to the wrong question which is what we frequently get with randomised controlled trials…timely approximation is probably better than the ultimate answer…*" The applications of decision-analytic models has been suggested to offer a potential solution to make these "*timely approximations*" in situations where all information needed in cost-effectiveness analysis is not available entirely from a single source but evidence has to be gathered and synthesised from multiple sources (Buxton et al. 1997, Drummond et al. 2005b).

1.2 Purpose of the study

This study focuses on the development of the applications of the decision-analytic models and the applicability of a set of Bayesian methods used in evidence synthesis and decision-analytic modelling in cost-effectiveness evaluations when the goal is to determine the most efficient use of available health care resources under conditions of uncertainty.

1.3 Structure of the study

This study is organised as follows: chapter 2 defines briefly the theoretical and methodological foundation of the study. Chapter 3 introduces the theoretical and methodological foundations of the study. Chapter 4 reviews applied methods and techniques used in this study. Chapter 5 outlines the specific aims of this study. Chapter 6 presents empirical case studies applying a set of Bayesian methods to synthesise the available evidence and to model decision problems under conditions of uncertainty in order to provide relevant information to health care decision-makers. Lessons learnt from the case studies are discussed and proposals for further studies are given in chapter 7. The conclusions are drawn in chapter 8.

2 THEORETICAL AND METHODOLOGICAL FOUNDATION OF THE STUDY

This study applies a multidisciplinary approach which combines economic theory, decision theory, and the principles of evidence-based medicine (EBM). The emphasis is placed on the applications of decision theory that are used to reach the objectives derived from welfare theory. Figure 1 depicts schematically the theoretical foundation and the scope of the study.

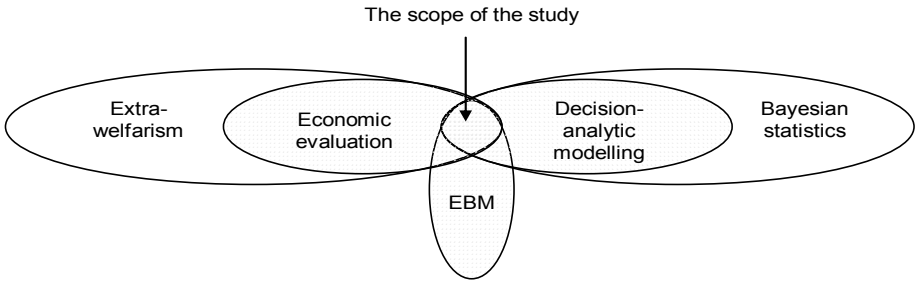


Figure 1. Theoretical foundation and the scope of the study; EBM = Evidence-Based Medicine

The rationales for the use of economic evaluation methods to produce normative recommendations about the reallocation of limited resources are traditionally based on welfare economics. In welfare economics, the objective of the economic evaluation is conventionally preferred to be the maximisation of overall utility for the society i.e. the reallocation of resources is done based on their relative desirability after normative assessment. (Weinstein & Manning 1997, Garber & Phelps 1997) However, the current study takes a somewhat more pragmatic perspective, which is formally known as extra-welfarism or alternatively as the decision-maker approach. The objective of this particular approach is that it should be a pragmatic aid to decision-making, not a complete prescription for social choice (i.e. the results of economic evaluations are intended to inform decision-makers rather than to define what decisions should be made). The general focus of extra-welfarism is the maximisation of the level of health (e.g. in the terms of QALYs) given a health care budget, or minimise cost for a given health outcome. (Brouwer & Koopmanschap 2000)

Figure 2 depicts the process of evidence synthesis and decision-analytic modelling applied for economic evaluation. An essential element of the process is to identify all relevant evidence in order to reduce bias and uncertainty in the assessment of cost-effectiveness. This element is consistent with the principles of evidence-based medicine (EBM), where systematically and comprehensibly synthesised clinical evidence is used to aid clinicians decide on appropriate treatment for their patients under specific clinical conditions and/or circumstances (Sackett et al. 1996). When one conducts an economic evaluation, however, it is not simply clinical evidence that is required. In addition, evidence relating to other factors e.g. costs and health-related quality of life is required. The additional evidence could be

collected e.g. from cohort studies, discharge registries or cross-sectional surveys. A list of potential sources of evidence for each data component of interest is given in section 4.4.1.

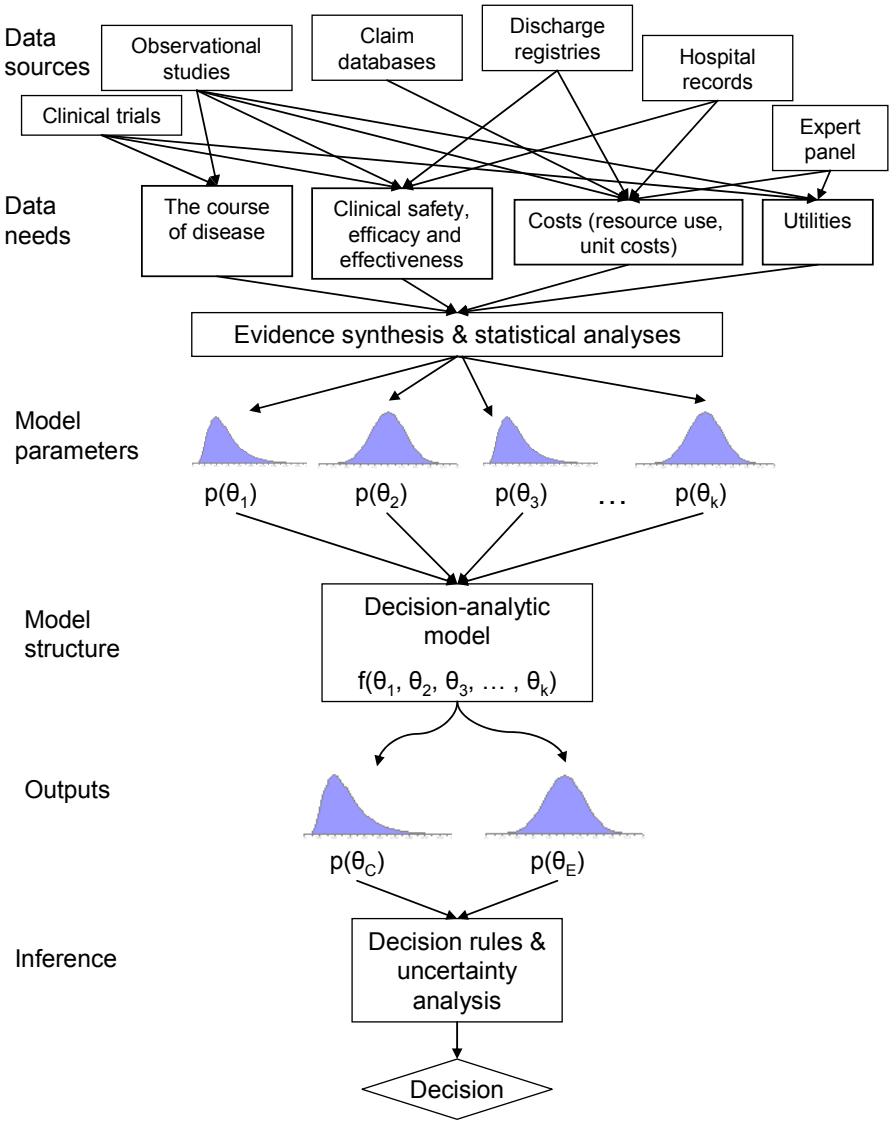


Figure 2. Schematic presentation of the process of evidence synthesis and decision-analytic modelling for economic evaluation

The applications of decision-analytic models may improve decision-making under uncertainty by a clear structuring of the decision problem and an explicit analysis of the expected consequences of different

decisions. The decision-analytic modelling has its theoretical foundations in statistical decision theory and it applies the axioms of expected utility theory (see e.g. Raiffa 1970 and Howard 1968 for further discussions).

The concept of expected value has a central role in economic evaluation, since the choice of a preferred decision is done on the basis of expected values (i.e. decision rules applied in economic evaluation are based on the comparison of differences between competitive health technologies in terms of their expected mean costs and effects). The use of expected values is based on the idea that, when the risks of investment are spread over the entire population, then the share of risks that falling on any one individual is so small that it does not have any influence on the decision (Arrow & Lind 1970).

Fundamentally, uncertainty around the mean values of estimates arises, because they are the functions of model parameters that are not known with absolute certainty i.e. there is uncertainty about true mean values of model parameters (cf. standard error of the mean). It should be noticed that this uncertainty differs from the uncertainty due to variability in individual outcomes (cf. standard deviation) and heterogeneity (i.e. differences in expected consequences that can, in part, be explained).

In the context of Bayesian statistics, the concept of probability is generalised to represent the degree of belief about the true values of model parameters given available evidence. This view of probability is essential in the decision-analytic modelling, since it allows the volume and precision of the evidence available for each of the model parameters to be reflected in prior probability distributions assigned to these parameters (cf. Figure 2). (Claxton 1999)

After synthesising the available evidence and assigning the prior distributions for uncertain model parameters, simulation methods are used to incorporate parameter uncertainty through the developed decision-analytic model so that repeated simulations provide a joint distribution for the incremental costs (ΔC) and effects (ΔE) (cf. figure 2). Figure 3 illustrates a hypothetical example about a situation, where a new treatment is more costly ($\Delta C>0$) and more effective ($\Delta E>0$) compared to the current practice. The scatter-plot depicts the joint distribution of ΔE and ΔC i.e. $p(\Delta E, \Delta C)$ based on hypothetical data. Distributions for any functions of ΔC and ΔE, such as ICER and INB (chapter 4.2), can be derived from $p(\Delta E, \Delta C)$.

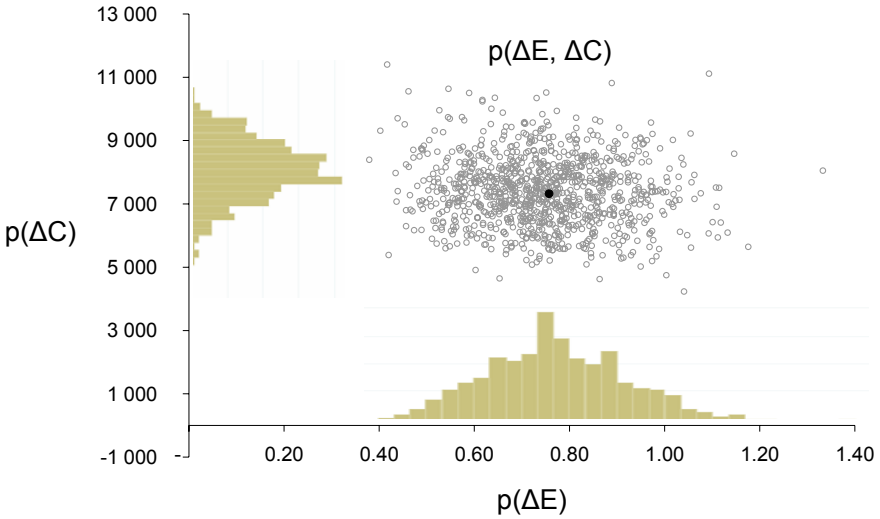
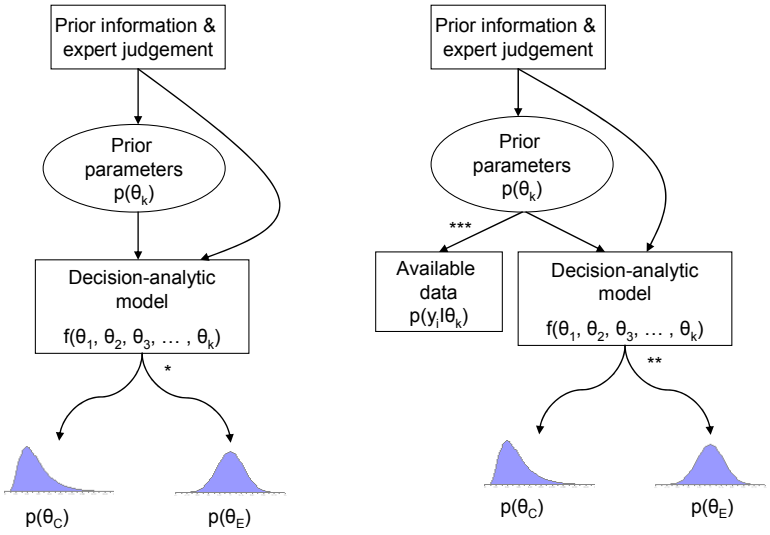


Figure 3. Marginal distributions of ΔC and ΔE. The scatter-plot depicts the joint distribution of ΔE and ΔC

There are two commonly applied simulation algorithms to model uncertainty in cost-effectiveness analysis -- the two-stage approach, which applies the Monte Carlo simulation algorithm (chapters 6.1 and 6.2 will provide examples about the applications of this approach) and the comprehensive decision modelling approach, which applies the Markov Chain Monte Carlo simulation algorithm (chapters 6.2 and 6.4 will provide examples about the applications of this approach). The two-stage approach distinguishes the processes of evidence synthesis and decision-analytic modelling from one another (cf. Figure 2), whereas the comprehensive decision modelling approach unifies the processes of evidence synthesis (i.e. statistical data analysis, such as meta-analyses, meta-regressions, etc.) and decision-analytic modelling into a single coherent model that can be used to simulate the expected costs and effects of alternative health technologies without any intermediate summary steps. (Spiegelhalter et al. 2000)

Figure 4 depicts schematically how these two approaches differ from another. When the two-stage approach is applied, the values of model parameters are usually based on subjective judgements, published studies, data analysis or some short of combination of these three. The volume and quality of this gathered evidence are illustrated in the model by specified prior probability distributions. The parameters for these specified prior probability distributions are estimated from data, derived from published studies, or defined by using some specified assumptions. Computation routines for the two-stage approach are presented in detail in chapter 4.5.1. The comprehensive decision modelling approach differs mainly from the two-stage approach in that, rather than incorporating uncertainty forward from specified prior probability distributions on parameters, it incorporates uncertainty in available patient- or study-level data back to parameters (i.e. initial prior distributions are updated by the Bayes theorem to poste-

rior probability distributions), and then propagates posterior uncertainty forward to model outputs. In other words, the comprehensive decision-analytic modelling approach provides a framework to estimate simultaneously the conditional posterior parameters that are based on specified prior evidence, expert judgements, and available trial- and patient-level data.



* Simulations are based on the prior probability distributions of model parameters
** Simulations are based on the joint posterior distribution of model parameters $p(\theta_1, \theta_2, \theta_3, \dots, \theta_k \mid y_i)$
*** The direction of arrow indicates the conditional independency of y_i i.e. $p(y_i|\theta_k)$

Figure 4. Two-stage (on the left-hand side) vs. comprehensive approach (on the right-hand side) to model uncertainty in decision-analytic models (adopted and modified from Spiegelhalter et al. 2004, 311).

In the comprehensive decision modelling approach, inferences about the parameters of interest are made by integrating out the unknown parameters over the joint posterior distribution of all model parameters. High-dimensional integrals, however, can not usually be solved in a closed form due to analytical problems and therefore MCMC sampling techniques (section 4.5.2) are needed to solve these problems (Gilks et al. 1996). Computation routines for the two-stage approach are presented in detail in chapter 4.5.2.

Relatively few empirical studies applying the comprehensive decision modelling approach in cost-effectiveness analysis have been published to date but the increasing number of application studies can be expected to appear in the near future. Published empirical applications applying the comprehensive decision modelling approach include e.g. the study by Cooper et al. (2003), where a comprehensive decision model was developed to evaluate the cost-effectiveness of a particular medication in the treatment of breast cancer. In addition, the studies by Iglesias & Claxton (2006) and Vergel et al.

(2007) provide more recent empirical examples about the comprehensive decision analyses that were applied to evaluate the cost-effectiveness of particular interventions in the UK setting.

The use of the comprehensive decision modelling approach allows also the iterative approach to cost-effectiveness analysis, since whenever new evidence that may affect a decision becomes available a decision-analytic model can be updated by using the current posterior evidence as a prior evidence for new evidence[1] (cf. Figure 12) (Fenwick et al. 2006). The iterative approach to cost-effectiveness analysis takes into account the dynamic nature of the operational environment in health care, where the cost-effectiveness of health technologies may change over time e.g. due to generic competition, learning effects, patient case mix, etc. Therefore, it is important to update evidence periodically to assess whether the cost-effectiveness of the health technology has changed sufficiently to necessitate modifying the original decision. (Banta & Thacker 1990, Sculpher et al. 1997, Fenwick et al. 2000, Laking et al. 2002, Sculpher et al. 2006, Fenwick et al. 2006). Figure 5 depicts the stages in the iterative approach to the cost-effectiveness analysis of health technologies.

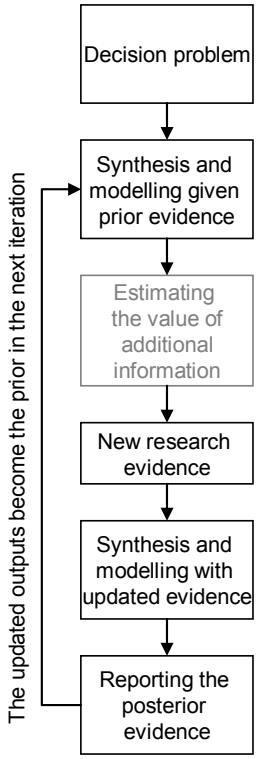


Figure 5. Iterative approach to the economic evaluation of health technologies

[1] This is possible due to the sequential use of Bayes theorem (see Spiegelhalter et al. 2004, 79)

In the iterative framework, the results from a decision-analytic model serve as inputs for policy-makers when deciding for example whether certain health technologies should be financed from the public funds. Actually, this decision has two separate but related aspects, namely the optimal public financing decision given existing evidence and acquiring more evidence to provide information for the decision on public financing. Therefore, two questions need to be answered: I) Which health technology is cost-effective (based on expected mean values) given the existing evidence? II) Is the acquisition of additional evidence to reduce the expected cost of decision uncertainty cost-effective? (Fenwick et al. 2006) Therefore, the current study concentrates on a set of methods that can be used to answer these two questions. For example, an answer to the first question can be given by applying methods, such as cost-effectiveness acceptability curves, which convert parameter uncertainty into decision uncertainty. Recent developments have made it possible to evaluate the value of additional evidence in monetary terms. These value of information (VOI) methods can be used to determine an upper bound to the value of additional evidence or to indicate what kind of additional evidence would be most valuable in reducing uncertainty surrounding the decision (Claxton 1999).

3 CONCEPT OF UNCERTAINTY IN EVIDENCE SYNTHESIS AND DECISION-ANALYTIC MODELLING

The credibility of decision-analytic models rests on their validity. The overall validity of models depends on how the selection of model structures, the identification and incorporation of evidence into the models, and the testing of the models' consistency are conducted. Consistency relates to many factors e.g. questions about the mathematical logic of models and how consistent the results produced by the models are with other available evidence. (Philips et al. 2006)

One systematic approach to identify, handle and document uncertainty is to categorise the different types of uncertainty according to general model development phases introduced by Howard (1968). Figure 6 depicts the model development phases and the different types of uncertainty related to each development phase.

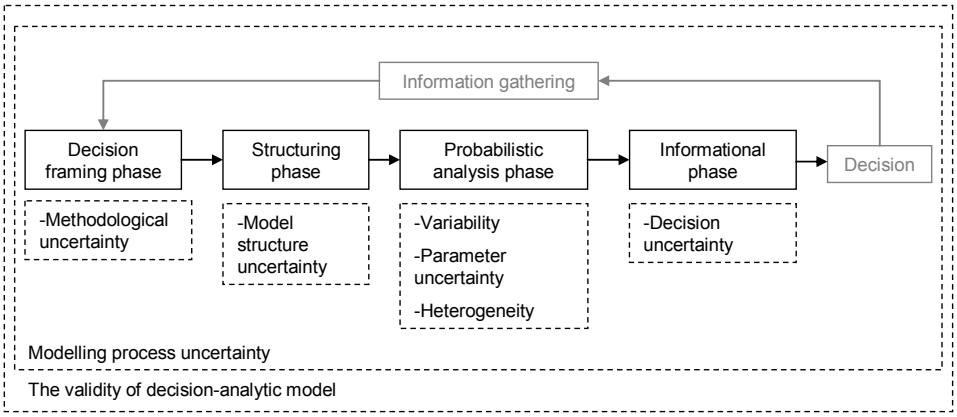


Figure 6. Sources of uncertainty classified according to model development phases. Summarised from Briggs 2000, Spiegelhalter & Best 2003 and Briggs et al. 2006

One potential source of uncertainty that relates to the model consistency is modelling process uncertainty, which arises e.g. due to errors relating to data incorporation or programming syntax. Internal consistency can be ensured by conducting logical checks and programming the model in an alternative software package. In some cases, modelling process uncertainty can be only handled by commissioning the same decision-analytic model from different teams of analysts (cf. health technology appraisal processes commissioned by NICE in UK). (Philips et al. 2006)

In the decision framing phase, *methodological uncertainty* may arise due to disagreement among researchers about the most appropriate analytical methods. Uncertainty may relate to e.g. the most preferred way to incorporate time preference or productivity losses/savings into the economic evaluation. (Briggs 2000, Briggs et al. 2006)

In the model structuring phase, *uncertainty at a structural level* (i.e. the appropriate qualitative structure of the model is uncertain) may arise e.g. due to lack of knowledge concerning to relationships between parameters (Briggs 2000, Spiegelhalter & Best 2003, and Briggs et al. 2006). According to a recent literature review, the most commonly faced sources of structural uncertainty are related to: I) inclusion or exclusion of potentially relevant comparators, II) inclusion or exclusion of potentially relevant events, III) statistical models to estimate specific parameters and IV) lack of evidence (Bojke et al. 2006). Usually structural uncertainty is handled by developing alternative model structures with different assumptions and thereafter computing results for each alternative specification. The decision about whether one model structure is superior compared to some other structure is commonly left to decision-makers. However, it has been demonstrated that the assessment of credibility of the alternative model structures may be challenging. (Drummond et al. 2005a).

The probabilistic analysis phase includes multiple types of uncertainty. The explanation of *variability* (i.e. the difference in outcomes that occur between patients by chance) is not the main object in decision-analytic modelling, since we are usually not interested about random variability in outcomes although uncertainty about individual outcomes may affect decision making through consideration of equity (Groot Koerkamp et al. 2007). While variability is defined as change in outcomes between homogeneous patients, *heterogeneity* relates to differences between patients that can, in part, be explained. Strictly speaking, heterogeneity is not a source of uncertainty as it relates to differences that can be explained e.g. using patient characteristics, such as age and sex (cf. Russell & Sisk 2000, where the effect of age on cost-effectiveness has been explored). However, even if the mean parameter values per subgroups are estimated, it should be noted that variability within subgroups will remain. Furthermore, the latent variables cause a potential source of uncertainty, because they obviously cannot be used to explain a proportion of overall variability between patients.

Parameter uncertainty relates to the volume and precision with which an input parameter used in a decision-analytic model is estimated. Uncertainty arises because the input parameters have definite values but they cannot be known with certainty. Sometimes parameter uncertainty has also been termed as *second-order uncertainty* to distinguish it from *first-order uncertainty*, which again is a synonym for variability (Stinnett & Paltiel 1997). Parameter uncertainty has a central role in evidence synthesis and decision-analytic modelling, since a decision-analytic model produces the expected costs and health outcomes of a health technology via the complex functions of uncertain input parameters (cf. Figure 2) (Doubilet et al. 1985, Briggs 2000, Fenwick et al. 2000, Claxton et al. 2001, Briggs et al. 2002, Spiegelhalter & Best 2003).

The use of simulation methods converts parameter uncertainty into *decision uncertainty.* Proper quantification of decision uncertainty provides the starting point for assessing the value of additional evidence.

Additional evidence may be valuable because it reduces the expected costs of making an incorrect decision. (Claxton 1999)

4 REVIEW OF APPLIED METHODS

4.1 Markov-models in decision-analytic modelling

The decision-analytic models embrace a variety of mathematical techniques but the two main forms used in the economic evaluation of health care technologies are decision trees and state transition models (Karnon 2003, Brennan et al. 2006). The decision trees are visual representations of selected treatment options and the consequences that may follow from each option. Each treatment option is followed by branches representing the possible events with their respective probabilities (Pauker & Kassier 1987, Detsky et al. 1997). However, the decision trees are relatively restricted in their abilities to describe the consequences of different decision in situations where they occur over time. In these particular cases, the state transition models (i.e. the Markov models) are suited to illustrate the long-term outcomes associated with different decisions. In recent years, the Markov models have become increasingly used vehicles in economic evaluations, since they allow researchers to construct flexible applications to reflect disease progression using constant, time-dependent, and discrete processes. (Sonnenberg & Beck 1993, Briggs & Sculpher 1998, Bala et al. 2006)

In the decision-analytic modelling, the Markov models are frequently applied to describe the course of diseases in the presence of particular treatments under conditions of uncertainty. Different consequences are modelled as transitions from one previously specified health state to another. These transitions between health states can be graphically represented using state transition diagram shown in Figure 7. Transitions between specified health states will occur until all members of a hypothetical cohort have entered an absorbing state, such as death, or the time horizon covered by the model is reached.

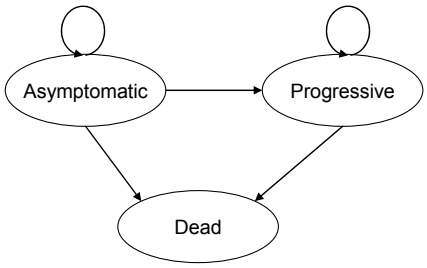


Figure 7. Markov model illustrated as a state transition diagram. Circles correspond to health states and arrows correspond to possible transitions from one health state to another.

The state transition models are divided into Markov chain and Markov process models depending on how the transition probabilities between specified health states are handled during the modelling. Markov chain models have constant probabilities of transitions between states whereas transition probabilities in Markov process models may be allowed to vary according to another model variable (e.g. time or different prognostic factors).

The fundamental assumption (i.e. the Markovian assumption) of Markov models is that the probability of moving out of a health state is not dependent on the health states that a patient may have experienced previously (i.e. the Markov models do not have any memory). The use of time-dependent transition probabilities and/or so called tunnel states is a way to work around the Markovian assumption. Another fundamental assumption is that the health states are mutually exclusive (i.e. an individual can be in one and only one health state at any given point in time). (Sonnenberg & Beck 1993)

The state transitions between the health states illustrated in the Figure 7 are given in a corresponding matrix form as follows:

$$S_{t=0} = [1, 0, 0]$$

[1]
$$P_t = \begin{bmatrix} p_{11t} & p_{12t} & 1-(p_{11t}+p_{12t}) \\ 0 & p_{22t} & 1-p_{22t} \\ 0 & 0 & 1 \end{bmatrix}$$

=

From…	To… Asymptomatic	Progressive	Death
Asymptomatic	P_{11t}	P_{12t}	$1-(P_{11t}+P_{12t})$
Progressive	0	P_{22t}	$1-P_{22t}$
Death	0	0	1

where $S_{t=0}$ is the vector of initial state probabilities at the time t=0. In this particular case, $S_{t=0}$ defines that 100% of the hypothetical cohort of interest assumed to be asymptomatic at the time t=0. P_t is the transition probability matrix for a cycle t≥1. (Spiegelhalter et al. 2003). More specifically, p_{11t} is a probability to stay in the asymptomatic state at the time t≥1, whereas p_{12t} is a probability to move from the asymptomatic state to the progressive state at the time t≥1, A residual $1-(p_{11t}+p_{12t})$ is a probability to move from the asymptomatic state to the death state. P_{22t} and its residual $1-P_{22t}$ are probabilities to stay in the progressive state and to move from the progressive state to the death state, respectively. The death state is handled as an absorbing state and therefore its transition probability is always one.

The initial state probabilities are often handled as the number of patients in the model. This analytical approach is called Markov cohort modelling. When the cohort modelling approach is applied, the initial state vector $S_{t=0}$ is usually defined e.g. [1000, 0, 0], which indicates that the size of the hypothetical cohort of interest is assumed to be 1000 and all individuals are assumed to be asymptomatic at the beginning of the Markov chain (Sonnenberg & Beck 1993). The results of the Markov cohort modelling have been found to be robust regardless of the number of individuals assumed in the hypothetical cohort (Cooper et al. 2003).

Figure 8 illustrates the first two cycles of hypothetical cohort simulations, where hypothetical health technologies T_1 and T_0 are compared using the model structure depicted in Figure 7. T_0 is the current treatment for a particular condition, whereas T_1 is a possible new treatment for that same condition. The initial state vector $S_{t=0}$ is set to be [1000, 0, 0]. The proportions of individuals in different health states at time t+1 are obtained by multiplying the initial state vector $S_{t=0}$ by the transition probability matrix P_t $\Rightarrow S_{t-1} \times P_t = S_{t+1}$. Markov estimation entails successive multiplication of S_{t-1} and P_t until all individuals into the cohort have entered an absorbing state or the time horizon covered by the model has been reached. (Spiegelhalter et al. 2003) As mentioned above, in the Markov chain analysis all probabilities in P_t are assumed to be constant, whereas in the Markov process analysis the values of P_t are allowed to vary according to other model variables.

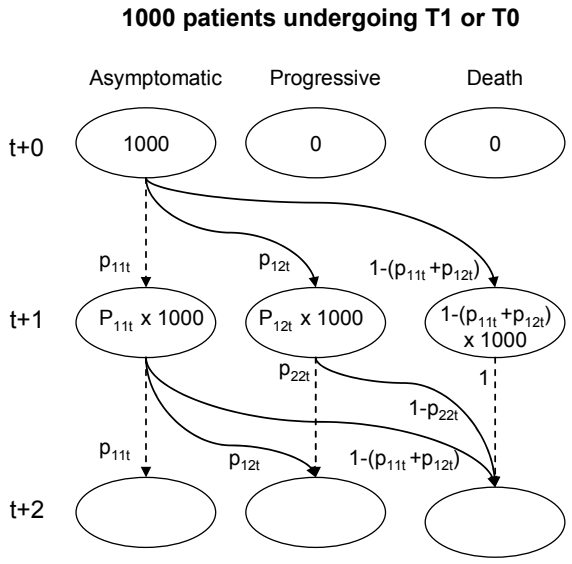


Figure 8. Illustration of the first two cycles of cohort simulations for hypothetical health technologies T_1 or T_0

In order to perform the cost-effectiveness analysis, one needs to examine how the results of Markov cohort analysis are altered by the introduction of the new treatment T_1. To estimate the incremental cost-effectiveness of the treatment T_1 as compared to T_0, the average costs and health outcomes for the hypothetical cohorts over the modelled time horizon should be estimated for both treatment groups. If the time horizon of interest is over a year, discounting is usually applied to generate the present value of the future costs and health outcomes (Cairns 2001).

The expected costs and health outcomes for the hypothetical cohorts of individuals receiving the treatment T_1 and T_0 are given by equations:

[2] Expected (mean) costs E(C): $\sum_{t=1}^{T}(1+r)^{-t}\sum_{s}^{S}\pi_{s,t}C_s$

[3] Expected (mean) effects E(E): $\sum_{t=1}^{T}(1+r)^{-t}\sum_{s}^{S}\pi_{s,t}E_s$

where T is the total length of time horizon, r is the discount rate, S is the number of health states in the specified model (cf. figure 8), $\pi_{s,t}$ is the number of individuals in health state s at time t, C_s and E_s are the mean cost and effect of health state s. (Bala & Mauskopf 2006). If the size of the hypothetical cohort is >1, then the expected costs and health outcomes can be obtained by dividing the cumulative costs and health outcomes by the size of the cohort (Cooper et al. 2003).

4.2 Decision rules and the statistical analysis of uncertainty in incremental cost-effectiveness analysis

4.2.1 Decision rules in cost-effectiveness analysis

Assume that the objective of decision-making is to decide whether or not to replace T_0 by T_1. T_1 would obviously be chosen if it demonstrated improved expected health outcomes at reduced expected costs and would obviously not be chosen if expected costs increased but expected health effects decreased. The more difficult decisions arise when T_1 improves health effects at increased cost or has reduced health effects but at reduced cost (O'Brien et al. 1994). These four situations are summarized in Table 1.

Table 1. Possible decisions based on incremental (mean) costs and health effects (O'Brien et al. 1994)

Incremental costs (ΔC)	Incremental health effects (ΔE)	Decision
$C_{T1}-C_{T0}<0$	$E_{T1}-E_{T0}>0$	Dominance. T_1 can be accepted as it is both cheaper and more effective
$C_{T1}-C_{T0}>0$	$E_{T1}-E_{T0}<0$	Dominance. T_1 is rejected as it is both more expensive and less effective than existing therapy.
$C_{T1}-C_{T0}>0$	$E_{T1}-E_{T0}>0$	Trade-off. Magnitude of the additional cost of T_1 should be considered relative to its additional health outcomes.
$C_{T1}-C_{T0}<0$	$E_{T1}-E_{T0}<0$	Trade-off. Magnitude of the cost-saving of T_1 should be considered relative to its reduced health outcomes.

Those four states of world mentioned in table 1 are equivalent to the four quadrants of the cost-effectiveness plane (see Figure 9), which has been advocated as a way of providing a graphical interpretation of cost-effectiveness results (Anderson et al. 1986, Black 1990). The cost-effectiveness plane is a two-dimensional space with the x-axis being the mean difference in expected health outcome (ΔE) and the y-axis being the mean difference in expected cost (ΔC). To aid reference, the points of the compass are sometimes applied to label the four quadrants on the cost-effectiveness plane.

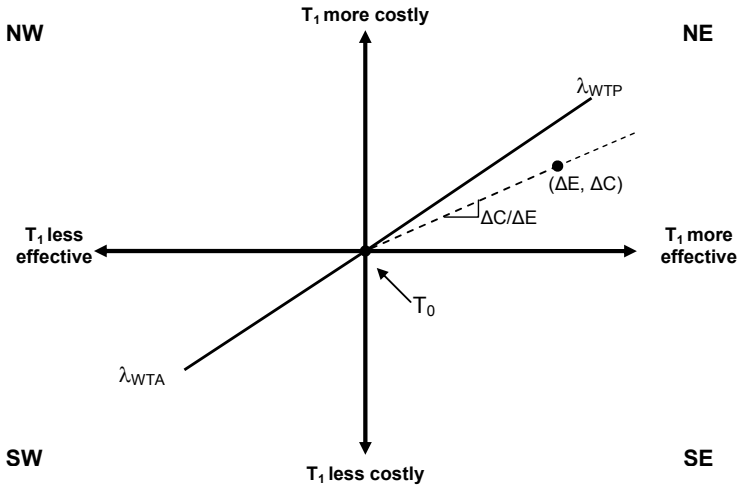


Figure 9. Cost-effectiveness plane

Decision problems associated with trade-offs represented in Table 1 and Figure 9 (the North-West and South-West quadrants) lead to a conventional incremental cost-effectiveness analysis, where the incremental cost-effectiveness ratio (ICER=ΔC/ΔE) is compared to a `threshold´ (Eichler et al. 2004) or `ceiling´ (Fenwick et al. 2006) value (usually denoted by λ) of cost-effectiveness that is considered to represent some maximum acceptable limit on what health care decision makers are prepared to pay for a unit of health outcome (ΔC/ΔE<λ)[2]. In figure 9, the slope of the dotted line extending from the origin (T_0) through a ΔC/ΔE estimate represents the ICER of T_1 relative to T_0. The steeper the slope of the line ΔC/ ΔE, the greater is the additional cost at which additional units of health output are gained by T_1 relative to T_0, and the less attractive T_1 becomes (O'Brien et al. 1994). The ICER can be interpreted as the additional investment of resources needed for each additional unit of health improvement expected to result from investing in T_1 rather than T_0.

[2] Conventionally, the constant returns of scale and perfect divisibility of compared treatments for homogenous population are assumed when incremental cost-effectiveness analysis is applied. Elbasha and Messonnier (2004) and Lord et al. (2006) have discussed the consequences if these fundamental assumptions are violated.

In Figure 9, the threshold value line λ_{WTP} on the NE quadrant of the cost-effectiveness plane is assumed to reflect the shadow price per unit of health output in the absence of a market (Weinstein & Zeckhauser 1973, Johannesson & Weinstein 1993, Karlsson & Johannesson 1996)[3] and the threshold value line λ_{WTA} on the SW quadrant of the cost-effectiveness plane is assumed to reflect the minimum amount that decision-makers would be willing to accept in exchange for foregoing an incremental gain in health. When the calculated ICER is below the level that decision-makers deem acceptable for producing the health outcome, then T_1 can be considered as being cost-effective and hence be recommended to be approved.

4.2.2 The incremental cost-effectiveness approach

Prior to the early 1990's, the *deterministic* analysis (i.e. cost and health outcome estimates were considered as point estimates) of the ICER estimate was the usual approach taken to deal with these problems. Uncertainty about the expected costs and health effects and their differences (i.e. ΔC and ΔE) was handled through conventional sensitivity analysis methods. In conventional sensitivity analyses, the magnitudes of the model parameters were varied within (arbitrarily) selected ranges and the effects on the results were observed. For example, in one-way sensitivity analysis, the magnitude of a single parameter was varied between its lower and upper bounds. In threshold analysis, the parameters were varied up to the point at which the expected optimal decision would change. Respectively, the analysis of extremes allowed researchers to create optimistic and pessimistic scenarios based on the most optimistic or pessimistic model parameter estimates. (Briggs et al. 1994) The use of conventional sensitivity analysis methods for handling uncertainty in economic evaluations has been reviewed elsewhere (Briggs et al. 1994, Briggs & Sculpher 1995, Agro et al. 1997).

In 1994, O'Brien et al. (1994) pointed out that ideally data should be wholly *stochastic* sampled from a random sample of the patient population, where both costs and effects are determined from the same patients under study. This finding launched the rapid development of statistical methods to characterise the uncertainty around ICERs. In addition, these developments have also influenced methodologies applied in the decision-analytic modelling, raising the question about how to formally depict uncertainty around p(ΔE, ΔC).

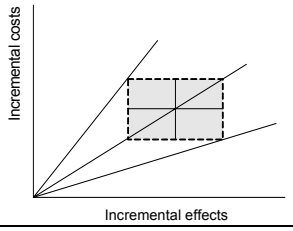
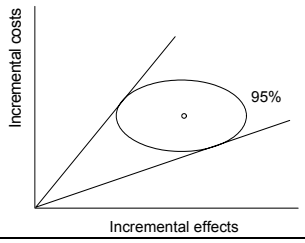
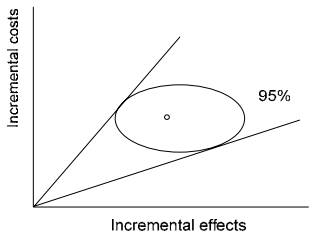
O'Brien and colleagues in 1994 showed that in the situation where T_1 is more effective and more costly (i.e. the NE quadrant of the cost-effectiveness plane) than T_0, it is possible to characterize uncertainty using 95% confidence intervals for the numerators and denominators of cost-effectiveness ratio separately, but not for the ratio itself. One of main problems related to characterizing uncertainty is that the ICER as a ratio statistic has a discontinuous distribution, which poses problems for defining the stan-

[3] Strictly speaking, from the decision-maker's perspective, λ can be viewed as the shadow price of a fixed health care budget (i.e. the highest ICER that decision makers could afford given a fixed health care budget). Whereas from a welfarist perspective, λ is an estimate of societal willingness to pay for an additional unit of health outcome.

dard error for the ratio. This problem is also known as “the close-to-zero problem” (i.e. the neighbourhood of zero in the denominator makes a formula for the variance of the ICER intractable).

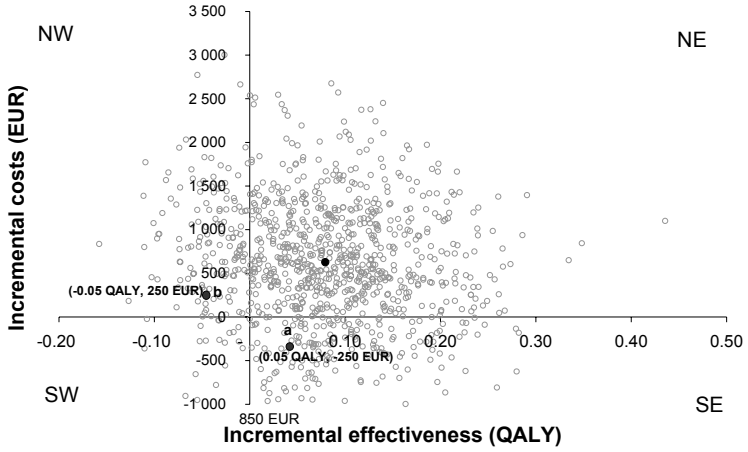
O’Brien and colleagues (1994) introduced for the first time also the close-to-zero problem associated with confidence interval for ICER in a clear manner and it greatly influenced the development of statistical methods for handling uncertainty in cost-effectiveness analysis at the end of 1990’s. Table 2 summarises the developments in defining confidence intervals for ICER, when simulated cost-effectiveness data is available. (Polsky et al. 1997, Tambour & Zethraeus 1998, Briggs et al. 1999, Heitjan et al. 1999)

Table 2. Summary of different developments to estimate confidence intervals for cost-effectiveness ratios, when simulated cost-effectiveness data is available

Author(s)	Name of method	Approach	Limitations	Graphical presentation on the cost-effectiveness plane
O’Brien et al. (1994)	The Box method	* Approximate the separate confidence intervals for the ΔC and the ΔE.	* Yields to an interval that is much wider than the true 95% confidence interval.	
O’Brien et al. (1994)	The Taylor series expansion	*Approximate the variance of the ΔC/ ΔE ratio using sample estimates of the means and variances.	* Assumption of a normal distribution may be justified only in the case of large samples. Does not give clear answer to the close-to-zero problem.	
Van Hout et al. (1994)	The confidence ellipse	* Assume the elliptical shape for p(ΔE, ΔC) * Formula for the ellipse is derived assuming that the ΔC and ΔE follow a joint normal distribution. * Allow the covariance between the numerator ΔC and the denominator ΔE.	* Gives only an approximation of the 95% confidence interval for the ratio, when estimation is done based on the slope of tangents to the 95% probability ellipse.	
Chaudhary & Stearns (1996) Willan & O´Brien (1996)	Fieller’s theorem	* Parametric method based on the assumption that the ΔC and the ΔE follow a joint normal distribution * Takes into account the potential skewness in the sampling distribution of the ΔC/ ΔE estimator, and therefore may not be symmetrically positioned around the ΔC/ ΔE point estimate	* Assumption of joint normality may not apply (e.g. costs are often positively skewed), particularly when sample sizes are small.	

Stinnett and Mullahy (1998), however, pointed out that problems arise when parts of the simulated replicate pairs lie in the dominance quadrants (i.e. in the SE and SW quadrants). The corresponding situation where both cost and effect differences are insignificant is illustrated in Figure 10[4].

A)



B)

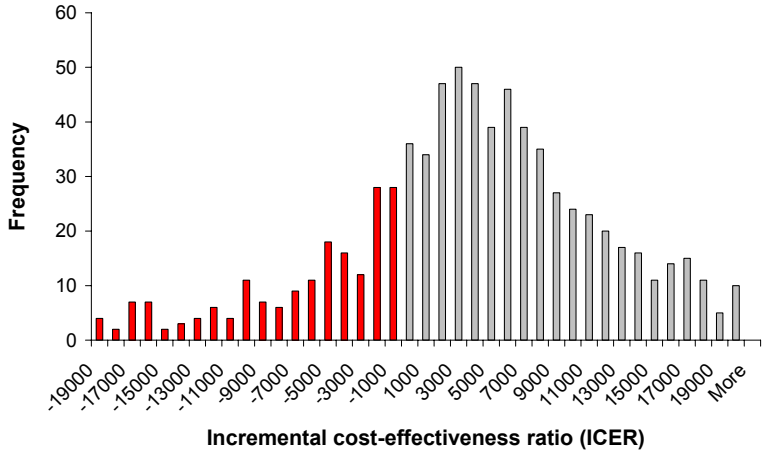


Figure 10. Joint distribution of simulated ΔC and ΔE replicate pairs that lie in more than one quadrant on the cost-effectiveness plane (A) and the corresponding empirical sampling distribution of the ICER presented as a histogram (note that the histogram is truncated into the range from -20 000 to 20 000 (B) in order to present the distribution more clearly).

[4] Figures are based on a hypothetical data generated by a Monte Carlo simulation.

When the parts of the joint distribution lie in the dominance quadrants, the estimation and interpretation of negative ratios can be ambiguous. Two negative ratios with equal values obtained from these quadrants have totally opposite meanings. For example, in Figure 10 replicates **a** (-250 EUR/0.05 QALY) and **b** (250 EUR/-0.05 QALY) have the same ICER estimate (-5000 EUR/QALY) but they have totally contrasting interpretations. Thus, the negative ICER estimate has no meaningful interpretation unless it is presented in the context of the quadrant in the cost-effectiveness plane to which it corresponds (see histogram in figure 10).

A more general problem is that the replicates in the negative quadrants have no meaningful ordering. In the NE and SW quadrants, low ICER estimates are preferred to high ICER estimates. However, no such simple decision rule exists in the negative quadrants. Furthermore, the problem of ICER ordering also applies to positive ratios in different quadrants. The decision rule in the NE quadrant is to adopt T_1 if the ICER estimate is below the λ_{WTP}. Whereas the decision rule in the SW quadrant is to adopt T_1 if the ICER estimate is above $\lambda_{WTA.}$ (O'Brien et al. 2002)

Stinnett and Paltiel (1997) highlighted the important fundamental problem of taking patient-level average ratios when estimating ICERs -- the mean of ratios is not equal to the ratio of the means:

[4] $$\frac{\overline{C}_{T1}}{\overline{E}_{T1}} - \frac{\overline{C}_{T0}}{\overline{E}_{T0}} \neq \frac{\left(\overline{C}_{T1} - \overline{C}_{T0}\right)}{\left(\overline{E}_{T1} - \overline{E}_{T0}\right)}$$

The consequence of this finding is that the incremental ratio cannot be constructed from the difference between the average cost-effectiveness ratios in each arm of a decision-analytic model (i.e. mean ratios from the empirical distributions of ICERs are meaningless, by contrast the ratio of means must be used, in simulation analyses).

4.2.3 The net benefit approach

In response to the problems encountered with the ICER approach, an alternative named as a net benefit statistic approach was introduced by Claxton and Posnett (1996), Tambour et al. (1998) and Claxton (1999) on the monetary scale and Stinnett and Mullahy (1998) on the effect scale. The equivalence and optimality of ratio-based and net benefit-based approaches to health care resource allocation have been described by Laska et al. (1999) and Craig and Black (2001).

The incremental net benefit statistic on the health scale ($\Delta\mathrm{NHB}(\lambda)$) is derived by rearranging the standard cost-effectiveness decision rule as follows:

[5] $$\Delta\mathrm{NHB}(\lambda) = \Delta E - \Delta C / \lambda$$

With this approach, the net effect is given in effectiveness units and not in monetary units as the net cost of program T_1 is divided by the value of one effectiveness unit (λ).

If the net gain in health output is instead multiplied by the value of the effectiveness unit, the net benefit is given in monetary terms. The incremental net monetary benefit of T_1 compared to T_0 as:

[6] $$\Delta\mathrm{NMB}(\lambda) = \lambda\Delta E - \Delta C > 0$$

Incremental net benefits can be seen as a linear combination of two asymptotically normal variables. Figure 11 shows incremental net benefit statistics both on the cost and effect scales as a function of λ, since usually the value of λ is unknown by the analyst when the analysis is performed. For the incremental net benefits on the cost scale, the slope of the line is the difference in effects. The point where the line meets the vertical axis represents the difference in costs, and where it crosses the horizontal axis is the ICER (i.e. $\Delta\mathrm{NMB}(\lambda) = 0$).

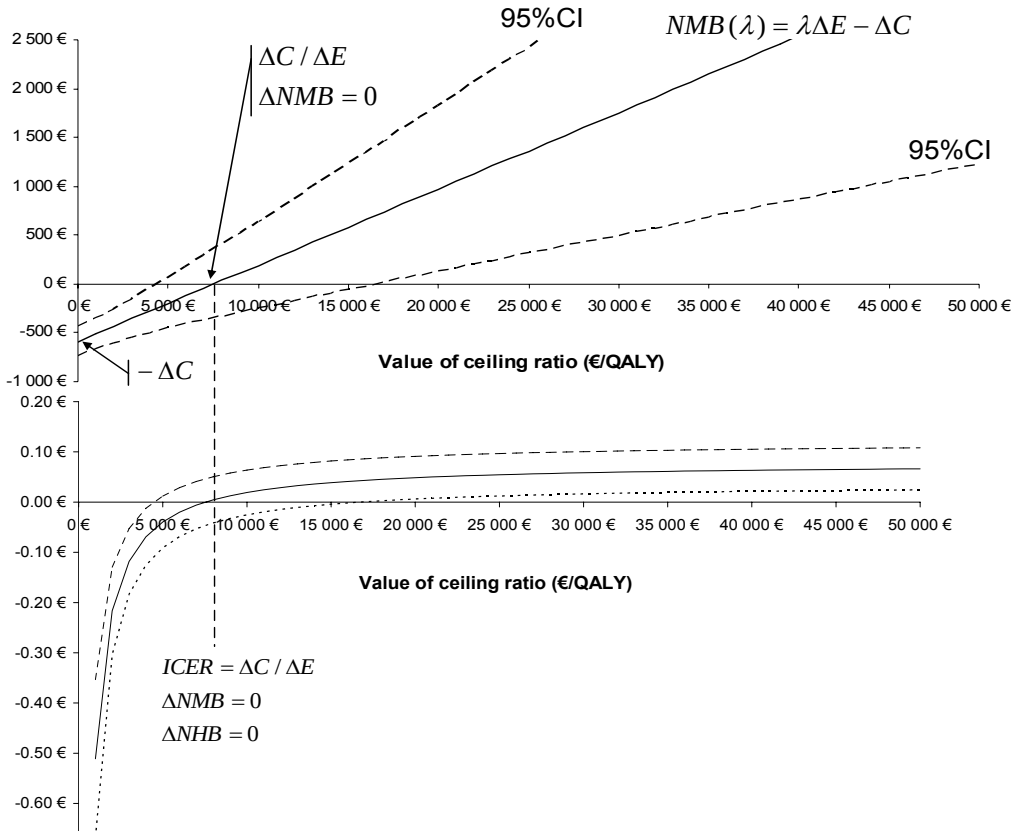


Figure 11. Net benefits on the cost (above) and effect (below) scales with the 95% confidence intervals for the data presented in figure 10 (the design of figure is adapted from Briggs 2001, 198).

In contrast to the ratio sampling distribution, the sampling distributions of $\Delta\text{NHB}(\lambda)$ or $\Delta\text{NMB}(\lambda)$ are continuous, and therefore the derivation of confidence intervals is also somewhat more straightforward when the net benefit approach is adopted. For net benefits, a (1-α)% confidence interval can be determined in the standard fashion as follows:

[7] $$\text{NB}(\lambda) \pm \text{z}_{\alpha/2}\sqrt{\sigma^2_{\text{NB}}}$$

, where NB(λ) is the estimated net benefit measure with variance σ^2_{NB}, and $z_{\alpha/2}$ is the critical value from the standard normal distribution. The variance for net-benefit measure can be defined as follows:

[8] $$\text{var}\left(\Delta\text{NHB}(\lambda)\right) = \text{var}\left(\Delta\text{E}\right) + \frac{1}{\lambda^2}\text{var}\left(\Delta\text{C}\right) - \frac{2}{\lambda}\text{cov}\left(\Delta\text{E}, \Delta\text{C}\right)$$

in terms of the net health benefits, or:

[9] $$\text{var}\left(\Delta\text{NMB}(\lambda)\right) = \lambda^2\text{var}\left(\Delta\text{E}\right) + \text{var}\left(\Delta\text{C}\right) - 2\lambda\text{cov}\left(\Delta\text{E}, \Delta\text{C}\right)$$

for the net health monetary benefit measure.

The decision rules in the net benefit framework are defined as follows:

[10] $$\begin{cases}\text{if } \Delta\text{NB}(\lambda) > 0, \text{ then prefer } \text{T}_1 \\ \text{if } \Delta\text{NB}(\lambda) < 0, \text{ then prefer } \text{T}_0\end{cases}$$

Thus, if the defined confidence intervals of incremental net benefits exclude 0 in the selected value of λ, then the decision makers should prefer T_1 over T_0. The use of 95% confidence intervals has been illustrated in Figure 11 (Briggs 2001)

As mentioned above the use of average ICER estimates are theoretically misleading (Stinnett & Paltiel 1997). Net benefit statistics on the other hand have the useful property:

[11] $$\begin{aligned}\Delta NMB(\lambda) &= \lambda\Delta E - \Delta C \\ &= \lambda(E_{T1} - E_{T0}) - (C_{T1} - C_{T0}) \\ &= (\lambda E_{T1} - C_{T1}) - (\lambda E_{T0} - C_{T0}) \\ &= NMB(\lambda)_{T1} - NMB(\lambda)_{T0}\end{aligned}$$

4.3 Bayesian methods for cost-effectiveness analysis

4.3.1 Basic concepts of Bayesian approach

There are good textbooks introducing the formalities used in Bayesian statistics (see e.g. Gelman et al. 1995, Parmigiani 2002, Spiegelhalter et al. 2004). However, a brief overview of the nature and principles underlying Bayesian methods is given before their use in the cost-effectiveness analysis context is considered.

The key element of Bayesian inference is the concept of subjective probability, where statements involving the use of probability are taken to represent *a degree of belief* about the event of interest[5]. The subjective interpretation of probability removes the need to associate probability with observable events, which allows one to make quantitative judgements about the likelihood of an assertion being correct in circumstances where there is no reasonable long-run frequency interpretation (i.e. decision-makers can approximate the probability that a decision that they are making is optimal). (Gelman et al. 1995)

The Bayesian approach makes a much wider use of probability distributions than traditional statistical methods. In the Bayesian statistics, all unknown parameters of interest (θ) are treated as *random variables* (rather than unknown constants) with probability distributions *p()*, which are used to describe the state of our knowledge about unknown parameters θ and the plausibility of different observed values of θ. (Gelman et al. 1995)

The full Bayesian setup requires one to define a specific parametric form for the unknown prior parameters (Gelman et al. 1995). In particular the shapes of probability distributions have a central role in Bayesian statistics because they are intended to represent the nature of parameters and the plausibility of different mean values (Briggs et al. 2006). For example, the sampling distributions of treatment costs are usually highly skewed to the right due to fact that only a few patients have high total costs and, in these cases, a gamma or a log-normal distribution can be fitted for mean cost data (cf. Hallinen et al. 2006, Nixon & Thompson 2004).

Bayesian inference about unknown parameters of interest is derived using Bayes theorem to condition on the values of observed quantities. One of the key advantages of Bayesian approach is that the Bayes' theorem allows one to explicitly incorporate prior evidence into the inference about some quantity of interest. This prior evidence is expressed in a mathematical form as a prior probability distribution

[5] Traditionally, when a random event is repeated a large number of times independently and under identical conditions, the probability of an event is approached by estimating its relative frequency of occurrence.

(i.e. the prior distribution is a quantification of the current state of understanding about the unknown quantity of interest). Prior information is then synthesized with the information in the collected data (also called likelihood) to produce the posterior distribution, which expresses one's belief about the value of the unknown parameter of interest after seeing the data. (Gelman et al. 1995)

The mathematical mechanism for evidence synthesis in Bayesian statistical inference can be defined as follows:

[12] $$p(\theta|y_1,...,y_n) = \frac{p(y_1,...,y_n|\theta)p(\theta)}{p(y_1,...,y_n)}$$

$p(\theta)$ is the prior distribution of the unknown quantity of interest, which describes our state of knowledge about θ before seeing any additional data. $P(y_1,\ldots,y_n|\theta)$ is the likelihood function (i.e. collected data conditional to prior evidence). $P(y_1,\ldots,y_n)$ is the marginal distribution of the data and it can be derived from $p(y_1,\ldots,y_n|\theta)$ and $p(\theta)$ by integrating the numerator over the support of θ. $P(\theta|y_1,\ldots,y_n)$ is the posterior distribution, which describes our state of knowledge about θ after observing additional data $y_1,\ldots,y_n$. (Gelman et al. 1995)

Figure 12 illustrates how the Bayes theorem can be used to combine two information sources in the hypothetical case of ΔNB(λ). It can be seen in Figure 12, that the posterior distribution is a compromise between the prior distribution and the given data (the likelihood distribution). For large samples, however, the impact of the choice of prior probability will diminish and the location of the posterior distribution is controlled by the given data (i.e. the posterior distribution will move towards the expected mean value of given data when the sample size increases). The increasing number of observed cases decreases the variance around the posterior mean and therefore, it is more precise than the prior mean alone. However, it should be noticed that in particular situations (e.g. when prior and likelihood evidence are markedly diverging from each other), the posterior variance can be similar to or even larger than the prior variance. According to the Bayesian philosophy, any inference a researcher desires to make is derived from the posterior distribution. (Gelman et al. 1995)

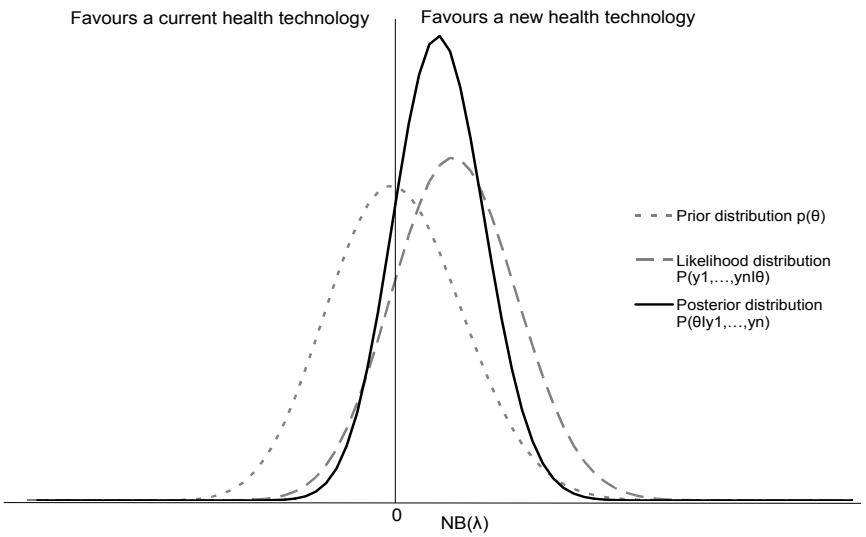


Figure 12. Prior distribution and evidence from the new data are synthesized to produce the posterior distribution

4.3.2 Cost-effectiveness acceptability curves

The net benefit approach to the cost-effectiveness analysis (as described in section 4.2.3) offers a potential solution to the problems related to the ICER approach. As considered above, T_1 can be implemented if $\Delta NB(\lambda) > 0$ and its confidence interval excludes 0. In the Bayesian framework, it is also possible to estimate the posterior probability that the incremental net benefit will be positive $p(\Delta NB(\lambda) > 0)$ (i.e. probability that T_1 is cost-effective, given the available evidence, for a range of λ values).

The probability of cost-effectiveness can be presented in the form of a cost-effectiveness acceptability curve (CEAC) as introduced originally by Van Hout et al. (1994). Van Hout et al. (1994) presented the use of CEACs in the context of frequentist statistics but a number of authors, however, have pointed out that the probabilistic interpretation of CEACs is only possible in the Bayesian framework (O'Hagan et al. 2000, Fenwick et al. 2001).

In practice, the probability of cost-effectiveness is estimated based on the posterior distribution of the incremental net benefits and the CEAC is constructed by plotting the relative proportion of the incremental net benefits that are positive (i.e. ΔNB(λ)>0) for a range of λ values. Further details of the methods for deriving CEACs for decisions involving two or more treatments are provided elsewhere (Fenwick et al. 2001, Fenwick et al. 2004, Fenwick et al. 2006).

The application of the Bayesian principles in evidence updating in the net benefit and CEACs context has been illustrated by Briggs 2001. If the prior distribution of the incremental net benefits is assumed to be vague (i.e. the prior distribution is uniform and prior ignorance is assumed) the results of a Bayesian analysis are similar to the results of a traditional analysis, since both analyses are based on the likelihood for the data. The interpretation of results, however, is much more flexible and coherent from the Bayesian point of view.

Figure 13 illustrates a posterior CEAC plotted under the assumption of a vague prior for the data presented in figure 10. Figure 13 shows that in this hypothetical example correspondence between the ICER estimate and the 50% point occurs because the distribution of ΔNB(λ) is asymptotically normally distributed. However, it might also possible that the 50% point on the CEAC will not correspond to the point estimate of the ICER due to a skewed distribution of ΔNB(λ).

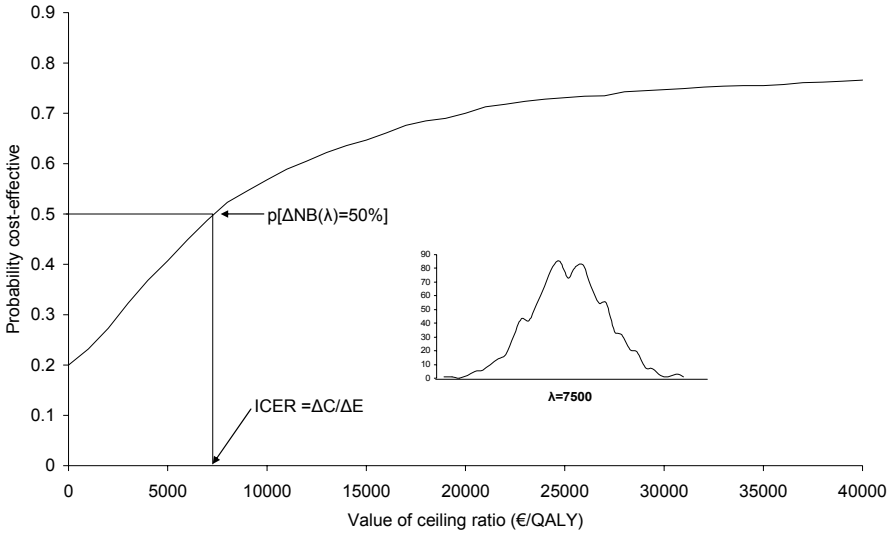


Figure 13. Posterior cost-effectiveness acceptability curve (CEAC) for the data presented in figure 10. The nested figure depicts the shape of the empirical distribution of ΔNB, when λ=7500 EUR/QALY.

The cost-effectiveness acceptability curve provides an indication of the uncertainty of T_1 being cost-effective in relation to T_0. However, it is important to point out that just because T_1 has a posterior probability of more than 50% of being cost-effective, the T_1 need not be the preferred choice. Depending on the shape of the posterior distribution (i.e. as mentioned above, the distribution of the ΔNB(λ) might not be symmetrically distributed), the point estimate for the ΔNB(λ) of T_1 may be positive for a value of λ, while the posterior probability of the treatment being cost-effective is less than 50% (Fenwick et al. 2001). The cost effectiveness acceptability frontier (CEAF) has been proposed to demonstrate this

phenomenon (Fenwick et al. 2000, Fenwick et al. 2001). In a recent study, Barton et al. (2007) have discussed discrepancies and their causes between CEAC and CEAF in detail.

4.4 Bayesian methods for evidence synthesis

4.4.1 Identifying evidence for decision-analytic models

When a decision-analytic model is developed, the evidence requirements consist of information on probabilities, effect sizes, adverse events, adherence, resource items, unit costs, and utility values. The probabilities represent transitions between the decision-analytic model's health states conditional to treatment responses. When only baseline transition probabilities (i.e. the natural course of a disease) are available the treatment effects, such as relative risks (RR), are used to illustrate the effects on diseases. The occurrence of adverse events may be one of the main cost drivers from an economic viewpoint and they may also have negative effect on patients' quality of life. Resource use and health state utility values are related to particular health states and conditions in a decision-analytic model. (Fayers & Hand 1997, Drummond 1998, Baltussen et al. 1999, Revicki & Frank 1999, Backhouse 2002, Claxton et al. 2002).

Since the overall validity of decision-analytic models depends on the quality of evidence put into the models, it is critical to identify and assess the sources and quality of model inputs to make the models more credible. Developed assessment methods, such as a hierarchy of data sources introduced by Cooper et al. (2005), offer applicable approach to identify and assess the quality of evidence according to their sources (Table 3). As shown in Table 3, data sources are ranked on an increasing scale from 1 to 6, where the most appropriate source is assigned a rank of 1. The potential biases of different data sources are discussed in detail elsewhere (e.g. see Nuijten 1998, Evans & Crawford 2000, Jordan & Lau 2003, Weinstein et al. 2003).

Table 3. Hierarchy of data sources for decision-analytic models (adapted and modified from Cooper et al. 2005)

Rank	Model input
	Clinical effect size, adverse events, and complications
1+	Meta-analysis of RCTs with direct comparison between comparator therapies, measuring final outcomes
1	Single RCT with direct comparison between comparator therapies, measuring final outcomes
2+	Meta-analysis of RCTs with direct comparison between comparator therapies, measuring surrogate outcomes Meta-analysis of placebo-controlled RCTs with similar trial populations, measuring the final outcomes for each individual therapy
2	Single RCT with direct comparison between comparator therapies, measuring the surrogate outcomes Single placebo-controlled RCTs with similar trial populations, measuring the final outcomes for each individual therapy
3+	Meta-analysis of placebo-controlled RCTs with similar trial populations, measuring the surrogate outcomes
3	Single placebo-controlled RCTs with similar trial populations, measuring the surrogate outcomes for each individual therapy
4	Case control or cohort studies
5	Non-analytic studies (e.g. case reports, case series)
6	Expert opinion
	Resource use
1	Prospective data collection or analysis of reliable administrative data for specific study
2	Recently published results of prospective data collection or recent analysis of reliable administrative data: same jurisdiction
3	Data without source specifications from previous economic evaluations: same jurisdiction
4	Recently published results of prospective data collection or recent analysis of reliable administrative data: different jurisdiction
5	Data source not known: different jurisdiction
6	Expert opinion
	Unit costs
1	Cost calculations based on reliable databases or data sources conducted for specific study: same jurisdiction
2	Recently published cost calculations based on reliable databases or data course: same jurisdiction
3	Data source not known: same jurisdiction
4	Using charge (price) rather than cost when societal perspective was adopted
5	Recently published cost calculations based on reliable databases or data sources: different jurisdiction
6	Data source not known: different jurisdiction
	Utilities
1	Direct utility assessment for the specific study from a sample either: (a) of the general population, or (b) with knowledge of the disease(s) of interest, or (c) of patients with the disease(s) of interest Indirect utility assessment from specific study from patient sample with disease(s) of interest, using a tool validated for the patient population
2	Indirect utility assessment from a patient sample with disease(s) of interest, using a tool not validated for the patient population
3	Direct utility assessment from a previous study from a sample either: (a) of the general population, or (b) with knowledge of the disease(s) of interest, or (c) of patients with the disease(s) of interest Indirect utility assessment from previous study from patient sample with disease(s) of interest, using a tool validated for the patient population
4	Data source not known: method of elicitation unknown
5	Patient preference values obtained from a visual analogue scale
6	Delphi panels, expert opinion

Based on the hierarchy of data sources, it is clear that the meta-analysis of RCTs is the most recommended method to summarize available clinical evidence. However, before a meta-analysis can be conducted, a systematic literature review is needed to identify all relevant studies. Major steps for preparing and performing the systematic literature review can be outlined briefly as follows (Higgins & Green 2006):

1. *Formulating a problem*
   The review objects are clarified on the basis of the decision problem defined in the beginning of the evaluation process and a research protocol is written.
2. *Locating and selecting studies*
   Comprehensive and documented search using a range of bibliographical databases and separate scientific journals is implemented. Methods to design and conduct a systematic literature review are documented in detail elsewhere (Higgins & Green 2006).
3. *Critical appraisal of studies and collecting data*
   Selected articles are read to assess their validity and the relevant information is then extracted from those selected articles.
4. *Estimating the effect size and its precision*
   The selected study results are pooled quantitatively using meta-analysis methods to obtain an overall effect size (Sutton et al. 2004). The exploration of issues that may affect the pooled results, such as the variability of effect sizes between studies (Thompson & Higgins 2002) and the imputation of missing values (Furukawa et al. 2006) have central roles in this part of the meta-analysis. Sections 4.4.3 and 4.4.4 introduce briefly the methods to estimate a pooled effect size using meta-analysis and explore the variability of effect sizes between studies in the Bayesian context.

4.4.2 Incorporating the quality of evidence into meta-analyses

Since evidence from the high quality data sources is assumed to be more valid than evidence from other sources, the use of quality scores to weigh evidence based on the features of particular data source has been proposed (Spiegelhalter & Best 2003). Explicit quality assessment scales, however, are mainly available for RCTs and less for other data sources. Therefore, the quality assessments are usually done for evidence that relates to clinical effect sizes. (Moher et al. 1995, Moher et al. 1996)

The quality scales that provide a summary numeric score of quality can be formally applied to weigh evidence used in the decision-analytic models. The scores can be used to summarise the quality of evidence e.g. in terms of pertinence, validity, and precision after the critical appraisal of the selected data sources (Tan et al. 2003). The case study described in chapter 6.2 will provide an example about the incorporation of the quality of clinical evidence into a decision-analytic model.

4.4.3 Synthesising evidence using meta-analysis

Meta-analysis is methodology for quantitatively synthesizing the results from the systematic literature review to obtain an estimate of the summary effect size across trials and is well suited to improving the quality of data used in the decision-analytic models. In meta-analysis, the effect sizes are measured in terms of quantitative units, which can be classified as belonging to one of three data types: binary (e.g.

yes/no), continuous (e.g. change in LDL-C), or ordinal scale (e.g. a disease severity scale). (Sutton et al. 2001).

Two models that are most commonly used in the meta-analysis of study results are the fixed-effects and random-effects models. The fixed-effects approach is based on the assumptions that each observed study results is estimating a common unknown overall pooled effect, whereas the random-effects approach assumes that each single observed study results is estimating its own unknown underlying effect, which in turn is estimating a common population mean. (Sutton & Abrams 2001) The difference between fixed- and random-effects approaches is illustrated graphically in Figure 14.

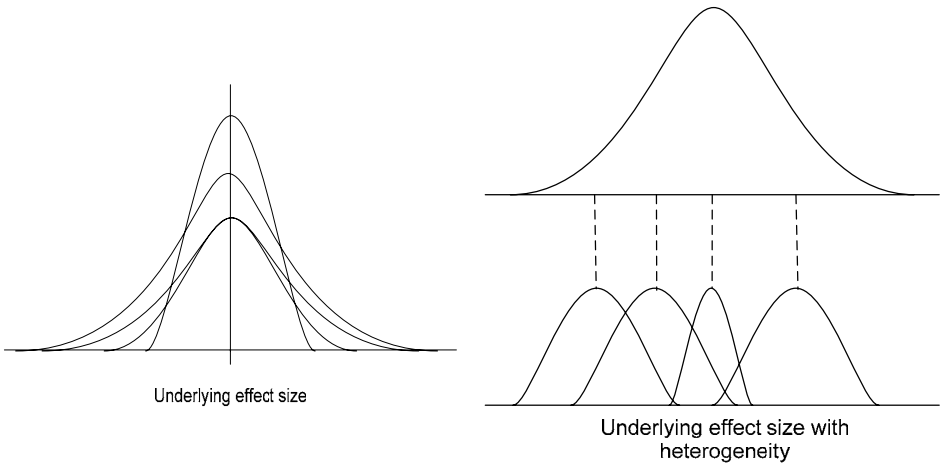


Figure 14. Fixed- (on the left-hand side) versus random-effects (on the right-hand side) approaches to meta-analysis (adapted from Normand 1999). Fixed effects estimates assume that there is a singly underlying population treatment effect, which will be reflected most accurately by larger studies with more statistical power (i.e. studies with infinitely large sample size would yield an identical result), whereas random effects estimates take into account the heterogeneity among studies (i.e. individual studies are assumed to estimate different underlying treatment effects, which in turn are estimating a common population mean).

When meta-analysis is used to combine effect sizes, an inverse variance-weighted method is usually applied. In the fixed-effect model, each study (*i*) result (y_i) is weighted according to the inverse of its variance ($1/v_i$) before the weighted average estimate ($\hat{\theta}_{ES}$) is estimated, since different studies estimate the true effect size with varying degrees of precision. The weighted average is estimated as follows:

[13] $$\hat{\theta}_{ES} = \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i}, \text{ where } w_i = 1/v_i \text{ and } i = 1,...,n$$

In using the random-effects approach, the weighted average estimate ($\hat{\theta}$) is estimated as above but the weights are given by $w_i = 1/(v_i + \tau^2)$, where τ^2 is an estimate of the between-study variance (i.e. the value of τ^2 is assumed to be zero, when the fixed effect approach is applied). (Schmid 2001)

The fixed-effects approach has been criticized as being unrealistic because medical studies are rarely or never identical replications of one another. Therefore, the random-effects approach has been suggested to reflect potential within- and between-trial (i.e. often termed heterogeneity in the meta-analysis literature) variation in the observed effect sizes due to differences in the trial designs, methods and patient characteristics. (Thompson 1993)

4.4.4 Hierarchical model structures in meta-analysis

The Bayesian approach offers a natural framework for the random-effects modelling. The random-effects meta-analysis model can be specified using a following hierarchical structure, which assumes normality *N(-,-)*:

[14]

$y_i \sim N(\theta_i, v_i)$	at level 1 (within-study variation)
$\theta_i \sim N(\mu, \tau^2)$	at level 2 (between-study variation)
$\mu \sim [-,-]$	Prior distribution for the average effect size of interest
$\tau^2 \sim [-,-]$	Prior distribution for the between-study variance

where [-,-] indicates a prior distribution to be specified (Schmid 2001, Sutton & Abrams 2001). To be fully Bayesian, prior distributions for the population parameters μ and τ^2 are needed; whereas, the values of v_i are assumed to be known in advance (i.e. the values are derived from each study *i*).

When normality is assumed, the prior distribution of the average effect size of interest is usually specified as $\mu \sim Norm(0, 0.000001)$ and the prior distribution of between-study variance is assumed to follow an inverse gamma distribution (i.e. $1/\tau^2 \sim gamma(0.001, 0.001)$). The inverse gamma distribution is the most frequently applied distribution for variance, since it produces a distribution, which is approximately uniform but has a spike of probability mass close to zero (i.e. the prior expectation before seeing the data is that there is no variation between studies). (Lambert et al. 2005) The further issues related to the specification of proper prior distributions have been discussed elsewhere (Smith et al. 1995, Sutton

& Abrams 2001, Lambert et al. 2005). After the specification of prior distributions, the aim is to discover the joint posterior distribution of the θ_i, μ and τ^2 given the data.

When the hierarchical specification for the random-effects models is used, a phenomenon, called *shrinkage*, emerges as an accessory product of this particular approach. This phenomenon implies that a better effect estimate for a single study (*i*) is gained by *borrowing strength* from other studies. This means that the posterior mean of the individual effect size (y_i) is based on the data from all other studies included in the meta-analysis. In practice, the posterior mean is a synthesis or "compromise" between the single observed effect size and the average treatment effect size. When there is little between-study variation in effect sizes, the posterior mean narrows down towards the average treatment effect size. Instead, when the variation is large the posterior mean stays closer to the single observed study effect. Ultimately, this leads to narrower credible intervals of any particular effect (cf. Figure 15). (Schmid 2001)

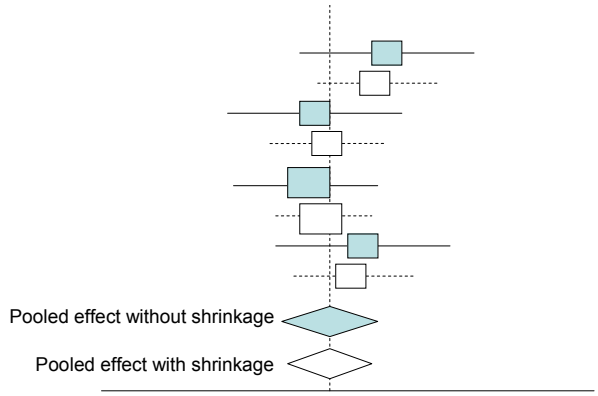


Figure 15. Shrinkage plot for hypothetical meta-analysis. Dotted lines depict the effect of shrinkage phenomenon.

4.4.5 Applying hierarchical linear models to explain heterogeneity in meta-analysis

The random-effects approach takes into account the variation in the underlying effect size between studies, but it does not provide a method of exploring or explaining the reasons why study results vary. In the case of meta-analysis, both the patient and the study characteristics can cause variation in the pooled results. Studies being pooled may differ with respect to many factors e.g. patient inclusion criteria, study design, the length of follow-up, treatment regimens or doses, etc. Therefore, it may be valuable to investigate subsets of patients and studies within the studies being pooled. In practice, the stratification of the studies can be undertaken according to study or patient characteristics. (Sutton et al. 2004)

Another way to explore variation, especially when the number of studies being combined is relatively large, is meta-regression, which can be seen as an extension of subgroup analyses. The meta-regression is the application of regression analysis, where study-level covariates (i.e. potential 'effect size modifiers') are used to explain the variability of treatment effects between studies (Thompson & Higgins 2002). These study-level summary results describe only between-study variation and therefore the meta-regression can only be used to detect study characteristics that differ across studies (Schmid et al. 2004).

The meta-regression can be arranged by representing the study means θ_i as functions of regression covariates X_i, when θ_i vary across studies as $\beta_0 + \beta_i X_i$. In meta-regression, more precise studies (i.e. studies with a larger sample size) have a greater influence in the analysis, because studies are weighted by the precision of their respective effect estimate (w_i). In addition, it is not reasonable to assume that all of the variation is explained by the study-level covariates, and therefore that w_i should be equal to the inverse sum of the within study variation and the residual between-trial variation (i.e. $w_i = 1/(v_i + \tau^2)$). The advantages and disadvantages of the meta-regression approach in meta-analysis have been discussed in detail elsewhere. (Thompson & Higgins 2002)

Formally, the meta-regression $y_i \sim N(\beta_0 + \beta_i X_i, v_i + \tau^2)$ can be evaluated using an extended hierarchical structure, where the aim is to discover the joint posterior distribution of the θ_i, β_0, β_i given the data:

[15]

$$
\begin{aligned}
&y_i \sim N(\theta_i, v_i) && \text{at level 1 (within-study variation)}\\
&\theta_i \sim N(\mu_i, \tau^2) && \text{at level 2 (between-study variation)}\\
&\mu_i = \beta_0 + \beta_i (X_i - \overline{X}) && \text{A linear predictor for the average effect of interest}\\
&\beta_i \sim [-,-] && \text{Prior distributions for the trial-level covariates}\\
&\tau^2 \sim [-,-]
\end{aligned}
$$

where [-,-] indicates again a prior distribution to be specified. If non-informative prior distributions are used, inference depends almost entirely on the data (Schmid 2001). The common effect parameter is replaced by a linear predictor $\beta_0 + \beta_i (X_i - \overline{X})$, where β_0 is the intercept term of the regression slope, β_i are the regression coefficients associated with trial-level covariates, and X_i is the trial-level covariate value from the *i*th trial. The trial-level covariates are centred around their own mean values $(X_i - \overline{X})$ to speed up convergence during the simulation process. The parameter τ^2 now represents the residual heterogeneity variance not explained by the trial-level covariates. The case study of chapter 6.3 will provide an example of the applications of meta-analysis and meta-regression in practice.

4.5 Incorporation of uncertainties into decision-analytic models

4.5.1 Computation routines for the two-stage approach

In the two-stage analysis, all model parameters are varied simultaneously, taking the parameter uncertainty of all model parameters into account at the same time. The two-stage analysis is commonly conducted using the Monte Carlo simulation, which is the most widely used method in situations, where closed form solutions are very complicated or impossible. The Monte Carlo simulation is based on the idea that fundamentally any probability distribution may be expressed in a cumulative form (i.e. a cumulative distribution function). A cumulative curve *y(x) = F(x)* is typically scaled from 0 to 1 on the vertical axis, with vertical axis values *(y)* representing the cumulative probability up to the corresponding horizontal axis value *(x)* (Figure 16). New random number from selected probability distribution can be obtained by drawing a random number *(y)* between 0 and 1 and then inverting $x = F^{-1}(y)$. Repeating this process a large number of times generates a random sample from the selected probability distribution (Briggs et al. 2002).

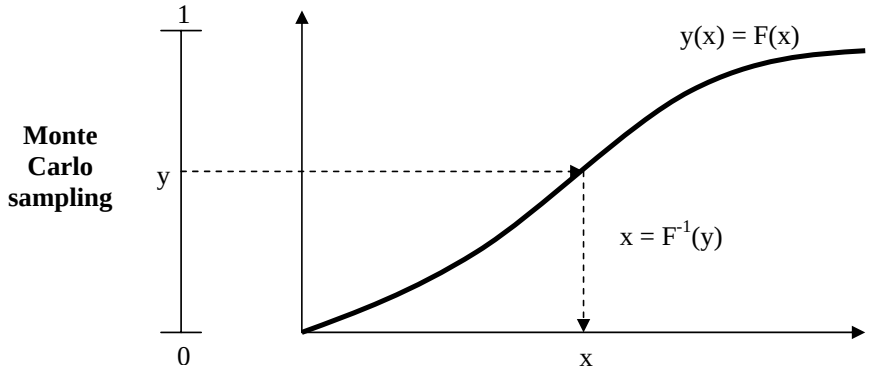


Figure 16. Generating random draws from a parametric distribution using Monte Carlo sampling (adapted from Briggs et al. 2002).

The selection of prior distributions assigned to model parameters is suggested to be based on the nature of the available data and the logical bounds of the parameter (e.g., transition probabilities are bounded on the interval [0, 1]). Due to the logical bounds, an assumption of normality for the model parameters is no longer valid in all cases, and transformations (Doubilet et al. 1985) or other types of distributions provided by Bayesian statistical theory are therefore needed (Briggs et al. 2006, Thompson & Nixon 2005). The parameterisation of the prior distributions can be done by applying the methods of moments (Briggs et al. 2006).

After the prior distributions for all relevant parameters are fitted, probabilistic sensitivity analysis can be employed applying a cohort analysis approach (Sonnenberg & Beck 1993), which is arranged to simulate the values of interest, such as expected costs and effects for both treatment alternatives T_1 and T_0 as follows:

1. Values of θ_k are drawn from their prior distributions $p(\theta_k)$ describing the uncertainty surrounding the value of each θ_k. *k* is the number of probabilistic prior distributions in the specified model (cf. figure 2).
2. Values of θ_k are held constant, while hypothetical cohorts of patients are run through the model according to the initial state vector and the specified transition matrix (see chapter 4.1).
3. Steps 1-2 are repeated a large number of times to generate empirical distributions for the overall expected costs and effects.

Consequently, the decision-analytic model is evaluated by averaging parameters of interest over a large number of iterations, which allows one to account for uncertainty in the model parameters. The simulation results can be interpreted and illustrated using the methods reviewed in chapter 4.2 and section 4.3.2.

4.5.2 Computation routines for the comprehensive decision modelling (MCMC) approach

Bayesian Markov Chain Monte Carlo (MCMC) methods have been applied earlier in situations where no closed form solutions are available. In addition, MCMC simulation methods offer a natural framework to build comprehensive decision-analytic models. The computation routines needed in Bayesian MCMC simulations are available e.g. in the software package WinBUGS, developed by the MRC Biostatistics Unit, University of Cambridge, UK.

WinBUGS applies one particular MCMC simulation method known as Gibbs sampling, which is a special case of the Metropolis-Hastings algorithm (Gilks et al. 1996). Gibbs sampling is a method for sampling from distributions over at least two dimensions and it is applicable in situations, where the joint distribution of model parameters is not known explicitly, but the conditional distribution of each parameter is known.

The concept of Gibbs sampling is illustrated graphically in Figure 17, where sampling is done from the joint posterior distribution of θ_1 and θ_2 (a). Starting from an initial point (θ_1^0, θ_2^0), a sample θ_1^1 is made from the conditional distribution $p(\theta_1 | \theta_2^0)$ (b). The subsequent sample is drawn from the conditional distribution $p(\theta_2 | \theta_1^1)$ (c), which after the initial point moves to the new state (θ_1^1, θ_2^1). This movement completes the iteration cycle (d).

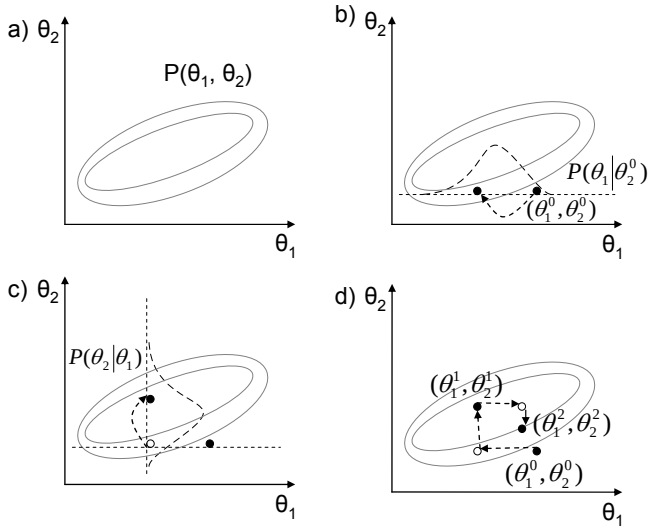


Figure 17. Simplified graphical example of Gibbs sampling procedure (adapted and modified from Mackay 2005, 370)

After model specification, Gibbs sampling can be conducted by applying the following procedure:

- Set initial values $\theta^0 = (\theta_1^0, \theta_2^0, ..., \theta_k^0)$, where k is the number of probabilistic distributions in the specified model (cf. figure 2). The value of θ_k can be any reasonable value from the parameter space.
- Repeat for j = 1,2,…,m, where m is the total number of samples. A complete nested loop (j) drawn from the k distributions is given by:

$$\theta_1^j \sim p(\theta_1 \mid \theta_2^{j-1}, \theta_3^{j-1}, \dots, \theta_k^{j-1})$$
$$\theta_2^j \sim p(\theta_2 \mid \theta_1^j, \theta_3^{j-1}, \dots, \theta_k^{j-1})$$
$$\theta_3^j \sim p(\theta_3 \mid \theta_1^j, \theta_2^j, \dots, \theta_k^{j-1})$$
$$\vdots$$
$$\theta_{k-1}^j \sim p(\theta_{k-1} \mid \theta_1^j, \theta_2^j, \theta_3^j, \dots, \theta_k^{j-1})$$
$$\theta_k^j \sim p(\theta_k \mid \theta_1^j, \theta_2^j, \theta_3^j, \dots, \theta_{k-1}^j)$$

Repeating samples a sufficient number of times ensures that the Gibbs sampler is truly sampling from the true conditional posterior distributions (i.e. convergence is reached) and that all areas of the joint posterior distribution are explored. If the Gibbs sampler is allowed to run for a sufficiently long period after convergence, it produces a complete sample from the conditional posterior distribution of θ_k.

Convergence can be checked formally e.g. in WinBUGS using the Gelman-Rubin diagnostic tool, which works by running several chains simultaneously, but starting from different initial values (θ_k^0). (Spiegelhalter et al. 2003) The simulations before convergence (i.e. burn-in samples) are discarded and final inferences are based on a sample of simulations after convergence. After achieving convergence, it is possible to return the expectations of interest, such as the posterior means of expected costs and effects, from particular conditional posterior distributions.

4.6 Estimating the value of additional evidence

Additional evidence is valuable because it reduces the expected costs of uncertainty around a decision. The expected costs of uncertainty can be determined jointly by the probability that a decision based on the existing information will be wrong (i.e. a wrong option is selected) and the consequences of that wrong decision. In health economic evaluation, the consequences associated with uncertainty are measured in the terms of NB(λ) foregone when the decision made upon the basis of existing evidence is incorrect. Expressing NB(λ) in monetary terms provides an explicit monetary valuation of the costs of uncertainty that can be compared with the cost of collecting additional evidence to determine the value of additional research. (Claxton 1999, Claxton et al. 2001)

Expected value of perfect information (EVPI) is a decision theoretic framework that provides useful information for decision-makers regarding uncertainty around the decision (Claxton 1999). Briefly, EVPI is a measure of the maximum value that can be placed on additional information and it is simply the difference between the expected value of the decision made with perfect information and the decision made on the basis of the available evidence.

Formally, when the decision given the available evidence is made, the decision-makers are faced with the fact that the value of NB(λ) will be uncertain and a decision must be made before it is known what particular values the uncertain model parameters (θ_k) in a decision-analytic model will take (cf. Figure 2). Therefore, the optimal decision given the existing information is to choose a treatment (T_i) that generates the maximum net benefit in monetary terms (i.e. $\max_{Ti} E_{\theta k} NB[T_i, \theta_k]$). With perfect information, it would be possible that decision-makers could select a treatment that maximises the net benefit for a particular value of θ_k. However, the true values of θ_k are unknown, so the expected value of perfect in-

formation is found by averaging the maximum net benefit over the joint distribution of θ_k (i.e. $E_\theta max_{Ti} NB(T_i, \theta_k)$). (Claxton 1999, Claxton et al. 2001)

When the simulation approach is employed, the calculation of EVPI is rather straightforward. As noted above, the appropriate decision given existing information is to choose the treatment with the highest average net benefit ($max_{Ti} E_{\theta k} NB[T_i, \theta_k]$). If we would have access to perfect information, we would choose the treatment with the highest net benefit at each iteration of the model (i.e. each iteration is assumed to represent the actualisation of a particular state of world). Determining the optimal strategy at each iteration and calculating the average net benefit from these choices gives the maximum net benefit with perfect information ($E_\theta max_{Ti} NB(T_i, \theta_k)$). The expected value of perfect information for an individual patient is therefore the average net benefit with perfect information minus the average net benefit of the treatment chosen given the available evidence (Claxton 1999, Claxton et al. 2001):

[16] $$EVPI = E_\theta max_{Ti} NB(T_i, \theta_k) - max_{Ti} E_\theta NB(T_i, \theta_k)$$

Since health care allocation decisions are always related to number of patients who could benefit from a new technology, it is worthwhile to calculate the EVPI for population based on some assessment of the effective life-time of that particular health technology, the expected number of patients over this period (I_t) and a discount rate (r).

[17] $$\text{Population EVPI} = EVPI * \sum_{t=1}^{t} \frac{I_t}{(1+r)^t}$$

If the population EVPI is low, obtaining more information is not very meaningful for the decision-makers as the uncertainty around the optimal decision given the existing information is small. On the other hand, when the population EVPI is high, it is potentially meaningful to require more evidence as this will add to what we know about the decision problem. Hence, when the population EVPI is high, substantial uncertainty surrounds the decision on financing of the treatments from the public funds. (Claxton 1999, Claxton et al. 2001) It should remember that the values of population EVPI analyses are only really important when compared to cost of additional research.

4.7 References

Agro K, Bradley C, Mittmann N, Iskedjian M, Ilersich AL & Einarson T. 1997. Sensitivity analysis in health economic and pharmacoeconomic studies. An appraisal of the literature. Pharmacoeconomics 11(1): 75-88.

Anderson J, Bush J, Chen M & Dolenc D. 1986. Policy space areas and properties of benefit-cost/utility analysis. Journal of the American Medical Association 255(6): 794-795.

Arrow K & Lind R. 1970. Risk and uncertainty. Uncertainty and the evaluation of public investment decisions. American Economic Review 60: 364-378.

Backhouse M. 2002. Use of randomised controlled trials for producing cost-effectiveness evidence. Potential impact of design choices on sample size and study duration.
Pharmacoeconomics 20(15): 1061-1077.

Bala M & Mauskopf J. 2006. Optimal assignment of treatments to health states using a Markov decision model. A introduction to basic concepts. Pharmacoeconomics 24(4): 345-354.

Baltussen R, Leidl R & Ament A. 1999. Real world designs in economic evaluation. Bridging the gap between clinical research and policy-making. Pharmacoeconomics 16(5): 449-458.

Banta H & Thacker S. 1990. The case for reassessment of health care technology. Once is not enough. Journal of the American Medical Association 264(2): 235-240.

Barton G, Briggs A & Fenwick E. 2007. Optimal decisions: the role of the CEAC, CEAF, and EVPI. Value in Health. *in press*.

Black W. 1990. The CE plane: a graphic representation of cost-effectiveness. Medical Decision Making 10: 212-214.

Bojke L, Claxton K, Palmer S & Sculpher M. 2006. Defining and characterising structural uncertainty in decision analytic models. CHE Research Paper 9, University of York.

Brennan A, Chick S & Davies R. 2006. A taxonomy of model structures for economic evaluation of health technologies. Health Economics 15: 1295-1310.

Briggs A. 2000. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 17(5): 479-500.

Briggs A. 2001. Handling uncertainty in economic evaluation and presenting the results. In M. Drummond and A. McGuire, ed. Economic evaluation in health care. Merging theory with practice. New York: Oxford University Press.

Briggs A, Goeree R, Blackhouse G & O'Brien B. 2002. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Medical Decision Making 22(4): 290-308.

Briggs A, Mooney C & Wonderling D. 1999. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. Statistics in Medicine 18(23): 3245-3262.

Briggs A & Sculpher M. 1995. Sensitivity analysis in economic evaluation: a review of published studies. Health Economics 4(5): 355-71.

Briggs A & Sculpher M. 1998. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 13(4): 397-409.

Briggs A, Sculpher M & Buxton M. 1994. Uncertainty in economic evaluation of health care technologies: the role of sensitivity analysis. Health Economics 3: 95-104.

Briggs A, Sculpher M & Claxton K. 2006. Decision modelling for health economic evaluation. New York, Oxford University Press.

Brouwer W & Koopmanschap M. 2000. On the economic foundations of CEA. Ladies and gentleman, take your positions! Journal of Health Economics 19: 439-459.

Buxton M. 1997. The pros and cons of modelling in economic evaluation. OHE Briefing 33.

Buxton M, Drummond M, Van Hout B, Prince R, Sheldon T, Szucs T & Vray M. 1997. Modelling in ecomomic evaluation: An unavoidable fact of life. Health Economics 6(3): 217-227.

Cairns J. 2001. Discounting in economic evaluation. In Drummond M and McGuire A., ed. Oxford University Press. Economic evaluation in health care: merging theory with practice.

Chaudhary M & Stearns S. 1996. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomised trial. Statistics in Medicine 15: 1447-1458.

Claxton K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 18(3): 341-364.

Claxton K, Neumann P, Araki S & Weinstein M. 2001. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. International Journal of Technology Assessment in Health Care 17(1): 38-55.

Claxton K & Posnett J. 1996. An economic approach to clinical trial design and research priority-setting. Health Economics 5(6): 513-524.

Claxton K, Sculpher M & Drummond M. 2002. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). The Lancet 360 (9334): 711-715.

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

Cooper N, Coyle D, Abrams K, Mugford M & Sutton A. 2005. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. Journal of Health Service Research & Policy 10(4): 245-250.

Craig B & Black M. 2001. Incremental cost-effectiveness ratio and incremental net-health benefit: two sides of the same coin. Expert Review of Pharmacoeconomics & Outcomes Research 1(1): 37-46.

Detsky A, Naglie G, Krahn M, Redelmeier D & Naimark D. 1997. Primer on Medical Decision Analysis: Part 2--Building a Tree. Medical Decision Making 17(2): 126-135.

Doubilet P, Begg C, Weinstein M, Braun P & McNeil B. 1985. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Medical Decision Making 5(2): 157-177.

Drummond M. 1998. Experimental versus observational data in the economic evaluation of pharmaceuticals. Medical Decision Making 18(2): s12-s18.

Drummond M, Barbieri M & Wong J. 2005. Analytic choices in economic models of treatments for rheumatoid arthristis: what makes a difference? Medical Decision Making 25: 520-533.

Drummond M, Sculpher M, Torrance G, O'Brien B & Stoddart G. 2005. Methods for the economic evaluation of health care programmes. New York: Oxford university press.

Eichler H-G, Kong S, Gerth W, Mavros P & Jönsson B. 2004. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value in Health 7(5): 518-528.

Elbasha E & Messonnier M. 2004. Cost-effectiveness analysis and health care resource allocation: decision rules under variable returns to scale. Health Economics 13(1): 21-35.

Evans C & Crawford B. 2000. Expert Judgement in Pharmacoeconomic Studies. Guidance and Future Use. Pharmacoeconomics 17(6): 545-553.

Fayers P & Hand D. 1997. Generalisation from phase III clinical trials: survival, quality of life, and health economics. The Lancet 350: 1025-27.

Fenwick E, Claxton K & Sculpher M. 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Economics 10(8): 779-87.

Fenwick E, Claxton K, Sculpher M & Briggs A. 2000. Improving the efficiency and relevance of health technology assessment: the role of iterative decision analytic modelling. York: Centre for Health Economics.

Fenwick E, O'Brien B & Briggs A. 2004. Cost-effectiveness acceptability curves - facts, fallacies and frequently asked questions. Health Economics 13(5): 405-415.

Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K & Sutton A. 2006. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Medical Decision Making 26(5): 480-496.

Furukawa T, Barbui C, Cipriani A, Brambilla P & Watanabe N. 2006. Imputing missing standard deviations in meta-analyses can provide accurate results. Journal of Clinical Epidemiology 59: 7-10.

Garber A & Phelps C. 1997. Economic foundations of cost-effectiveness. Journal of Health Economics 16: 1-31.

Gelman A, Carlin J, Stren H & Rubin D. 1995. Bayesian data analysis. London, Chapman & Hall.

Gilks W, Richardson S & Spiegelhalter D. 1996. Markov Chain Monte Carlo in practice. Suffolk, Chapman & Hall.

Groot Koerkamp B, Hunink M, Stijnen T, Hammitt J, Kuntz K & Weinstein M. 2007. Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. Medical Decision Making 27(2): 101-111.

Hallinen T, Martikainen J, Soini E, Suominen L & Aronkytö T. 2006. Direct costs of warfarin treatment among patients with atrial fibrillation in a Finnish health care setting. Current Medical Research & Opinion 22(4): 683-692.

Heitjan D, Moskowitz A & Whang W. 1999. Problems with interval estimates of the incremental cost-effectiveness ratio. Medical Decision Making 19: 9-15.

Higgins J & Green S. 2006. Cochrane handbook for systematic reviews of interventions 4.2.5 [updated May 2005].

Hjelmgren J, Berggren F & Andersson F. 2001. Health economic guidelines - similarities, differences and some implications. Value in Health 4(3): 225-250.

Howard R. 1968. The foundations of decision analysis. IEEE transactions on system science and cybernetics 4(3): 211-219.

Iglesias C & Claxton K. 2006. Comprehensive decision-analytic model and Bayesian value-of-information analysis. Pentoxifylline in the treatment of chronic venous leg ulcers. Pharmacoeconomics 24:465-478.

Johannesson M & Weinstein M. 1993. On the decision rules of cost-effectiveness analysis. Journal of Health Economics 112: 459-467.

Jordan H & Lau J. 2003. Linking pharmacoeconomic analyses to results of systematic review and meta-analysis. Expert Review of Pharmacoeconomics Outcomes Research 3(4): 441-448.

Karlsson G & Johannesson M. 1996. The decision rules of cost-effectiveness analysis. Pharmacoeconomics 9: 113-120.

Karnon J. 2003. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. Health Economics 12: 837-848.

Laking G, Price P & Sculpher M. 2002. Assessment of the technology for functional imaging in cancer. European Journal of Cancer 38: 2194-2199.

Lambert P, Sutton A, Burton P, Abrams K & Jones D. 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine 24: 2401-2428.

Laska EM, Meisner M, Siegel C & Stinnett A. 1999. Ratio-based and net benefit-based approaches to health care resource allocation: proofs of optimality and equivalence. Health Economics 8(2): 171-4.

Lord J, Laking G & Fischer A. 2006. Non-linearity in the cost-effectiveness frontier. Health Economics 15(6): 565-577.

Mackay D. 2005. Information theory, inference, and learning algorithms, Cambridge University Press.

Moher D, Jadad A & Tugwell P. 1996. Assessing the quality of randomised controlled trials. Current issues and future directions. International Journal of Technology Assessment in Health Care 12(2): 195-208.

Moher D, Jadad AR, Nichol G, Penman M, Tugwell P & Walsh S. 1995. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. Controlled Clinical Trials 16(1): 62-73.

Nixon R & Thompson S. 2004. Parametric modelling of cost data in medical studies. Statistics in Medicine 23(8): 1311-1331.

Normand S-L. 1999. Meta-analysis: formulating, evaluating, combining, and reporting. Statistics in Medicine 18: 321-359.

Nuijten M. 1998. The selection of data sources for use in modelling studies. Pharmacoeconomics 13(3): 305-316.

O'Brien B, Gertsen K, Willan A & Faulkner A. 2002. Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Economics 11(2): 175-180.

O'Brien BJ, Drummond MF, Labelle RJ & Willan A. 1994. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Medical Care 32(2): 150-63.

O'Hagan A, Stevens J & Montmartin J. 2000. Inference for the cost-effectiveness acceptability curve and cost-effective ratio. Pharmacoeconomics 17(4): 339-349.

Parmigiani G. 2002. Modeling in medical decision making. A Bayesian approach, West Sussex: John Wiley & Sons.

Pauker S & Kassier J. 1987. Decision analysis. New England Journal of Medicine 316(5): 250-258.

Philips Z, Bojke L, Sculpher M, Claxton K & Golder S. 2006. Good practice guidelines for decision-analytic modelling in health technology assessment. A review and consolidation of quality assessment. Pharmacoeconomics 24(4): 355-371.

Polsky D, Glick H, Willke R & Schulman K. 1997. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Economics 6: 243-252.

Raiffa H. 1970. Decision analysis. Introductory lectures on choices under uncertainty, Addison-Wesley Publishing Company.

Revicki D & Frank L. 1999. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. Pharmacoeconomics 15(5): 423-434.

Russell L & Sisk J. 2000. Modeling age differences in cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 16(4): 1158-1167.

Sackett D, Rosenberg W, Gray J, Haynes R & Richardson W. 1996. Evidence-based medicine: what it is and what it isn't. BMJ 312: 71-2.

Schmid C. 2001. Using Bayesian inference to perform meta-analysis. Evaluation & Health Professions 24(2): 165-189.

Schmid CH, Stark PC, Berlin JA, Landais P & Lau J. 2004. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. Journal of Clinical Epidemiology 57(7): 683-697.

Sculpher M, Claxton K & Akehurst R 2005. It's just evaluation for decision-making: recent developments in, and challenges for, cost-effectiveness research. In Smith P, Ginnelly L and Sculpher M, ed. Health policy and economics: opportunities and challenges. Maidenhead: Open University Press.

Sculpher M, Claxton K, Drummond M & McCabe C. 2006. Whither trial-based economic evaluation for health care decision making? Health Economics 15: 677-687.

Sculpher M, Drummond M & Buxton M. 1997. The iterative use of economic evaluation as part of process of health technology assessment. J Health Serv Res Policy 2(1): 26-30.

Smith T, Spiegelhalter D & Thomas A. 1995. Bayesian approaches to random-effects meta-analysis: a comperative study. Statistics in Medicine 14: 2685-2699.

Sonnenberg F & Beck J. 1993. Markov models in medical decision making: a practical guide. Medical Decision Making 13: 322-338.

Spiegelhalter D, Abrams KE & Myles J. 2004. Bayesian approaches to clinical trials and health-care evaluation. West Sussex: John Wiley & Sons, Ltd.

Spiegelhalter D & Best N. 2003. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Statistics in Medicine 22(23): 3687-3709.

Spiegelhalter D, Myles J, Jones D & Abrams K. 2000. Bayesian methods in health technology assessment: a review. Health Technology Assessment 4(34).

Spiegelhalter D, Thomas A, Best N & Lunn D. 2003. WinBUGS user manual.

Stinnett A & Mullahy J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 18(2): S65-S80.

Stinnett A & Paltiel A. 1997. Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. Medical Decision Making 17(4): 483-9.

Sutton A & Abrams K. 2001. Bayesian methods in meta-analysis and evidence synthesis. Statistical Methods in Medical Research 10: 277-303.

Sutton A, Abrams K & Jones D. 2001. An illustrated guide to the methods of meta-analysis. Journal of Evaluation in Clinical Practice 7(2): 135-148.

Sutton A, Abrams K, Jones D, Sheldon T & Song F. 2004. Methods for meta-analysis in medical research. Chichester: John Wiley & Sons, Ltd.

Tambour M & Zethraeus N. 1998. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. Health Economics 7(2): 143-147.

Tambour M, Zethraeus N & Johannesson M. 1998. A note on confidence intervals in cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 14(3): 467-471.

Tan S-B, Dear K, Bruzzi P & Machin D. 2003. Strategy for randomised clinical trials in rare cancers. BMJ 327: 47-49.

Thompson S. 1993. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. Statistical Methods in Medical Research 2(2): 173-192.

Thompson S & Higgins J. 2002. How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine 21: 1559-1573.

Thompson S & Nixon R. 2005. How Sensitive Are Cost-Effectiveness Analyses to Choice of Parametric Distributions? Medical Decision Making 25(4): 416-423.

Van Hout B, Al M, Gordon G & Rutten F. 1994. Costs, Effects and C/E-ratios alongside a clinical trial. Health Economics 3: 309-319.

Vergel Y, Palmer S, Asseburg C, Fenwick E, de Belder M, Abrams K & Sculpher M. 2007. Is primary angioplasty cost effective in the UK? Results of a comprehensive decision analysis. Heart 93:1238-1243.

Weinstein M & Manning W. 1997. Theoretical issues in cost-effectiveness analysis. Journal of Health Economics 16: 121-128.

Weinstein M & Zeckhauser R. 1973. Critical ratios and efficient allocation. Journal of Public Economics 2: 147-157.

Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C & Luce BR. 2003. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices. Value in Health 6(1): 9-17.

Willan A & O'Brien B. 1996. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. Health Economics 5: 297-305.

5 AIMS OF THE STUDY

The general aim of this study was to develop the applications of decision-analytic models and to explore the applicability of a set of Bayesian methods for evidence synthesis and decision-analytic modelling in health economic evaluations. Specifically, this was done by:

1. Developing the applications of Markov models to reflect the courses of diseases and their expected costs and health outcomes in the presence of particular treatments under conditions of uncertainty.
2. Using a set of Bayesian methods to synthesize available evidence whilst reflecting on their imprecision, heterogeneity and quality.
3. Using optional approaches to incorporate parameter uncertainty into decision-analytic models.
4. Using different approaches to represent and interpret the cost-effectiveness results.
5. Using the value of information methods to estimate the expected costs of decision uncertainty if wrong decisions are made and to assess the value of additional evidence that would reduce decision uncertainty.

The first study question relates to challenges in developing a well-specified probabilistic decision-analytic model under conditions of methodological and structural uncertainty. The second question relates to challenges to synthesise evidence for model parameters, when uncertainty arises due to variability, heterogeneity, and the imprecision of estimates. The third study question relates to alternative approaches to capture parameter uncertainty in a decision-analytic model in such a way that decision uncertainty can be reflected and illustrated to decision-makers. The fourth study question relates to the additional possibilities to represent and interpret the probabilistic cost-effectiveness results. Finally, the fifth study question relates to challenges in defining the value of additional evidence that would reduce the uncertainty around the specific decision.

6 CASE STUDIES

6.1 Modelling the cost-effectiveness of a family-based program in mild Alzheimer's disease employing the two-stage approach[6]

6.1.1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease characterized by gradual loss and progressive deterioration of memory and other cognitive functions, and decline in functional capacity. The reported prevalence in Western countries varies between 0.5–3.0% in individuals aged 65–74 years, between 4.1-18.7% in the age group 75–84, and between 13.1–47.2% in persons aged over 85 years (Sulkava et al. 1985, Evans et al. 1989, Bachman et al. 1992, Skoog et al. 1993, Ott et al. 1995, Polvikoski et al. 2001). There are approximately 80 000 patients with moderate or severe dementia in Finland. The number of patients is expected to increase to 128 000 by the year 2030. AD imposes a heavy economic burden on the social and health care system. The economical impact of the disease is due to the fact that patients with AD utilize health services at higher rates than age-matched controls. There are few systematic studies on the costs of illness due to dementia in Finland, but the estimates have varied from 0.8 to 1.3 billion EUR per year. One recent study from Sweden suggests that the costs of dementia were 3.4 billion US dollars in the year 1996 with annual average costs of 23 600 US dollars per patient (Wimo et al. 1997). The total cost of caring for persons with severe AD has been estimated to be 2.25 times higher than for patients with mild or moderate disease (Hu et al. 1986). The largest increase in dementia costs occurs when the patient has to be institutionalized. However, international comparisons are difficult because of different methods of financing and organizing care. The decisions to connect an individual to institutional care vary significantly from country to country. In Finland, 40 to 50% of all demented patients are in long-term institutional care.

The clinical features of AD appear insidiously and develop according to a uniform pattern from the earliest symptoms to severe dementia and death. Initial symptoms are episodic memory loss and an impaired ability to learn new facts, reflecting the pathology in the entorhinal cortex and hippocampus (Braak & Braak 1991). As the disease progresses, other cognitive symptoms, such as executive dysfunction, aphasia, agnosia and apraxia appear, resulting in decline of functional capacities. Patients with moderate or severe dementia have many psychiatric and behavioral symptoms, which greatly increase caregiver burden and stress. Finally, the patient becomes bedridden and incontinent. Death occurs on average 6–8 years after the appearance of the initial symptoms (Jost & Grossberg 1995).

[6] This chapter has been published in: Martikainen J, Valtonen H & Pirttilä T. Potential Cost-effectiveness of a Family-Based Program in Mild Alzheimer's Disease Patients. Eur J Health Econom 2004;5:136-42. Reproduced with permission.

The introduction of new therapies for AD may result in significant economic, clinical and social ramifications. However, it is not clear how these treatment effects translate into quality of life of the patients and caregivers, and whether they can postpone the institutionalization of the patient. Extrapolating long-term effects from short-term clinical trials is difficult, although some studies do suggest that treatment can postpone nursing home care (Knopman et al. 1996). Interventions supporting community care have shown promising results in postponing institutionalization (Brodaty et al. 1997, Eloniemi-Sulkava et al. 2001, Mittelman et al. 1996).

6.1.2 Objectives

The objective of the present study was to examine ex ante by a simulation the long-term cost-effectiveness of a cognitive-behavioural family intervention (CBFI) program in helping the informal caregivers (spouses or adult children) to postpone the need to transfer AD patient to a nursing home. The actual ongoing CBFI program trial is designed to be an additional service for AD patients and their informal caregivers, and its results can later be compared to the simulation results in order to evaluate the applicability of the simulation method. Thus, the two alternative forms of treatment are the current practice, or the current practice combined with the CBFI program. The current practice consists different forms of community services (meals on wheels, cleaning services etc.) and periodical institutional care (1-2 weeks/period), while the informal caregivers are able to rest. The AD patients and their informal caregivers can obtain these services from the public or private sector, since also private sector services are national insurance schemes. The service-mix to be offered to the AD patient and his/her informal caregiver is decided based on *e.g.* their living conditions and current physical and mental health status. The overall goal is to make it possible that the AD patient can live at home as long it is reasonable. In the CBFI program, the main emphasis is on supporting the patients' and their caregivers' capabilities in coping with the disease. The CBFI program consists of short courses in rehabilitation centres with the comprehensive support of dementia family care coordinators. The courses include physical and recreational training for AD patients, and psychological as well as educational support and counselling for the caregivers.

The CBFI program is planned to be funded by the Finnish Social Insurance Institution. Thus, the study has been designed to obtain information about the average outcomes across the group of patients receiving the support services mentioned above and thereby to inform policy decisions at the societal level. Therefore, we are not particularly concerned with the variability of cost-effectiveness due to patients´ and their caregivers´ characteristics (so called first-order uncertainty), but rather with the lack of complete knowledge regarding the true values of incremental costs and effectiveness between different support service-mixes (so called second-order uncertainty). In a stochastic decision analysis, the effect of first-order uncertainty can be minimized by increasing the sample size and the lack of complete knowledge with respect to true values of model parameters can be characterized by a second-order

Monte Carlo simulation. (Stinnett & Paltiel 1997) Therefore, in this particular case, the second-order Monte Carlo simulation with a Bayesian approach was applied to answer the question: given existing information, should the new rehabilitation program be implemented as a cost-effective option?

6.1.3 Methods and data

Model inputs. A basic Markov model was derived from the publication of by Neumann et al. (1999) and the second-order simulation followed the study principles of Claxton et al. (2001). Neumann et al. (1999) applied a deterministic Markov model to illustrate the cost-effectiveness of donepezil-treatment for mild and moderate AD patients. Claxton et al. (2001) utilized the same data and model as Neumann et al. (1999) but they used a probabilistic approach to illustrate parametric uncertainty more accurately. They both used the AD model, where disease stages were classified as mild, moderate and severe, based on the Clinical Dementia Rating Scale (CDR) classification (Morris 1993). Transition probabilities, including stage-to-death transitions, were derived from Neumann et al. (1999) and Neumann et al. (2001) (Table 4). However, it should be noted that the probabilities from Neumann et al. (1999, 2001) articles do not sum up to 1 and therefore some modification was needed. Furthermore, the model was modified by excluding the transition from moderate state to mild, because that transition was assumed to be biologically implausible. A schematic presentation of the modified model is presented in Figure 18.

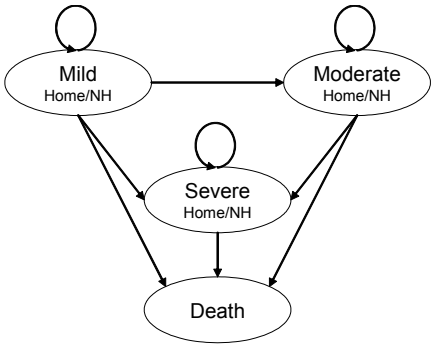


Figure 18. Schematic structure of the modified AD model. (NH= Nursing home)

The model included community (COM) and nursing home (NH) settings of care. Baseline transition probabilities between COM and NH were derived from Neumann et al. (1999, 2001) (Table 4). The transition probabilities between and inside the model stages are presented in Table 4.

The effect of the CBFI program on the delay of NH admission was based on estimates from Mittelman et al. (1996) study, where a total of 206 subjects were recruited in the study over a 3.5-year period. The total follow-up time of the study subjects was 8 years. The median time before NH admission was 1203

(95%CI 944 to 1412) days in the treatment group and 874 (95%CI 684 to 1064) days in the control group after adjusting for informal caregivers´ sex. The relative risk from a Cox proportional hazard model of NH admission after adjusting for informal caregivers´ sex, patient age, and patient income was 0.65 (95%CI 0.45 to 0.94) in the intervention group. This means that those AD patients receiving additional support program had an approximately 35% lower risk of NH admission than the AD patients without that support program. However, Mittelman et al. study did not examine changes in quality of life of the study subjects. Therefore, we wanted to include preference weights for different disease stages. Country-specific preference weights were unavailable, thus the weights were derived from Neumann et al. (1998) (Table 4). The effect of the CBFI program on the informal caregivers´ health-related quality of life (HRQOL) was constructed assuming that it depended on the AD patients´ disease progression. In other words, when the AD patient's health state decreases it was assumed to simultaneously have an impact on the caregiver's HRQOL, since the caregiver's role has been associated with psychological and physical morbidity, even an increased risk of mortality (Dunkin et al. 1998, Schultz et al. 1999). However, the model assumed that the caregivers remained alive over the five year care-giving period.

The organisation and consequently the utilization of the elderly care varies greatly in the Finnish municipalities. This means in fact, that in spite of the uniform legislation there is no standard 'present practice' covering the whole country, but instead, a number of differing local practices. In our case, the community care resource utilization of the AD patients with the informal caregivers was estimated according to those prevailing in two Finnish municipal health centres (Table 4). These municipal health centres were selected to represent the local organisation of elderly care services among AD patients in the eastern Finnish municipalities. The values of these utilized resources were mainly collected from the list of health service unit costs in Finland (Heikkinen et al. 2001).

The cost of the CBFI program was derived from the information from the rehabilitation centre, where the CBFI program is intented to be implemented initially. Costs related to travelling to the rehabilitation centre were also considered since Finland is geographically large and travelling distances can be quite long (Table 4).

According to the Finnish Social Insurance Institute, the reimbursed average costs of dementia medication were 1034€ per patient in the year 2000 in Finland (Table 4). The effects and use of the medication were assumed to be equal in both groups. All costs were counted in the monetary values for the year 2001. At the baseline, both costs and effects were discounted by a 5% annual discount rate.

Table 4. Parameters in the modified AD model

Annual transition probabilities	**Value**	**Source**
Stage to stage		
Mild to mild	0.615	**, ***
Mild to moderate	0.322	**, ***
Mild to severe	0.042	**, ***
Mild to dead	0.021	**, ***
Moderate to moderate	0.608	**, ***
Moderate to severe	0.339	**, ***
Moderate to dead	0.053	**, ***
Severe to severe	0.847	**, ***
Severe to dead	0.153	**, ***
Community to nursing home		
Mild	0.038	**, ***
Moderate	0.110	**, ***
Severe	0.259	**, ***
Quality of life weights		
Patients		
Mild		
Community	0.68	**
Nursing home	0.71	**
Moderate		
Community	0.54	**
Nursing home	0.48	**
Severe		
Community	0.37	**
Nursing home	0.31	**
Caregivers		
Mild		
Community	0.86	**
Nursing home	0.86	**
Moderate		
Community	0.86	**
Nursing home	0.88	**
Severe		
Community	0.86	**
Nursing home	0.88	**
Other parameters		
The effect of the CBFI-program (RR)	0.65	****
Annual discount rate – costs (%)	0%-5%	
Annual discount rate – benefits (%)	0%-5%	

Resource utilisation	**Value**	**Source**
Visit to neurologist (visits per year)		
Mild	2	†
Moderate	4	†
Severe	2	†
Interval care (Days per year)		
Mild	31	‡
Moderate	50	‡
Severe	85	‡
Day care (Days per year)		
Mild	43	‡
Moderate	5	‡
Severe	3,5	‡
Home nursing (Visits per year)		
Mild	28	‡
Moderate	18	‡
Severe	5,5	‡
Home help service (Visits per year)		
Mild	211	‡
Moderate	300	‡
Severe	37	‡
Meals on wheels (# of meals)		
Mild	47	‡
Moderate	38	‡
Severe	5	‡
Unit costs (EUR)		
Neurologist assessment	180	*****
Nursing home day	122	*****
Interval care	122	*****
Day care	59	*****
Home nursing	36	*****
Home help service	35	*****
Meals on wheels	5	*****
Other costs (EUR)		
Travelling	61	*****
Medication due to AD	1034	*
CBFI-program		
1st year	1000	
2nd year	1000	

Sources:

† Expert panel
‡ Two Finnish municipal health centres
* The Finnish Social Insurance Institute´s database
** Neumann et al. 1999
*** Neumann et al. 2001
**** Mittelman et al. 1996
***** Heikkinen et al. 2001

The existing information of the model parameters was originally characterized as point estimates. Thus, the model was deterministic and the only possible way to examine the effect of parametric variation on the model outputs was a basic sensitivity analysis (Briggs et al. 1994). However, we wanted to empha-

size the uncertainty related to the true values of the parameters and its impact on the model outputs more accurately by assigning prior distributions to characterize the uncertainty surrounding the model parameters (Claxton et al. 2001). Thus, the prior distributions, which represent beliefs concerning the true values of the model parameters were based on existing information, *i.e* the Bayesian approach (Bland & Altman 1998). Fundamentally, the majority of the decision models are informal applications of Bayesian reasoning (Briggs 1999).

All stage transition probabilities were characterized as beta distributions, because the beta distribution has applicable properties. Firstly, the beta(*r,n*) distribution is a continuous distribution, which obtain values between 0 and 1 and secondly, its parameters *r* and *n* can be thought of as counts of the event of interest and *n* as the total sample size of the specific patient cohort (Claxton et al. 2001, Briggs 2000). The parameters *r* and *n* were estimated by adjusting transition probabilities according to hypothetical cohort of 1,000 patients.

Health care resource utilization estimates were characterized as being gamma distributed since the distribution of resource utilization data is often positively skewed and it is logically bound to be ≥ 0. However, the mean resource estimates, which are located far from 0, were assumed to have a normal distribution. Unfortunately, reliable standard error information of the mean resource estimates was unavailable except for the travelling costs (Heikkinen et al. 2001). Therefore, artificial standard error estimates were calculated by multiplying the mean resource utilization estimates by value of 0.5 (Briggs et al. 2002). This was assumed to introduce a quite large variance into the mean resource estimates (Briggs 2000). Unit costs, which were derived from the list of health service unit costs, were handled as fixed prices assuming that they reflect the true opportunity costs of these consumed resources.

The health state utilities for the AD patients and their informal caregivers were characterized as being normally distributed. Since standard errors ($\widehat{SE}$) were not available, the uncertainty ranges for the health state utilities were set to be $\pm$ 0.1 and $\widehat{SE}$ was calculated using the equation 18.

[18] $$\hat{SE} = \frac{UB - LB}{2 * 1.96}$$

where *UB* is the upper and *LB* is the lower boundary for the health state utilities uncertainty ranges (Briggs 2000).

Uncertainty related to the relative risk estimate was illustrated using a log normal distribution. The log scale gives a natural constraint to the relative risk estimate over the interval from 0 to infinity. The mean relative risk estimate was derived from the study of Mittelman et al. (0.65) but since the CBFI program was not totally identical with the program, which was applied in Mittelman et al. study (1996) we created more sceptical effect scenario assuming an uncertainty range from 0.45 to 1.2, which would indicate

also that there was an insignificant effect difference between the CBFI program and the current practice.

After assigning the distributions to the parameters, the second-order Monte Carlo simulation was used to produce the empirical distributions for the mean costs and effects from the constructed probabilistic model. During the second order simulation, all of the model's parameters were randomly sampled and held constant while the 1,000 individuals of the *i*th cohort were cycled through the model. The distribution of the patients in each sample represents the results of a single artificial trial (first-order uncertainty). Then the mean costs and effects were calculated for this *i*th cohort. After every *i*th cohort, the model's parameters were sampled and calculated again. This estimation process was repeated 1,000 times (Briggs et al. 2002). After the final cohort, the average costs and effects were calculated based on these sampled mean estimates (second-order uncertainty).

Handling uncertainty. In the cost-utility analysis (CUA), the incremental cost-effectiveness ratio ($\hat{ICER}$) comparing the current practice (CP) with the CBFI program is typically defined as:

[19] $$\hat{ICER} = \frac{\overline{C}_{CBFI} - \overline{C}_{CP}}{\overline{E}_{CBFI} - \overline{E}_{CP}} = \frac{\Delta\overline{C}}{\Delta\overline{E}} < \lambda$$

where $\overline{C}_i$ and $\overline{E}_i$ represent sample means for the costs and effects. Lambda (λ) represents the decision makers' maximum willingness to pay for a QALY gained. Simulation results are often presented as a scatter-plot of the mean differences in cost and QALYs gained between the comparators. This graph is also known as a cost-effectiveness plane, where the x and y-axis divide the graph into four separate quadrants. In our case, these quadrants represent the following scenarios for the CBFI program in comparison with the current practice: (I) the CBFI program preferred to the current practice if and only if $\Delta\overline{C} / \Delta\overline{E}$ < λ, (II) the CBFI program preferred to the current practice, (III) the CBFI program preferred to the current practice if and only if $\Delta\overline{C} / \Delta\overline{E}$ > λ, and (IV) the current practice preferred to the CBFI program (Glick et al. 2001).

Recent methodological advantages have generated a new decision rule by rearranging the $\hat{ICER}$-formula as follows:

[20] $$\begin{aligned} &\frac{\Delta\overline{C}}{\Delta\overline{E}} < \lambda \\ &\Delta\overline{C} < \lambda\Delta\overline{E} \\ &\frac{1}{\lambda}\Delta\overline{C} < \Delta\overline{E} \\ &\Delta\overline{E} - \frac{1}{\lambda}\Delta\overline{C} > 0 \end{aligned}$$

The rearranged formula is the incremental net health benefit (ΔNHB) decision rule introduced by Stinnett and Mullahy (1998). The equivalence and optimality of ratio-based and net benefit-based ap-

proaches to health care resource allocation was described by Laska et al. (1998). The ΔNHB approach is a relatively new framework for handling uncertainty in CUA studies, but the ΔNHB approach permits one to avoid many of the problems encountered with CE-ratios when establishing confidence intervals for cost-effectiveness ratios (Glick et al. 2001). The advantage of ΔNHB is that the joint distribution of cost and effectiveness is asymptotically normally distributed. The disadvantage of ΔNHB has been proposed to be the fact that the value of λ is unknown. To avoid this problem, the net-benefit statistics can be plotted as a function of λ (Stinnett & Mullahy 1998). From the Bayesian point of view, this function can be interpreted to reveal the probability that an intervention is cost-effective for given value of λ (ΔNHB>0) (Claxton et al. 2001). Thus, when decision uncertainty is combined with varying societal willingness to pay (λ) per gained QALY, the cost-effectiveness information can be presented as an acceptability curve. The acceptability curve is a graphical representation, which helps decision makers to define the uncertainty of decisions and to acquire more information if necessary (Van Hout et al. 1994).

6.1.4 Results

The Markov model with the second-order simulation generated the cost-effectiveness estimates for the current practice and the CBFI program. The mean estimates of costs and QALYs gained with 95% uncertainty ranges are presented in Table 5. The 95% uncertainty ranges were obtained by selecting the 2.5^{th} and 97.5^{th} percentile points of the ranked vector of 1,000 simulation replicates. The mean results of the economic analysis indicate that the CBFI program is potentially cost-saving and it is superior to (less costly and more effective) the current practice when the program is used as an additional support service for patients with mild AD patients and their informal caregivers (Table 5). However, based on the 95% uncertainty ranges, the differences between mean estimates are not statistically significant. Furthermore, it seems that the CBFI program lightly decreases the quality of life of informal caregivers. However, the difference between the quality of life is not statistically significant.

Table 5. Mean estimates of costs and effects with 95% uncertainty ranges (in parenthesis)

Estimate	CBFI-program	Current practice	Difference
Patient			
Mean costs (EUR)	43,933 (19,785 to 71,026)	46,925 (19,073 to 75,740)	-2992
Mean QALYs	1.88 (1.71 to 2.06)	1.87 (1.72 to 2.05)	0.01‡
ICER	-	-	Cost-saving*
Caregiver			
QALYs	3.13 (2.94 to 3.32)	3.14 (2.95 to 3.32)	-0.01‡

‡ Results are approximate due to rounding – the model uses full accuracy
* The CBFI program is more effective and less costly

The results of 1,000 simulations from the AD model with 95% uncertainty ranges are presented in Figure 19. As shown in Figure 19, the 95% uncertainty ranges cover more than one quadrant of the cost-

effectiveness plane effecting the significant uncertainty surrounding the $\hat{ICER}$-estimate. From a decision-making point of view the use of frequentist interpretation rules do not provide information on the likelihood that the CBFI program is cost-effective. On the cost-effectiveness plane, the high concentration of points in quadrant II (71.3% of iterations) indicates that the CBFI program is potentially cost-saving offering more QALYs with lower costs but the dispersion of points in quadrants I (3.8% of iterations), III (19.6% of iterations), and IV (5.3% of iterations) indicates that there is also some uncertainty surrounding the mean estimates of costs and QALYs gained.

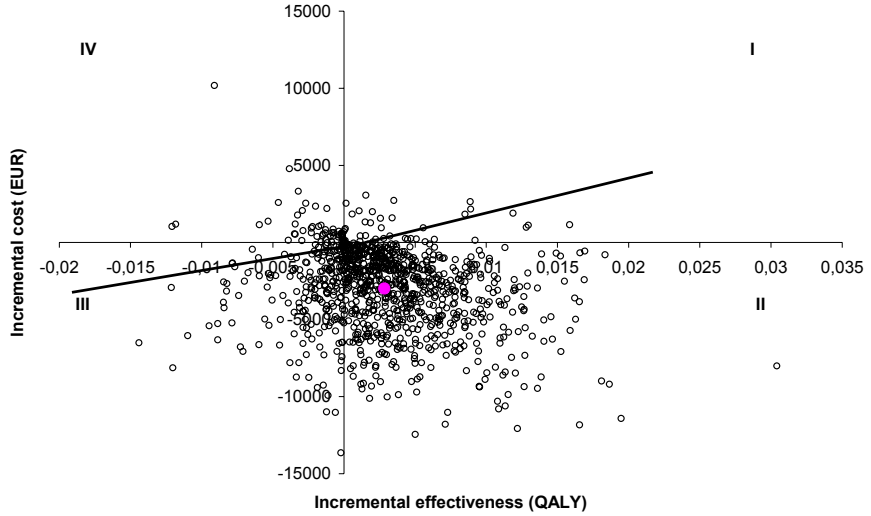


Figure 19. Scatter-plot of mean differences in cost and QALYs gained between the CBFI program and the current practice.

The acceptability curve approach was used to emphasize the uncertainty related to different levels of willingess to pay per gained QALY. The value of λ was varied from 1 EUR to 100,000 EUR and the probability that ΔNHB would be positive was calculated at various values of λ to obtain the acceptability curve. The acceptability curve for the CBFI program as compared to the current practice is presented in Figure 20. As shown in Figure 20, the probability that the CBFI program is cost-effective for the AD patients is over 0.9 at all values of willingness to pay per gained QALY. Therefore, based on the existing information there is a less than 10% chance that the CBFI program is not an optimal option for adoption into wide-scale use in Finland.

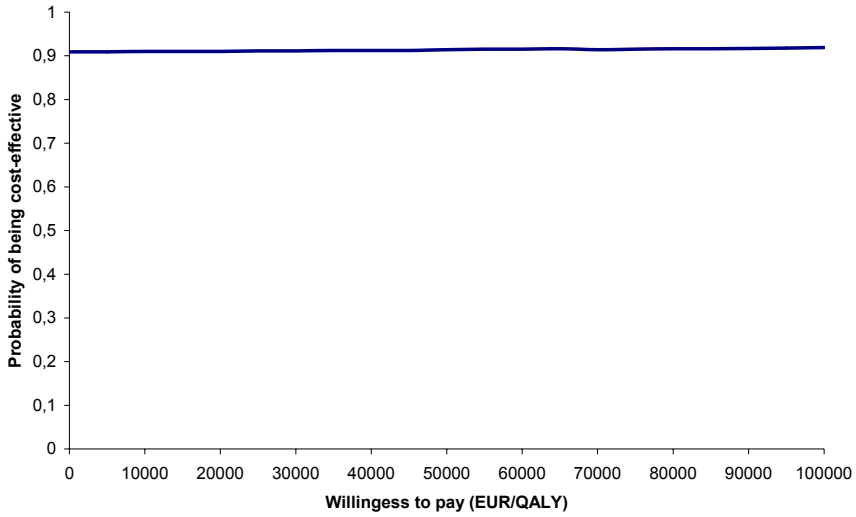


Figure 20. Acceptability curve of the CBFI program as compared to the current practice.

6.1.5 Conclusions

Based on the current information, the CBFI program is a potentially cost-saving option and it has the highest probability of being optimal, since the probability that the CBFI program will provide greater net benefits than the current practice is over 0.9 and the error probability is less than 0.1 at all of the values of willingess to pay per gained QALY. Furthermore, the caregiver's HRQOL was insensitive to the AD patient's disease stage and settings of care. However, there is some evidence that generic preference-weighted instruments may not capture all of the differences in the burden suffered by the caregiver, though these can gathered by specific caregiver burden instruments (Bell et al. 2001). This indicates that more research is needed to assess better the HRQOL estimate in AD caregivers. In addition to those that mentioned above, the QALY-weights of AD patients are both a measurement and a conceptual problem (what do we mean by quality of life for a patient with severe dementia?). The patients´ ability to respond to questions themselves is limited in the moderate and severe stages of the disease, and the primary caregivers were asked to complete questionnaire as proxy respondents. Therefore, more information about the preferences of patients is also needed.

There are also some other limitations, which should be taken into account in the evaluation of this study. Firstly, there were several problems with the data availability. The sources of uncertainty related to the majority of the model parameters have been well defined by Neumann et al. (1999). Secondly, the cost consequences of the CBFI program depend on the present organization of the care of the elderly and this varies greatly in the Finnish municipalities. In fact, the 'efficient current practice' is not

known. This is one additional source of uncertainty. The more institutional-based the care is, the more likely the CBFI program is to be cost-efficient. Thirdly, indirect costs were not included in our model since the majority of informal caregivers are pensioners who take care of their spouses. Thus, production losses were assumed to be relatively small compared to the direct costs. Furthermore, the informal caregivers are not paid for the help they provide apart from some small payments related to the caregiving, which are provided by social security system.

From the methodological point of view, the stochastic decision modelling with the Bayesian approach permits a better and more powerful characterization of uncertainty surrounding the model's parameters than the deterministic modelling does. Furthermore, the ΔNHB approach with the acceptability curve enables avoidance of the problems of interpretation for confidence intervals when uncertainty covers more than one quadrant of the cost-effectiveness plane (Briggs 2000). Moreover, the stochastic modeling based on existing information is a useful tool to demonstrate the need for supplemental information. If one wishes to estimate the monetary value of additional information, it can be defined by using sophisticated quantitative methods such as the expected value of perfect information (EVPI) (Claxton et al. 2001).

In the future, it will be important to collect more information about the cost-effectiveness (CE) of interventions for AD, including pharmacological and rehabilitation programs. In fact, it might be most rational to investigate the CE of treatments that are combinations of pharmacological and non-pharmacological therapies. For example, acetylcholinesterase inhibitors have been shown to significantly improve cognition and daily functioning of AD patients (Birks et al. 2001, Dooley & Lamb 2000, Olin & Scheider 2002). Moreover, recent studies suggest that these drugs may also decrease some behavioral symptoms (Feldman et al. 2001, Tariot et al. 2000).

6.1.6 References

Bachman DL, Wolf PA, Linn R, Knoefel JE, Cobb J, Belanger A, D'Agostino RB & White LR. 1992. Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham study. Neurology 42: 115-119.

Bell CM, Araki SS & Neumann PJ. 2001. The association between caregivers burden and caregiver health-related quality of life in Alzheimer´s disease. Alzheimer Disease & Associated Disorders 15: 129-136.

Birks J, Grimley Evans J, Lakovidou V & Tsolaki M. 2001. Rivastigmine for Alzheimer's disease (Cochrane review). The Cochrane Library, Issue 4.

Bland JM & Altman DG. 1998. Bayesians and frequentists. British Medical Journal 317: 1151.

Braak H & Braak E. (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathologica 82: 239-259

Briggs AH. 1999. A Bayesian approach to stochastic cost-effectiveness analysis. Health Economics 8: 257-61.

Briggs AH. 2000. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 17: 479-500.

Briggs AH, Goeree R, Blackhouse G & O´Brien BJ. 2002. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Medical Decision Making 22: 290-308.

Briggs A, Sculpher M & Buxton M. 1994. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health Economics 3: 95-104.

Brodaty H, Gresham M & Luscombe G. 1997. The Prince Henry Hospital dementia caregiver training program. International Journal of Geriatric Psychiatry 12: 183-192

Claxton K, Neumann PJ, Araki S & Weinstein MC. 2001. Bayesian value-of-information analysis: An application to a policy model of Alzheimer's disease. International Journal of Technology Assessment in Health Care 17: 38-55.

Dooley M & Lamb HM. 2000. Donepezil: a review of its use in Alzheimer's disease. Drugs and Aging 16: 199-226.

Dunkin JJ & Anderson-Hanley C. 1998. Dementia caregiver burden: a review of the literature and guidelines for assessment and intervention. Neurology 51: S53-S63.

Eloniemi-Sulkava U, Notkola I-L, Hentinen M, Kivelä S-L, Sivenius J & Sulkava R. 2001. Effects of supporting community-living demented patients and their caregivers: a randomized trial. Journal of the American Geriatrics Society 49: 1282-1287.

Evans DA, Funkenstein HH, Albert MS, Scherr PA, Cook NR, Chown MJ, Hebert LE, Hennekens CH & Taylor JO. 1989. Prevalence of Alzheimer's disease in a community population of older persons. Journal of American Medical Association 262: 2551-2556.

Feldman H, Gauthier S, Hecker J, Vellas B, Subbiah P & Whalen E. 2001. A 24-week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease. Neurology 57: 613-20.

Glick HA, Briggs AH & Polsky D. 2001. Quantifying stochastic uncertainty and presenting results of cost-effectiveness analyses. Expert Review in Pharmacoeconomics Outcomes Research 1: 25-36.

Heikkinen K, Hujanen T & Rusama H. 2001. Terveydenhuollon yksikkökustannukset Suomessa vuonna 2000. Aiheita 23. National Research and Development Centre for Welfare and Health.

Hu T, Huang L & Cartwright WS. 1986. Evaluation of the costs of caring for the senile demented elderly: a pilot study. Gerontologist 26: 158-163.

Jost BC & Grossberg GT. 1995. The natural history of Alzheimer´s disease: a brain bank study. Journal of the American Geriatrics Society 43: 1248-1255.

Knopman D, Schneider L, Davis K, Talwalker S, Smith F, Hoover T & Gracon S. 1996. Long-term tacrine treatment: effects on nursing home placement and mortality. Neurology 47: 166-177.

Laska EM, Meisner M, Siegel C & Stinnett AA. 1998. Ratio-based and net benefit-based approaches to health care resource allocation: proofs of optimality and equivalence. Health Economics 8: 171-174.

Mittelman MS, Ferris SH, Shulman E, Steinberg G & Levin B. 1996. A family intervention to delay nursing home placement of patients with Alzheimer's disease. Journal of the American Medical Association 276: 1725-1731.

Morris JC. 1993. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43: 2412-14.

Neumann PJ, Araki SS, Arcelus A, Longo A, Papadopoulos G, Kosik KS, Kuntz KM & Bhattacharjya A. 2001. Measuring Alzheimer´s disease progression with transition probabilities. Estimates from CERAD. Neurology 57: 957-964.

Neumann PJ, Hermann RC, Kuntz KM, Araki SS, Duff SB, Leon J, PA Berenbaum PA, Goldman PA, Williams LW & Weinstein MC. 1999. Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. Neurology 52: 1138-1145.

Neumann PJ, Hermann RC & Weinstein MC. 1998. Measuring QALYs in dementia. In: Wimo A, Jonsson B, Karlsson, Winblad B. (Eds.) Health Economics in Dementia. Sussex: John Wiley & Sons, p.359-370.

Olin J & Schneider L. 2002. Galantamine for Alzheimer's disease (Cochrane review). The Cochrane Library, Issue 1.

Ott A, Breteler MMB, Harskamp van F, Claus JJ, Cammen van der TJM, Grobbee DE & Hofman A. 1995. Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. British Medical Journal 310: 970-973.

Polvikoski T, Sulkava R, Myllykangas L, Notkola I-L, Niinistö L, Verkkoniemi A, Kainulainen K, Kontula K, Perez-Tur J, Hardy J & Haltia M. 2001. Prevalence of Alzheimer's disease in very elderly people. A prospective neuropathological study. Neurology 56: 1690-1696.

Schultz R & Beach SR. 1999. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. Journal of the American Medical Association 282: 2215-2219.

Skoog I, Nilsson L, Palmertz B, Andreasson L-A & Svanborg A. 1993. A population-based study of dementia in 85-year-olds. New England Journal of Medicine 328: 153-158.

Stinnett AA & Mullahy J. 1998. Net health benefits: a new framework for the analysis in cost-effectiveness analysis. Medical Decision Making 18(Suppl): S68-S80.

Stinnett AA & Paltiel AD. 1997. Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. Medical Decision Making 17: 483-489.

Sulkava R, Wikström J, Aromaa A, Raitasalo R, Lehtinen V, Lahtela K & Palo J. 1985. Prevalence of severe dementia in Finland. Neurology 35: 1025-1029.

Tariot PN, Solomon PR, Morris JC, Kershaw P, Lilienfeld S & Ding C. 2000. A 5-month, randomized, placebo-controlled trial of galantamine in AD. Neurology 54: 2269-2276.

Van Hout B, Al MJ, Gordon GS & Rutten FF. 1994. Cost, Effects and C/E-ratios alongside a clinical trial. Health Economics 3:309-319

Wimo A, Ljunggren G & Winblad B. 1997. Costs of dementia and dementia care – a review. Int J Geriatr Psychiatr 12: 841-856.

6.2 Modelling the cost-effectiveness of temozolomide in the treatment of recurrent glioblastoma multiforme - incorporating the quality of clinical evidence into a decision-analytic model[7]

6.2.1 Introduction

Glioblastoma multiforme (GBM) is the most malignant glioma (grade IV) and tends to progress and recur in brain tissue despite aggressive treatment (Newlands et al. 1997) The median survival time for GBM patients is only about 12 months. GBM is treated with combinations of surgery, radiation therapy and chemotherapy based on patient performance status. There is no accepted treatment for GBM and, usually, responses are poor regardless of the chosen therapy method. GBM is considered an incurable disease, which prompts the need for new endpoints to measure treatment efficacy such as improvement of neurological symptoms and quality of life, in addition to overall survival benefit.

Temozolomide (TMZ) is an oral alkylating agent with demonstrated efficacy as therapy for GBM and anaplastic astrocytoma (Newlands et al. 1997) TMZ resembles dacarbazine but has several additional benefits such as oral administration and enhanced ability to cross the blood-brain barrier. TMZ forms its active metabolite 5-(3-methyl-ltrazeno) imidazole-4-carboxamide spontaneously in physiological pH without requiring liver metabolism for activation.

Phase II studies of patients with malignant gliomas have shown TMZ improves disease and symptom control but phase III studies showing marked survival benefit were not published at the time of writing this manuscript. The most promising studies have been done with the combination of TMZ and radiation therapy. (Stupp et al. 2002)

Cost-effectiveness studies of various treatments for malignant gliomas are scarce. This probably relates to problems in conducting adequate clinical trials of diseases in which, only fairly small patient populations are available to trials at any one time.

The objective of this study was to evaluate the cost-effectiveness of TMZ for the treatment of GBM from a societal perspective. The cost-effectiveness of TMZ was compared to that of PCV chemotherapy by calculating the following figures for both treatments: cost per life-month, cost per progression-free life-month and cost per QALY. Since there were no available data comparing the cost-effectiveness of TMZ, and PCV, a decision modelling approach was applied.

[7] This chapter has been published in: Martikainen J, Kivioja A, Hallinen T & Vihinen P. Economic Evaluation of Temozolomide in the Treatment of Recurrent Glioblastoma Multiforme. Pharmacoeconomics 2005;23(8):803-815. Reproduced with permission.

The parameter values used in decision models are often gathered from a variety of sources and, by their very nature, contain some level of uncertainty. In our modelling process, this uncertainty related to parameter values was characterised by assigning probability distributions to all parameters for which there was uncertainty in their true values (i.e. a Bayesian approach was applied). (Briggs et al. 2002, Doubilet et al. 1985) Accordingly, the probability that TMZ is cost-effective as compared to the current practice (PCV-chemotherapy) at different levels of willingness-to-pay per gained effects was represented by calculating acceptability curves. Additionally, the value of new information for reducing uncertainty related to the choice of treatment with TMZ or PCV was evaluated using the expected value of perfect information (EVPI) approach.

6.2.2 Methods

Model The disease process of malignant gliomas was characterized with a Markov model (Sonnenberg & Beck 1993, Briggs & Gray 1998) (Figure 21). The Markov model included three health states: 'progression-free', 'progression' and 'death'. It was assumed that all patients underwent primary treatments, such as surgery and radiotherapy, before first relapse and, therefore, patients who had not undergone primary treatment were not included in the model. The progression-free state in the model describes the time from the beginning of chemotherapy to the second relapse. The progression state describes the time from the second relapse to death. Death was modelled as an absorbing state. Transitions back to the progression-free state from the progression state were also not considered possible in the model. In our model, the length of a cycle was one month.

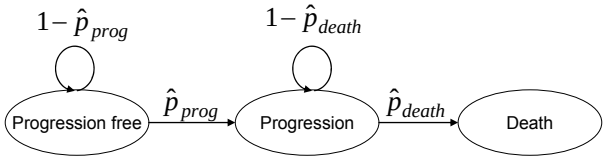


Figure 21. State transition diagram of high-grade gliomas

6.2.3 Data sources and handling of uncertainty

Clinical outcomes data. For the purposes of this study, a systematic review of published scientific articles found in PubMed was conducted in June 2003. The search term used was "(PCV OR temozolomide) AND (GBM OR glioblastoma)" and this search resulted in 109 hits. Of these 109 papers, 23 were considered relevant, based on their abstracts, and were evaluated thoroughly. Of these, only those studies that fulfilled all of the following criteria were selected for calculations of endpoints, overall survival (OS) and progression-free survival (PFS):

1. Disease and disease stage had to be GBM at first relapse
2. Results had to be presented as OS (or 6-month survival, OS-6) or PFS (or 6-month progression-free survival, PFS-6)
3. Chemotherapy had to include either TMZ (≥ 100 mg/m^2/d) or some part of PCV chemotherapy.

In the case of GBM, six studies were considered adequate for inclusion in the measurement of efficacy (Table 6). We assessed the quality (in terms of pertinence and validity) of the studies included in the review by using a methodology suggested by Tan et al. (2003) The members of the research team scored the papers for quality independently and, in unclear cases, a consensus decision was made. After the pertinence (cancer type, treatment, endpoint) and validity scores were given, they were multiplied to obtain a study weight. The weights were used to adjust the study sizes to reflect the uncertainty associated with the methodological restrictions or limited pertinence and validity of the studies. This adjusting was done simply by multiplying the study size with the study weight. The weights and adjusted study sizes of the studies are presented in Table 6 (the description of the criteria used for weighing the studies is available from the authors).

Table 6. Weights and adjusted study sizes of the included efficacy studies of temozolomide- and PCV-treatments for GBM.

Study	Cancer type	Treatment	Endpoint	Pertinence score*	Validity score	Weight˚	Unadjusted study size	Adjusted study size
Temozolomide								
Brada et al. (2001)	1	0.5	1	0.5	0.2	0.1	138	13.8
Yung et al. (2000)	1	0.8	0.9	0.8	1	0.8	112	89.6
Brandes et al. (2003)	0.9	0.9	1	0.9	0.4	0.36	22	7.92
					Total		**272**	**111.32**
PCV								
Kappelle et al. (2001)	1	0.5	1	0.5	0.2	0.1	63	6.3
Boiardi et al. 2001	1	0.5	1	0.5	0.4	0.2	27	5.4
Yung et al. 2000	1	0.8	0.9	0.8	1	0.8	113	90.4
Brandes et al. 2003	0.9	0.9	1	0.9	0.4	0.36	32	11.52
					Total		**235**	**113.62**

*Pertinence score is the minimum value from three previous columns (Cancer type, Treatment and Endpoint).
˚ Weight is pertinence score multiplied with validity score.

The transition probabilities (values associated with the arrows in Figure 21) needed in the model were estimated from the median PFS and median OS times reported in the selected articles. This was done using the following equations (Miller & Homan 1994):

[21] $$\hat{p}_{prog} = \frac{\sum_{i=1}^{n}\left(\left(1-0.5^{\frac{1}{PFS_i}}\right)\times SS_{adj_i}\right)}{\sum_{i=1}^{n} SS_{adj_i}}$$

[22] $$\hat{p}_{death} = \frac{\sum_{i=1}^{n}\left(\left(1-0.5^{\frac{1}{OS_i-PFS_i}}\right)\times SS_{adj_i}\right)}{\sum_{i=1}^{n} SS_{adj_i}}$$

where $\hat{p}_{prog}$ = probability of transition from progression-free state into progression state, $\hat{p}_{death}$ = probability of transition from progression state into death state, SS_{adji} is the adjusted study size from study *i*, and PFS_i and OS_i are the median estimates from the study *i*.

The uncertainty related to these probabilities was characterised by expressing the parameter values in the model as beta distributions. (Gelman et al. 1995) The beta distribution is commonly parameterized as *beta(α,β)*, where *α* is the number of patients transferred to a state during one cycle and *β* is calculated as the total sample size of the treatment group minus the value of *α*. As none of the selected studies included information on the number of events, we obtained the approximate values of parameter *α* by using the following equations:

[23] $$SS_{adj} \times \hat{p}_{prog} = \alpha_{adj\_prog}$$

[24] $$SS_{adj} \times \hat{p}_{death} = \alpha_{adj\_death}$$

where equation 23 gives α_{adj_prog} for transitions to *progression* state, equation 24 gives α_{adj_death} for transitions to *death* state, SS_{adj} is the pooled adjusted study size and α_{adj} is the adjusted number of patients transferring from one state to another.

After calculating the adjusted α parameters, the β parameters were obtained by simply subtracting α_{adj} from SS_{adj}. Beta distributions for disease progression and death based on these estimated parameter values are presented in Figures 22 and 23. The use of the 'quality' adjusted parameters α and β within the beta distribution reflected the uncertainty associated with the methodological restrictions or the limited pertinence and validity of the studies in the decision model. A study with high pertinence and validity had a high weight and a large adjusted study size, thus contributing more to the estimates of clinical

efficacy. The mean of the beta distribution is E(β) = α /α+β, where the sum α+β has an applicable property to characterise uncertainty, since the larger the sum, the smaller the variance around this mean E(β). Thus, uncertainty related to the validity of trials could be taken into account by applying the combination of sample size adjustment method (Tan et al. 2003) and the beta distribution when pooling data for modelling purposes.

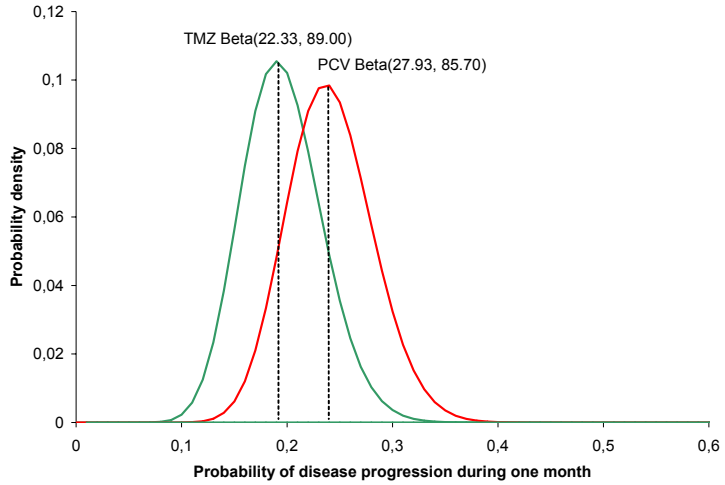


Figure 22. Probability of disease progression during one month

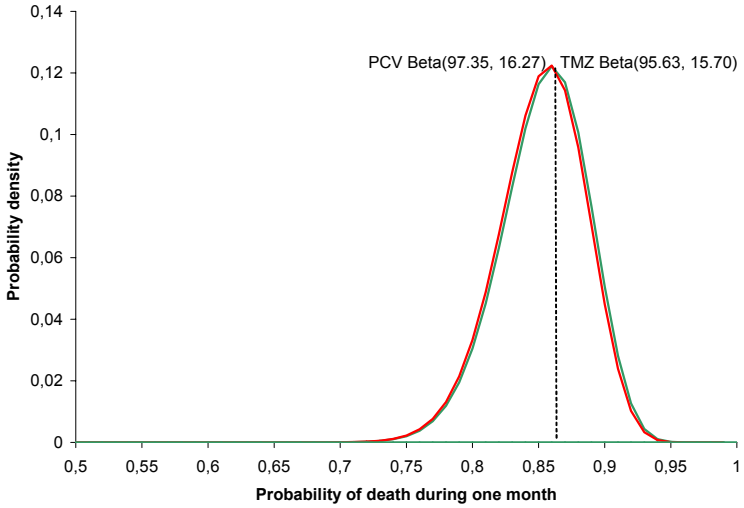


Figure 23. Probability of death during one month

Utility data. In order to perform cost-utility analysis, utility scores for patients in the model's different health states were needed for the patients treated with TMZ or PCV. There was no preference-based

information directly available on the quality of life of patients with high-grade gliomas after the first and the second relapse at the time of our study. Therefore, the quality of life estimates were gathered using proxy respondents. Utility scores were obtained using a visual analogue scale (VAS) method. (Torrance et al. 2001, Brazier et al. 2003) The questionnaires were sent to eight leading neuro-oncologists, of whom 6 responded. The top anchor of the VAS scale ("perfect health") was defined analogously with the Health Utilities Index Mark 3 (HUI3) system (Health Utilities, Inc) (Feeny et al. 1995) and the bottom anchor of the VAS scale was defined as 'dead', which was assumed to be the worst imaginable state. The utilities for health states were defined regardless of the treatment. The neuro-oncologists were not informed of the chemotherapies being compared and were not paid for participating in the study. Adverse events related to particular chemotherapy were also defined without reference to the therapy in question to enable blind estimates of utilities.

Uncertainty related to the utility scores was characterised using the beta distributions, since the utility scores are also bounded between 0 and 1. To obtain the beta parameters based on mean $(\overline{\mu})$ and standard error (σ) estimates from the utility data, the following reformulations were made: $\alpha = r$ and $\alpha + \beta = n$. Subsequently, we applied the following equations (Briggs et al. 2002):

[25] $$n = \frac{\overline{\mu}(1-\overline{\mu})}{\sigma^2} - 1 \qquad r = \overline{\mu} n$$

The parameters of the beta distributions used in the model are shown in Table 7.

Table 7. Utility parameters

Disease stage*	Mean (standard error)	α	β
S1	0.55 (0.06)	39.08	31.76
S2	0.41 (0.06)	26.05	37.48
S3	0.43 (0.10)	9.14	12.24
S4	0.31 (0.07)	13.67	30.42
S5	0.14 (0.09)	2.13	12.72

*Disease stages are: S1: 45-year-old glioma patient who is going to the first surgery; S2: Same patient after surgery and radiotherapy (60 Gy) on the tumor area. The symptoms have recurred; S3: Chemotherapy alternative TMZ; S4: Chemotherapy alternative PCV; S5: Progression.

Resource use and cost data. Data on the use of health care resources were collected mainly from hospital databases. The resource use associated with TMZ treatment was gathered from a cohort (n=16), which consisted of high-grade GBM patients treated with TMZ in Helsinki and Turku University Hospitals between 1998 and 2000. The resource use associated with PCV treatment was gathered from a cohort of patients (n = 10), who were treated in Turku University Hospital between 1998 and 2000. Resource utilisation associated with post-chemotherapy state was assumed to be equal in both groups.

The unit costs of the resource items were derived from the list of Finnish health service unit costs. (Hujanen 2003) All chemotherapy costs were derived from the official database of medicine prices in Finland, except for lomustine, for which the costs are based on the wholesale price paid by the Kuopio University Hospital Pharmacy. The wholesale price of lomustine was transformed to retail price using the official formula on Council of States decree on price list. (Council of States act on price list for drugs 2002) All prices were used in Finland in 30th October 2003. When calculating costs of chemotherapy, the treatment with TMZ was assumed to be given during 5 days in 28-day cycles at dosages of 150–200 mg/m^2. Respectively, PCV treatment was assumed to be administered in 42-day cycles (lomustine 110 mg/m^2 per oral on day 1; procarbazine 60 mg/m^2 per oral on days 8-21; vincristine 1.4 mg/m^2 intravenously on days 8 and 29). All patients were assumed to have an average body surface area of 1.73m^2. The number of administered chemotherapy cycles correlated with the time that the patient spent in the progression free health-state. Unit costs of all resource items are shown in Table 8.

Table 8. Unit costs of resource items

Item	Unit Cost (€)‡
Oncologist visit	160
Hospital day	338
Laboratory visit	15
MRI scan	480
TMZ-chemotherapy cost (per month)	2093
Anti-emetics related to TMZ chemotherapy (per month)	76
PCV-chemotherapy cost (per month)	179
Anti-emetics related to PCV chemotherapy (per month)	162
Travelling cost (per visit)	30

‡ Unit cost rounded to the nearest €

An empirical examination showed that the sample distributions of the collected resource items were positively skewed. This is a typical phenomenon for cost data, and thus the uncertainty of the resource use was modelled using a gamma distribution. (Briggs & Gray 1998) The methods of moments were applied to obtain mean (α) and variance (β) parameters for the gamma distributions (Briggs et al. 2002)

The impact of TMZ and PCV treatments on the probability of requiring antiemetics was represented as a beta distribution. It was assumed that a patient suffered from serious (grade 3-4) nausea or vomiting would be given 5-HT_3-receptor antagonists. If a patient suffered from milder (grade 1-2) nausea or vomiting or did not experience any nausea, they were assumed to be treated with metoclopramide. The occurrence and severity of nausea or vomiting were obtained from reported clinical trials. (Brada et al. 2001, Levin et al. 1985, Levin et al. 2003, Medical Research Council Brain Tumour Working Party 2001, Buckner et al. 2003, Prados et al. 1999, Yung et al. 1999, Yung et al. 2000) On the basis of clinical trial data, our model assumed that an average of 6.06% of the patients treated with TMZ and 16.13% of the PCV patients suffered from severe nausea or vomiting, and, therefore, required 5-HT_3-receptor antago-

nists. The rest of the patients in both treatment groups were assumed to use metoclopramide, regardless of whether or not they experienced mild nausea or vomiting or even nausea at all, in which case they would probably not have received any antiemetics in a real-life treatment scenario. However, the generalisation in this assumption was not considered crucial, as it would not favour TMZ. The parameters for beta distribution were calculated using the equations proposed by Briggs et al. (2002). The mean amounts of all resources used and their distribution parameters are presented in Table 9.

Table 9. Means and standard errors of utilised resources per month and the associated distribution parameters

Resource item	TMZ		PCV		Post chemotherapy	
	Mean (SE)	Distribution parameters	Mean (SE)	Distribution parameters	Mean (SE)	Distribution parameters
Oncologist visit	0.78 (0.13)	Gamma(5,5.63)[#]	2.42 (0.41)	Gamma(14.36, 5.94)	0.15 (0.07)	Gamma(0.33, 2.15)
Hospital days	1.20 (1.10)	Gamma(1.3,1.1)	6.85 (2.51)	Gamma(18.72, 2.73)	10.9 (4.17)	Gamma(28.58, 2.62)
Laboratory visits	0.90 (0.26)	Gamma(3.17, 3.5)	2.55 (0.34)	Gamma(19.03, 7.47)	1.20 (0.33)	Gamma(4.3, 3.6)
MRI scans	0.23 (0.11)	Gamma(0.47,2.06)	0.25 (0.04)	Gamma(1.44,5.85)	0.47 (0.14)	Gamma(1.54, 3.28)
Antiemetics		Beta(10.77;166.86)*		Beta(7.44;38.70)		

[#] Gamma(α,β); α = mean^2/standard error; β = mean/standard error
* Beta(α;β); α=r / β=n-α

All costs were counted in the monetary values for the year, 2001. Both future costs and effects, which occurred over a one-year period, were discounted with a 5% annual discount rate. A monthly discount rate was obtained from the annual discount rate.

Probabilistic sensitivity analysis. The cost-effectiveness model was analysed using a second-order Monte Carlo simulation. In the second-order Monte Carlo simulation, cohorts consisting of 1000 patients were run through the model 1000 times using Treeage DATA Pro software (Treeage Software Inc. Williamstown, USA). Each of these 1000 simulations consisted of the following steps:

1. Parameter values were drawn from their prior distributions
2. Expected costs and effects were estimated for each cohort
3. Steps 1-2 were repeated a large number of times to generate empirical distributions for the overall expected costs and effects.

Evaluation of cost-effectiveness. The results from the decision model were summarised as an incremental cost-effectiveness ratio (ICER). The ICER summarises the additional resource consumption needed for an increase in an additional unit of effectiveness as follows:

[26] $$\hat{ICER} = \frac{\overline{C}_{TZM} - \overline{C}_{PCV}}{\overline{E}_{TZM} - \overline{E}_{PCV}} = \frac{\Delta\overline{C}}{\Delta\overline{E}} < \lambda$$

where $\overline{C}_{TZM}$ = population mean costs of TMZ, $\overline{C}_{PCV}$ = population mean costs of PCV, $\overline{E}_{TZM}$ = population mean effectiveness of TMZ, $\overline{E}_{PCV}$ = population mean effectiveness of PCV, $\Delta\overline{C}$ = incremental population mean costs, $\Delta\overline{E}$ = incremental population mean effectiveness and λ = maximum societal willingness-to-pay per each additional unit of effectiveness.

Acceptability curves were applied to represent uncertainty in cost-effectiveness results. An acceptability curve represents the probability of cost-effectiveness as a function of willingness-to-pay per additional unit of effectiveness (λ) and helps decision makers to define the uncertainty associated with decisions and acquire more information, if necessary. (Stinnett & Mullahy 1998, O'Hagan et al. 2000) Based on the Monte Carlo simulation results, the acceptability curves were determined as the proportion of the (ΔE , ΔC) points where the TZM was cost-effective as conditional to λ.

Value of additional research. Information from additional research is valuable because it reduces the uncertainty surrounding a clinical decision. A net monetary benefit (NMB) approach was applied to estimate the value of additional research. (Tambour et al. 1998) Expressing expected net benefits in a monetary scale gives an explicit monetary valuation of the costs of uncertainty that can be compared with the cost of collecting further information. The net benefits can be expressed on a monetary scale by transforming the ICER into NMBs as follows (Tambour et al. 1998):

[27]
$$\frac{\Delta\overline{C}}{\Delta\overline{E}} < \lambda$$
$$\lambda\Delta\overline{E} - \Delta\overline{C} > 0$$

In the current study, the effectiveness was measured in the terms of QALYs. The expected costs of uncertainty can be interpreted as an EVPI because perfect information (e.g. an infinite sample of patients) would eliminate all uncertainty surrounding the decision. The theoretical background of the EVPI approach and its applications in health economics have been previously presented by Claxton et al. (Claxton 1999; Claxton et al. 2001) and others (Karnon 2002).

In a Bayesian decision theoretic framework, the appropriate decision on treatment selection given existing (prior) information is to choose the treatment with the highest average NMB ($\lambda\overline{E} - \overline{C} > 0$).To obtain the highest average NMB with perfect information, we determined the optimal strategy after each iteration and calculated the average NMB with perfect information from these choices. (Fenwick et al. 2000) The overall EVPI was estimated as follows:

[28] Overall EVPI = $N\overline{M}B_{\pi} - N\overline{M}B_{\varepsilon}$

$N\overline{M}B_{\pi}$ = the average NMB of the treatment chosen with perfect information

$N\overline{M}B_{\varepsilon}$ = the average NMB of the treatment chosen with the existing information.

From the societal decision makers' point of view, it could be valuable to know the overall value of information at a population level, which might be affected by the adoption decision. This population level information could be utilised in decisions concerned with whether to fund additional research to reduce uncertainty relating to the decision. Furthermore, the value of information methodology provides an opportunity for the decision makers to prioritise research in a more systematic and coherent fashion than is usually done.

The overall value of information at a population level for the high-grade glioma patients was determined by multiplying the overall EVPI with the number of patients that would be affected by the information over the effective lifetime of the TMZ chemotherapy:

[29] $$\text{Population EVPI} = EVPI * \sum_{t=1}^{T} \frac{I_t}{(1 + r)^t}$$

where I = incidence in period, t = period, T = total number of periods for which information from additional research would be useful and r = discount rate. (Claxton 1999, Claxton et al. 2001)

The exact incidence figures of GBM in Finland could not be found from any official statistics. According to the Central Brain Tumour Registry of the United States, the incidence of GBM in 1995-1999 was 3.24 per 100 000 person-years in the United States. (CBTRUS 2002) Calculations based on this American incidence figure and current estimates of the Finnish population suggest an annual incidence of around 168 new cases in Finland. The present prevalence of the high-grade gliomas was assumed to be around 100 cases. Furthermore, we assumed that the information from research would be valuable for 10 years and the discount rate was adjusted to be 5%. EVPI calculations were executed using a Visual Basic macro for Excel™.

6.2.4 Results

Cost-effectiveness. Cost-effectiveness estimates were obtained by performing 1000 second-order simulations of the Markov model. The cost-effectiveness of TMZ, compared to PCV, was measured as the incremental cost-effectiveness ratio (ICER). The results of 1000 simulations performed for all effect variables are plotted in the cost-effectiveness plane in Figures 24-26.

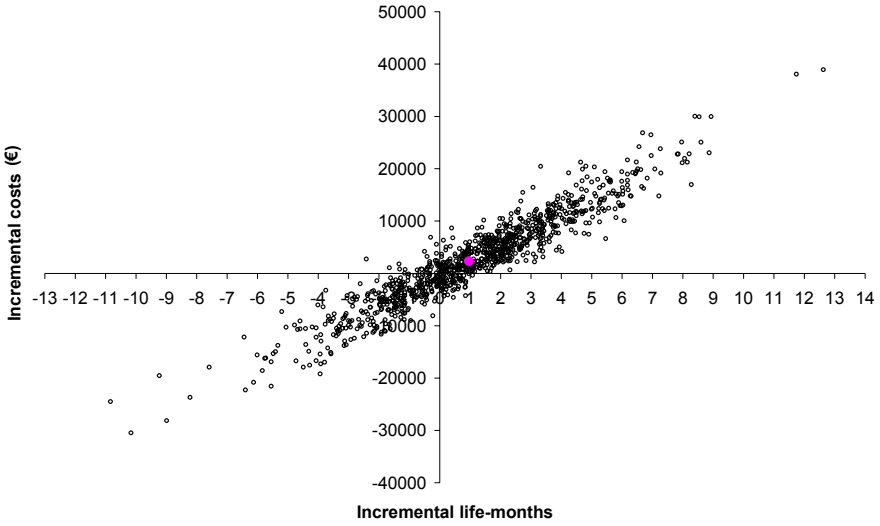


Figure 24. Scatter-plot of mean cost and effect differences (life-months)

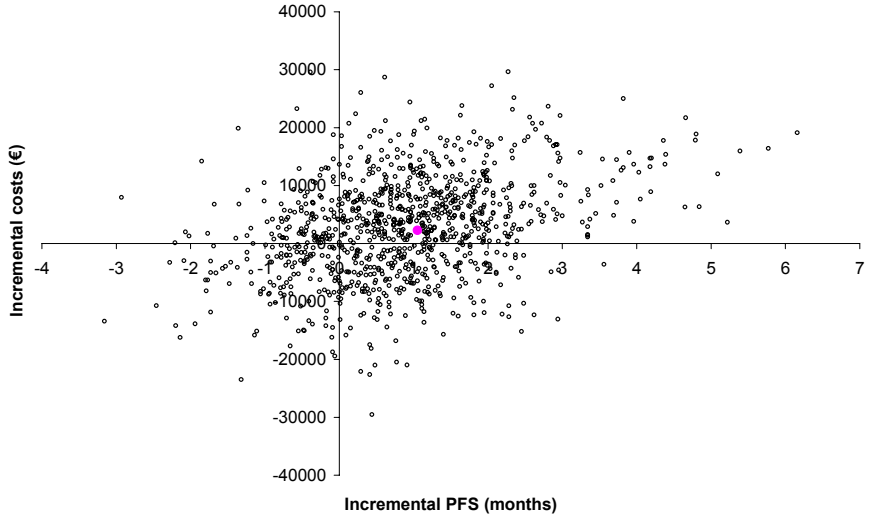


Figure 25. Scatter-plot of mean cost and effect differences (progression-free life-months)

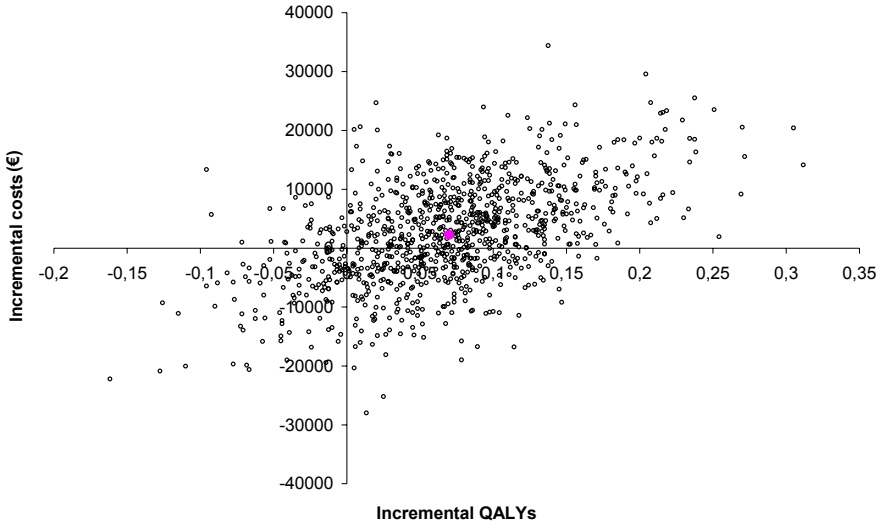


Figure 26. Scatter-plot of mean cost and effect differences (QALYs)

The mean values from the 1000 simulations for costs and effects with TMZ and PCV treatments are shown in Table 10. Additional costs for each extra life-month gained with TMZ treatment are € 2367 (i.e. € 28 404 per gained life-year), each extra progression free life-month gained with TMZ treatment are € 2165 (i.e. € 25 980 per gained progression-free life-year), and each extra QALY gained with TMZ treatment are € 32 471.

Table 10. Mean and median effects and costs of 1000 simulations

	Overall survival (months)		Progression free survival (months)		Quality adjusted life months	QALYs	Mean costs (€)
	Mean	Median	Mean	Median	Mean	Mean	
TMZ	12.07	11.92	4.74	4.61	2.98	0.25	35 380
PCV	11.11	10.96	3.69	3.59	2.14	0.18	33 107

The acceptability curve approach was used to explore the uncertainty related to different levels of societal λ for each gained extra life-month, extra progression-free life-month and extra quality adjusted life-month. The value of λ was varied between 0 and € 100 000. The acceptability curves measured for extra life-months, extra progression-free months, and extra quality adjusted life-months for TMZ, compared to PCV, is presented in Figure 27.

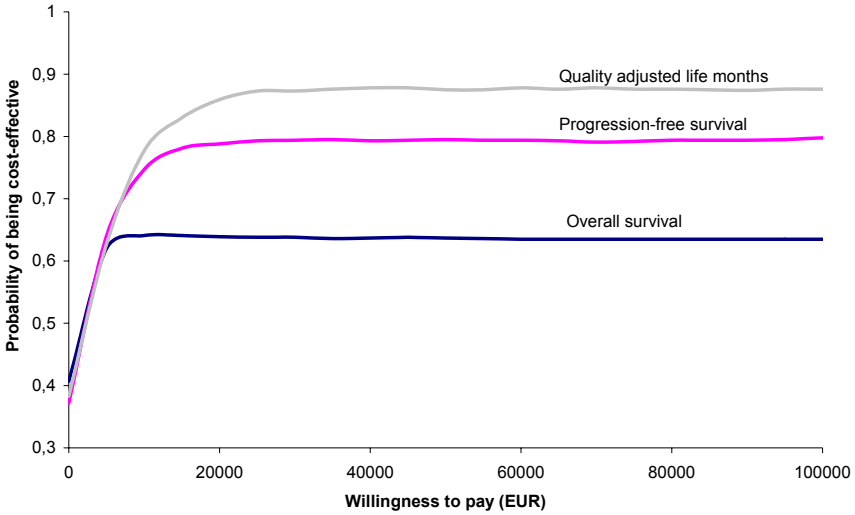


Figure 27. Acceptability curves for TMZ compared to PCV conditional to measured endpoint

As is shown in Figure 27, the probability of TMZ being more cost-effective than PCV is over 60 per cent at all values of willingness-to-pay per gained life-month above € 5000. Similarly, the probability of TMZ being more cost-effective than PCV is over 75 per cent at all values of willingness-to-pay per gained progression-free life-month above € 10 000 and about 85 per cent at all values of willingness-to-pay per gained quality adjusted life-month above € 20 000.

Value of additional research. The costs of uncertainty and the potential efficiency of additional research was characterised using the EVPI approach. Based on incidence and prevalence estimates the discounted Finnish high-grade glioma population who could benefit from the TMZ treatment was estimated to be 1410 until the year 2012. The population EVPI as a function of λ is presented in Figure 28.

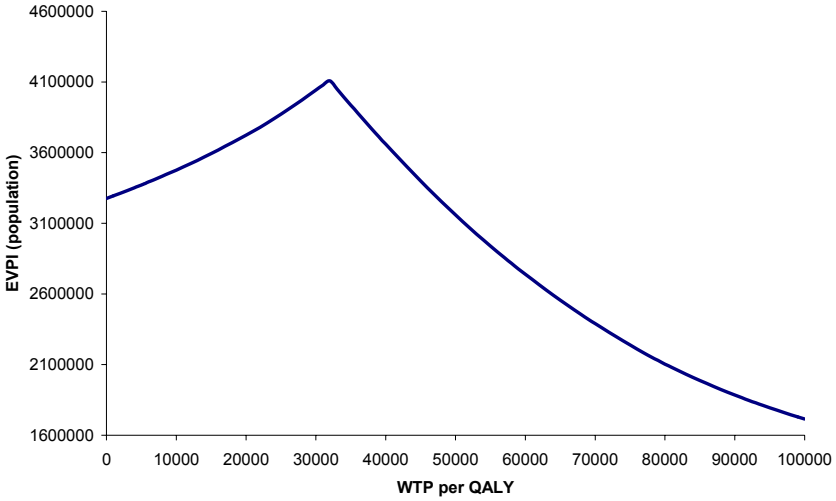


Figure 28. The population EVPI as a function of willingness to pay (λ) per additional QALY

As shown in Figure 28, the EVPI for the population treatment decision reaches its peak when the monetary benefit is 0 (this occurs when λ is € 32 471). At this point, the uncertainty of the decision is as high as it can be since the incremental net monetary benefit (INB) of 0 indicates that the mean benefits of the two treatments are equal and, in some cases (when the distribution of INB is symmetric), it also implies a probability of 0.5 of making a wrong choice between the treatments. However, in the current case, the distribution of INB at this point is slightly positively skewed and the probability interpretation mentioned above does not hold. The population EVPI at this point is approximately € 4.1 million. This represents the maximum value of acquiring information and, if the fixed costs of proposed research are below this EVPI value, additional research is potentially cost-effective. As the value of λ increases, the TMZ becomes more cost-effective (this happens because the TMZ has positive health effects over the PCV and, with a higher value placed on that benefit, then the TMZ option will be favoured) and, hence, the uncertainty surrounding the decision decreases.

6.2.5 Discussion

In the current study, we have demonstrated the cost-effectiveness of TMZ in the treatment of GBM. There are few, if any, published cost-effectiveness studies of various treatments for GBM. Thus, it is difficult to compare the results of the current study to other studies associated with such treatments. Only one review by Dinnes et al. (2001) on the cost-effectiveness of TMZ compared to PCV was found, providing a speculative cost estimate of £42 920 (€ 64 223[1]) per gained quality adjusted life year with TMZ treatment. As our estimate of each extra quality adjusted life-years was € 32 471 in the Finnish health care system, the estimate by Dinnes et al. (2001), although studied within the British health care

system, is almost two times higher. The difference in the cost-effectiveness estimates seems to be attributable to the differing costs of treatments since Dinnes et al. (2001) estimated that the difference of benefits (QALY) between TMZ and PCV is 0.09, whereas in this study the difference in QALYs gained was only 0.07. Due to lack of data, Dinnes et al. (2001) did not take into account the costs incurred after the progression of disease. Their estimate of the costs of PCV treatment before progression is far lower than the costs found in our study. These differences are due to the following factors: assumption that antiemetics are used for 5 days for TMZ patients and for 3 days for PCV patients, omission of hospital days for both treatments, omission of laboratory tests for PCV treatment and adding oncology visits only to TMZ treatment. However, according to our resource use data, more laboratory tests, oncology visits and hospital days were needed for PCV-treated patients compared to TMZ-treated patients. Also the assumption that the need for antiemetics is only related to CCNU in PCV-treated patients is questionable since nausea and vomiting are also side-effects of procarbazine (Yung et al. 2000), which is taken during 14 days in a treatment cycle as well as vincristine (according to the summary of product characteristics).

There are some problematic issues related to this study. First, the resource use data in this study was collected from only two university hospitals. The resource use can vary from one hospital to another, but we consider these two institutions to represent a reasonable estimate of the resources used in Finland in the treatment of high-grade gliomas. On the other hand, these data were collected from patients treated between the same timeframe (1998-2000), which, for its part, increases the validity of these estimates.

Secondly, the effectiveness of the two treatment options were obtained from only three (TMZ) and four (PCV) published reports, with total adjusted sample sizes of 111 and 113 subjects, respectively. The two comparative studies, which, because of their adjusted sample sizes, were also crucially decisive in the estimates of treatment effectiveness were by Yung et al. (2000) and Brandes et al. (2003) In the case of overall survival, Yung et al. (2000) report an OS that is 1.5 months better in the TMZ-treated group than in procarbazine-treated group. Furthermore, Brandes et al. (2003) suggest the OS of TMZ-treated patients is 2.2 months greater than that of PCV-treated patients. Although Brandes et al. (2003) calculated the OS starting from surgery, this does not distort the results since results from this study are accounted for in both treatment options. In the current study, the difference in OS is again in favour of TMZ-treated patients. The magnitude of the estimate is only one month and, therefore, it can be stated that the OS estimates do not overrate the situation for TMZ when compared to PCV.

Also, when the effectiveness indicator is PFS, Yung et al. (2000) report a difference of 2.9 months in favour of TMZ (vs. procarbazine alone) and Brandes et al. (2003) 3.8 months (vs. PCV). In the current study, the PFS estimate for TMZ is approximately one month better than that obtained with PCV. Therefore, in this study, the PFS estimates do not give undeserved credit to TMZ.

When the effectiveness was measured as QALYs, there are some more severe limitations to this study. First, the patients did not evaluate their own well-being. Instead, 6 neuro-oncologists were were asked, based on descriptions of average patient symptoms, to estimate the patients' quality of life on a scale of 0 to 1. Their evaluations of the QOL were considerably lower for the patients in the preoperative state of their treatment than patients themselves have elsewhere reported (0.55 vs. 0.84) (Salo et al. 2002) This was to be expected, since physicians are known to evaluate the health state of their patients as being worse than the patients themselves (Spangers & Aaronson 1992, Addington-Hall & Kaira 2001). Also, it has been previously noted that QOL scores are higher when measured with a 15D-questionnaire than a VAS-scale (Rissanen et al. 1995). Secondly, the disease-states that were described in the questionnaire were the views of a single, although experienced, neuro-oncologist. These states could be described differently by a different expert. Nevertheless, the gained QALY estimates in this study are, again, not as high as previously assumed by Dinnes et al. (2001) (0.07 vs. 0.09).

Uncertainty related to model's inputs were handled simultaneously using probabilistic sensitivity analysis (PSA). PSA offers a method to translate a parameter uncertainty into a decision uncertainty, which can be characterised using acceptability curves. PSA seems to be a valid method to handle uncertainty when the model's inputs are gathered from heterogeneous sources, despite the fact that correlations between the inputs are often unavailable. The EVPI is a developing approach to prioritise further research and it can offer valid information for decision makers in the near future.

6.2.6 Conclusions

There is a high probability that TMZ is more cost-effective than PCV in high-grade glioma patients when the society is willing to pay at least € 2035 for each month's worth of effectiveness, regardless of the indicator used (life-month, progression free life-month, or quality adjusted life-month). When quality of life aspects are taken to account, TMZ, in addition to prolonging survival, becomes even more cost-effective as compared to PCV in the treatment of high-grade glioma patients.

6.2.7 References

Addington-Hall J & Kalra L. 2001. Measuring quality of life - Who should measure quality of life? British Medical Journal 322(9): 1417-20.

Boiardi A, Silvani A, Eoli M, et al. 2001. PCV chemotherapy for recurrent glioblastoma multiforme. Neurology 56(12): 1782.

Brada M, Hoang-Xuan K, Rampling R, et al. 2001. Multicenter phase II trial of temozolomide in patients with glioblastoma multiforme at first relapse. Annals of Oncology 12(2): 259-66.

Brandes AA, Vastola F, Basso U, et al. 2003. A Prospective Study of Glioblastoma in the Elderly. Cancer 97(3): 657-62.

Brazier J, Green C, McCabe C, et al. 2003. Use of visual analog scale in economic evaluation. Expert Review in Pharmacoeconomics & Outcomes Research 3(3): 293-302.

Briggs AH, Goeree R, Blackhouse G, et al. 2002. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Medical Decision Making 22(4): 290-308.

Briggs A & Gray A. 1998. The distribution of health care costs and their statistical analysis for economic evaluation. Journal of Health Service Research & Policy 3(4): 233-45.

Briggs A & Sculpher M. 1998. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 13(4): 397-409.

Buckner JC, Gesme D, O'Fallon JR, et al. 2003. Phase II trial of procarbazine, lomustine, and vincristine as initial therapy for patients with low-grade oligodendroglioma or oligoastrocytoma: efficacy and associations with chromosomal abnormalities. Journal of Clinical Oncology 21(2): 251-5.

CBTRUS. Statistical report: Primary brain tumours in the United States, 1995-1999. 2002. Published by the Central Brain Tumour Registry of the United States.

Claxton K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 18(3): 341-64.

Claxton K, Neumann PJ, Araki S, et al. 2001. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. International Journal of Technology Assessment in Health Care 17(1): 38-55.

Dinnes J, Cave C, Huang S, et al. 2001. The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review. Health Technology Assessment 5(13). Available at: http://www.ncchta.org/fullmono/mon513.pdf

Doubilet P, Begg CB, Weinstein MC, et al. 1985. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Medical Decision Making 5(2): 157-77.

Feeny D, Furlong W, Boyle M, et al. 1995. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics 7(6): 490-502.

Fenwick E, Claxton K, Sculpher M, et al. 2000. Improving the efficiency and relevance of health technology assessment: the role of iterative decision analytic modelling. York: Centre for Health Economics.

Gelman A, Carlin J, Stren H, et al. 1995. Bayesian data analysis. London: Chapman & Hall.

Hujanen T. 2003. Terveydenhuollon yksikkökustannukset Suomessa vuonna 2001. Helsinki: Stakes Aiheita 1/2003

Kappelle AC, Postma TJ, Taphoorn MJB, et al. 2001. PCV chemotherapy for recurrent glioblastoma multiforme. Neurology 56(1): 118-20.

Karnon J. 2002. Planning the efficient allocation of research funds: an adapted application of a non-parametric Bayesian value of information analysis. Health Policy 61(3): 329-47.

Levin VA, Hess KR, Choucair A, et al. 2003. Phase III Randomized Study of Postradiotherapy Chemotherapy with Combination □-Difluoromethylornithe-PCV versus PCV for Anaplastic Gliomas. Clinical Cancer Research 9: 981-90.

Levin VA, Wara WM, Davis RL, et al. 1985. Phase III comparison of BCNU and the combination of procarbazine, CCNU, and vincristine administered after radiotherapy with hydroxyurea for malignant gliomas. Journal of Neurosurgery 63: 218-23.

Medical Research Council Brain Tumour Working party. 2001. Randomized trial of procarbazine, lomustine, and vincristine in the adjuvant treatment of high-grade astrocytoma: a Medical Research Council trial. Journal of Clinical Oncology 19(2): 509-18.

Miller D & Homan S. 1994. Determining transition probabilities: confusion and suggestions. Medical Decision Making 14(1): 52-8.

Newlands ES, Stevens MF, Wedge SR, et al. 1997. Temozolomide: a review of its discovery, chemical properties, pre-clinical development and clinical trials. Cancer Treatment Reviews 23(1): 35-61.

O'Hagan A, Stevens JW & Montmartin J. 2000. Inference for the cost-effectiveness acceptability curve and cost-effectiveness ratio. Pharmacoeconomics 17(4): 339-49.

Prados MD, Scott C, Sandler H, et al. 1999. A phase 3 randomized study of radiotherapy plus procarbazine, CCNU, and vincristine (PCV) with or without BUdR for the treatment of anaplastic astrocytoma: a preliminary report of RTOG 9404. International Journal of Radiation Oncology • Biology • Physics 45(5): 1109-15.

Rissanen P, Sintonen H & Pekurinen M. 1995. 15D Terveyteen liittyvän elämänlaatumittarin, visuaalisen analogiamittarin ja koetun terveydentilan mittarin arvot aikuisikäisessä normaaliväestössä. Sosiaalilääketieteellinen aikakauslehti 32(3): 207-11.

Salo J, Niemelä A, Joukamaa M, et al. 2002. Effect of brain tumour laterality on patients' perceived quality of life. The Journal of Neurology, Neurosurgery, and Psychiatry 72: 373-7.

Sonnenberg FA & Beck JR. 1993. Markov models in medical decision making: a practical guide. Medical Decision Making 13: 322-38.

Sprangers M & Aaronson N. 1992. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. Journal of Clinical Epidemiology 45(7): 743-60.

Stinnett AA & Mullahy J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 18: S68-S80.

Stupp R, Dietrich P-Y, Ostermann Kraljevic S, et al. 2002. Promising survival for patients with newly diagnosed glioblastoma multiforme treated with concomitant radiation plus temozolomide followed by adjuvant temozolomide. Journal of Clinical Oncology 20(5): 1375-82.

Tambour M, Zethraeus N & Johannesson M. 1998. A note on confidence intervals in cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 14(3): 467-71.

Tan SB, Dear KBG, Bruzzi P, et al. 2003. Strategy for randomised clinical trials in rare cancers. British medical Journal 327(7405): 47-9.

Torrance GW, Feeny D & Furlong W. 2001. Visual analog scales: do they have a role in the measurement of preferences for health states? Medical Decision Making 21(4): 329-34.

Valtioneuvoston asetus lääketaksasta [Council of States act on price list for drugs], 2002

Yung WK, Albright RE, Olson J, et al. 2000. A phase II study of temozolomide vs. procarbazine in patients with glioblastoma multiforme at first relapse. British Journal of Cancer 83(5): 588-93.

Yung WK, Prados MD, Yaya-Tur R, et al. 1999. Multicenter Phase II trial of temozolomide in patients with anaplastic astrocytoma or oligoastrocytoma at first relapse. Journal of Clinical Oncology 17(9): 2762-71.

6.3 Synthesising evidence and modelling the cost-effectiveness of plant stanol esters in the prevention of coronary heart disease employing the comprehensive decision modelling approach[8]

6.3.1 Introduction

Coronary heart disease (CHD) is a major cause of death in many Western countries and is becoming an important cause of morbidity and mortality worldwide (WHO 2003) In Finland, CHD is the leading cause of death for both men and women and it accounts for a substantial share of health care costs. In 2003, the total health care costs attributable to CHD were estimated to be 286 million euros. (Petersen et al. 2005)

One of the major risk factors that predispose an individual to CHD is an elevated serum cholesterol concentration which is a risk which can be modified by changes in diet. Incorporating foods enriched with plant stanols or sterols into the daily diet can substantially enhance the cholesterol-lowering effect of diet. Controlled trials have demonstrated that daily intake of 2 grams of stanols or sterols can reduce low-density lipoprotein (LDL) cholesterol by about 10 %. (Law 2000, Katan et al. 2003) Considered from both the viewpoint of the individual patient and the whole population, this reduction is significant. A meta-analysis of cohort studies showed that a long term reduction in serum cholesterol concentration of 0.6 mmol/l (i.e. about 10 %), lowers the risk of ischemic heart disease by 50% at age 40, falling to 20% at age 70. (Law et al. 1994)

The cost-effectiveness of prevention strategies for CHD has been widely studied. The evaluated interventions have included primary prevention with statins (Caro et al. 1997), smoking cessation counselling (Cummings et al. 1989), dietary advice (Lindgren et al. 2003) and exercise (Lindgren et al. 2003, Hatziandreu et al. 1988). However, the cost-effectiveness of plant stanol and sterol esters has remained unevaluated though the use of spreads containing plant stanol or sterols as part of a healthy diet could potentially help to reduce the incidence of CHD and in that way impact on the associated health care costs.

In the present study, we have evaluated the cost-effectiveness (€/QALY) of plant stanol ester incorporated in spread in the prevention of CHD without and with the combination of HMG-CoA reductase inhibitors (statins). In the current study, cardiovascular prevention is understood as a reduction of absolute risk for CHD, irrespective of clinical stage. We assessed the cost-effectiveness of plant stanol ester

[8] This chapter has been published in: Martikainen J, Ottelin A-M, Kiviniemi V & Gylling H. Plant Stanol Esters are Potentially Cost-Effective in the Prevention of Coronary Heart Disease in Men: Bayesian Modelling Approach. Eur J Cardiovasc Prev Rehabil 2007;14:265-272. Reproduced with permission.

enriched daily diet compared to normal daily diet from the society's point of view. However, productivity changes due to the intervention are not considered in our analysis.

6.3.2 Methods

Clinical efficacy. We carried out two meta-analyses, which were based on systematic literature reviews. The first concentrated on studies examining total cholesterol reduction with plant stanol ester added to the spread. The second review was carried out to obtain information about reducing total cholesterol with the combination of plant stanol ester added spread and statin drug treatment. We selected the change in total cholesterol as a primary endpoint in meta-analyses, since all our CHD risk prediction algorithms applied in a model included total cholesterol as a risk factor.

We identified the trials from Medline and Cochrane collaboration database (to December 2004), and previous review articles (Law 2000, Katan et al. 2003). We included all randomised placebo-controlled trials, irrespective of participants' sex, age or disease. Participants in most trials were healthy with above average lipid concentrations. We excluded trials that used free stanols or some other product than spread, or participants that were children or had ileostoma. Trials using plant stanol ester and statin treatment in combination were utilised only in the second meta-analysis. The efficacy of plant stanol ester was defined as the reduction in serum total cholesterol concentration, expressed as the change from the placebo period in the treated group in cross-over studies or versus placebo groups in studies with parallel design.

Due to perceived differences in the trial designs, methods and patient characteristics, we applied a Bayesian random effects model to estimate the summary measure for the placebo adjusted effect size for total cholesterol. (Schmid 2001, Smith et al. 1995) To adopt a full Bayesian approach, we specified prior distributions for the overall pooled effect size parameter and the between-trial heterogeneity parameter. We used non-informative prior distributions, however, to ensure that the data from the trials dominated the final inferences. Furthermore, we used the random-effects meta-regression approach to determine whether other recorded factors, such as trial size, dose, trial duration, age, and the baseline cholesterol level of participants would modify the effect size for total cholesterol (Thompson & Higgins 2002).

Decision model. We constructed a discrete-state discrete-time Markov model to estimate the expected costs and health outcomes in terms of gains in quality adjusted life years (QALYs) associated with plant stanol ester enriched daily diet and normal daily diet among hypothetical cohorts of Finnish men and women at a specific age who were initially without established CHD. We carried out the estimations for men and women separately at four different initial ages (i.e. 30, 40, 50 and 60 years) at which the regular use of STA as a part of daily diet was assumed to be started.

The structure of the Markov model was based on a previously published CHD model (Cook et al. 2004).The simplified flow chart of the Markov model is illustrated in Figure 29. Our Markov model used an annual cycle length. Each year, the hypothetical cohort of men or women without established CHD were at risk of having of a fatal non-cardiovascular event, a fatal CHD event, an acute non-fatal CHD event, or they might survive to the next year without the occurrence of any CHD event. After that year, the subjects' risk factors were updated based on Finnish age- and sex-specific risk factor profile data and the Markov model was run again until all subjects in the cohort were entered into terminal states or until 100 years of age was reached. The age of 100 years was determined to be the maximum allowed subject age in the Markov model.

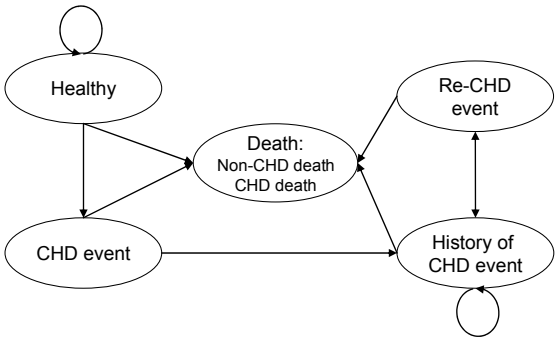


Figure 29. Simplified illustration of the Markov model for outcomes. CHD, Coronary Heart Disease. Transition probabilities conditional to age and sex between defined health states were derived from life-tables and risk functions based on FINRISK and 4S studies.

Assessing the risks of CHD and non-CHD events. We used data derived from the FINRISK (Laatikainen et al. 2002a, Laatikainen 2002b) and Health 2000 (Aromaa & Koskinen 2000) -studies to determine age- and sex-specific risk factor profiles (including total serum cholesterol level [mmol/l], systolic blood pressure [mmHg], high-density lipoprotein cholesterol level [mmol/l], smoking prevalence [%], and diabetes prevalence [%]) for the cohorts of interest in the Markov model. We estimated the age- and sex-specific annual risk of non-CHD death cause by subtracting down the total mortality (derived from Finnish standard all-cause mortality life-tables by age and sex) by the fraction of deaths due to cardiovascular diseases (ICD-10: I20-I25, I46, R96, and R98). (Statistics Finland 2002, National Public Health Institute 2004)

We estimated the annual total risk of initial non-fatal CHD event (ICD-10: I20-I21) or CHD death (ICD-10: I20-I25) using the modified (i.e. the risk of events were estimated per annum instead of over 10-year periods) FINRISK risk function (Bhopal et al. 2005). We used age- and sex-specific probabilities obtained from the National Cardiovascular Disease Register to allocate the total FINRISK risk function predictions into the two health states (hospitalisation due to non-fatal CHD event or CHD death) in the Markov model (National Public Health Institute 2004). After the initial occurrence of the non-fatal CHD event, the model was programmed to transit subjects to the secondary prevention part of the Markov

model (i.e. the history of CHD event health state). We estimated the annual risk of the subsequent events using risk function derived from the 4S study (Johannesson et al. 1997). Finally, we combined the pooled estimates of reduction in total serum cholesterol (mmol/l) from the meta-analyses with FIN-RISK and 4S risk functions to estimate reduction in the annual risk of CHD events in subjects using stanol ester.

Costs. We estimated all costs from a societal perspective including the direct costs of prevention and morbidity. All costs were estimated in year 2001 euros. When CHD developed in subjects, the model tracked the costs due to hospitalisation and other treatments. The average age- and sex-specific costs of hospitalisation and outpatient care were based on a Finnish dataset comprising 9 226 patients. (Häkkinen et al. 2002)

We programmed the Markov model assuming that after the discharge subjects received a prescription for chronic CHD or dyslipidemia medication. In year 2001, the average annual medicine cost of treating chronic CHD was 163€ in both men and women based on National Agency for Medicines' and the Social Insurance Institution's databases (2003). The average annual costs for treating dyslipidaemia associated with chronic CHD were 438€ in men and 416€ in women. For modelling purposes, we estimated the average annual medicine costs. The average annual medicine costs were estimated by weighting the average costs by the proportion of recorded patients in both medicine categories in Finland. The weighted average annual costs of medicines were 239€ and 225€ in men and women, respectively.

In the model, plant stanol ester was assumed to be incorporated into a spread. The cost of STA spread (Raisio Benecol®) and corresponding regular spread (Raisio Keiju® light) were obtained from a survey by The Finnish Consumer Agency and the state provincial offices. (Finnish Consumer Agency 2004) The average price for plant stanol ester spread was estimated to be 13.2€ per kilo and the price of recommended daily dose (2 g) 120€ per year. The price of using plant stanol ester spread was compared with the price of using regular spread that is equivalent to the plant stanol ester spread but lacks the added plant stanols. The additional cost of using the recommended dose of plant stanol ester spread per day was estimated to be 97€ per year.

Quality of life. We obtained the age- and sex-specific quality of life for the population of interest without acute CHD events from a Finnish EQ-5D survey (n=2 374). (Ohinmaa et al. 1996) When a CHD event occurred in subjects, the model tracked the decreases in quality of life due to the CHD event. The disutility due to the initial or subsequent CHD event was estimated to be approximately -0.078 (SD 0.245) in men and -0.127 (SD 0.247) in women as measured by the EQ-5D instrument from a dataset of 615 patients who underwent coronary artery bypass grafting (CABG) or percutaneous transluminal coronary angioplasty (PTCA) in Finland. (Kattainen 2004)

We assumed that after 12 months, the postoperative quality of life had linearly improved to the same level as in the age- and sex-matched general population (Kattainen 2004). The use of STA as such was not assumed to affect the quality of life.

Discounting. In the base case scenario, we discounted costs and quality adjusted life years at 3.5% per annum to generate the present value of future costs and health benefits. (NICE 2004) In terms of sensitivity analyses, the results are presented also when costs and health benefits are discounted at 0% and 5% (as recommend by the Finnish health economic guidelines) per annum.

Uncertainty analysis. We performed the meta-analyses and uncertainty simulations within a Bayesian modelling framework. The Bayesian modelling framework offers a coherent approach to synthesise all available sources of evidence into a single model. (Spiegelhalter & Best 2003) The advantages of the Bayesian modelling framework have been discussed in more detail elsewhere (Cooper et al. 2004, Samsa et al. 1999). We estimated the joint posterior distribution of the model parameters by simulation using the Gibbs sampler programmed in WinBUGS software (version 1.4, MRC Biostatistics Unit, Cambridge, UK). Final posterior parameter estimates were based on a total of 10000 Markov Chain Monte Carlo (MCMC) samples. The first 2000 samples were discarded to ensure stability of the posterior sampling procedure. Results are reported with 95% credibility intervals (CrI), analogous to confidence intervals from a frequentist approach.

The decision as to whether the use of stanol ester incorporated in spread is cost effective in the prevention of CHD depends on the decision makers' maximum willingness to pay for additional QALY gained with the use of stanol ester added spread. To illustrate this decision uncertainty, we constructed cost-effectiveness acceptability curves for the cost per QALY gained (Fenwick et al. 2004).

6.3.3 Results

Clinical efficacy. The primary clinical outcome was a reduction in total serum cholesterol levels (mmol/l) (Table 11). The first meta-analysis included a total of 19 randomised trials. All these trials were placebo controlled and their design was either parallel or crossover. The trials included a total of 1192 subjects from Europe, Australia, Canada, and Japan with the trial durations ranging from 3 to 52 weeks, and stanol doses of 0.8 to 3.4 g/day.

Table 11. Description of the studies included in the meta-analysis

Reference	No of subjects in treatment group/ placebo group	Mean age	Duration (weeks)	Dose (g/day)	Placebo adjusted reduction in total cholesterol (mmol/l)
Plat & Mensink (2000)	70/42	33	8	4	-0,39
Niinikoski et al. (1997)	12/12	37	5	3	-0,50
Hallikainen & Uusitupa (1999)	38/17	43	8	2,3	-0,61
Westrate & Meijer (1998)	95 *	45	3,5	2,7	-0,37
Miettinen et al. (1995)	7/8	45	9	1	-0,31
Vanhanen et al. (1994)	34/33	46	6	3,4	-0,37
Vanhanen (1994)	7/8	47	6	0,8	-0,32
				2	-0,50
Hallikainen et al. (2000a)	34 *	49	4	2	-0,57
Miettinen et al. (1995)	51/51	50	52	1,8	-0,44
				2,6	-0,62
Gylling et al. (1997)	22 *	51	7	3	-0,55
Hallikainen et al. (2000b)	22 *	51	4	0,8	-0,17
				1,6	-0,45
				2,3	-0,69
				3	-0,76
Jones et al. (2000)	15 *	52	3	1,8	-0,33
Gylling & Miettinen (1999)	21 *	53	5	2,4	-0,46
Nguyen et al. (1999)	77/76	53	8	2	-0,27
	71/76			3 US	-0,42
	74/76			3 EU	-0,31
Andersson et al. (1999)	19/21	55	8	2	-0,42
Noakes et al. (2002)	46 *	57	3	2,5	-0,46
Gylling & Miettinen (1994)	11 *	58	6	3	-0,36
Plat et al. (2000)	39 *	31	4	2,5	-0,32
Homma et al. (2003)	33/34	47	4	2	-0,34
	34/34			3	-0,29
Vanhanen. (1994)	7/7	55	6	1,5 [a]	-0,17
Gylling & Miettinen (1996)	8 *	60	7	3 [a]	-0,24
Gylling et al. (1997)	10 *	52	12	3 [b]	-0,56
Blair et al. (2000)	71/77	56	8	3 [c]	-0,41
Gylling & Miettinen (2002)	11 *	67	8	2,25 [d]	-0,54

*cross-over studies. [a], combined with pravastatin 40 mg/day. [b], combined with simvastatin 10-20 mg/day. [c], combined with lovastatin, pravastatin, simvastatin or atorvastatin, doses/day not available. [d], combined with simvastatin 20 mg/day.

The random-effect meta-analysis indicated mean difference of 0.362 mmol/l (95%CrI 0.31 to 0.41) when comparing the effect of plant stanol ester and placebo in lowering the total cholesterol level. A corresponding mean percentage change in total cholesterol level was 6% (95%CrI 5.1 to 6.8). Based on the results of random effect meta-regression, none of the selected trial-level covariates were significant confounders. Thus, these results ensured the appropriateness of data synthesis for the pooled analysis.

The second meta-analysis included five trials of reducing total cholesterol with the combination of plant stanol ester and statin treatment (Table 12) (Vanhanen 1994, Gylling & Miettinen 1996, Gylling et al. 1997, Blair et al. 2000, Gylling & Miettinen 2002). Of the studies, three trials were placebo controlled and in two studies, the effect of STA and statin treatment was compared with the baseline diet. The studies included 191 subjects from Europe and USA. The statins used in the studies were simvastatin, pravastatin, atorvastatin and lovastatin at doses of 10-40 mg/day. Pooled results from the combination treatments showed that the plant stanol ester was able to decrease total cholesterol levels more effi-

ciently when combined with statin treatment in comparison to a single drug or plant stanol ester therapy alone. The mean placebo-adjusted decrease in total cholesterol level was 0.385 mmol/l (95%CrI 0.18 to 0.61). The corresponding mean percentage change in total cholesterol was 6.2% (95%CrI 2.9 to 9.9).

Cost-effectiveness. The primary economic outcome was cost (€) per QALY. Table 12 shows the mean cost per QALY estimates in different age groups for men and women. The use of different discount rates for both costs and benefits markedly altered the cost per QALY gained by 19-56 % (9-46 %) this being dependent on age and sex. The costs of plant stanol ester were partly offset due to savings in the health care sector. Discounted lifetime cost savings in the health care sector ranged approximately from 20.4€ to 124€ per subject conditional on age and sex, if the cost of plant stanol ester was ignored. These saving estimates are real savings to health care providers, since the cost of the plant stanol ester spread is actually paid by the consumer.

Table 12. Cost per quality adjusted life year gained (€/QALY) with plant stanol esters as compared to normal diet

Subgroup	Base case (3.5%)	0% discount rate for costs and benefits	5% discount rate for costs and benefits
Men			
30	20 999	9 742	28 101
40	14 554	8 959	17 399
50	10 106	7 254	10 772
60	7 436	6 010	8 104
Women			
30	112 151	49 090	163 255
40	75 289	39 683	99 738
50	50 043	31 733	60 857
60	34 327	25 602	38 889

Figures 30 and 31 show the cost-effectiveness acceptability curves in different age groups for men and women in the base case. An incremental cost-effectiveness ratio (ICER) of about 30 000€ - 50 000€ per QALY gained is likely to be considered cost-effective by the National Institute for Health and Clinical Excellence (NICE) in UK. (NICE 2005) If the decision makers' maximum willingness to pay per QALY gained is assumed to be this upper value, then the use of plant stanol ester is cost-effective for men in all age groups. For women at the age of 50 years and younger there is a less than 15% probability, whereas at the age of 60 years and older there is an over 90% probability, that the use of STA would be considered cost-effective, if the NICE's scale of willingness to pay is used.

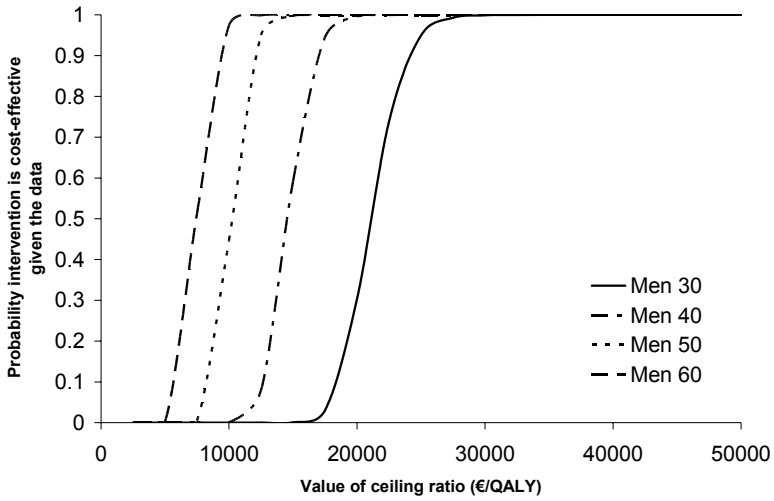


Figure 30. Cost effectiveness acceptability curves for men applying 3.5% discount rate. Cost effectiveness acceptability curve shows probability that plant stanol ester in spread is cost effective as compared to daily diet with regular spread for a range of decision makers' maximum willingness to pay (a ceiling ratio) for a quality adjusted life year (QALY).

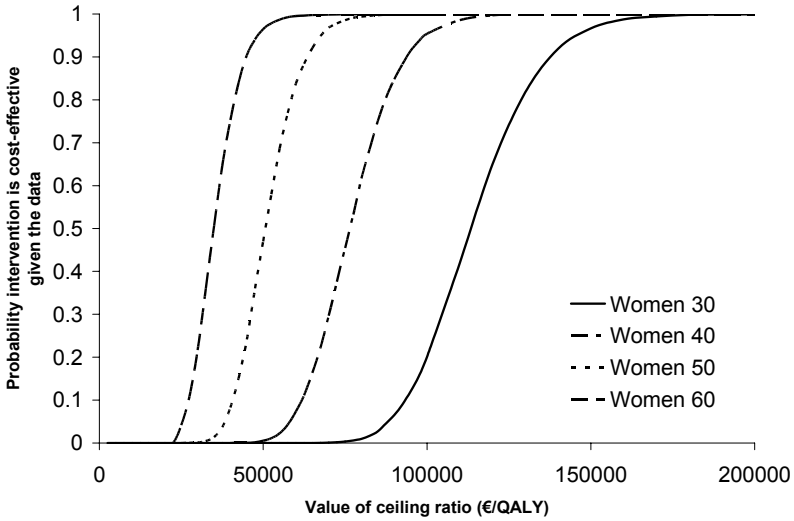


Figure 31. Cost effectiveness acceptability curves for women applying 3.5% discount rate

6.3.4 Discussion

Our study suggests that the use of plant stanol ester incorporated in spread is a cost-effective option in the prevention of CHD in all adult males and in women in older age-groups with average total serum cholesterol levels of 5 mmol/l or higher. The study results show that the regular use of plant stanol ester in the prevention of CHD yields an incremental cost-effectiveness ratio (ICER) which ranges from 7 436€ to 112 151€, conditional on age and sex. The cost-effectiveness of plant stanol ester increases in older age-groups, since age is the single strongest predictor of CHD risk (Avins and Browner 1998). Actually, it seems that, at least in Finland, plant stanol ester spread is mostly used by those age groups where the cost-effectiveness is highest (i.e. the average user is aged 55 years or over) (Anttolainen et al. 2001) The use of plant stanol ester reveals significantly lower ICERs for men than for women in all age groups. However, no difference in the cholesterol-lowering response to plant stanol administration between genders has been reported (Vanhanen et al. 1993, Miettinen & Vanhanen 1994, Vanhanen et al. 1994). The gender difference in ICER is probably due to the fact that CHD is markedly more common in men than in women (Jousilahti et al. 1999). For both sexes, CHD risk increases with age but the increase is sharper for women which can be also seen in the results of this analysis.

For the purposes of this study, the adherence to using plant stanol ester spread was not examined and the consumed amount was assumed to be 25 grams daily (2 grams stanol /day). The adherence in normal practice is likely to be less than perfect. (Luoto et al. 2004) However, changing from regular spread to plant stanol ester spread as the daily spread is a very minor change in dietary habits when compared to many other interventions recommended to reduce CDH risk factors (e.g. smoking cessation, increasing physical activity) that require an active effort to modify the lifestyle. The effect of somewhat poorer adherence may be balanced, to some extent, by the fact that 5% of Finnish adult population uses butter as their spread and 16% use mixtures of butter and plant oils. (Laatikainen 2002b) When used as a substitute for butter, plant stanol ester spread will evoke an even greater reduction in cholesterol levels.

The current analysis was based on a Finnish health care setting and, therefore, the study results might not be directly transferable to other countries. Cost-effectiveness estimates might vary between countries e.g. due to differences in CHD risk profiles, health care resource use, unit costs, and health state utilities. However, we believe that the study results are at least somewhat relevant to other Western Countries having populations with intermediate CHD mortality rates. (Kromhout 2001)

The study results are based on the assumption that changes in serum cholesterol levels could be converted to changes in the incidence of CHD events via the CHD risk equations. It is generally the case that economic models are able to predict fairly accurately the incidence of CHD events experienced by hypothetical individuals fulfilling selected entry criteria. (Morris 1997) However, in the future, it would be

valuable if the long-term cost-effectiveness of plant stanol ester could be proved also in a controlled and randomised trial.

6.3.5 Conclusions

Based on the results presented here, the recommendation that plant stanol ester incorporated in spread be used as a part of daily diet instead of regular spread could be viewed as cost-effective public health policy in the prevention of CHD in all adult males and in older age-groups of women with total serum cholesterol levels of 5 mmol/l or greater. The use of plant stanol ester in spread could also be seen as a potentially cost effective option in the prevention of CHD when compared to previously published cost-utility estimates (University of Oxford 2005).

6.3.6 References

Andersson A, Karlström B, Mohsen R & Vessby B. 1999. Cholesterol-lowering effects of a stanol ester-containing low-fat margarine used in conjunction with a strict lipid-lowering diet. European Heart Journal Supplements 1 (Suppl S): 80-90.

Anttolainen M, Luoto R, Uutela A, Boice JD, Jr, Blot WJ, McLaughlin JK, et al. 2001. Characteristics of users and nonusers of plant stanol ester margarine in Finland: an approach to study functional foods. Journal of the American Dietetic Association 101: 1365-8.

Aromaa A & Koskinen S, ed. 2002. Health and functional capacity in Finland. Baseline results of the Health 2000 health examination survey. Publications of National Public Health Institute B3/2002. Helsinki: National Public Health Institute.

Avins AL & Browner WS. 1998. Improving the prediction of coronary heart disease to aid in the management of high cholesterol levels: what a difference a decade makes. Journal of The American Medical Association 279: 445-9.

Bhopal R, Fischbacher C, Vartiainen E, Unwin N, White M & Alberti G. 2005. Predicted and observed cardiovascular disease in South Asians: application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data. J Public Health (Oxf) 27: 93-100.

Blair SN, Capuzzi DM, Gottlieb SO, Nguyen T, Morgan JM, Cater NB. 2000. Incremental reduction of serum total cholesterol and low-density lipoprotein cholesterol with the addition of plant stanol ester-containing spread to statin therapy. American Journal of Cardiology 86:46-52.

Caro J, Klittich W, McGuire A, Ford I, Norrie J, Pettitt D, et al. 1997. The West of Scotland coronary prevention study: economic benefit analysis of primary prevention with pravastatin. British Medical Journal 315: 1577-82.

Cook JR, Yin D, Alemao E & Drummond M. 2004. Development and validation of a model to project the long-term benefit and cost of alternative lipid-lowering strategies in patients with hypercholesterolaemia. Pharmacoeconomics 22 (Suppl 3): 37-48.

Cooper NJ, Sutton AJ, Abrams KR, Turner D & Wailoo A. 2004. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. Health Economics 13: 203-26.

Cummings SR, Rubin SM & Oster G. The cost-effectiveness of counseling smokers to quit. 1989. Journal of The American Medical Association 261: 75-9.

Fenwick E, O'Brien BJ & Briggs A. 2004. Cost-effectiveness acceptability curves -facts, fallacies and frequently asked questions. Health Economics 13: 405-15.

Finnish Consumer Agency. 2004. Kuluttajaviraston hintavertailuja 1/2004: Terveysvaikutteisina markkinoidut elintarvikkeet. Helsinki: Kuluttajavirasto.

Gylling H, Radhakrishnan R & Miettinen TA. 1997. Reduction of serum cholesterol in postmenopausal women with previous myocardial infarction and cholesterol malabsorption induced by dietary sitostanol ester margarine: women and dietary sitostanol. Circulation 96: 4226-31.

Gylling H & Miettinen TA. 1994. Serum cholesterol and cholesterol and lipoprotein metabolism in hypercholesterolaemic NIDDM patients before and during sitostanol ester-margarine treatment. Diabetologia 37: 773-80.

Gylling H & Miettinen TA. 1996. Effects of inhibiting cholesterol absorption and synthesis on cholesterol and lipoprotein metabolism in hypercholesterolemic non-insulin-dependent diabetic men. Journal of Lipid Research 37:1776-85.

Gylling H & Miettinen TA. 1999. Cholesterol reduction by different plant stanol mixtures and with variable fat intake. Metabolism 48:575-80.

Gylling H & Miettinen TA. 2002. LDL cholesterol lowering by bile acid malabsorption during inhibited synthesis and absorption of cholesterol in hypercholesterolemic coronary subjects. Nutrition, Metabolism & Cardiovascular Diseases 12: 19-23.

Hatziandreu EI, Koplan JP, Weinstein MC, Caspersen CJ & Warner KE. 1988. A cost-effectiveness analysis of exercise as a health promotion activity. American Journal of Public Health 78: 1417-21.

Hallikainen MA, Sarkkinen ES, Gylling H, Erkkila AT & Uusitupa MI. 2000a. Comparison of the effects of plant sterol ester and plant stanol ester-enriched margarines in lowering serum cholesterol concentrations in hypercholesterolaemic subjects on a low-fat diet. European Journal of Clinical Nutrition 54: 715-25.

Hallikainen MA, Sarkkinen ES & Uusitupa MI. 2000b. Plant stanol esters affect serum cholesterol concentrations of hypercholesterolemic men and women in a dose-dependent manner. Journal of Nutrition 130: 767-76.

Hallikainen MA & Uusitupa MI. 1999. Effects of 2 low-fat stanol ester-containing margarines on serum cholesterol concentrations as part of a low-fat diet in hypercholesterolemic subjects. American Journal of Clinical Nutrition 69: 403-10.

Homma Y, Ikeda I, Ishikawa T, Tateno M, Sugano M & Nakamura H. 2003. Decrease in plasma low-density lipoprotein cholesterol, apolipoprotein B, cholesteryl ester transfer protein, and oxidized low-density lipoprotein by plant stanol ester-containing spread: a randomized, placebo-controlled trial. Nutrition 19: 369-74.

Häkkinen U, Idänpään -Heikkilä U, Keskimäki I, Rauhala A, Klaukka T & Teitto E. 2002. Akuutin sydäninfarktien hoitokäytäntöjen, kustannusten ja vaikuttavuuden vertailu. Stakes, Aiheita 9/2002. Helsinki: National Research and Development Centre for Welfare and Health (STAKES).

Johannesson M, Jonsson B, Kjekshus J, Olsson AG, Pedersen TR & Wedel H. 1997. Cost-effectiveness of simvastatin treatment to lower cholesterol levels in patients with coronary heart disease. Scandinavian Simvastatin Survival Study Group. New England Journal of Medicine 336: 332-6.

Jones PJ, Raeini-Sarjaz M, Ntanios FY, Vanstone CA, Feng JY & Parsons WE. 2000. Modulation of plasma lipid levels and cholesterol kinetics by phytosterol versus phytostanol esters. Journal of Lipid Research 41: 697-705.

Jousilahti P, Vartiainen E, Tuomilehto J & Puska P. 1999. Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. Circulation 99: 1165-72.

Katan MB, Grundy SM, Jones P, Law M, Miettinen T, Paoletti R, et al. 2003. Efficacy and safety of plant stanols and sterols in the management of blood cholesterol levels. Mayo Clinic Proceedings 78: 965-78.

Kattainen E. 2004. A longitudinal research of the coronary artery bypass grafting and percutaneous transluminal coronary artery angioplasty patients' health-related quality of life. (PhD thesis). Kuopio: Kuopio University Publications. E. Social Sciences, University of Kuopio.

Kromhout D. 2001. Epidemiology of cardiovascular diseases in Europe. Public Health Nutrition 4:441-57.

Laatikainen T, Tapanainen H, Alfthan G, Salminen I, Sundvall J, Leiviskä J, et al. 2003a. FINRISKI 2002. Tutkimus kroonisten kansantautien riskitekijöistä, niihin liittyvistä elintavoista, oireista, psykososiaalisista tekijöistä ja terveyspalvelujen käytöstä. Tutkimuksen toteutus ja tulokset 1. Perusraportti. Publications of the National Public Health Institute B7/2003. Helsinki: National Public Health Institute.

Laatikainen T, Tapanainen H, Alfthan G, Salminen I, Sundvall J, Leiviskä J, et al. 2003b. FINRISKI 2002. Tutkimus kroonisten kansantautien riskitekijöistä, niihin liittyvistä elintavoista, oireista, psykososiaalisista tekijöistä ja terveyspalvelujen käytöstä. Tutkimuksen toteutus ja tulokset 2. Taulukkoliite. Publications of the National Public Health Institute B7/2003. Helsinki: National Public Health Institute.

Law M. 2000. Plant sterol and stanol margarines and health. British Medical Journal 320:861-4.

Law MR, Wald NJ & Thompson SG. 1994. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischemic heart disease? British Medical Journal 308: 367-72.

Lindgren P, Fahlstadius P, Hellenius ML, Jonsson B & de Faire U. 2003. Cost-effectiveness of primary prevention of coronary heart disease through risk factor intervention in 60-year-old men from the county of Stockholm -a stochastic model of exercise and dietary advice. Preventive Medicine 36: 403-9.

Luoto R, Simojoki M, Uutela A, Boice JD, Jr, McLaughlin JK & Puska P. 2004. Consistency of use of plant stanol ester margarine in Finland. Public Health Nutrition 7: 63-8.

Miettinen TA, Puska P, Gylling H, Vanhanen H & Vartiainen E. 1995. Reduction of serum cholesterol with sitostanol-ester margarine in a mildly hypercholesterolemic population. New England Journal of Medicine 333: 1308-12.

Miettinen TA & Vanhanen H. 1994. Dietary sitostanol related to absorption, synthesis and serum level of cholesterol in different apolipoprotein E phenotypes. Atherosclerosis 105:217-226.

Morris S. 1997. A comparison of economic modelling and clinical trials in the economic evaluation of cholesterol-modifying pharmacotherapy. Health Economics 6:589-601.

National Public Health Institute. Cardiovascular Disease Register, statistical database. http://www.ktl.fi/cvdr/ (accessed 24 Jun 2004).

National Agency for Medicines and the Social Insurance Institution. 2003. Finnish statistics on medicines 2002. Helsinki: National Agency for Medicines and the Social Insurance Institution (Kela).

National Institute for Clinical Excellence (NICE). 2004. Guide to the Methods of Technology Appraisal. Oxford: Radcliff Medical Press.

National Institute for Clinical Excellence (NICE). Expert opinion regarding experience with NICE in UK. http://www.nice.org.uk (accessed 22 Aug 2005).

Nguyen TT, Dale LC, von Bergmann K & Croghan IT. 1999. Cholesterol-lowering effect of stanol ester in a US population of mildly hypercholesterolemic men and women: a randomized controlled trial. Mayo Clinic Proceedings 74: 1198-206.

Niinikoski H, Viikari J & Palmu T. 1997. Cholesterol-lowering effect and sensory properties of sitostanol ester margarine in normocholesterolemic adults. Scandinavian Journal of Nutrition 41: 9-12.

Noakes M, Clifton P, Ntanios F, Shrapnel W, Record I & McInerney J. 2002. An increase in dietary carotenoids when consuming plant sterols or stanols is effective in maintaining plasma carotenoid concentrations. American Journal of Clinical Nutrition 75: 79-86.

Ohinmaa A & Sintonen H. 1996. Quality of Life of Finnish Population Measured by EuroQol. In: Badia X, Herdman M, Segura A, eds. EuroQol Planery Meeting Discussion papers, 3-5 Oct 1996, Barcelona. Barcelona: Catalan Institute of Public Health.

Petersen S, Peto V, Rayner M, Leal J, Luengo-Fernandez R & Gray A. 2005. European cardiovascular disease statistics 2005. London: British Heart Foundation.

Plat J, Kerckhoffs DA & Mensink RP. 2000. Therapeutic potential of plant sterols and stanols. Current Opinion in Lipidology 11: 571-6.

Plat J & Mensink RP. 2000. Vegetable oil based versus wood based stanol ester mixtures: effects on serum lipids and hemostatic factors in non-hypercholesterolemic subjects. Atherosclerosis 148: 101-12.

Plat J, van Onselen EN, van Heugten MM & Mensink RP. 2000. Effects on serum lipids, lipoproteins and fat soluble antioxidant concentrations of consumption frequency of margarines and shortenings enriched with plant stanol esters. European Journal of Clinical Nutrition 54: 671-7.

Samsa GP, Reutter RA, Parmigiani G, Ancukiewicz M, Abrahamse P, Lipscomb J, et al. 1999. Performing cost-effectiveness analysis by integrating randomized trial data with a comprehensive decision model: application to treatment of acute ischemic stroke. Journal of Clinical Epidemiology 52: 259-71.

Schmid CH. Using Bayesian inference to perform meta-analysis. 2001. Evaluation & the Health Profesions 24: 165-89.

Smith T, Spiegelhalter D & Thomas A. 1995. Bayesian approaches to random-effects meta-analysis: a comparative study. Statistics in Medicine 14: 2685-99.

Spiegelhalter D & Best N. 2003. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Statistics in Medicine 22: 3687-709.

Statistics Finland. 2002. Causes of Death 2001. Helsinki: Publications of Statistics Finland.

Thompson SG, Higgins JP. 2002. How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine 21: 1559-73.

University of Oxford Health Economics Research Centre. *Making difficult choices: decision making in health care*. http://www.herc.ox.ac.uk/research/diff.shtml (accessed 1Dec 2005).

Vanhanen HT, Blomqvist S, Ehnholm C, Hyvonen M, Jauhiainen M. Torstila I, et al. 1993. Serum cholesterol, cholesterol precursors, and plant sterols in hypercholesterolemic subjects with different apoE phenotypes during dietary sitostanol ester treatment. Journal of Lipid Research 34: 1535-44.

Vanhanen H. 1994. Cholesterol malabsorption caused by sitostanol ester feeding and neomycin in pravastatin-treated hypercholesterolaemic patients. European Journal of Clinical Pharmacology 47: 169-76.

Vanhanen HT, Kajander J, Lehtovirta H & Miettinen TA. 1994. Serum levels, absorption efficiency, faecal elimination and synthesis of cholesterol during increasing doses of dietary sitostanol esters in hypercholesterolaemic subjects. Clinical Science (London) 87: 61-7.

Weststrate JA & Meijer GW. 1998. Plant sterol-enriched margarines and reduction of plasma total- and LDL-cholesterol concentrations in normocholesterolaemic and mildly hypercholesterolaemic subjects. European Journal of Clinical Nutrition 52: 334-43.

World Health Organization. 2003. The World Health Report 2003: Shaping the Future. Geneva: World Health Organization.

6.4 Economic evalution of sunitinib malate in the treatment of cytokine-refractory metastatic renal cell carcinoma (mRCC) - the comprehensive decision modelling approach[9]

6.4.1 Introduction

Renal cell carcinoma (RCC) is the most common kidney cancer among adult population. Since there are few, if any early-warning signs or symptoms, the diagnosis is usually delayed. Thus, a great proportion of diagnoses are made when RCC is already at a locally advanced or metastatic stage (mRCC). (Motzer et al. 1996) RCC is most frequently diagnosed between the ages 50 and 70, although it can occur at any age (Martel and Lara 2003). The incidence in men is higher than that in women (Bray et al. 1995).

RCC has a poor if any response to chemo- or radiation therapy and low response rates for cytokines interleukin-2 (IL-2) or interferon-α (IFN-α). Even in the first-line treatment, the response rate for these agents is only in the range of 10 to 20%. (Motzer et al. 1996, Motzer & Russo 2000) The prognosis for mRCC is poor, the five-year survival has been estimated to be below 10% (Motzer et al. 1996)

There is no generally approved standard treatment for mRCC. Cytokine therapy (IFN-α or IL-2) is thought to be the most effective treatment, and is commonly used as a first-line therapy, even though only a small number of patients respond to these agents (Motzer et al. 1996). For second-line therapy, the treatment options are even more limited (Motzer et al. 2004). Until recently there has been no effective therapy available for those mRCC-patients who fail to respond, who are unable to tolerate cytokine therapy, or whose disease progresses after an initial response (Motzer et al. 2006a). However, a growing understanding of the biological process underlying different malignancies including RCC has offered possibilities for new treatment alternatives.

Tyrosine kinases play important roles in the regulation of cellular proliferation and survival through numerous pathways. In cancer cells, tyrosine kinases are disregulated in several ways and this disturbance has been implicated in cancer progression. (Krause & Van Etten 2005) Sunitinib is a novel multitargeted tyrosine kinase inhibitor which inhibits these malignant processes through multiple pathways and it has been shown to have anti-tumour and anti-angiogenetic activity (Motzer et al. 2006a). A dose of 50 mg is administered orally on 28 consecutive days followed by a 2 weeks rest period. The treatment is continued in 6-week cycles until progression in disease or intolerable adverse effects occur.

[9] This chapter has been published in: Purmonen T*, Martikainen JA*, Soini EJ, Kataja V, Vuorinen R-L, Kellokumpu-Lehtinen P-L. Economic Evaluation of Sunitinib malate in the Second Line Treatment of Metastatic Renal Cell Carcinoma (mRCC). Clinical Therapeutics, *in press*. Reproduced with permission.
* Authors share equal contribution

Sunitinib has shown high objective response rates in patients with cytokine-refractory mRCC compared with other therapies (Motzer et al. 2006b, Staehler et al. 2006).

This study compares the costs and outcomes among patients with mRCC actively treated with sunitinib to those receiving second line treatment as currently practiced in the Finnish health care setting. The present second line treatment is usually Best Supportive Care including palliative BioChemoTherapy (BSC+BCT, referred subsequently to as BSC). The difference in costs and outcomes between these two treatment arms represents the incremental impact of the use of sunitinib for second-line treatment of mRCC. The viewpoint of the analysis is societal; however indirect costs are not included.

6.4.2 Methods

Disease model. Economic evaluation was conducted using a probabilistic Markov state-transition model. A model with three disease states was used to describe the natural history of patients with mRCC, who have experienced failure of prior cytokine-based therapy. The simplified structure of the model is shown in Figure 32. The transitions through the states were assumed to occur in one-month cycles (i.e. 30 days) and the model was run up to 5 years since the life expectancy for mRCC patients is relatively short.

The Bayesian modelling approach was applied to enable the simultaneous estimation of all inputs in the model (including transition probabilities, resource use, unit costs, and utilities), sensitivity analysis for data and model specifications, and evaluation of the model (Cooper et al. 2004). Uncertainty was propagated into the model using prior probability distributions, which were specified from the prior evidence identified from clinical trials, the literature and the local sample. The model was constructed and evaluated using a Markov Chain Monte Carlo (MCMC) simulation implemented in the WinBUGS software (version 1.4, MRC Biostatistics Unit, Cambridge, UK).

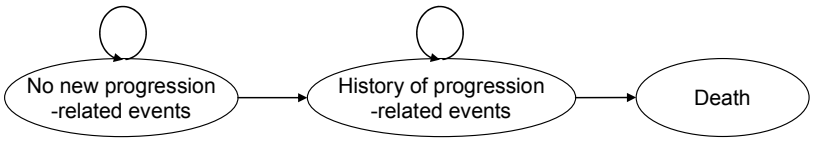


Figure 32. Structure of the Markov-model of mRCC

Transition probabilities. The efficacy of sunitinib in the treatment of mRCC was gathered from recently published trials (Motzer et al. 2006b, Motzer et al. 2006c). Both studies were single-arm, multi-center, open-label, phase II trials. A pooled analysis (n=168) of these trials was used to determine the median progression-free survival (PFS) (Motzer et al. 2006b). Since the median overall survival (OS) was not

attained in both trials, the information from a single trial (n=63) was utilized (Motzer et al. 2006c). Detailed information on survival estimates and patient characteristics in sunitinib-trials can be found in the original reports (Motzer et al. 2006b, Motzer et al. 2006c).

Since both sunitinib-trials (Motzer et al. 2006b, Motzer et al. 2006c) were single arm trials, a comparator arm was needed in order to perform the incremental cost-effectiveness analysis. Therefore, the data from the local sample (n=39) was gathered from the medical records of two Finnish university hospitals and this was utilized to represent survival and resource use in the BSC-arm in the comparison.

Information about the BSC was collected on a structured form by clinical experts. The sample was collected cumulatively one year at a time from patients who had deceased during the years 1996-2006. According to expert opinion, the treatment practice had not changed during these years and, thus, the sample was coherent.

To ensure better consistency with the sunitinib-trials' exclusion criteria, patients with a history of brain metastases, other cancers or serious cardiac events were not included in the sample. In addition, patients with a poor general condition were excluded, since their condition would not have permitted active treatment. The baseline characteristics of the local sample are illustrated in Table 13. Prior nephrectomy was not an inclusion criterion in both sunitinib-trials (Motzer et al. 2006b, Motzer et al. 2006c), and thus, the patients in the local data sample were not excluded, even if nephrectomy had not been done.

Table 13. Characteristics of Finnish mRCC-population from two university hospitals

Characteristics (n=39)	
Sex, No. (%)	
Men	27 (69)
Women	12 (31)
Age, median, years	68
Prior nephrectomy, No. (%)	33 (85)
Prior systemic cytokine-based treatment, No. (%)	
Interferon-alpha	39 (100)
Interleukin-2	0 (0)

In the sunitinib-trials, progression was defined by the RESICT-criteria or by death due to RCC. However, in daily clinical practice, it is neither possible nor meaningful to use imaging (e.g. X-ray, CT or MRI) to define progressions in a stage IV disease. In the local data sample, a progression was assumed to occur, when there had been a major change in treatment (e.g. starting radiation therapy, sur-

gical operation, hospitalization or a change in cytostatic treatment). The medical expert collecting the information made the retrospective expert decision if the change in treatment line was connected to progression. Progression-free survival (PFS) was defined as the time between cytokine failure and progression in the disease. Death due to RCC was also considered as a progression of the disease.

Estimating progression-free and overall survival times from the local sample. Survival analysis was performed to estimate the overall (OS) and the progression-free survival (PFS) after cytokine failure by fitting a Weibull model to data from the local sample (n=39). The probability that an individual survives from the time origin to a point in time beyond t, was estimated by the Weibull survival function S(t):

[30] $$S(t) = P(T \geq t) = \exp(-\lambda t^{\gamma})$$

where the lambda (λ) parameter gives the scale of the distribution and the gamma (γ) parameter defines the shape of the distribution. The scale parameter was parameterized as $\exp(\beta_0)$ and the shape parameter was estimated from the data.

Correlation between the PFS and the OS were also modelled assuming that the PFS may influence the OS, but that there was no inverse effect of the OS influencing the PFS. Hence, correlation structure was modelled using a Weibull regression model:

[31] $$S(t|PFS) = \exp(-\exp(\beta_0 + \beta_1(PFS - \overline{PFS})t^{\gamma})$$

where the PFS was used as a covariate. The PFS was centered to mean ($\overline{PFS}$) to speed up convergence. To obtain a fully Bayesian approach, the model specification was completed by adding prior distributions on β, λ and γ. Non-informative prior distributions for the model parameters were assumed without any prior expectations about the magnitude of parameter values. This assumption was used to ensure that the data from the local sample dominated the final inferences.

Goodness-of-fit of survival models. A point estimate of the deviance (i.e. -2log(likelihood)) was used to assess the goodness-of-fit of the Weibull models for the data from the local sample (Nixon and Thompson 2004). The estimated Weibull survival curves and the corresponding product-limit survival times from the local sample are presented in Figure 33.

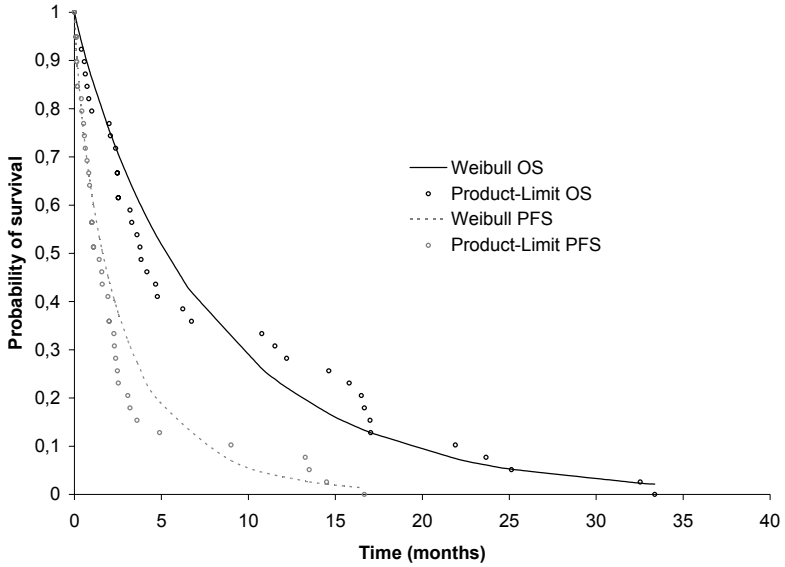


Figure 33. Goodness-of-fit of Weibull survival estimates for the data from the local sample (n=39)

Transition probabilities between health states. Since the mean survival times were not attained or available in sunitinib-trials, the median values were used to estimate monthly transition probabilities in both study arms. For the sunitinib-treatment arm, the median survival estimates were obtained from the published trials, whereas for the BSC-treatment the corresponding figures were estimated directly from the collected data. The median survival estimates (Table 14) were converted to monthly transition probabilities using the following formula:

[32] Risk of an event (1 month) = [1-(0.5)^(1/median time to event)]

The estimated base-case transition probabilities of disease progression were 0.0811 in the sunitinib-arm and 0.3841 in the BSC-arm. The corresponding monthly risks of death due to mRCC were 0.0408 and 0.1649. In the analysis, these transition probabilities were assumed to remain constant over time (i.e. the OS- and PFS-curves were assumed to follow a declining exponential distribution in both groups).

Table 14. Median survival times of mRCC-patients with BSC- and sunitinib-treatments

	Median	95% CI	SE
Overall survival, months			
Sunitinib (n=63)	16,4	10,8-NA	(NA)
Best Supportive Care (n=39)	3,83	(2.16-5.51)	0.85
Progression-free survival, months			
Sunitinib (n=168)	8,2	(7.8-10.4)	0.66
Best Supportive Care (n=39)	1,43	(0.7-2.17)	0.37

NA= Not yet attained

Characterizing uncertainty related to transition probabilities. The uncertainty related to transition probabilities in the sunitinib-arm was characterized by expressing the parameter values in the model as beta distributions. The beta distributions were parameterized as beta (α, β), where α was defined as the number of patients transferring to a new state during one month and β was defined to be the sample size of the sunitinib-trials minus the value of α. The values of α were approximated based on the estimated monthly probabilities of disease progression and death.

In the BSC-arm, the uncertainty related to transition probabilities was propagated directly from the Weibull survival models. Thus, no further distribution assumptions were needed. Since there were multiple possibilities to transitions from “No new progression-related events” -health state in the model, the associated transition probabilities were normalized to ensure that they summed up to 1 during the simulation process.

Resource use and costs. In order to define the monthly costs of BSC in the Finnish setting, the health care resource utilization and medication use of each patient (n=39) was collected for the entire follow-up period. The recommended unit prices for health care services were case-mix adjusted for the Finnish regional price differences (Hujanen 2003) and real-valued to the year 2005 with the official health care price index (Statistics Finland 2006).

The consumption and cost of cancer medications, additional IFN-α-products, analgesics and bisphosphonates were specified. It was assumed that medical costs not related to RCC were equal in both treatment arms and therefore were not collected. A medical specialist collecting the data made the expert decision about whether or not a particular medicine was related to the treatment of RCC.

The costs of medication were calculated using the most economical generic product prices in the official price list when applicable. For medications no longer on the consumer market, the last existing price was used. Prices from previous years were not converted to 2005 currency, since medications in Finland do not follow the general price index. Prices for products with special sales permit were gathered directly from the distributor and transformed to retail prices using the official formula (2006). Medi-

cations administered in hospital care were assumed to be included in the cost of a treatment day and, thus, were not included in the costs of medication in order to avoid double counting of costs. If the duration or dose of medication was not mentioned, the defined daily doses from the year 2006 (National Agency for Medicine 2006) were used, and the duration was assumed to be one month. Travel costs were allocated to all separate outpatient visits and radiation therapy treatment days.

The resource use for the follow-up time is presented in Table 15. The resource use among BSC-patients is very heterogenic. Thus, an average cost per follow-up day from the whole population was used in the analysis and this comprised an average treatment cost of 1339 €/month. The price without VAT was used in all costs.

Table 15. Resource utilization in local mRCC-patient sample (n=39)

Number of medications used during the follow-up No. (%)				
	IFN-α	Cancer medication	Bisphosphonates	Analgesics
0	33 (85)	14 (36)	32 (82)	9 (23)
1	6 (15)	12 (33)	5 (13)	12 (31)
2	0 (0)	7 (18)	2 (5)	6 (15)
3 or more	0 (0)	5 (13)	0 (0)	12 (31)

Number of imaging examinations used during the follow-up No. (%)				
	X-ray	CT	Ultrasound	MRI
0	12 (31)	24 (61)	25 (64)	36 (92)
1	11 (28)	10 (26)	8 (20)	3 (8)
2	3 (8)	2 (5)	3 (8)	0 (0)
3	3 (8)	1 (3)	1 (3)	0 (0)
4 or more	10 (25)	2 (5)	2 (5)	0 (0)

Number of health care service units utilized during the follow-up No. (%)				
	Ward care days (University hospital)	Ward care days (Health care centre)	Outpatient visits (all levels)	Radiation therapy days
0	10 (26)	19 (49)	14 (36)	23 (59)
1-5	6 (15)	2 (5)	15 (38)	2 (5)
6-10	6 (15)	2 (5)	6 (15)	6 (15)
11-19	12 (32)	3 (8)	3 (8)	3 (8)
20 or more	5 (13)	13 (34)	1 (3)	5 (13)

Since no published information exists on the use of sunitinib in current Finnish practice, an expert panel of four clinicians treating mRCC-patients was used to estimate an average treatment protocol for those patients. The monthly treatment costs for sunitinib-treatment arm were defined according to this protocol. Patients were assumed to be seen by an oncologist and also to have laboratory tests taken twice in

the first cycle and once in all latter cycles. Imaging was utilized once in the two first cycles and then on every other cycle. One third of the use of imaging was assumed to be CTs and two thirds to be ultra-sound or X-ray assessments. The cost of medication and estimated costs for treatment of adverse events were added to the monthly costs.

The mean monthly costs are illustrated in table 16. The more active resource use at the beginning is due to the follow-up of the treatment tolerability and entry medical examinations. After the termination of sunitinib-treatment, the monthly costs are assumed to be equal to BSC-treatment arm. The costs were assumed to be gamma distributed. Since no information was available on the variance of mean monthly costs in the sunitinib-treatment, the variance was assumed to be proportionally equal to the variance of the mean monthly costs in the data from the local sample. The parameters of gamma distributions were estimated using the methods of moments (Briggs et al. 2002).

Table 16. Mean monthly costs per patient in sunitinib- and BSC-treatments

SUNITINIB-arm		Cost (€)	SE	Distribution
Health care utilization	Month 1	545[b]	114[x]	Gamma (2600, 4.77)
	Months 2-3	324	68[x]	Gamma (1546, 4.77)
	Months 3 =>	201	42[x]	Gamma (959, 4.77)
Drug costs per month		3748		Uniform (3748, 4061)
BSC-arm[*]		1339*	281[x]	Gamma(6389, 4.77)

b) Includes treatment of adverse events

*) Total costs within the follow-up divided by cumulative days alive (x30)

x) Proportionally equal to SE of mean monthly costs in the data from the local sample

Utilities. The utility values were obtained on day 1 and day 28 of every 6-week cycle in the sunitinib-trial using EQ-5D. The average (SE) utility before new progression was 0.764 (0.026) and decreased to 0.731 (0.061) after the progression (data on file, Pfizer). The utility of a deceased patient was defined to be zero. Since utility data was unavailable for the BSC arm, utility values in different health states were assumed to be equal in both study arms and to be beta-distributed. The parameters of beta distributions were obtained using the methods of moments (Briggs et al. 2002).

Sensitivity analysis. The sensitivity of selected discount rates and time horizon was considered using different discount rates and time horizons in the analysis. In the base-case analysis for five years, both costs and QALYs were discounted using a discount rate of 5 per cent.

Since the survival data used in the current model was not collected in a parallel study setting, it was considered to be a potential source of uncertainty. When the observed covariates between sunitinib-trials and the local sample were compared, the differences in the average age were considered to be the most important source of uncertainty, and thus, the impact of age difference was studied.

Model evaluation. To evaluate the Markov model, a cohort of 1000 artificial mRCC patients with evidence of metastases was entered into the model. It was also assumed that mRCC patients had failed to benefit from cytokine therapy because of intolerance or disease progression before they entered into the model. This Markov model was evaluated simultaneously with the Weibull survival models.

Final posterior parameter estimates were based on a total of 50 000 MCMC samples. The first 10 000 samples were discharged to ensure stability of the posterior sampling procedure. Furthermore, the model convergence was confirmed by checking the trace plots of the samples, autocorrelation samples, and the Monte Carlo standard error statistics in WinBUGS. The code for the cost-effectiveness model is given in Appendix 1.

Cost-effectiveness. Cost-effectiveness and cost-utility analyses were conducted. The parameter uncertainty was converted into decision uncertainty using cost-effectiveness acceptability curves (CEACs). CEACs illustrate the probability of cost-effectiveness at different willingness-to-pay threshold values.

Value of information analysis. Additional research evidence is valuable because it reduces the expected costs of uncertainty surrounding an implementation decision. If a wrong decision has been made, the expected costs can be determined as a product of the probability of wrong decision and its consequences. In the field of health care, the consequences of wrong decision can be determined in the terms of net monetary benefits (NMB) lost due to a wrong decision. The expected costs of uncertainty surrounding the adoption decision were estimated by applying the expected value of perfect information (EVPI) approach. Briefly, the EVPI for an individual patient ($EVPI_{patient}$) is simply the difference between the expected value of the decision made with perfect information about the uncertain model parameters and the decision made on the basis of existing information. When a probabilistic Markov model is employed, the calculation of EVPI is rather straightforward as the EVPI is obtained by taking the average of the maximums in each iteration of the simulation. (Claxton 1999)

Population EVPI. To estimate the EVPI at the population level for the mRCC patients, the $EVPI_{patient}$ was multiplied by the number of patients with mRCC that would be affected by the information over the effective lifetime of the sunitinib-treatment:

$$\text{Population EVPI} = EVPI_{patient} * \sum_{t=1}^{T} N_t * (1+r)^{-t} \quad [33]$$

where T is the total length of the assumed effective lifetime, which in this particular case was assumed to be either 5 or 10 years. N_t is the number of patients that may potentially benefit from the sunitinib-treatment at time t. (Claxton 1999)

The number of new mRCC-cases was estimated based on statistics obtained from the Finnish cancer registry. According to the registry, the incidence of new RCC cases was 710 in the year 2004. Using the assumptions that 50% of the cases will eventually develop to mRCC and that 50% of those would receive active second-line treatment, the number of mRCC-patients affected by second-line sunitinib-treatment amounts to 914 and 1 440 at the 5 and 10 year time period, respectively. A discount rate (r) of 5% was used in the calculations.

6.4.3 Results

According to the simulation results, treatment with sunitinib compared with Finnish current treatment (BSC) prolonged life expectancy by approximately 1 year and progression-free time by 6.3 months. It also resulted with 0.72 additional QALYs compared with BSC. However, in the 5-year time period, it was responsible for around 30 880 € incremental costs per patient. The base-case results indicated that cost per progression-free month (PFM) gained was 4 883 €, cost per life-year gained (LYG) was 30 250 €, and cost per an additional QALY was 42 877 €. When parametric uncertainty was taken into account, it was observed that the incremental QALYs ranged from 0.5 to 0.9 while the mean difference in costs ranged from around 24 000 € to 38 000 € (Figure 34). According to our results, sunitinib is more costly but also more effective than BSC in all situations.

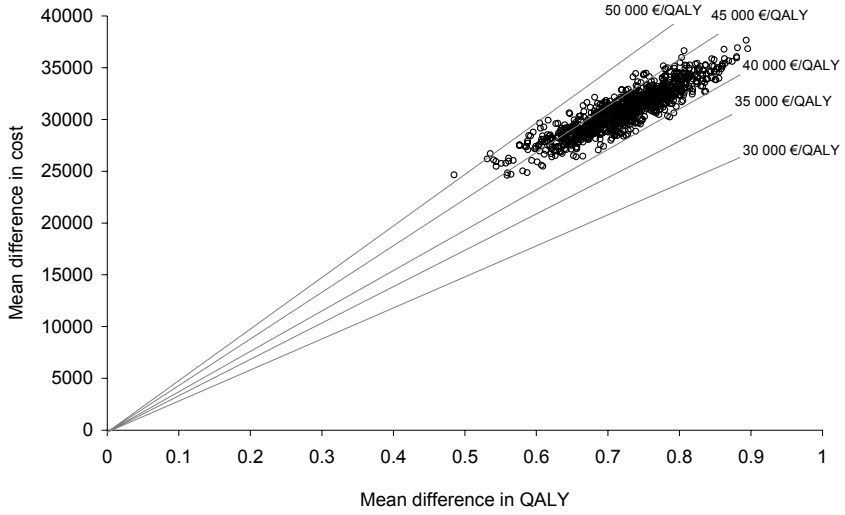


Figure 34. Cost-effectiveness plane. Base case probabilistic sensitivity analysis for 5 years

The uncertainty related to ICER is depicted as a cost-effectiveness acceptability curve (Figure 35) which is established through the probabilistic sensitivity analysis (PSA) carried out in the second-order Monte Carlo simulation framework. It seems that when society's willingness-to-pay rises above 40 000 € / QALY, the probability of sunitinib's cost-effectiveness increases rapidly. For example, with a 45 000

€ / QALY willingness-to-pay threshold, sunitinib has approximately an 88% probability of being cost-effective compared to BSC in a 5-year period in the Finnish setting.

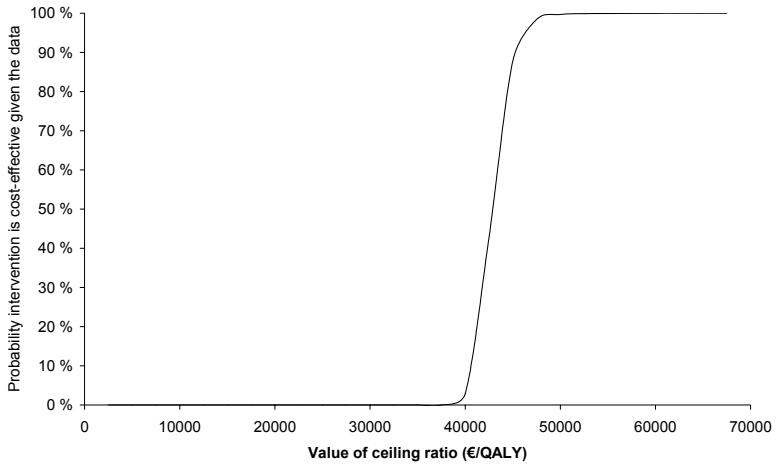


Figure 35. Cost-effectiveness acceptability curve of sunitinib versus BSC

The population EVPI for the decision between sunitinib- and BSC -treatments was 607 000 euros at a willingness to pay level of 42 500 €. Very low population EVPI values were associated with willingness-to-pay values in the range of 0 € to 37 500 € per QALY gained. This indicates that if a decision maker's willingness-to-pay for an extra QALY is clearly less than 37 500 €, then the requirement for further information is unlikely to be cost-effective. This is the same, when the decision maker's willingness-to-pay for an extra QALY is more than 52 500 €. The results from the EVPI analysis are depicted in Figure 36.

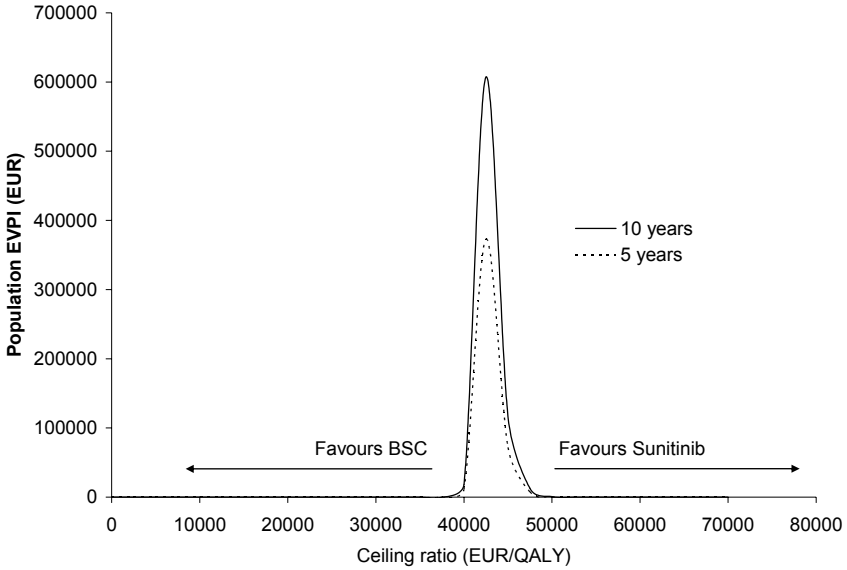


Figure 36. Expected value of perfect information for Finnish mRCC-patients

Sensitivity analyses. The discounting of costs and effects did not have any significant effect on the results. Most of the sunitinib-related costs are incurred in the first year because of the relatively short time to disease progression. The incremental cost-effectiveness ratio for an additional QALY decreased after the first year and rapidly leveled off to the base-case level.

No statistically significant difference either in the OS or the PFS was found when comparing the patient groups under and over 60 years in the BSC-arm. Results from analysis using survival data from patients under 60 years in the BSC-arm were in line with results obtained from the base case analysis. The sample size of patients under 60 years was only 12 and therefore cannot be used as a reliable source of information for the analysis. However, this analysis revealed that the average age in this population did not bias the results.

6.4.4 Conclusions

The incremental cost-effectiveness ratio in our study was 42 877 € per QALY gained in a 5-year time period, which indicates that sunitinib-treament is potentially cost-effective when compared with the current Finnish treatment practice. There are only a limited number of studies concerning the cost-effectiveness of novel cancer treatments in the Finnish setting. However, in a recent study, the cost per QALY gained for temozolomide in the treatment of glioblastoma multiforme was 32 471 € which is

of a similar magnitude as the value calculated in this study (Martikainen et al. 2005). These figures are also comparable to those stated by NICE (Devlin & Parkin 2004).

Despite the rather good reliability of the data sources, this study has some limitations. The first limitation in this study concerns the comparability of the patient populations. Differences in patient characteristics, health status and severity of disease at baseline may raise doubts about the comparability of Finnish data and data from clinical sunitinib-trials. However, in order to obtain the costs and effectiveness related to the Finnish current care, unselected sampling with comparable criteria to the sunitinib-trials was carried out in two university hospitals. Thus, a real Finnish situation in mRCC-treatment was obtained and in this respect the survival difference due to different baseline characteristics in our data and clinical trials should have only a minor impact on the main questions related to survival and costs. Furthermore the subgroup analysis revealed no significant difference in survival times in the different age groups.

Another limitation of this study is that the results are dependent on the expected survival times. However, the survival estimates from the Finnish data sample are comparable to those stated in earlier studies. It has been previously reported that median survival for recurrent or metastatic RCC ranges from around 2 to 13 months, depending on the patient's risk factors (Elson et al. 1988). Survival estimates from 29 consecutive clinical trials involving 251 patients with conventional second-line mRCC treatment showed the median OS and PFS to be 10.2 and 2.4 months, respectively (Motzer et al. 2004). In a previous Finnish study concerning first-line treatment the median OS was 37.8 weeks with vinblastine monotherapy (n=81) and 67.6 weeks when the treatment was the combination of INF-α2a and vinblastine (n=79). The corresponding figures for PFS were 9 weeks and 13 weeks. (Pyrhönen et al. 1999)

Nonetheless it has to be stated, that the patients confronted in daily clinical practice may be in a weaker initial health-state than patients participating in clinical trials. There may also be a selection of patients with more slowly progressive disease in clinical trials of second-line therapy, because patients with poor health status would not be included in the population receiving such treatment [5]. Thus, the expected survival in clinical studies may be longer than that encountered in the average mRCC-population. For these reasons, a weak general condition was one of the exclusion criteria in our study. However, it may be challenging for the investigators to identify those patients relying on patient files. Nevertheless, the data used in this study was collected from patients suffering from mRCC and, thus, can be considered reliable. At the time of writing this article, a Phase III trial of sunitinib versus IFN-α in first-line treatment of mRCC was published. Sunitinib showed marked improvement in the progression-free survival and was associated with a higher objective response. This provides more evidence on the efficacy of sunitinib also when compared to active treatment. (Motzer et al. 2007)

The final limitation of this study concerns the fact that the treatment protocol of the sunitinib-arm was solely based on expert opinion and does not necessarily depict the real practice. However, since there is no consensus on what represents current practise for a patient treated with sunitib exists, this assumption was used as the best available information. The assumption that sunitinib-treatment would be discontinued immediately after observed disease progression may also not reflect the clinical practice in all settings. Currently there is also no evidence to suggest that switching from sunitinib to another tyrosine kinase -inhibitor after the onset of progression would be effective (Staehler et al. 2006).

Despite the possible limitations of this study, sunitinib can be considered a suitable alternative as a second-line therapy of mRCC. The limitations of the study settings were noticed and carefully assessed in the analyses. The assumptions made in this study were confirmed by clinical experts treating mRCC-patients and were conservative rather than over-optimistic.

6.4.5 References

Bray F, Sankila R, Ferlay J & Parkin DM. 1995. Estimates of cancer incidence and mortality in Europe in 1995. European Journal of Cancer 38: 99-166.

Briggs AH, Goeree R, Blackhouse G & O'Brien BJ. 2002. Probabilistic Analysis of Cost-Effectiveness Models: Choosing between Treatment Strategies for Gastroesophageal Reflux Disease. Medical Decision Making 22: 290-308.

Claxton K. 1999. The irrelevance of inference: a decision making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 18: 341-64.

Cooper NJ, Sutton AJ, Abrams KR, et al. 2004. Comprehensive decision analytical modeling in economic evaluation: a Bayesian approach. Health Economics 13: 203-26.

Devlin N & Parkin D. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 13: 437-52.

Elson PJ, Witte RS & Trump DL. 1988. Prognostic factors for survival in patients with recurrent or metastatic renal cell carcinoma. Cancer Research 48(24): 7310-13.

Hujanen T. 2003. Terveydenhuollon yksikkökustannukset Suomessa vuonna 2001, Stakes publications 1/2003, Helsinki.

Krause D & Van Etten R. 2005. Tyrosine Kinases as Targets for Cancer Therapy. New England Journal of Medicine 353: 172-87.

Lääkelaitos. 2006. Lääkkeiden luokitus (ATC) ja määritellyt vuorokausiannokset (DDD), Lääkelaitoksen julkaisuja, Edita Prisma Oy, Helsinki.

Martel CL & Lara PN Jr. 2003. Renal cell carcinoma: current status and future directions. Critical Reviews in Oncology/ Hematology 45:177-90.

Martikainen JA, Kivioja A, Hallinen T & Vihinen P. 2005. Economic Evaluation of Temozolomide in the Treatment of Recurrent Glioblastoma Multiforme. Pharmacoeconomics 23(8): 803-15.

Motzer RJ, Bander NH & Nanus DM. 1996. Renal-Cell Carcinoma. New England Journal of Medicine 335(12): 865-75.

Motzer RJ, Bacik J, Schwartz LH, et al. 2004. Prognostic factors for survival in previously treated patients with metastatic renal cell carcinoma. Journal of Clinical Oncology 22(i3): 454-63.

Motzer RJ, Hoosen S, Bello CL & Christensen JG. 2006a. Sunitinib for the treatment of solid tumours: a review of current clinical data. Expert Opinion on Investigational Drugs 15(5): 553-61.

Motzer RJ, Hutson TE, Tomczak P, et al. 2007. Sunitinib versus Interferon Alfa in Metastatic Renal-Cell Carcinoma, New England Journal of Medicine 356(2): 115-24.

Motzer RJ, Michaelson MD, Redman BG, et al. 2006c. Activity of SU11248, a Multitargeted Inhibitor of Vascular Endothelial Growth Factor Receptor and Platelet-Derived Growth Factor Receptor, in Patients With Metastatic Renal Cell Carcinoma. Journal of Clinical Oncology 24(1): 16-23.

Motzer RJ, Rini BI, Bukowski RM, et al. 2006b. Sunitinib in Patients with Metastatic Renal Cell Carcinoma. Journal of The American Medical Association 295(21): 2516-24.

Motzer RJ & Russo P. 2000. Systemic therapy for renal cell carcinoma. Journal of Urology 163(2): 408-17.

Nixon RM & Thompson SG. 2004. Parametric modelling of cost data in medical studies. Statistics in Medicine 30: 1311-31.

Pyrhönen S, Salminen E, Ruutu M, et al. 1999. Prospective Randomized trial of Interferon Alfa-2a Plus Vinblastine Versus Vinblastine alone in Patients with Advanced Renal Cell Carcinoma. Journal of Clinical Oncology 17(9): 2859-67.

Staehler M, Haseke N, Schoeppler G, et al. 2006. Modern Therapeutic Approaches in Metastatic Renal Cell Carcinoma. EAU-EBU Update Series (2006)

Statistics Finland and Local Finland. 2006. Price Indexes 2006. [online] Available from URL: http://www.kunnat.net/k_perussivu.asp?path=1;29;374;36984;43148;52537 [Accessed 29.8.2006]

7 FINDINGS AND DISCUSSION

7.1 Applicability of Markov models

The conducted case studies proved that the Markov models offer a clear and coherent mathematical structure to combine all relevant evidence and to assess the consequences, such as expected costs and effects, of different decision options in advance. The Markov models are particularly suited to the modelling of disease progression over time. In addition, the decision-analytic models are relatively inexpensive vehicles to utilise, especially, when compared to trial-based studies. However, it should be noted that the decision-analytic models are complementary rather than substitute vehicles for the trial-based cost-effectiveness analyses (Claxton et al. 2002). For example, the decision-analytic models can be applied when there is a need to assess all relevant evidence (chapters 6.1-6.4), to link intermediate outcomes (e.g. change in serum cholesterol levels) to final endpoints (e.g. change in QALYs) (chapter 6.3), to make results applicable to the decision-making context due to a gap between clinical trial evidence and the requirements for a decision (chapters 6.2 and 6.4), or to estimate cost-effectiveness for specific subgroups (chapter 6.3). However, there is a range of issues that affect the applicability of Markov models to inform decision-making in practice. Therefore, some practical considerations that are common across model-based cost-effectiveness studies are discussed below.

One important aspect relates to the clinical validity of a developed model, since it is important to ensure that the health state definitions and the transitions used in the decision-analytic models hold clinical validity in the face of current understanding. Therefore, when the development of a model structure is guided by previous published model structures, it is essential that clinical experts are used to validate the applied model structure. For example, in case study in chapter 6.1, the applied model structure was modified after consulting clinical experts, since they considered that one transition between health states used in the previous cost-effectiveness model was biologically implausible due to the nature of the disease (i.e. transition from the moderate disease state back to the mild disease state was considered to be biologically implausible, since the Alzheimer's disease is a neurodegenerative disease characterised by progressive deterioration in memory, and it was omitted from the model).

The decision-analytic models are meant to be an aid to decision making under uncertainty at a particular point in time (Sculpher et al. 2000). Consistency between a decision problem and a model structure is evidently a basic requirement in decision-analytic modelling but the attainment of this requirement may sometimes be challenging due to the rapid flow of new scientific evidence affecting the current understanding about the standard of care. In the case study of chapter 6.2, for example, the initial model structure was developed to model the decision problem about the cost-effectiveness of temozolomide as a second-line treatment (i.e. after surgery and radiotherapy) for patients with recurrent glioblastoma multiforme. However, just before the publication of the cost-effectiveness report a new RCT with prom-

ising results was published. This new study indicated that when temozolomide was administered as a first-line adjuvant treatment with radiotherapy statistically significant survival benefit with minimal additional toxicity was achieved as compared to radiotherapy alone (Stupp et al. 2005). This clinical finding changed the current standard of care in a way which diminished the usefulness of the developed cost-effectiveness model to inform decision-making in the changed situation. This finding highlights the importance of the timing of cost-effectiveness evaluations and emphasises the need for the models to adapt to new evidence as it becomes available (cf. the iterative approach introduced in chapter 2).

The decision-analytic models primarily purpose is to bring together all relevant evidence into consideration and to reveal the relation between the model parameters and outcomes in a transparent way rather than to make very accurate predictions about possible consequences in the future. The optimal balance between a model's accuracy and transparency is hard to determine due to trade-off between these two concepts; when the model is made more accurate, the complexity of the model increases simultaneously, which decreases the decision-makers' ability to understand it. Therefore, the published guidelines for decision-analytic modelling have recommended that transparency should have a somewhat greater weight in model development. (Weinstein et al. 2003). However, since transparency does not determine the model's accuracy (i.e. even if we could replicate the model's results, replication would not tell us the model's accuracy), it is always reasonable to assess the external consistency of a developed model e.g. by comparing the model's results, such as the time to disease progression, to previously published results (Eddy 2006).

The application of Markov process approach increases flexibility in the decision-analytic modelling. The case studies of chapters 6.1, 6.2, and 6.4 were developed assuming that the transition probabilities between the defined health states would stay constant over time. In fact, the case study of chapter 6.3 proved that covariate- and time-dependency can be built into Markov transition probabilities in a way that the transition probabilities vary according to the patients' underlying risk profiles and time in the model (i.e. the risk of events increases as the cohort ages). In addition, the case study of chapter 6.3 showed that the Markov assumption can be relaxed by building temporary tunnel states into the model. In this particular case, this was done by modelling the occurrence of myocardial infarction as a temporary health state, in which the members of cohort spend only one year, which after they move into a health state depicting patients with a history of myocardial infarction event.

The models should be kept as simple as possible, while capturing all essential parts of disease processes including the effects of health technologies, in order to aid their comprehension by decision-makers (Buxton et al. 1997). However, this may entail some justified simplifications. For example, the quite complex structure of the model in the case study of chapter 6.3 was developed trying to keep a chain of evidence (i.e. a health technology ⇨ intermediate outcomes ⇨ main health outcomes) as

transparent as possible. Therefore, possible additional preventive effects e.g. those related to the risk of stroke (Law et al. 2003) were omitted.

Finally, the conduction of logical checks and programming the models in alternative software packages ensures internal consistency (i.e. the mathematical logic of the models). For example, in the case study described in chapter 6.4, the model was developed using both WinBUGS and Microsoft Excel™ to identify any possible logical and programming errors.

7.2 Applicability of evidence synthesis methods

The identification, selection, and critical appraisal of evidence for the decision-analytic models are the most time consuming parts of the model development process. When systematic literature reviews are conducted, the definition of inclusion criteria is the critical driver of time-consumption, since imprecise or too extensive criteria may lead to a substantial increase in the time required to conduct a single systematic review. However, if the inclusion criteria are too strict, there is the increased risk that all relevant evidence will not be identified. Therefore, it is important that the inclusion criteria are derived consistently from the decision problem.

The critical appraisal of identified studies is the essential phase of evidence synthesis. Structured quality assessment forms were found to be helpful in the evaluation of the identified articles, since they standardise and make the review process more transparent, especially when there is more than one reviewer. In addition, the filled forms provide a historical record of decisions occurring during the review process.

Once the critical appraisal of selected studies is done, evidence is extracted from the selected studies by trained reviewers. For example, in case study described in chapter 6.3, the data extraction was done by one principal reviewer and verified by other members of research team, whereas, in case study described in chapter 6.2, evidence was extracted and evaluated independently by each member of research team. In both cases, however, disagreements were resolved by consensus. It is recommended that at least two reviewers are used to extract evidence from the selected studies, since multiple data extraction reduces errors in the evidence extraction process (Buscemi et al. 2006).

Well defined evidence extraction tables are helpful in the organisation of extracted evidence. In the present case studies, evidence was first extracted into the evidence tables developed in Microsoft Excel™ in its original form and then transformed as needed. The advantage of spreadsheet-based tables is that they can be programmed to calculate data conversions (e.g. mg/dl to mmol/l) automatically for data reported in various formats.

The quality of reporting may cause problems in the evidence extraction, since published studies frequently do not report all of the information required for meta-analysis. For example, in the case study of chapter 6.3, missing data was first tried to locate by requesting from the authors of the study. However, we did not manage to obtain all missing data and therefore data imputation methods were applied to estimate replacement estimates in some studies, where all required evidence was not reported in study reports (Higgins & Green 2006).

Finally, meta-analyses and meta-regressions are conducted to summarise and analyse the collected evidence. The present study proved that the analyses are relatively easy to conduct with WinBUGS-software, since models can be specified by using either a graphic- or text-based model description (Spiegelhalter et al. 2003). Appendix 2 depicts the graphic- and corresponding text-based model descriptions for the meta-analysis applied in the case study in chapter 6.3 as an example.

Generally, cost-effectiveness evaluations tend to focus on a time period at or around the implementation of new health technology, when experience and evidence about its clinical and economic consequences may still be relatively limited. In the conducted case studies, it was often possible to obtain the resource use data from published sources or hospital administration systems. When data was collected from hospital administration systems, patient records were first identified from the patient administration systems according to defined inclusion criteria, after which data were extracted from patient records by using structured data extraction forms. Since the resource use may vary from one hospital to another, the resource use data were collected from separate hospitals to obtain a sufficient estimate for resource use (see chapters 6.2 and 6.4 for example). In the near future, the extraction of resource use from the hospital administration systems would be expected to become easier and less time-consuming, with the introduction of electronic patient records.

It is not always possible to obtain all evidence from published sources or hospital administration systems. In these cases, the use of expert opinions is generally accepted (Sculpher et al. 2000). In the case study described in chapter 6.4, evidence about the use of sunitinib in current Finnish practice was not available and therefore an expert panel of clinicians treating mRCC-patients was asked to determine an average treatment protocol for those patients in advance. Expert opinions were also applied to define the number of scheduled visits (chapter 6.1) and as proxy respondents to define utility values for cancer patients (chapter 6.2). When expert opinions are used to define the scheduled administration patterns of drugs, the elicited information can be assumed to be quite reliable but when the experts are used as proxy respondents to define patients' utilities, then caution is needed. If the expert opinions are the only available source of evidence for utilities, it may be more valuable to try reflect the uncertainty inherent in the proxy answers rather than to attempt to achieve a consensus statement about the "true" utility value.

Finally, the present study showed that evidence used in meta-analysis and decision-analytic models requires very often adaptation or translation to other value scales, which increases the number of the additional sources of methodological and process uncertainty. For example, additional assumptions are needed e.g. when median time to progression estimates are converted to the monthly transition probabilities (cf. chapter 6.4) or when the standard error of mean is estimated based on the assumed 95% confidence interval (cf. chapter 6.1).

7.3 Applicability of different approaches to parameter uncertainty

The probabilistic approach was applied in all conducted case studies. The case studies described in chapters 6.1 and 6.2 applied the two-stage approach, where evidence synthesis and cost-effectiveness modelling were conducted as separate processes. In contrast, the case studies of chapters 6.3 and 6.4 applied the comprehensive modelling approach, where evidence synthesis and cost-effectiveness modelling processes are conducted simultaneously. The use of the probabilistic decision-analytic models enables a more realistic representation of uncertainty in the model's outcomes. Furthermore, the probabilistic models correctly estimate expected costs and effects under conditions of parameter uncertainty; even though decision-analytic models are non-linear (i.e. the model outputs are multiplicative functions of input parameters) (Ades et al. 2005). This is a particularly important feature in the case of the Markov models, which are non-linear due to the transition matrix.

A common criticism related to the probabilistic approach is that the choice of prior distributions for model parameters is essentially arbitrary. However, the present study revealed that the choice of distribution for an individual parameter is guided mainly by the nature of that parameter and by assumptions commonly employed to estimate confidence intervals in statistics (Briggs et al. 2006, 84). In addition, the number of statistical probability distributions needed in such models is relatively small. Table 17 summarises and justifies commonly used statistical probability distributions for model parameters.

Table 17. Summary of statistical distributions used in the probabilistic decision-analytic models.

Parameter type	Distribution	Justification
Transition probabilities	Beta (binomial data)	Returns values within the logical constraints [0,1].
	Dirichlet (multinomial data)	The multivariate generalization of the beta distribution. It will generate the exactly same results as a series of conditional beta distributions (see Briggs et al. 2003).
Baseline clinical data	Normal distribution	Justification rests on the *central limit theorem*, which states that the sampling distribution of mean will be normally distributed with sufficient sample size.
Resource use data	Log-normal / gamma	Positively skewed distribution required with values above zero.
Unit costs	Fixed	Assumes that fixed unit cost reflect the true opportunity cost of consumed resource.
	Normal	Assumes that unit costs are located far from 0 and the sampling distribution of mean is normally distributed.
Relative risks	Lognormal	Relative risk ratios are estimated on the log scale, which justifies the use of lognormal distribution.
Utilities	Beta	Returns values within the logical constrains [0,1].

The results of the case study described in chapter 6.2 indicated that it is possible to incorporate the quality of clinical evidence into the decision-analytic model by applying probability distributions that reflect uncertainty associated with the efficacy parameters. A justification for the use of statistical distributions rests on the assumption that evidence with poor quality makes a model parameter less precise. However, the quality of evidence is a multidimensional concept and therefore it may be difficult to capture quality fully in a single score. Furthermore, the fundamental relationship between the precision of evidence and the quality of evidence needs to be paid further attention, since the stance that is taken on different types of evidence is not necessarily a statistical issue, but a question of expert judgement (i.e. should we include only randomised studies, or allow also other types of studies that may include additional sources of bias?) (Ades & Sutton 2006).

Some model parameters are deterministic in nature and hence there is no need to specify statistical distributions for them. Discount rates, for example, are handled as deterministic, because (methodological) uncertainty arises due to lack of consensus about the most appropriate value for a discount rate, not due to the imprecision of the parameter estimate. In addition, most of the model parameters that describe the characteristics of a patient cohort, such age and sex, are handled as deterministic variables. (Briggs 2000) Uncertainty related to these deterministic parameters can be depicted using simple univariate sensitivity analysis. For example, in Table 12 an applied discount rate is varied and incremental cost-effectiveness ratio are re-estimated to see how the applied discount rate affects on the results.

There are some limitations that may affect the usefulness of the probabilistic models in general. First, insufficient evidence may prevent the specification of proper prior distributions for the model parameters (e.g. a standard error of mean is not reported) and thereby additional assumptions are needed (e.g. the standard error is the same value as the mean), which may increase the levels of uncertainty. For example, at the moment the Finnish resource use and unit cost list do not provide information about the precision (i.e. the standard errors) of mean resource use estimates. Therefore, univariate analyses are still required to help understand the relative importance of individual parameters. Second, the probabilistic methods can create a misleading impression about the accuracy of the results, which may detract attention away from the considerations of model structure uncertainty and the quality of evidence. Third, the imprecise and insufficient appreciation of the probabilistic approach by the decision-makers may diminish the implementation of these methods. However, this problem may be purely an educational issue and it might be solved by arranging further training on this topic.

When the two-stage and comprehensive approaches to decision-analytic modelling are compared, several advantages can be found in the comprehensive approach. Firstly, it effectively integrates statistical evidence synthesis and parameter estimation with probabilistic decision-analytic model into a single unified framework. Secondly, the comprehensive approach enables the use of coherent Bayesian methods for updating prior distributions with available data; even in situations where priors and likelihoods are not conjugate distributions. Thirdly, the use of the comprehensive approach removes need to assume parametric distributional shapes for the posterior probability distributions. (Spiegelhalter 2004) Fourthly, MCMC simulation from the joint posterior distribution of model parameters will incorporate and propagate the dependency structure of model parameters (as a result of explicitly defined evidence structure), rather than assuming independency between the model parameters (Spiegelhalter 2004, 335, Ades et al. 2006). Fifthly, the comprehensive approach permits the incorporation of informative prior evidence directly in a decision-analytic model. However, the incorporation of informative prior distributions is not a necessary requirement in MCMC simulation, since non-informative prior distributions can be used when there is no relevant prior evidence available (Cooper et. al. 2004).

There are also some disadvantages relating to the comprehensive approach. First, the comprehensive approach is much more complex than the two-stage approach and its implementation requires full MCMC software, which is not very user friendly at the moment. Second, the comprehensive decision-analytic models may be computationally expensive in terms of the computer time required in simulation. For example, in the case study described in chapter 6.3 the first version of the model took over 24 hours to compute. However, reprogramming managed to reduce the computer time markedly and finally the generation of 10 000 samples took only approximately 15 minutes on a PC with a Pentium 4 CPU 2.66GHz-processor using 1.5 GB of RAM.

7.4 Applicability of different approaches to represent and interpret the cost-effectiveness results

Once the simulations of decision-analytic models have been undertaken, the results of simulation are usually reported with means and 95% credibility intervals (95% CrI) (see Tables 5 and 10 for example). The credibility intervals are analogous to confidence intervals from the frequentist approach but they allow a more flexible interpretation than the conventional confidence intervals. Hypothesis testing is not meaningful in the modelling context. In a cost-effectiveness analysis, the emphasis is placed on the estimation of mean cost and effect differences between health technologies of interest, since the incremental cost-effectiveness ratio or the corresponding net benefit estimate is estimated based on the mean differences.

The point estimates, such as ICER and ΔNB(λ), do not take into account uncertainty that is related to these factors. However, the modelling results can be presented on the cost-effectiveness plane, which depicts graphically the joint uncertainty around the mean cost difference (ΔC) and the mean effect difference (ΔE) (see Figures 24-26 for example). The graphical approach is illustrative, since it clearly defines the location of joint density on the four separate quadrants with totally different meanings. The interpretation of graphics is relatively easy when the joint density lies in only one quadrant (see Figure 34 for example), whereas the situation where the joint density is spread over all four quadrants is much more complicate to interpret. However, in such situations where the cost and effect differences are insignificant, it is possible to estimate the posterior probability that the joint density lies in the particular quadrant (see the results of the case study described in chapter 6.1 as an example) to aid decision-making.

When the aim is to illustrate uncertainty relating to the different levels of willingness to pay (λ), the credibility intervals (or the confidence intervals in the frequentist approach) and the cost-effectiveness acceptability curves (CEACs) are applicable to reflect that uncertainty. However, the 95% credibility intervals are only valid if all simulated replications are only in one quadrant of the cost-effectiveness plane, whereas the CEACs offer a solution to reflect uncertainty also in situations where the joint density lies more than one quadrant. Since the threshold value for λ itself is unknown, the CEACs are usually drawn as a function of λ. The use of CEACs allows the presentation of decision uncertainty in terms of probability that a new health technology is considered as being cost-effective for a given value of λ. One minus this probability reflects the error probability, that is, the chance that an inefficient decision will be made on the basis of the available evidence.

In addition to being used to depict parameter uncertainty, CEACs can be applied to illustrate heterogeneity and methodological uncertainty related to the cost-effectiveness results. For example, figures 30 and 31 depict multiple CEACs, which are used to reflect the same treatment choice between alternative health technologies in the presence of heterogeneity (i.e. the probability of cost-effectiveness may vary

according to age and sex). Similarly, the multiple CEACs can be used to depict methodological uncertainty that arise due to the disagreement of proper discount rates, time horizons or clinical endpoints. For example, in the case study in chapter 6.3, multiple CEACs were used to reflect the probability of cost-effectiveness conditional on a particular treatment endpoint (see figure 27 for example). Furthermore, they can also be applied to reflect the effect of additional evidence (i.e. prior evidence is updated by additional data) on the probability of cost-effectiveness in the Bayesian context (Briggs 2001).

Recent debate, however, has raised some concerns related to the capability of CEACs to reflect uncertainty in cost-effectiveness analysis. The CEACs have been claimed to be insensitive to any change of the joint density in the NW and SE quadrants of the cost-effectiveness plane and to a radial shift of the joint density in the NE and SW quadrants. (Groot Koerkamp et al. 2007) These limitations have been acknowledged, but the CEACs are still considered as useful in representing uncertainty, especially, when combined with alternative ways to represent the cost-effectiveness results, such as the cost-effectiveness planes and the EVPI analyses (Fenwick & Briggs 2007, Schwartz 2007).

7.5 Applicability of the value of information methods

When the cost-effectiveness of a new health technology is uncertain, the decision-makers worry about the expected costs of uncertainty (i.e. the probability of an inefficient decision multiplied by the consequences of that wrong decision). Any additional evidence that can reduce the expected cost of uncertainty around the decision is naturally valuable. However, when the decision about the acquisition of additional evidence is done, the expected benefit of additional evidence with expected cost of that additional evidence needs to be compared.

The concept of expected value of perfect information (EVPI) has been developed to provide a formal analysis of the expected benefit of additional evidence (Claxton 1999, Claxton et al. 2002). The applications of the EVPI approach were developed in the case studies in chapters 6.2 and 6.4. The results of these applications reflect the value of additional research at the population level. For example, Figure 36 illustrates the population EVPI. At the cost-effectiveness threshold of 42 500 euros per QALY, the population EVPI is 607 000 euros. This point, which is equal to the expected ICER, represents the maximum value of acquiring additional evidence and, if the fixed costs of proposed research are below this EVPI value, additional research is considered to be potentially cost-effective. When the value of λ is low (i.e. much less than 42 500 euros), the technology is not expected to be cost-effective and additional evidence is unlikely to change that decision (the EVPI is very low). Similarly, when the value of λ is higher (i.e. much higher than 42 500 euros) the technology does appear to be more cost-effective and hence, the uncertainty surrounding the decision decreases markedly. This happens because the expected ICER is much lower than the value of λ. This example shows that the population EVPI can be

used to define sufficient evidence to adopt a new technology by determining when it is inefficient to collect additional evidence (Sculpher & Claxton 2005). However, in reality, the object functions of society and commercial parties may differ from each other, which may lead to the different definitions of sufficient evidence. For example, the commercial value of additional evidence can be determined as the value of increased turnover due to the increased probability of regulation and reimbursement approvals, whereas the societal value of additional evidence can be determined as the amount of health gain e.g. measured in terms of QALYs.

The present study revealed some limitations related to the EVPI analyses. The EVPI analysis places the emphasis on the evaluation of the precision rather than the quality of evidence, meaning that the accuracy of EVPI estimations depend heavily on the quality of evidence and the validity of a developed decision-analytic model. The concept of EVPI increases also the number of uncertain parameters in the analysis. For example, the population relevant to a particular decision is in itself uncertain. Uncertainty also relates to epidemiological parameters, such as the incidence and prevalence of a disease. Furthermore, the future price levels and the effective lifetime of technologies in comparison are uncertain. For example, in the case study described in chapter 6.4 the effective lifetime of technology was assumed to be either 5 or 10 years, since no more precise estimates were available. In addition, the results of the EVPI analyses are conditional on the unknown level of λ and hence, the results have to be represented as a function of λ. A further problem related to the implementation of the concept of EVPI into practice is the fact that decision-makers may consider the concept of EVPI as being too complex. Recently experiences from UK indicate that the problems of conducting EVPI analyses may not be primarily technical or methodological but rather related to be the policy environment, where decisions related to reimbursement and research decisions are made in separate remits (Claxton & Sculpher 2006).

7.6 Future research indicated by the case studies

The applied case studies indicated a range of targets for further research related to both evidence synthesis and decision-analytic modelling. When the decision-analytic models are developed, one of the major limitations that affects the validity of developed models is an absence of head-to-head trials comparing the technologies of interest in a particular cost-effectiveness analysis. Since this is a common problem in the cost-effectiveness analyses, further applications to enable indirect comparison via a common comparator (e.g. placebo) are needed. Recent developments in medical statistics may offer a potential solution for the indirect comparisons in decision-analytic modelling (see Nixon et al. 2007 for example).

The present study as well as other recent developments (Claxton et al. 2002, Briggs et al. 2002) has mainly concentrated on modelling the random error of individual data sources rather than estimating

uncertainty that arises when evidence for model parameters is weak. The combination of data from multiple sources may produce more precise estimates due to the increased sample size but the combination of different types of studies may increase heterogeneity across the studies. However, recent developments related to generalised evidence synthesis may offer a solution to combine evidence from different sources into a single model parameter, while reflecting the potential sources of bias e.g. due to confounding variables and patient selection (Ades & Sutton 2006). Unfortunately, one disadvantage of this approach is that it increases the complexity of the analysis, which may cause problems with the presentation and interpretation of the results.

In addition, further developments to assess and reflect the quality of evidence used in the decision-analytic models are needed. The assessment of the quality of evidence is challenging, since as mentioned above, the concept of quality is multidimensional and therefore it is hard to encapsulate into one score that gives a quality weight for each study. Recent developments, however, may offer a potential basis for attempts to incorporate the quality of evidence into the decision-analytic models in the near future (Spiegelhalter et al. 2003, Braithwaite et al. 2007)

This study applied the Markov modelling approach. However, in the future it is important to try to develop new approaches that can be used to relax assumptions needed in the Markov modelling. For example, one critical constraint in a Markov model is that each patient can be only one health state at a time, which leads to the requirement for multiple distinct health states to represent all possible combinations (i.e. distinct health states are needed to model the subsequent course of disease conditional on a patient's previous history). One promising approach that relaxes the assumptions used in the Markov modelling is discrete event simulation (DES), which is widely used in the operational research (Caro 2005). Some cost-effectiveness analysis applications have been already published (see for example Caro 2005 and Heeg et al. 2005).

Further developments to enhance the handling of model uncertainty and transparency are also to be welcomed. The rapid development of information and communication technology may lead to the opportunity to utilise internet-based applications offering a more transparent approach to the decision-analytic modelling in the near future. For example, Hubben et al. (2007) have demonstrated that web-based user-interfaces enhance the usefulness of decision-analytic models in the support of decision-making[10]. However, they did not manage to develop a model that permitted incorporation of structural changes as a part of the uncertainty analysis. In addition, the use of web-based models may raise a range of intellectual property right issues that need to be resolved in advance.

At the moment, there are relatively few published studies that have validated the cost-effectiveness results by comparing the modelling results directly to the trial-based results (see Morris 1997 for exam-

[10] The interface is publicly accessible at http://pcv.healtheconomics.nl

ple). One reason for that is inherently true that this kind of validation will only be meaningful if the conditions under which a trial is actually implemented closely reflect those assumed in the decision-analytic model. However, since the case study described in chapter 6.1 was conducted before the actual clinical study was started, it would be interesting to ascertain the validity of the developed model to identify a cost-effective option by comparing the previously published modelling results to the results obtained from the clinical study after they are made available.

The present study did not consider the recently developed concepts of the expected value of partial perfect information (EVPPI), the expected value of sample information (EVSI), and the expected net benefits of sampling (ENBS) methods. The EVPPI identifies the type of additional evidence which is most valuable for a decision. EVSI is defined as the difference in net benefits between the baseline population EVPI and the posterior population EVPI estimated using updated probability distributions. Finally, one way of defining ENBS is that it represents the difference between EVSI and the cost of obtaining the additional data (i.e. sampling cost). (Ades et al. 2004) However, further research is needed to establish the feasibility of these advanced methods in practice. In addition, one further challenge that relates to the net benefit approach itself is the absence of explicit cost-effectiveness threshold(s) (λ). Therefore, further research is also needed to establish the explicit threshold value(s) or at least explicit criteria that affect the maximum willingness to pay for additional QALY in Finland (cf. e.g. Devlin & Parkin 2004, Dakin et al. 2006).

Finally, further consensus work is needed to develop Finnish standards for reporting the cost-effectiveness results. For example, in recent Finnish publications, the presentation of cost-effectiveness results on the cost-effectiveness plane (cf. Linna et al. 2002, Räsänen et al. 2006, Kellokumpu-Lehtinen et al. 2007) has differed somewhat from international recommendations (cf. Drummond et al. 2005, 40) and this may lead to misunderstandings in the interpretation of the cost-effectiveness results.

7.7 References

Ades A, Claxton K & Sculpher M. 2006. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Economics 15:373-381.

Ades A, Lu G & Claxton K. 2004. Expected value of sample information calculations in medical decision modeling. Medical Decision Making 24: 207-227.

Ades A, Lu G & Higgins J. 2005. The Interpretation of Random-Effects Meta-Analysis in Decision Models. Medical Decision Making 25(6): 646-654.

Ades A & Sutton A. 2006. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. J R Statist Soc A 169: 5-35.

Braithwaite RS, Roberts MS & Justice AC. 2007. Incorporating Quality of Evidence into Decision Analytic Modeling. Ann Intern Med 146(2): 133-141.

Briggs A. 2000. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 17(5): 479-500.

Briggs A 2001. Handling uncertainty in economic evaluation and presenting the results. Economic evaluation in health care. Merging theory with practice. M. Drummond and A. McGuire. New York, Oxford University Press.

Briggs A, Ades A & Price M. 2003. Probabilistic Sensitivity Analysis for Decision Trees with Multiple Branches: Use of the Dirichlet Distribution in a Bayesian Framework. Medical Decision Making 23(4): 341-350.

Briggs A, Goeree R, Blackhouse G & O'Brien B. 2002. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Medical Decision Making 22(4): 290-308.

Briggs A, Sculpher M & Claxton K. 2006. Decision modelling for health economic evaluation. New York, Oxford University Press.

Buscemi N, Hartling L, Vandermeer B, Tjosvold L & Klassen TP. 2006. Single data extraction generated more errors than double data extraction in systematic reviews. Journal of Clinical Epidemiology 59(7): 697-703.

Buxton M, Drummond M, Van Hout B, Prince R, Sheldon T, Szucs T & Vray M. 1997. Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life. Health Economics 6(3): 217-227.

Caro JJ. 2005. Pharmacoeconomic analyses using discrete event simulation. Pharmacoeconomics 23(4): 323-332.

Claxton K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 18(3): 341-364.

Claxton K & Sculpher M. 2006. Using value of information analysis to prioritise health research. Some lessons from recent UK experience. Pharmacoeconomics 24(11): 1065-1068.

Claxton K, Sculpher M & Drummond MF. 2002. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). The Lancet 360(9334): 711-715.

Cooper N, Sutton A, Abrams K, Turner D & Wailoo A. 2004. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. Health Economics 13(3): 203-226.

Dakin H, Devlin N & Odeyemi I. 2006. “Yes”, “no” or “yes, but”? Multinomial modelling of NICE decision-making. Health Policy 77: 352-367.

Devlin N & Parkin D. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 13: 437-452.

Drummond M, Sculpher M, Torrance G, O'Brien B & Stoddart G. 2005. Methods for the economic evaluation of health care programmes. New York: Oxford University Press.

Eddy D. 2006. Accuracy versus transparency in pharmacoeconomic modelling. Finding the right balance. Pharmacoeconomics 24(9): 837-844.

Fenwick E & Briggs A. 2007. Cost-Effectiveness Acceptability Curves in the Dock: Case Not Proven? Medical Decision Making 27(2): 93-95.

Groot Koerkamp B, Hunink M, Stijnen T, Hammitt J, Kuntz K & Weinstein M. 2007. Limitations of Acceptability Curves for Presenting Uncertainty in Cost-Effectiveness Analysis. Medical Decision Making 27(2): 101-111.

Heeg B, Buskens E, Knapp M, van Aalst G, Dries P, de Haan & van Hout. 2005. Modelling the treated course of schizophrenia: development of a discrete event simulation model. Pharmacoeconomics 23 Suppl. 1: 17-33.

Higgins J & Green S. 2006. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005].

Hubben GAA, Bos JM, Glynn DM, van der Ende A, van Alphen L & Postma MJ. 2007. Enhanced decision support for policy makers using a web interface to health-economic models--Illustrated with a cost-effectiveness analysis of nation-wide infant vaccination with the 7-valent pneumococcal conjugate vaccine in the Netherlands. Vaccine 25(18): 3669-3678.

Kellokumpu-Lehtinen P, Bergh J, Salminen E, Wiklund T, Lehtinen S, Aronen P & Sintonen H. 2007. Cost-effectiveness of intensive adjuvant chemotherapy for high-risk breast cancer: is tailored and dose-escalated chemotherapy with growth factor support (GFS) more costly and less effective than marrow-supported high-dose chemotherapy - results from a randomized study. Acta Oncol 46(2): 146-152.

Law MR, Wald NJ & Rudnicka AR. 2003. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ 326(7404): 1423-1430.

Linna M, Taimela E, Apajasalo M & Marttila RJ. 2002. Probabilistic sensitivity analysis for evaluating cost-utility of entacapone for Parkinson´s disease. Expert Rev. Pharmacoeconomics Outcomes Res. 2(2): 91-97.

Morris S. 1997. A comparison of economic modelling and clinical trials in the economic evaluation of cholesterol-modifying pharmacotherapy. Health Economics 6: 589-601.

Nixon R, Bansback N & Brennan A. 2007. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Statistics in Medicine 26(6): 1237-1254.

Räsänen P, Krootila K, Sintonen H, Leivo T, Koivisto A, Ryynänen O, Blom M & Roine R. 2006. Cost-utility of routine cataract surgery. Health Qual Life Outcomes 294: 74.

Sculpher M & Claxton K. 2005. Establishing the Cost-Effectiveness of New Pharmaceuticals under Conditions of Uncertainty—When Is There Sufficient Evidence? Value in Health 8(4): 433-446.

Sculpher M, Fenwick E & Claxton K. 2000. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. Pharmacoeconomics 17(5): 461-477.

Spiegelhalter D. 2004. Incorporating Bayesian ideas into health-care evaluation. Statistical Science 19(1): 156-174.

Spiegelhalter D, Thomas A, Best N & Lunn D. 2003. WinBUGS user manual.

Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO & the European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups and the National Cancer Institute of Canada Clinical Trials Group. 2005. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. N Engl J Med 352(10): 987-996.

Schwartz A. 2007. The Acceptability of Acceptability Curves: Comments on Groot Koerkamp and Others. Medical Decision Making 27(2): 96-97.

Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C & Luce BR. 2003. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices in Modeling Studies. Value in Health 6(1): 9-17.

8 CONCLUSIONS

Based on the study findings, the following conclusions can be drawn:

1. The Markov models offer flexible and modifiable mathematical structure to reflect the courses of diseases and their consequences in the presence of particular health technologies at both the patient group and the population level. The applications of Markov models are one potential way of informing decision-makers about the decision problems relating to the implementation of new health technologies, since they offer a systematic, clear structured and coherent framework to combine evidence from a range of sources and to assess the consequences of different decision options in advance. However, further developments are needed to improve the assessment of the effect of structural uncertainty on decision uncertainty.
2. A Bayesian approach to evidence synthesis offers a coherent and applicable approach to synthesising the available relevant evidence while still reflecting their imprecision and heterogeneity. The combination of data from several studies produces more precise estimates due to the increased sample size but the combination of different studies may simultaneously increase heterogeneity between the studies. Since the quality of evidence has obvious relevance to the validity of decision-analytic models, further developments are needed to reflect the quality of evidence used in the decision-analytic models. Furthermore, the present study proved that the data used in meta-analysis and decision-analytic models requires very often preparation and this can increase the number of additional sources of methodological and process uncertainty.
3. Both the two-stage approach applying Monte Carlo simulation and the comprehensive approach applying Markov Chain Monte Carlo (MCMC) simulation are valid ways to reflect the parameter uncertainty (i.e. the imprecision of parameter values) in the decision-analytic models. The main advantage of the comprehensive approach over the two-stage approach is that it provides on integrated approach to the decision-analytic modelling, where evidence synthesis and decision-analytic modelling are conducted simultaneously. This improves handling of the sources of uncertainty in the decision-analytic models. However, the disadvantage of the comprehensive approach is its additional complexity, which diminishes its applicability in practice. Therefore, more user friendly programs are needed to ensure that this approach can be applied more widely.
4. Incremental cost-effectiveness ratios, which are estimated as the ratio of the expected means, do not provide information about the joint uncertainty around the mean estimates. Therefore, alternative methods should be used to depict the joint uncertainty of expected costs and effects. Since single methods, such as the cost-effectiveness acceptability curves (CEACs) based on the net benefit estimates, have their own limitations in depicting the joint uncertainty, the simultaneous use of multiple approaches, such as the CEACs, the cost-effectiveness acceptability frontiers (CEAFs) and the expected value of perfect information (EVPI), to represent uncertainty in the cost-effectiveness analyses would be preferable. Furthermore, the conventional sensitiv-

ity analyses can be still recommended as ways of reflecting the methodological and structural uncertainty in the cost-effectiveness results.

5. The concept of expected value of perfect information (EVPI) is theoretically consistent but its applicability in practice is challenging due to the complexity of this concept. Therefore, more research is needed to determine how this useful concept can be better utilised in practice.

Appendix 1. WinBUGS code for the case study 6.4

```
# ECONOMIC EVALUATION OF SUNITINIB MALATE IN THE TREATMENT OF CYTOKINE-REFACTORY mRCC

model{

# SUNITINIB MALATE

        # Probability of progression

         tp_prog~dbeta(13.62, 154.37)

        #  Probability of death

         tp_death~dbeta(2.57, 60.429)

        # Transition probabilities

        for (i in 1:cycles) {

        TP_A_A[i]<- (1-tp_prog)/((1-tp_prog)+tp_prog+tp_death)
        TP_A_B[i]<- tp_prog/((1-tp_prog)+tp_prog+tp_death)
        TP_A_C[i]<- tp_death/((1-tp_prog)+tp_prog+tp_death)

        TP_B_B[i]<- 1-tp_prog
        TP_B_C[i]<- tp_prog
        }

        # Markov model

        # Number of individuals in each health state at the start

        pi[1,1]<- N
        pi[1,2]<- 0
        pi[1,3]<- 0

        for (i in 2:cycles){

                lambda[i,1,1]<- TP_A_A[i]
                lambda[i,1,2]<- TP_A_B[i]
                lambda[i,1,3]<- TP_A_C[i]

                lambda[i,2,1]<- 0
                lambda[i,2,2]<- TP_B_B[i]
                lambda[i,2,3]<- TP_B_C[i]

                lambda[i,3,1]<- 0
                lambda[i,3,2]<- 0
                lambda[i,3,3]<- 1
        }

        # Marginal probability of being in each state at time >1

        for (s in 1:S) {                        # Number of health states

                for(i in 2:cycles){

                pi[i, s]<- inprod(pi[(i-1),], lambda[i, ,s])

                }
        }
```

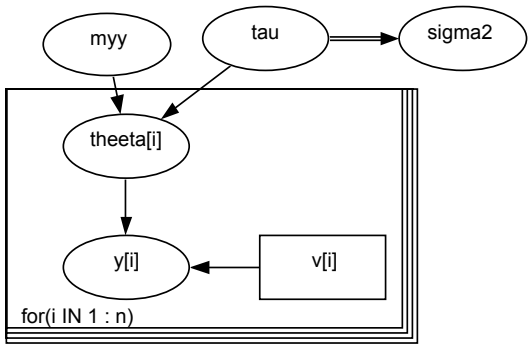
```
# Distributions for costs

for (i in 1:cycles){

C_sutent[i] ~ dunif(3748, 4061)
}

for (i in 1:2){

C_treat[i] ~dgamma(2600, 4.7761)
}

for (i in 3:4){

C_treat[i] ~ dgamma(1546, 4.7761)
}

for (i in 5:cycles){

C_treat[i] ~ dgamma(959, 4.7761)
}

# Distributions for utilities

for (i in 1:cycles){

U_stable[i] ~ dbeta(203.01, 62.71)

U_prog[i] ~  dbeta(37.90, 13.25)
}

# Costs in the treatment group

for (i in 1:cycles){

        c[i,1]<- (C_sutent[i] + C_treat[i])
        c[i,2]<- C_treat[i]
        c[i,3]<- 0

TotalCost[i]<- inprod(pi[i, ], c[i, ])/pow((1+discount), (i-1))
}

TotalCum.Cost<- sum(TotalCost[])

# Progression-free survival (PFS)

for (i in 1:cycles){

        PFS_S[i]<- inprod(pi[i, ], bl_PFS[])
}

TotalCum.PFS<- sum(PFS_S[])

# Overall survival (OS)

for (i in 1:cycles){

        OS_S[i]<- inprod(pi[i, ], bl_OS[])
}

TotalCum.OS<- sum(OS_S[])

# Quality adjusted survival

for (i in 1:cycles){

        u[i,1]<- U_stable[i]
        u[i,2]<- U_prog[i]
        u[i,3]<- 0

TotalUtility[i]<- inprod(pi[i, ], u[i, ])/pow((1+discount), (i-1))
```

```
}

TotalCum.U<- sum(TotalUtility[])

# Mean costs and benefits in the treatment group

mean.C<-TotalCum.Cost/ N
mean.PFS<-TotalCum.PFS/ N
mean.OS<- TotalCum.OS / N
mean.U<-TotalCum.U/ N

# BEST SUPPORTIVE CARE (BSC)

# Probability of progression in the BSC group

for(p in 1:39) {
        t.pfs[p] ~ dweib(gamma_1, mu1.pfs[p])
        mu1.pfs[p]<- exp(beta0.pfs)
    }

        beta0.pfs ~ dnorm(0.0, 0.0001)

        # Prior distribution for a shape parameter

        gamma_1 ~ dgamma(1, 0.0001)

        # Median survival

        median.pfs<- (log(2)/exp(beta0.pfs))

        #  Transition probability based on median survival time

        tp_prog_bsc <- 1-pow(0.5, (1/median.pfs))

# Probability of death after progression (OS - PFS) in the BSC group

        diff<- median.os - median.pfs
        tp_subprog_bsc <- 1-pow(0.5, (1/diff))

#  Probability of death
for(p in 1:39) {
        t.os[p] ~ dweib(gamma_2, mu2.os[p])
        mu2.os[p]<- exp(beta0.os+beta1.os*(t.pfs[p]-mean0.pfs))
    }

        beta0.os ~ dnorm(0.0, 0.0001)
        beta1.os ~ dnorm(0.0, 0.0001)

        # Prior distribution for a shape parameter

        gamma_2 ~ dgamma(1, 0.0001)

        # Median survival

        median.os<- (log(2)/exp(beta0.os+beta1.os))

        # Transition probability based on median survival time

        tp_death_bsc<- 1-pow(0.5, (1/median.os))

# Transition probabilities

for (i in 1:cycles) {

TP_A_A_bsc[i]<- (1-tp_prog_bsc)/((1-tp_prog_bsc)+tp_prog_bsc+tp_death_bsc)
TP_A_B_bsc[i]<- tp_prog_bsc/((1-tp_prog_bsc)+tp_prog_bsc+tp_death_bsc)
TP_A_C_bsc[i]<- tp_death_bsc/((1-tp_prog_bsc)+tp_prog_bsc+tp_death_bsc)


TP_B_B_bsc[i]<- 1-tp_prog_bsc
TP_B_C_bsc[i]<-  tp_subprog_bsc
```

```
}
# Markov model

pi_bsc[1,1]<- N                 # Number of individuals in each health state at the start
pi_bsc[1,2]<- 0
pi_bsc[1,3]<- 0

for (i in 2:cycles){

        lambda_bsc[i,1,1]<- TP_A_A_bsc[i]
        lambda_bsc[i,1,2]<- TP_A_B_bsc[i]
        lambda_bsc[i,1,3]<- TP_A_C_bsc[i]

        lambda_bsc[i,2,1]<- 0
        lambda_bsc[i,2,2]<- TP_B_B_bsc[i]
        lambda_bsc[i,2,3]<- TP_B_C_bsc[i]

        lambda_bsc[i,3,1]<- 0
        lambda_bsc[i,3,2]<- 0
        lambda_bsc[i,3,3]<- 1
}

# Marginal probability of being in each state at time >1

for (s in 1:S) {                # Number of health states

        for(i in 2:cycles){

        pi_bsc[i, s]<- inprod(pi_bsc[(i-1),], lambda_bsc[i, ,s])

        }
}

# Distributions for costs

for (i in 1:cycles){

        C_BSC[i] ~ dgamma(6389, 4.77)
}

# Costs in the BSC group

for (i in 1:cycles){

        c_bsc[i,1]<- C_BSC[i]
        c_bsc[i,2]<- C_BSC[i]
        c_bsc[i,3]<- 0

        TotalCost_bsc[i]<- inprod(pi_bsc[i, ], c_bsc[i, ])/pow((1+discount), (i-1))
}

TotalCum.Cost_bsc<- sum(TotalCost_bsc[])

# Progression-free survival (PFS)

for (i in 1:cycles){

        PFS_BSC[i]<- inprod(pi_bsc[i, ], bl_PFS[])
}

TotalCum.PFS_bsc<- sum(PFS_BSC[])

# Overall survival (OS)

for (i in 1:cycles){

        OS_BSC[i]<- inprod(pi_bsc[i, ], bl_OS[])
}

TotalCum.OS_bsc<- sum(OS_BSC[])
```

```
# Quality adjusted survival

for (i in 1:cycles){

        u_bsc[i,1]<- U_stable[i]
        u_bsc[i,2]<- U_prog[i]
        u_bsc[i,3]<- 0

TotalUtility_bsc[i]<- inprod(pi_bsc[i, ], u_bsc[i, ])/pow((1+discount), (i-1))
}

TotalCum.U_bsc<- sum(TotalUtility_bsc[])

# Mean costs and benefits in the BSC group

mean.C_bsc<-TotalCum.Cost_bsc/ N
mean.PFS_bsc<-TotalCum.PFS_bsc/ N
mean.OS_bsc<- TotalCum.OS_bsc / N
mean.U_bsc<-TotalCum.U_bsc/ N

# Incremental costs and benefits

d_C <- mean.C - mean.C_bsc
d_PFS <- mean.PFS - mean.PFS_bsc
d_OS <- mean.OS - mean.OS_bsc
d_QALY<- (mean.U - mean.U_bsc)/12

# Probability of cost effectiveness at K euros per progression-free months gained

for(k in 1:40) {

        INB.PFS[k] <- K[k] * d_PFS - d_C
        P.PFS[k] <- step(INB.PFS[k])

# Probability of cost effectiveness at K euros per life months gained

        INB.OS[k] <- K[k] * d_OS - d_C
        P.OS[k] <- step(INB.OS[k])

# Probability of cost utility at K euros per QALY

        INB[k] <- K[k] * d_QALY - d_C
         P.CUA[k] <- step(INB[k])
    }
}

# DATA

list(cycles=45, N=1000, S=3, bl_PFS=c(1,0,0), bl_OS=c(1,1,0), discount=0.004166, K=c(2500,…, 100000),t.os=c(4.67,…, 0.10),
t.pfs=c(0.87,…, 0.10))

# INITIALS

list(beta0.pfs=0, beta0.os=0, gamma_1=1, gamma_2=1)
```

Appendix 2. Graph- and text-based model descriptions for random-effects meta-analysis in WinBUGS.




```
model{
   # Description of model
      for (i in 1:n){
         v[i] <- 1/pow(SD[i],2)
         y[i] ~ dnorm(theeta[i],v[i])
         theeta[i] ~ dnorm(myy,tau)
      }
   # Prior distributions
      myy ~ dnorm(0,0.000001)
      tau ~ dgamma(0.001,0.001)
      sigma2<-1/tau
}
```

In the above figure, arrows represent a stochastic relationship, double arrows represents logical relationships, circles represent parameters, a square represents data and a large square plate represents loops (i= 1,..., n), where n is the number of studies included into the meta-analysis. In the specified model, between-study variation (v[i]) is handled as known and needed data is extracted from the selected studies. In addition, some data transformations are needed, since WinBUGS parameterises the normal distribution as N(μ, τ), where $\tau = 1/\sigma^2$.

Kuopio University Publications E. Social Sciences

E 133. He, Hong-Gu. Non-Pharmacological Methods in Children's Postoperative Pain Relief in China.
2006. 160 s. Acad. Diss.

E 134. Löfman, Päivi. Itsemääräämisen edistäminen. Osallistavan toimintamallin kehittäminen reumapotilaiden hoitotyöhön.
2006. 118 s. Acad. Diss.

E 135. Saaranen, Terhi. Promotion of school community staff's occupational well-being in co-operation with occupational health nurses – participatory action research in Eastern Finland in 2001–2004.
2006. 105 s. Acad. Diss.

E 136. Jansen, Ilona. An Archaeology of Philosophical Anthropology - A Reconstruction of the Historical Philosophical Background of Ethnography and Consequences for Nursing Science.
2006. 149 s. Acad. Diss.

E 137. Waldén, Anne. Muurinsärkijät - Tutkimus neurologisesti sairaan tai vammaisen lapsen perheen selviytymisen tukemisesta.
2006. 312 s. Acad. Diss.

E 138. Klemola, Annukka. Omasta kodista hoitokotiin. Etnografia keskipohjalaisten vanhusten siirtymävaiheesta.
2006. 145 s. Acad. Diss.

E 139. Willberg, Mirja. Millä perusteella kilpailuttamaan? Tutkimus sosiaali- ja terveyspalvelujen kilpailuttamisen ja sen toimijoiden tietoperustasta ja preferensseistä.
2006. 192 s. Acad. Diss.

E 140. Pirskanen, Marjatta. Nuorten päihteettömyyden edistäminen. Varhaisen puuttumisen malli koulu- ja opiskeluterveydenhuoltoon.
2007. 132 s. Acad. Diss.

E 141. Ensio, Anneli & Ryynänen, Olli-Pekka (toim.). Akuutin sydänpotilaan hoitoketjun arvioinnin kaksi näkökulmaa.
2007. 140 s.

E 142. Luukka, Katri. Vastavalmistuneiden lähihoitajien koulutuksenaikaiset merkitykselliset oppimiskokemukset vanhustyössä.
2007. 162 s. Acad. Diss.

E 143. Holopainen, Arja. Changes in meanings of teacherhood among finnish nurse teachers: a substantive theory of nurse teacherhood.
2007. 101 s. Acad. Diss.

E 144. Mäntynen, Raija. Kuntoutumista edistävä hoitotyö aivohalvauspotilaiden alkuvaiheen jälkeisessä moniammatillisessa kuntoutuksessa.
2007. 144 s. Acad. Diss.

E 145. Hautala-Jylhä, Pirjo-Liisa. Psychiatric Post-ward Outpatient Services – Between Hospital and Community.
2007. 112 s. Acad. Diss.

E 146. Kemppainen, Ulla. Ninth-grade adolescents´ health behavior in the Pitkäranta district (Russian Karelia) and in Eastern Finland.
2007. 149 s. Acad. Diss.