UNCERTAINTY IN MEDICAL DECISION MAKING

Knowing how little you know

Bas Groot Koerkamp

ISBN : 978-90-9024801-1
Cover design : Bas Groot Koerkamp
Layout : A.W. Everaers
Printed by : Print Partners Ipskamp

The printing of this dissertation was financially supported by the Departments of Epidemiology and Radiology of the Erasmus MC, Rotterdam, the Netherlands.

The research presented in this dissertation was financially supported by a grant (no. 904-66-091) from the Netherlands Organisation for Health Research and Development (ZonMW).

UNCERTAINTY IN MEDICAL DECISION MAKING

Knowing how little you know

Onzekerheid in medische besliskunde

Weten hoe weinig je weet

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 1 december 2009 om 13:30 uur

door

Bas Groot Koerkamp
geboren te Eindhoven

ERASMUS UNIVERSITEIT ROTTERDAM

PROMOTIECOMMISSIE

Promotores: Prof.dr. M.G.M. Hunink
Prof. M.C. Weinstein
Prof.dr. T. Stijnen

Overige leden: Prof.dr. E.W. Steyerberg
Prof.dr. W.B.F. Brouwer
Prof.dr. K. Claxton

Aan mijn ouders

CONTENTS

1

Introduction

GENERAL BACKGROUND

Medicine is decision making

Making decisions about the care of individual patients is fundamental to health care. For each patient, many decisions have to be made. In the emergency room, for example, a doctor should decide which patient to see first, decide whether an x-ray should be made of an injured ankle, and decide how this specific ankle fracture of this specific patient should be treated. Medical training is focused on acquiring the knowledge and experience to make such decisions. Other factors that are essential for patient care, including empathy and technical abilities, also involve decision making. For example, in the outpatient clinic, a trade-off is needed when one patient needs more time and empathy, but the waiting room is packed and the physician is an hour behind schedule. In the operating room, a surgeon must decide whether to proceed with a complicated laparoscopic procedure to remove a gall bladder, to convert to an open procedure, or to ask a more experienced surgeon for help.

Informal decision making is prone to error

In daily practice, most medical decisions are based on experience and judgment. Informally, an assessment is made of the probabilities and outcomes of each alternative, as well as the patient's preference for each outcome. Unfortunately, human judgment is fallible: people (including professionals) can make severe errors in estimating probabilities and outcomes.[1] Therefore, patients may benefit from a formal assessment of the probabilities and outcomes involved in a medical decision. Many decisions are nowadays resolved by such a formal assessment. For example, whether to make an x-ray of an injured ankle is resolved based on a decision rule.[2] On the other hand, a formal consideration of each individual decision with which a doctor is confronted seems infeasible.

Paradigms for formal assessment of decisions

Medical decision making (MDM) and evidence-based medicine (EBM) are separate paradigms that provide tools for formal assessment of medical decisions. They were developed because of concerns about human judgment, practice variation, and the proliferation of diagnostic and treatment options.[3] The mainstay of EBM is critical literature appraisal, starting with an answerable clinical question that is summarized in the mnemonic PICO: patient's problem, intervention, compare with alternative intervention, and outcome. The results of such an appraisal still demand considerable informal judgment on the part of the clinician: for example, results may not apply very well to an individual patient or studies may have conflicting results. Moreover, patient preferences, rare events, and health care costs are typically ignored in EBM, meaning that informal judgment on the part of the clinician still plays a role. MDM applies decision models to guide medical decisions and has

a strong foundation in decision theory.[4] Decision models can bring together all available evidence relevant for a decision; for example, disease incidence from population statistics, treatment effects from meta-analyses, patient preferences and rare complications from observational studies, and costs from medical claims databases. The model has no limit to the number of alternatives compared or to the length of follow-up. The aim of MDM is to perform a complete formal assessment of every aspect that is relevant for a decision. The main drawback of decision models is that building them is very time consuming. As a result, EBM has a higher acceptance in daily patient care and MDM in guideline development, health policy, and cost-effectiveness analysis.

Decision making and health care costs

Consideration of cost in addition to health benefits has more recently complicated decision making in health care. More beneficial health care interventions have become available than a health care system can afford. Priorities therefore have to be set. Most new interventions are beneficial but also more costly. Implementing such interventions requires increasing the overall health care budget or withholding other interventions. The latter seems fair only if the new intervention has a better value for money. The purpose of cost-effectiveness analysis is to provide information regarding the decision to implement new interventions by weighing the additional benefits against the additional costs. As a result, it may improve people's health by setting the appropriate priorities. As Stinnett noted, "investing in a cost-ineffective intervention is not simply an unwise use of money in some vague sense, but a foregone opportunity to achieve greater gains in people's health".[5] The cost-effectiveness of interventions is evaluated in clinical trials or in decision models. Trials have appeal because of a high internal validity[6], but only models can synthesize all available evidence.[7] The National Institute for Health and Clinical Excellence (NICE) makes recommendations on the adoption of health care interventions in the United Kingdom. In doing so it is required to explicitly consider cost-effectiveness.[8]

Decision making and uncertainty

Decision making is further complicated by uncertainty about probabilities and outcomes. For example, clinical trials often lack power to draw definitive conclusions. Even if arbitrary levels of significance are reached (typically a p-value < 0.05), there remains a finite possibility that the supposedly optimal intervention is not the "true" optimal intervention. However, while tests of hypotheses are relevant for exploring scientific phenomena, they are less useful in decision making. A decision has to be made, regardless of the amount of evidence and the extent of uncertainty.

Uncertainty is an even larger problem if clinical trials also consider health care costs, because the variation in costs typically exceeds the variation in health outcomes, requiring

larger sample sizes. When most uncertainty has been resolved at the decision level, uncertainty will remain at the patient level. For example, little doubt remains that a patient (in a good general condition) with a 7 cm aneurysm of the abdominal aorta is expected to be better off with elective repair of his aneurysm. However, it is still uncertain whether or not he will survive the intervention.

Various methods are used to present uncertainty. For clinical trials measuring efficacy, the consensus is that confidence intervals rather than p-values should be used to present the results. Confidence intervals don't blend magnitude and precision of the results, allowing for assessment of medical relevance in addition to statistical significance.[9] For cost-effectiveness analyses, the presentation of uncertainty is more complicated, and no consensus exists. The analysis and presentation of uncertainty in cost-effectiveness analyses is the first main focus of this dissertation.

Decision making and Bayesian methods

When clinicians evaluate the results of a new clinical trial, they informally consider other relevant evidence and judgment. For example, most clinicians will demand stronger experimental evidence for a homeopathic drug, of which efficacy other than placebo would conflict with chemistry than for a drug that was developed based on understanding of biochemical pathways. Bayesian methods enable an explicitly quantitative use of external evidence and judgment when interpreting data from a study.[10] This approach acknowledges that most decisions will not be based exclusively on the results of a single study. Bayes' theorem is a formula that can determine how evidence and uncertainty about the effect of interventions are changed by a new study. Currently, Bayes' theorem is not commonly used in interpreting the results of new studies. However, Bayesian methods are increasingly used in MDM. Moreover, Bayes' theorem is applied informally on a daily basis by doctors when they interpret diagnostic test results. For example, for a patient with a typical presentation of acute appendicitis, a moderate leucocytosis may add up to sufficient evidence to proceed to surgery. In a patient with non-specific abdominal pain, the same leucocytosis may not justify immediate surgery. Both the leucocytosis and the outcome of a new clinical trial are interpreted in light of what is already known.

Uncertainty and the need for more research

Clinical trials often show no statistically significant difference between the treatments compared. Two erroneous conclusions are common about the need for more research. Some authors conclude that the decision has been settled: the interventions are equivalent, and more research is not needed. However, it can and does occur that a clinically relevant difference is not found because the study did not have sufficient power to detect relevant differences.[11] In a famous quote, Altman warned that "absence of evidence is not evidence

of absence".[12] Other authors conclude the exact opposite — more research is needed — based on a similar non-significant difference. They seem convinced that there must be a difference and because it wasn't found with the current study, another (larger) study is needed. Phillips pointed out that the conclusion "more research is needed" requires some assessment of this research's expected benefit for future patients in relation to the cost of the research. He observed that studies in health care typically conclude that "more research is needed" without such an assessment.[13] Value of information (VOI) analysis evaluates uncertainty resulting in a formal assessment of the expected benefit and the cost of a proposed study.[14] VOI analysis is the second main focus of this dissertation.

Uncertainty and sample size calculations

When more research seems justified, the investigator should choose an optimal sample size for the proposed study. Decision uncertainty will decrease with increasing sample size, but the study costs will increase. Most sample size calculations are based on arbitrary values for the minimal clinically relevant difference in treatment effect, the level of significance (typically α=0.05) and desired power (typically 1-β=0.8). In practice the input parameters are often chosen to result in a predetermined sample size that primarily reflects feasibility and cost.[15] As an alternative to these arbitrary methods, VOI analysis can find the optimal sample size of a proposed study.

DEVELOPMENTS AND CHALLENGES

In this section we discuss the main developments regarding the analysis and presentation of uncertainty in cost-effectiveness analyses and VOI analysis. The analysis of uncertainty focuses on decision models. Developments regarding VOI analysis apply to both decision models and patient-level data from clinical trials.

Analyzing parameter uncertainty in decision models

The outcomes of a decision model depend on uncertain parameters, including probabilities, health state utilities, and costs. Deterministic sensitivity analysis can explore how the model outcomes change when individual parameters are varied across an often arbitrary range. The conclusions of a model may appear robust when the optimal intervention is insensitive to these analyses. However, a change in optimal intervention may prove more likely if several parameters are varied simultaneously. Unfortunately, deterministic sensitivity analysis allowing more than two parameters to vary simultaneously is rather cumbersome. Moreover, the analyses don't present how likely each model outcome is.

Probabilistic sensitivity analysis (PSA), also known as second-order Monte Carlo simulation, was introduced to evaluate uncertainty of all parameters simultaneously.[16, 17] Probability distributions are assigned to each parameter and propagated through the model resulting in probability distributions for the model outcomes. A distribution of the expected outcome presents the range of possible outcomes as well as the probability of each outcome. The National Institute for Health and Clinical Excellence (NICE) in the United Kingdom requires PSA for the assessment of health technologies.[18, 19]

Stochastic uncertainty can complicate the assessment of uncertainty in decision models. A PSA in a patient-level model, also known as microsimulation model, requires a computer-intensive two-level Monte Carlo simulation to obtain the distribution of the expected outcome.[20] The most common rationale for patient-level models is that events within the model influence subsequent probabilities, utilities, or costs. The simultaneous analysis of patient heterogeneity and parameter uncertainty poses another challenge.[21] Subgroups should be evaluated because the cost-effectiveness of an intervention may vary across subgroups.

Presenting uncertainty of cost-effectiveness analyses

Guidelines for cost-effectiveness analyses require that a measure of uncertainty is presented in addition to point estimates.[6] Typically the preferred measure of uncertainty is not specified because no consensus exists about the optimal measure. The cost-effectiveness plane (CE plane) may seem an attractive candidate because it presents the joint distribution of each intervention.[22] For the comparison of two interventions, the CE plane can present the incremental joint distribution. The main drawback of the CE plane is that it can overestimate decision uncertainty when more than two interventions are compared and — additionally — the outcomes of the interventions are positively correlated. While clinical trials typically compare two interventions, comparing more than two positively correlated interventions is the rule rather than the exception in decision models.

In the search for alternatives, separate hypothesis tests for costs and effects were quickly dismissed.[23] Instead, many methods were developed for inference about the incremental cost-effectiveness ratio (ICER).[24, 25] However, it was soon realized that uncertainty intervals for the ICER are only valid when the incremental distribution is limited to the first or third quadrant of the CE plane.[5] Inference about the ICER became less popular, because any decision requiring formal consideration has some uncertainty regarding the sign of either incremental cost or effect. Van Hout introduced the cost-effectiveness acceptability curve (CEAC) as an alternative to the discredited inference about the ICER.[26] The curve provides a graphical presentation of the probability that an intervention is optimal, across a range of values of the willingness-to-pay (wtp). CEACs are created using bootstraps of patient-level data or the results of a PSA in decision models. The main drawback of CEACs is that

they consider only the probability of selecting a suboptimal intervention and not the consequences.

Stinnett introduced the net benefit framework as another alternative to analyze and present uncertainty.[5] The net health benefit (NHB) is defined as: NHB = health effect – cost/willingness-to-pay. Standard statistical methods can now be applied to calculate point estimates and uncertainty intervals using patient-level data or PSA results of a decision model. For patient level data, the analyst can choose between assuming normality of the mean NHB — based on the central limit theorem — or non-parametric methods such as bootstrapping. The (incremental) NHB can be presented graphically including uncertainty boundaries across a range of values for the wtp, similar to a CEAC. The main drawback of this presentation is that it can overestimate decision uncertainty when more than two interventions are compared and — additionally — the outcomes of the interventions are positively correlated.

VOI analysis

The analysis of uncertainty is particularly relevant for the decision as to whether more quantitative research (e.g., a clinical trial) regarding the competing interventions is justified.[27] A future study could reduce uncertainty, which is expected to benefit future patients or reduce health care costs. A cost-benefit trade-off should be made — before performing a study — between the expected cost and the expected benefit of a proposed clinical study. The trade-off assumes that money spent on either health care or research aims to improve the health of patients and that the funds come from the same limited resources. VOI analysis can estimate the expected benefit of a proposed study using a decision model or the results of a clinical trial. The analysis involves integrating the probability of implementing a suboptimal intervention with the associated consequences in foregone health benefits or increased health care costs.

Investigators increasingly use VOI analysis to analyze and present uncertainty in medical decision making.[28] VOI analysis was introduced by Grundy[29] in the late fifties and developed by Raiffa and Schlaifer.[4] Howard noted in 1966 that: "Placing a value on the reduction of uncertainty is the first step in experimental design, for only when we know what it is worth to reduce uncertainty do we have a basis for allocating our resources in experimentation designed to reduce the uncertainty."[30] In 2002, Claxton introduced the philosophy of VOI analysis to the clinical audience of the Lancet.[14] Moreover, he demonstrated the feasibility of VOI analysis to guide the research priority setting of the National Health Service in the United Kingdom.[31] Ades clarified the mathematical notation and algorithms for the various VOI analyses.[32]

VOI analysis typically starts with estimating the total expected value of perfect information (total EVPI). It is the expected benefit per patient of a study with an infinite sample size, resulting in perfect information about all (total) uncertain parameters. Such a study would eliminate all uncertainty, but is of course hypothetical. The population EVPI is the total EVPI per patient multiplied by the number of future patients expected to benefit from the results of a proposed study. More research is not justified if the population EVPI is less than the fixed costs of a future study. If the population EVPI exceeds the fixed costs, more research is potentially justified.

Currently, VOI applications using patient-level data or complex decision models typically don't proceed beyond estimating the population EVPI.[33, 34] However, VOI analysis can also help to identify the study that maximizes the difference between the expected benefit for future patients and the expected cost of the study. This optimal study is characterized by its design (e.g., randomized controlled trial), the subset of sampled parameters (e.g., quality of life only, or a selection of cost parameters), the sample size, and the associated study costs. Partial VOI analysis is performed to select the optimal subset of sampled parameters, representing key parameters that are responsible for most of the decision uncertainty. The expected value of sample information (EVSI) is estimated to find the optimal sample size of a study. The required methods to find the optimal study — estimating the partial EVPI and partial EVSI — were only recently resolved and require considerable computational time. These methods are typically applied to hypothetical data or simple decision models. These initial examples often used closed form solutions, while increasingly complex applications require simulation methods.[28]

OUTLINE OF THIS DISSERTATION

The research presented in this dissertation is divided over eight chapters. In general, Chapters 2 to 6 have a more methodological character. Chapters 7 to 9 present applications of novel methodology to clinical problems.

In Chapter 2 we introduce cost-effectiveness analysis to a clinical audience of surgeons. New interventions in surgery typically provide small additional health benefits and are more expensive in comparison to current care. If a health care system cannot afford all beneficial new interventions, cost-effectiveness analysis can help set priorities. This article demonstrates the basics of evaluating and performing cost-effectiveness analyses. In addition, we discuss various challenges for the application of cost-effectiveness analysis, including challenges specific to surgical interventions.

Chapter 3 is a tutorial about uncertainty and patient heterogeneity in medical decision models. Parameter uncertainty, patient heterogeneity, and stochastic uncertainty of outcomes are increasingly important concepts in medical decision models. We demonstrate the various methods to analyze uncertainty and patient heterogeneity in a decision model using a real-life example. Differences and analogies between the analyses are pointed out, as well as practical issues. The scope includes nested Monte Carlo simulations that are required in patient-level models for PSA and in nonlinear models for partial VOI analysis.

Chapter 4 focuses on the combined analysis of uncertainty and patient heterogeneity in medical decision models. When more than one type of uncertainty and heterogeneity is analyzed, the correct algorithm to obtain the model outcomes of interest can be complicated. We distinguish eight model types, each dealing with a different combination of parameter uncertainty, patient heterogeneity, and stochastic uncertainty. The model outcomes of interest include the expected outcome, the distribution of the expected outcome reflecting lack of perfect knowledge, the distribution of the expected outcome reflecting patient heterogeneity, and the distribution of the individual outcome. The analyses — required to obtain the model outcomes — are expressed in equations, explained in stepwise algorithms, and demonstrated in examples. Nested Monte Carlo simulations are necessary for most analyses.

In Chapter 5 we demonstrate the limitations of acceptability curves for presenting uncertainty in cost-effectiveness analyses. Clinical journals increasingly illustrate uncertainty about the cost and effect of health care interventions using cost-effectiveness acceptability curves (CEACs). CEACs present the probability that each competing alternative is optimal across a range of values of the cost-effectiveness threshold. First, we discuss why uncertainty is of interest to policy makers, since it may not be obvious that the evaluation of uncertainty results in better decisions. Next, we demonstrate and explain the various limitations of CEACs.

In Chapter 6 we evaluate two different approaches for estimating the partial expected value of perfect information (partial EVPI) to identify key parameters in cost-effectiveness analysis. Both approaches have been described and applied in the literature, but they may result in different outcomes, and a different importance ranking of parameters. The objectives of this manuscript are to set out the correct methods to estimate partial EVPI and to explain and demonstrate why a generally recommended method is incorrect both conceptually and mathematically.

In Chapter 7 we apply VOI analysis to a decision model comparing various diagnostic tests for coronary heart disease. The objective is to design the optimal future study regarding

diagnostic tests for patients with chest pain. We estimated the partial EVPI for various study designs: observational studies for test characteristics and health state utilities, a cost study, and clinical trials measuring treatment effects. For the study measuring health state utilities we estimated the optimal sample size.

In Chapter 8 we apply VOI analysis to data from an economic clinical trial. Most published applications of VOI analysis have used decision models or hypothetical trial data. We explain how VOI analysis using patient-level data about costs and effects is simplified using several justifiable assumptions. As an example, we use data from a clinical trial comparing treatments of intermittent claudication. The objective of the VOI analysis is to design a future study that maximizes the difference between the expected benefit for future patients and the expected cost of the study. This optimal study is characterized by its design, the subset of sampled parameters, the sample size, and the associated study costs.

In Chapter 9 we apply VOI analysis to guide future outcomes research regarding the use of MR imaging for patients with acute knee trauma in the emergency room setting. We estimate the total EVPI reflecting the expected value of eliminating all decision uncertainty that remained after completion of a clinical trial. In addition, we identify the parameters that are responsible for most of the decision uncertainty, evaluate the expected benefit of various study designs, and estimate their optimal sample size.

Finally, Chapter 10 summarizes the previous chapters, and Chapter 11 is an epilogue containing some concluding thoughts and suggestions for future research.

2

Cost-effectiveness analysis for surgeons

Bas Groot Koerkamp
Y. Claire Wang
M.G. Myriam Hunink

Surgery. 2009 Jun;145(6):616-22.

Invited paper in Clinical Research issue of Surgery.

ABSTRACT

New interventions in surgery typically provide small additional health benefits and are more expensive in comparison to current care. If a health care system cannot afford all beneficial new interventions, cost-effectiveness analysis can help set priorities. In this setting, surgeons increasingly encounter policy makers questioning the cost-effectiveness of the treatments they provide. Especially when writing guidelines, surgeons should understand cost-effectiveness analysis to evaluate the value for money of their interventions.

The value for money of a health care intervention can be expressed by its incremental cost-effectiveness ratio (ICER). The ICER is the ratio of the additional cost and the additional effect of an intervention (e.g., laparoscopic cholecystectomy) in comparison to an alternative intervention (e.g., open cholecystectomy) for the same patient. These costs and effects can be evaluated in randomized controlled trials or in decision models. This article demonstrates the basics of evaluating and performing cost-effectiveness analyses.

The application of cost-effectiveness analysis to help set priorities in health care faces several challenges, including credibility, generalizability, and ethical implications. Additional challenges more specific to surgery include the learning curve for new surgical interventions and the gradual improvement of surgical technology. Adherence to guidelines for cost-effectiveness analyses could address some challenges; other challenges simply reflect the difficulty of making decisions under uncertainty. Despite these challenges, priorities have to be set. In the UK cost-effectiveness is explicitly considered for the adoption of interventions whereas in the US this is done more implicitly.

INTRODUCTION

"In 40 years of practice I never withheld a test or treatment that I thought would help my patients, even when I had to absorb the costs myself."[35] Like this surgeon, most surgeons want the best possible care for their patient, regardless of the costs. Surgeons, however, are increasingly likely to encounter value for money issues. Hospitals and insurance companies may interfere with their patient care when, for example, a PET scan is not reimbursed. In the development of novel technologies, such as laparoscopic procedures, surgeons often need to convince policy makers that the new interventions are good value for money. When writing guidelines, surgeons are expected to consider the value for money of their interventions.

What has changed in recent decades is that more beneficial health care interventions have become available than a health care system can pay for. Priorities therefore have to be set. Interventions that are beneficial by reducing mortality or morbidity and save money are straightforward: They should be implemented. Most new interventions, however, are beneficial but more costly. Implementing such interventions requires increasing the health care budget – at the expense of, for example, welfare or education – or withholding other interventions. The latter seems only fair if the new intervention has a better value for money. Cost-effectiveness analysis aims to inform the decision to implement new interventions by weighing the additional benefits against the additional costs. Its objective is not to cut health care costs. Cost-effectiveness analysis aims to improve the health of patients by setting the right priorities.

This article is organized as follows. First, we demonstrate the basic concepts of cost-effectiveness analysis, using two recent examples from the surgical literature. We discuss how to find, evaluate, and perform cost-effectiveness analyses, especially in the field of surgery. We then address several important challenges for cost-effectiveness analysis, some specific to the analysis of surgical interventions. Finally, we address the current use and impact of cost-effectiveness analysis in the United States and the United Kingdom.

COST-EFFECTIVENESS ANALYSIS - THE BASICS

Box 1 and 2 present summaries of two cost-effectiveness analyses that we use to illustrate some of the concepts in this section.

BOX 1: LUNG-VOLUME REDUCTION SURGERY

Methods

- Patients with severe emphysema
- Lung-volume reduction surgery vs. medical treatment
- Randomized controlled trial – 1218 patients
- Societal perspective
- Health effects: mortality, exercise capacity, quality-of-life
- Cost: medical and non-medical costs, based on resource use in trial and Medicare charges
- Time horizon: 3 years trial data plus 7 years modeled
- Discounting: both cost and effects

Cost-effectiveness analysis

- Surgery more effective: difference of 0.19 QALYs
- Surgery more costly: difference of 36,000$
- ICER of 190,000 $/QALY

Subgroups

- upper-lobe emphysema and low baseline exercise capacity: ICER of 98,000 $/QALY
- non-upper-lobe emphysema and low baseline exercise capacity: ICER of 330,000 $/QALY
- Substantial uncertainty for subgroup estimates

Model with 10 years time horizon

- at 10 years: ICER of 53,000 $/QALY
- Substantial uncertainty for 10-year estimates

Conclusion

For patients with severe emphysema – esp. with upper-lobe disease and low baseline exercise capacity – lung-volume reduction surgery may be cost-effective as compared to medical treatment.

BOX 2: PRIMARY HYPERPARATHYROIDISM (PHPT)

Methods

- Observation vs. pharmacologic vs. parathyroidectomy (PTX)
- Asymptomatic 60-year-old patient with PHPT
- Decision model
- Third-party-payer perspective
- Health effects: QALYs
- Costs: medical costs using charges
- Time horizon: remaining life time
- Discounting: both cost and effects

Cost-effectiveness analysis

- PTX more effective than observation: difference of 0.16 QALYs
- PTX more costly than observation: difference of $800
- ICER – PTX vs. observation: $5,000 per QALY
- Pharmacologic more effective than PTX: difference of 0.01 QALYs
- Pharmacologic more costly than PTX: difference of $176,000
- ICER – pharmacologic vs. PTX: $21,000,000 per QALY
- Sensitivity analysis: robust regarding costs and complications, sensitive to quality of life improvement after PTX

Conclusion

For 60-year-old patients with asymptomatic PHPT, PTX is very cost-effective as compared to observation and pharmacologic treatment is not cost-effective.

Health effects

When evaluating surgical interventions, we are typically interested in more than one health effect or outcome. For example, the study on lung-volume reduction surgery considered both mortality and the exercise capacity of the patients (box 1).[36, 37] The patient and the surgeon have the difficult task of aggregating and integrating these outcomes with the patient's preferences. Using quality-of-life as the outcome measure has the advantage of aggregating all health outcomes, except survival. Several methods are available to measure the quality-of-life.[38] The study on lung-volume reduction surgery also measured quality-of-life with a questionnaire at baseline, at 6 months, at 12 months, and yearly thereafter. The outcome quality-adjusted life expectancy (QALE) integrates the survival data with the quality-of-life and is expressed in quality-adjusted life years (QALYs). This aggregate measure is commonly used in cost-effectiveness analyses and facilitates the comparison of interventions.

Costs

For an unbiased comparison of the costs of interventions, all short-term and long-term costs should be included. The study on primary hyperparathyroidism (box 2) included the costs of surgery, diagnostic tests, hospital stay, complications, medications, physician's fees, and follow-up visits.[39] Consideration of some costs depends on the perspective from which the analysis is conducted. From the societal perspective, non-medical costs – including the costs of patient's time and travel, as well as productivity loss during illness – are relevant in addition to medical costs. However, they may not be relevant from the third-party payer's perspective. The study on lung-volume reduction surgery (box 1) was performed from the societal perspective and included both medical and non-medical costs. Even if a study represents your perspective, the costs may not be representative for your setting. The study on lung-volume reduction surgery (box 1), for example, used US costs that may differ from costs in other countries. The source of the cost data is also important. The study on primary hyperparathyroidism (box 2), for example, used charges to represent the costs. Although data on charges are more readily available, they typically overestimate actual costs and require adjustments using cost-to-charge ratios.

Cost-effectiveness

An intervention is considered cost-effective when it has an additional health benefit deemed worth the additional cost.[40] The value for money of a health care intervention is usually measured by its incremental cost-effectiveness ratio (ICER). The ICER is the ratio of the additional cost and the additional effect of an intervention (e.g., laparoscopic cholecystectomy) in comparison with an alternative (e.g., no surgery or open cholecystectomy) for the same patient. The cost-effectiveness of an intervention can be expressed in, for example, cost per life-saved, cost per life-year saved, or cost per averted cancer.

The study on primary hyperparathyroidism (box 2) estimated the cost per quality-adjusted life-year (QALY) of each intervention. The advantage of using QALYs to measure health effects is that it allows comparison of the value for money of preventive, diagnostic and therapeutic health care interventions for different indications or disease areas. For example, the cost-effectiveness of parathyroidectomy for asymptomatic primary hyperparathyroidism (box 2) can be compared directly with other commonly adopted interventions such as kidney dialysis. All health care interventions can be ranked according to their cost-effectiveness. In theory, the greatest health of a population is reached when interventions are funded from the top of the cost-effectiveness ranking downward until the health care budget is exhausted. The state of Oregon has attempted to establish a priority list for its Medicaid benefit package following such a framework, although the final plan differs from its original conception after much controversy.[41] Other factors than cost-effectiveness were important to policy makers; some of these issues will be discussed later.

A universally agreed-upon threshold value for the cost per QALY does not exist, despite that several figures – for example, $50,000 or $100 000 per QALY gained – are frequently cited. In evaluating interventions in the developing world, the World Health Organization gives the following general guidance: an ICER of less than the GDP (gross domestic product) per capita is very cost-effective, between 1 and 3 times GDP per capita is cost-effective, and more than 3 times GDP per capita is not cost-effective.[42] For the US in 2007 these values would be $46,000/QALY, between $46,000 and $138,000/QALY, and $138,000/QALY respectively.[43] Alternatively, the cost per QALY can be compared with established ICERs of generally accepted interventions. For example, the ICER for kidney dialysis and statins are often used to advocate adoption of interventions with lower or similar cost-effectiveness.

Study design

Surgeons often encounter cost-effectiveness analyses as part of a randomized controlled trial. Data on resource use (e.g., the number of hospital days) is collected for each patient together with health outcomes. The study on lung-volume reduction surgery (box 1) used this design. Randomization minimizes confounding on both cost and effectiveness estimates, resulting in high internal validity. Patients in clinical trials, however, often differ from the general patient population in demographics or disease severity and may receive care different from normal practice, resulting in lower external validity (or generalizability). Moreover, the follow-up period in trials is often insufficient to observe all relevant costs and effects. For example, endovascular repair of abdominal aortic aneurysms results in more re-interventions in the long-term than open repair. Not considering the cost and harm of these re-interventions can lead to a bias favouring endovascular repair. Another potential drawback of trials is that they are often under-powered to capture differences in rare complications[44] – for example, injury to the common bile duct in open versus laparoscopic cholecystectomy.

An alternative study design is to use a decision model to assess cost-effectiveness. These models aim to bring together all available evidence from various sources – for example, disease incidence from population statistics, treatment effects from meta-analyses, rare complications from observational studies, and costs from medical claims databases. The model has no limit for the number of alternatives compared or the length of follow-up. However, in building a model many assumptions – sometimes implicit – are made. Consequently, models that consider the same interventions may show large variation in incremental cost-effectiveness ratios, as has been demonstrated for screening for abdominal aortic aneurysms.[45] Advocates for economic trials continue to emphasize their internal validity, while advocates for models stress their external validity and consideration of all available evidence. Their relative value depends on the decision at hand and the available evidence.

Time horizon and discounting

Cost-effectiveness analyses must have an appropriate time horizon to include all relevant (immediate and future) costs and effects associated with the interventions. For example, the study on primary hyperparathyroidism (box 2) considered the lifetime costs of calcitriol when surgery inadvertently resulted in hypoparathyroidism. Such future costs should be discounted to "present values". The rationale behind discounting costs is that people would rather pay a bill in 10 years than today. Surgical interventions typically incur most costs up-front, whereas medical treatments may involve long-term cost. As a result, parathyroidectomy for primary hyperparathyroidism appears less cost-effective compared with medical treatment when costs are discounted than if costs are undiscounted. Analogously, future health effects are discounted because people also prefer immediate health gains over the same gains occurring in the future. For example, a screening program for colon cancer appears less cost-effective when the future health benefits are discounted than if discounting is omitted. Often analysts present both undiscounted and discounted results.

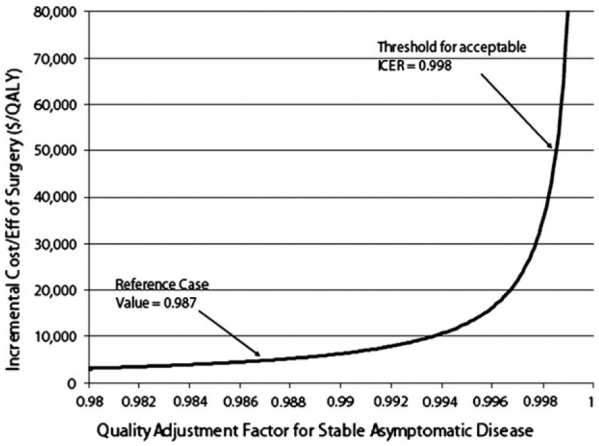

Figure Sensitivity analysis for the quality-of-life (or quality adjustment factor) for asymptomatic hyperparathyroidism. The reference case is the best estimate for the quality-of-life that was used for the main results. Surgery would be the preferred intervention for values for the quality-of-life below 0.998, assuming a threshold for acceptable ICERs of $50,000 per QALY. – reprinted with permission from Surgery 2006;140:874-81.

Uncertainty

Just like other clinical outcome measures, ICERs calculated from randomized controlled trials are imprecise because of the finite sample size. This uncertainty is often presented as the probability that an intervention is cost-effective, under different threshold criteria of cost-effectiveness. In decision models uncertainty is typically explored using sensitivity analysis. Key parameters are varied across a plausible range and the impact on the ICER is presented graphically. The study of primary hyperparathyroidism (box 2) performed

BOX 3: COST-EFFECTIVENESS ANALYSIS – CHECKLIST

Validity of the results

- Is the population of interest well-defined?
- Are all relevant alternatives well-defined and considered?
- Is the perspective stated and are all relevant costs included?
- Is the source of the effect provided?
- Is the measurement of the quality-of-life stated?
- Is the source of the costs provided?
- Is time horizon sufficient for all relevant costs and effects?
- Is discounting applied to both costs and effects?
- Regarding models: Are the structure, data, and assumptions disclosed and valid?

The results

Did the analysis consider;

- Incremental costs and effects for each strategy?
- Incremental cost-effectiveness ratio?
- Sensitivity analyses and subgroup analyses?

Applying results to my patient

- Does the study reflect my patient population?
- Does the study reflect my perspective and setting?
- Which intervention should I adopt?

sensitivity analyses and found that the conclusion is sensitive to the assumption on the quality-of-life improvement expected after parathyroidectomy in asymptomatic 60-year-old patients (Figure). However, the cost-effectiveness of parathyroidectomy is quite stable given a plausible range of costs and complication rate estimates. This indicates that further research yielding more precise estimates of the quality-of-life improvement would make the model more robust.

Checklist and read more

Box 3 presents a concise checklist for evaluating and performing a cost-effectiveness analysis. It summarizes the issues that we have discussed in this section. More extensive guidelines for reporting cost-effectiveness analyses have been published.[46, 47] Moreover, many textbooks are available about decision making in general, and about economic evaluation alongside clinical trials or using decision models.[38, 48-50] Eddy has written an excellent series in the Journal of the American Medical Association in which he explains cost-effectiveness analysis to his father, a retired surgeon.[35, 51-53] The Users' Guides on Evidence-based Medicine include 2 articles about economic analysis in clinical practice, which are also avail-

Table *Some recent cost-effectiveness analyses in surgery.*

Patients	Treatment	Alternative	ICER*
Hyperthyroidism and solitary thyroid nodule[67]	Thyroid lobectomy	Radio-active iodine	13,000
Morbid obesity with serious comorbidity[68]	Gastric banding	No surgery	16,000
Abdominal aortic aneurysm 6.5cm, age [8069]	Endovascular repair	Open surgical repair	17,000
Severe claudication[70]	Duplex ultrasound and angioplasty	Supervised exercise program	24,000
Rectal cancer[71]	Pre-operative radiotherapy	No pre-operative radiotherapy	29,000
Liver transplant waiting list[72]	Liver transplantation	Medical treatment	54,000
Asymptomatic men aged 65 to [7473]	Screening for abdominal aortic aneurysms	No screening	68,000
Severe ileocolonic Crohn's disease DDDISEDISEASEdisease[74]	Infliximab and Remicade	Surgery	78,000
Stage I or II breast cancer[75]	Patient choice†	Mastectomy	94,000
Heavy-smokers, eligible for lung resection[76]	Annual CT screening	No screening	140,000

ICER, Incremental Cost-effectiveness Ratio, expressed in dollars per quality-adjusted life year (QALY) gained.
*Adjusted to 2007 $ per QALY gained from the cost-effectiveness analysis Registry of Tufts Medical Center.
†Patient choice between breast conservation surgery with radiation treatment or mastectomy versus mastectomy.

able from the website of the Centre for Health Evidence in Canada.[54-56]

COST-EFFECTIVENESS ANALYSES IN THE SURGICAL LITERATURE

Cost-effectiveness analyses of surgical interventions are increasingly performed, although the absolute numbers are still low in comparison with other fields in health care.[57] Kruper et al. found only 110 cost-effectiveness analyses of surgical procedures between 1995 and 2004.[58] The Table presents some cost-effectiveness analyses of surgical interventions with their ICER.

Quality of cost-effectiveness analyses for surgical interventions

A good research study should comply with certain methodologic criteria to allow for evaluation of its validity. For example, most journals require authors to adhere to the CONSORT statement (www.consort-statement.org) for reporting trials. In the previous section we discussed similar methodologic checklists for cost-effectiveness analyses.[46, 47] A recent study evaluated cost-effectiveness analyses of surgical procedures using a list of Blackmore et al. with 10 basic methodologic principles.[59] On average these studies adhered to only 4 out of 10 principles: The perspective of the study is rarely defined, relevant

long-term costs are often not included, the outcome measure is sometimes unclear, and the source of the cost data is sometimes not explained.[58] These observations present room for improvements for future cost-effectiveness analyses in the surgical literature.

Finding a cost-effectiveness analysis

Cost-effectiveness analyses of surgical interventions can be found in Medline (www.pubmed.gov) using the Medical Subject Heading (MeSH) term "cost-benefit analysis", which includes "cost effectiveness" as an entry term. In addition, 2 databases are accessible on the internet free of charge. The Economic Evaluation Database of the British National Health Service (www.crd.york.ac.uk/crdweb) contains over 7000 abstracts of quality assessed economic evaluations. The same database is also available through the Cochrane Library. The Cost-effectiveness analysis Registry of Tufts Medical Center in Boston, MA (https://research.tufts-nemc.org/cear/default.aspx) provides access to a comprehensive database of cost-effectiveness ratios in the medical literature.

CHALLENGES FOR COST-EFFECTIVENESS ANALYSES

We will discuss several challenges for cost-effectiveness analysis of surgical procedures. Most of these challenges are also encountered with other types of studies and outside surgery. The last 2 challenges are more unique to the applications in surgery.

Credibility of the cost-effectiveness analysis

Cost-effectiveness analyses are often performed using decision models. Surgeons may perceive a decision model as a "black box". Lack of transparency regarding data sources and assumptions is sometimes responsible for this perception. Therefore, researchers should clearly state all data sources and assumptions in a comprehensive and transparent fashion, as recommended by guidelines for cost-effectiveness analysis. Sensitivity analyses are also vital to enhance the credibility. The unfamiliarity with decision models, as compared with, for example, regression analyses, may also contribute to the perceived lack of credibility.

Credibility concerns can also emerge when analysts using different decision models report diverging results for the same research question.[45] The diverging results are explained by differences in data sources and assumptions. A vast amount of evidence from clinical and laboratory research is typically available pertaining to a medical decision and multiple assumptions may be reasonable. Small changes in these assumptions may have a large impact on the estimated cost-effectiveness of an intervention. Often, the diverging results simply reflect the uncertainty and disagreement about which data sources and assump-

tions are most appropriate. However, the credibility of some cost-effectiveness analyses sponsored by the industry was recently challenged because they were more likely to report a favourable cost-effectiveness for the new intervention.[60]

Generalizability of the ICER

The results of a cost-effectiveness analysis may not be generalizable to the local setting of a surgeon. The analysis may be of limited use if the perspective (e.g., society or hospital) or the setting (e.g., UK or US cost data) is different. Sometimes this limitation can be overcome by presenting the results for multiple perspectives or by providing details on relevant inputs to allow readers to adjust the calculations based on their own settings. Another concern is that the results from a cost-effectiveness analysis may not apply to a specific patient. The cost-effectiveness of a surgical intervention depends on characteristics such as age, severity of disease, and co-morbidities of the target population. A cost-effectiveness analysis typically reports the ICER for a so-called "base-case" analysis, for example, a 60-year-old white male with no co-morbidities. The reported cost-effectiveness may be different for an individual patient undergoing the same operation in your practice. Generalizability can be improved by reporting separate ICERs for all relevant subgroups. For example, in the study of lung-volume reduction surgery, surgery was only cost-effective for some subgroups of patients but not others.[37]

Ethical Implications

The theoretical foundations of the cost-effectiveness analysis framework often provoke debates on equity, fairness and ethics. For example, cost-effectiveness analysis may favour interventions that benefit younger people over interventions that benefit the elderly. A life-saving treatment (e.g., an appendectomy) will be deemed more cost-effective for a patient with a long life expectancy since more QALYs can be gained compared with saving an elderly patient. Cost-effectiveness analysis may also favour interventions for diseases for otherwise healthy patients: More QALYs can be gained compared with saving a patient with multiple co-morbidities. Furthermore, although the QALY measure captures health improvements in both quantity and quality of life, it does not distinguish between large gains for a small number of individuals versus small gains for a large number of individuals. The distribution of costs and health burdens across the population is not considered, which apparently disregards any societal preference for equity.

Cost-effectiveness analysis, therefore, aims to inform decisions; a program's cost-effectiveness should not be the only determinant of its adoption decision. The National Institute for Health and Clinical Excellence (NICE) in the United Kingdom considers factors other than cost-effectiveness for adoption.[61] For example, an intervention is more likely to be adopted in spite of a high ICER given a high burden of disease, such as multiple sclerosis.

Variation in quality of surgical care

Variation in outcome across surgeons has been well demonstrated.[62] This variation is relevant for both costs and health effects. For example, the cost-effectiveness of parathyroidectomy depends on the rate of recurrent laryngeal nerve damage, which varies between surgeons. In addition, most new surgical interventions – for example, laparoscopic hemicolectomy - demonstrate a learning curve resulting in a period of suboptimal outcomes, possibly both in higher costs and in inferior clinical outcomes. Ideally, randomized controlled trials should start only after reaching the plateau of the learning curve to obtain the appropriate ICER for these interventions. On the other hand, one could argue that the higher costs and complication rates associated with the surgical training are real and should be considered in adoption decisions. The cost-effectiveness also depends on variation in the details of the surgical procedure. A precise description of a surgical intervention is important to replicate the study results in daily patient care.

Gradual advances in surgery

The gradual improvement of surgical technique and technology causes another challenge more specific to surgery. Advances in surgery often occur from a series of small modifications to existing techniques. The effect of small changes may not become apparent in randomized controlled trial settings. As a result, by the time long-term results of a new technology are available, the current surgical intervention may differ from the evaluated intervention. For example, the cost-effectiveness of endovascular repair of abdominal aneurysms depends greatly on the long-term rate of re-interventions. When these long-term results become available, the technology of stents may have improved, requiring fewer re-interventions. Consequently, the actual cost-effectiveness changes over time and remains uncertain.

DISCUSSION

New interventions in surgery typically provide small additional health benefits and are more expensive in comparison to current care. If a health care system cannot adopt all beneficial interventions, priorities have to be set by considering value for money. In this setting, surgeons are increasingly likely to encounter policy makers questioning the cost-effectiveness of their interventions. Surgeons must understand cost-effectiveness analysis to make guidelines for surgical diseases and interventions.

Several challenges exist for conducting and applying cost-effectiveness analysis in surgery. The challenge regarding generalizability, and to some extent credibility can be dealt with

by adherence to published methodologic guidelines. Other challenges do not represent a problem of cost-effectiveness analysis itself, but reflect the difficulty of making decisions under uncertainty. The ethical challenge emphasizes that cost-effectiveness analysis should not be the only factor in determining adoption of health care interventions. Despite these challenges, priorities have to be set.

The National Institute for Health and Clinical Excellence (NICE) makes recommendations on the adoption of health care interventions in the United Kingdom. In doing so, it is required to explicitly consider cost-effectiveness in addition to effectiveness.[8] The cost per QALY figure is used to inform, but not determine, their decisions. Medicare in the United States – providing coverage to the elderly and disabled – does not explicitly consider cost-effectiveness. Its reimbursement is limited to health care interventions that are "reasonable and necessary".[63] Efforts to explicitly consider cost-effectiveness in the United States have not been successful owing to methodologic challenges and the resistance from the pharmaceutical and medical-device industry.[64] Moreover, political risks are perceived because of a pervasive discomfort with medical decisions being influenced by anyone else than the patient and his or her physician.[65] Considering cost-effectiveness, however, is inevitable to achieve the greatest health care benefit from any given level of Medicare spending. More open consideration of cost-effectiveness may happen in the United States after political change or increasing pressure on health care resources.[66]

3

Uncertainty and patient heterogeneity in medical decision models

Bas Groot Koerkamp
Milton C. Weinstein
Theo Stijnen
Majanka H. Heijenbrok-Kal
M.G. Myriam Hunink

Med Decis Making. 2009: In press.

Honorary Mention for excellent scientific research paper. Master of Science in Clinical Epidemiology, Netherlands Institute for Health Sciences, Rotterdam, The Netherlands, 2003.

ABSTRACT

Parameter uncertainty, patient heterogeneity, and stochastic uncertainty of outcomes are increasingly important concepts in medical decision models. The purpose of this paper is to demonstrate the various methods to analyze uncertainty and patient heterogeneity in a decision model. We distinguish various purposes of medical decision modeling, serving various stakeholders. Differences and analogies between the analyses are pointed out, as well as practical issues. The analyses are demonstrated with an example comparing imaging tests for patients with chest pain. For complicated analyses step-by-step algorithms are provided. The focus is on Monte Carlo simulation and value of information analysis. Increasing model complexity is a major challenge for probabilistic sensitivity analysis and value of information analysis. We discuss nested analyses that are required in patient-level models, and in nonlinear models for analyses of partial value of information analysis.

INTRODUCTION

Uncertainty and patient heterogeneity are receiving increasing attention in medical decision modeling. The National Institute of Health and Clinical Excellence (NICE) in the UK has advocated the use of probabilistic sensitivity analysis to assess parameter uncertainty.[77] Value of information analysis has been introduced to estimate the expected benefit of future research, reducing parameter uncertainty.[14] At the same time, decision models have become increasingly complex, often requiring patient-level simulation which introduces stochastic uncertainty. Uncertainty about the model structure further complicates decision modeling. [78, 79] Finally, patient heterogeneity is relevant, for example, to identify subgroups for which a new intervention is cost-effective.

The objective of this paper is to demonstrate the various methods to analyze uncertainty and patient heterogeneity in a decision model. We illustrate these methods with an example comparing imaging tests for patients with chest pain. Tutorials on uncertainty or patient heterogeneity in decision modeling have typically focused on a single methodology: for example, probabilistic sensitivity analysis [16, 17, 80], or the expected value of perfect information[81, 82]. A broader view avoids confusion about similar methods, such as first-order and second-order Monte Carlo simulation. Moreover, appreciating value of information analysis, for example, is facilitated when its relation with other methods addressing parameter uncertainty is understood. Our focus is on Monte Carlo simulation and value of information analysis. The scope of this paper also includes various nested simulations to accommodate multiple levels of uncertainty and patient heterogeneity.

We will distinguish various purposes of medical decision modeling, serving various stakeholders. Before the advent of cost-effectiveness analysis, the patient and his doctor were the main stakeholders of medical decision modeling. Initially decision models were developed to improve the care of specific patients. Nowadays, many models in health care consider costs in addition to health effects and policy makers have become stakeholders too. NICE uses the cost-effectiveness outcomes of typically a base case cohort analysis to inform reimbursement decisions. We will demonstrate how analyses, other than the base case cohort analysis, using the same decision model, can serve additional purposes or stake holders.

In the next section we give a brief description of the decision problem, including a deterministic analysis, used as the primary example for the methods presented in this paper. In the sections that follow, we consider stochastic uncertainty, parameter uncertainty, model uncertainty, and patient heterogeneity. For each analysis we provide step-by-step instructions. We provide verbal explanations of the analyses to avoid the hurdle imposed by mathematical notation. In appendix 2 we present the relevant equations. We assume familiarity with Markov cohort analysis and probability distributions.[38, 83]

1. REAL LIFE EXAMPLE

1.1 Decision problem: diagnostic strategies for patients with chest pain

Conventional catheter coronary angiography (CA) is considered the reference standard test that can distinguish patients with coronary heart disease (CHD) from patients without CHD. Unfortunately, CA has considerable drawbacks: it has a risk of mortality and morbidity, and it is expensive. Multidetector computed tomographic angiography (CTA) is less expensive and has a minimal risk. Its test characteristics (sensitivity and specificity), however, are imperfect: CTA misclassifies both patients with CHD and those without CHD. The initial risk and cost of CA versus the harm and cost of misclassifying patients with chest pain is the main trade-off when choosing between these imaging tests.

Each year 400,000 patients in the USA newly present with chest pain that may be caused by coronary heart disease (CHD).[84, 85] It is important to identify patients with CHD, since they can benefit from a coronary arterial bypass graft (CABG) or a percutaneous coronary intervention (PCI). The current decision problem is to find the optimal imaging test to diagnose CHD in patients with chest pain.

The available evidence regarding the costs and effects of imaging tests for diagnosing CHD was synthesized into a Markov model from the health care system perspective. The model extrapolated the evidence on costs and effects over the entire remaining lifetime of patients. Although earlier versions of the model compared various imaging strategies[77, 86], for illustrative purposes we will now only compare CA with CTA. For comparison we will consider a third strategy, in which patients with chest pain receive medical therapy without an imaging test.

1.2 Decision model

In the model we assume that if a CTA test result was positive or uninterpretable, a CA followed. Patients with a positive CA receive a PCI for one or two vessel disease and a CABG for three-vessel disease and left main disease. Both treatments have an associated disutility and a short-term risk of mortality and myocardial infarction. The beneficial effects of treatment are three-fold: reduction in long-term mortality, reduction in long-term risk of myocardial infarction, and reduction of chest pain severity. Age- and gender-specific life tables were used to model the subsequent lifetime outcomes using Markov models.[87] Three chest pain states were distinguished in the model: no, mild, and severe chest pain. Moreover, each year patients may suffer a myocardial infarction or undergo a CABG or PCI, depending on the extent of the coronary heart disease and treatment history. We modeled the cost of tests and treatments, as well as the annual cost – depending on chest pain severity

and left ventricular ejection fraction – for patients with CHD. The model outcomes of each imaging test were quality-adjusted life expectancy (QALE) and expected lifetime costs. We applied a half-cycle correction for all analyses.[38] Tables 1 to 3 in appendix 3 present all model parameters with their 95% uncertainty interval and sources.

1.3 Deterministic analysis

We performed a Markov cohort analysis for 55-year old men (representing men aged 50 to 59) with atypical chest pain. Table 1 presents the results of this deterministic analysis, including incremental cost-effectiveness ratios (ICERs). Policy makers can conclude that for a willingness-to-pay (WTP) of less than $31,000/QALY both imaging tests are not cost-effective. CA is cost-effective if the WTP is $85,000/QALY or more. In between these threshold values, CTA is cost-effective.

Table 1 Results of the deterministic analysis of the example model. All costs are in US$, all effects in quality-adjusted life years.

Strategy	Cost	Incr Cost	Effect	Incr Effect	ICER
No imaging test	26953		12.96567		
CTA	34997	8044	13.22813	0.262	30649
CA	35153	156	13.22997	0.002	84836

Incr, incremental; ICER, Incremental Cost-Effectiveness Ratio; CTA, multidetector computed tomographic angiography. CA, conventional catheter coronary angiography.

2. STOCHASTIC UNCERTAINTY

Stochastic uncertainty – also known as first-order uncertainty or individual patient variability – represents the uncertainty in patient-level outcomes.[88] This uncertainty is entirely due to chance.[89] In decision models each chance node contributes to this uncertainty. For example, at a specific chance node a patient has an estimated probability of 3.2% that he will die from surgery. Stochastic uncertainty reflects the uncertainty related to the actual outcome - a patient may or may not fall within the 3.2% of patients that die – which should be distinguished from uncertainty around the 3.2% due to the limited sample of patients in which the value was estimated (i.e., parameter uncertainty) and from uncertainty about whether the 3.2% applies to this particular type of patient (i.e., patient heterogeneity).

We evaluated stochastic uncertainty using first-order Monte Carlo analysis (a.k.a. microsimulation).[38, 83] This analysis should not be confused with second-order Monte Carlo analysis dealing with parameter uncertainty, which we discuss in section 3.2. A first-order Monte Carlo analysis simulates subjects one-by-one. Probabilities at chance nodes and a random number generator result in a subject's path along the chance nodes. This path is called a

random walk or a "trial". Counters (a.k.a. tracker variables) can record the accumulated (quality-adjusted) lifetime and costs, as well as events along the subject's path. When the subject dies, the simulation restarts with a new subject. We performed 10,000 random walks in the example model. Using the results of this analysis we can calculate, for example, that the probability that a 55-year-old man with atypical chest pain lives at least another 10 years is 82%. Patients are typically interested in such outcomes in addition to expected outcomes.

While the analysis of stochastic uncertainty can serve patients as stake holders it is not the most common rationale for its use [90]. In so-called patient-level models, a first-order Monte Carlo simulation (i.e., microsimulation) is necessary to estimate expected outcomes.[20] First-order Monte Carlo simulation allows modeling of the influence of patient history on subsequent events. For example, when a subject suffers a myocardial infarction, his future mortality rate will increase. Although Markov cohort models can include patient history in the definition of health states, this requires expanding the number of states which can become unwieldy.[38]

If a stakeholder is uninterested in individual patient outcomes, then the analysis of stochastic uncertainty in patient-level models only contributes noise to the expected outcomes. Instead of a single Markov cohort analysis, many trials are required to obtain a precise estimate of the expected outcomes. Evaluating more trials of the model will improve precision but also requires more computing time. The precision of these expected outcomes could be assessed by the standard error of the mean (*sem*). The *sem* is calculated as the standard deviation of the trial outcomes divided by the square root of the number of trials. Gaussian process modeling has been suggested as a time-efficient alternative to first-order Monte Carlo simulation [91], but such methodology is still in the developmental stage. Griffin et al. discussed how patient-level modeling can be avoided in certain circumstances.[90]

3. PARAMETER UNCERTAINTY

3.1 Deterministic sensitivity analysis

Deterministic sensitivity analysis evaluates the influence of uncertainty in one or more parameters on the expected outcomes. In univariable (one-way) sensitivity analysis, the outcome of each strategy is calculated over a justifiable range of one parameter, for example, the 95% uncertainty interval based on the results of a study. A tornado diagram presents the results of many one-way sensitivity analyses. In a tornado diagram a horizontal bar represents the range of expected outcomes at the decision node (across all strategies) given

the range of each selected estimated parameter (e.g., the 95% uncertainty interval). The tornado shape arises by ordering the bars by width, starting with the widest at the top. A mark can indicate where the optimal strategy changes across the range of a parameter.

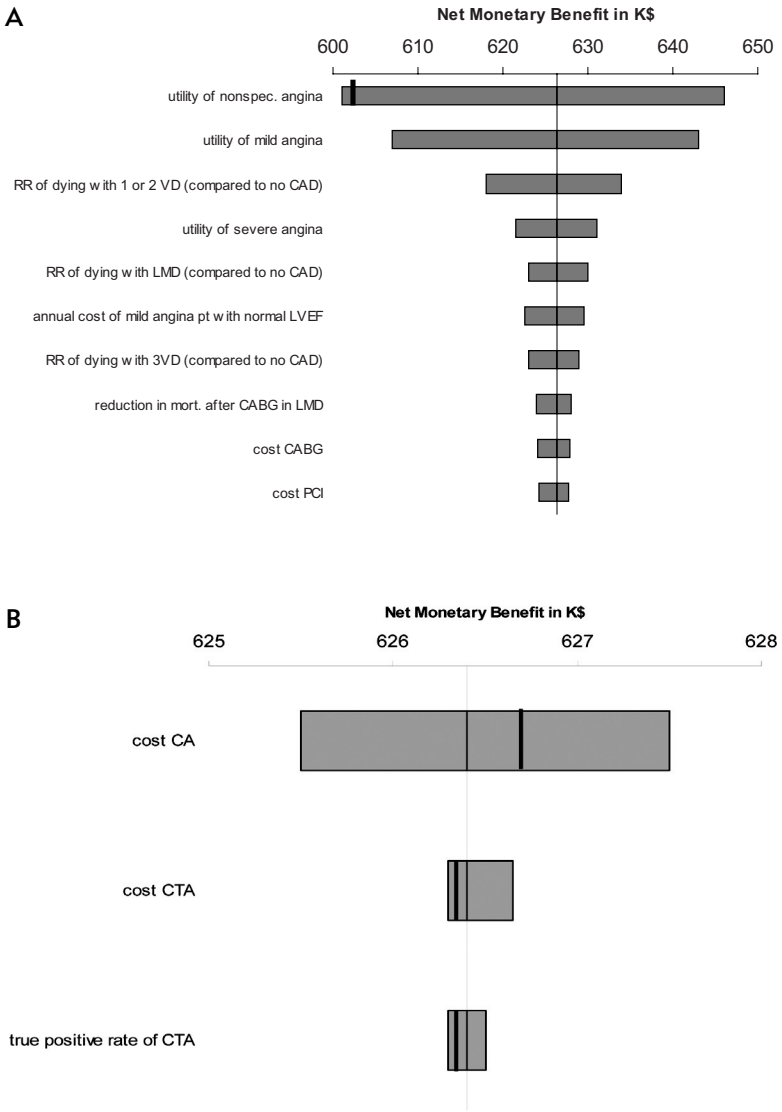

Figure 1 Tornado diagram for the ten most influential estimated parameters (A). "Influential" means that they have the highest impact on the expected outcome at the decision node. Of the 80 assessed parameters, 3 additional parameters changed the optimal strategy within their range (B).
RR, relative risk; VD, vessel disease; CAD, coronary arterial disease; LMD, left main disease; LVEF, left ventricular ejection fraction; CABG, coronary arterial bypass graft; PCI, percutaneous coronary intervention; WTP, willingness-to-pay; CTA, computed tomographic angiography."

Table 2 Results of the probabilistic sensitivity analysis (columns 1-5) and expected value of information analysis (columns 1-8).

Sample	NMB (CTA)	NMB (No test)	NMB (CA)	sample best	sample max	baseline max	opportunity loss
1	$598,104	$588,775	$597,540	CTA	$598,104	$598,104	$0
2	$608,478	$606,464	$608,743	CA	$608,743	$608,478	$265
3	$609,489	$606,648	$602,758	CTA	$609,489	$609,489	$0
4	$624,283	$624,815	$624,614	no test	$624,815	$624,283	$532
5	$636,035	$626,805	$633,813	CTA	$636,035	$636,035	$0
6	$635,034	$625,744	$636,382	CA	$636,382	$635,034	$1,348
7	$638,400	$630,043	$637,928	CTA	$638,400	$638,400	$0
8	$637,761	$628,196	$632,105	CTA	$637,761	$637,761	$0
9	$655,188	$640,486	$655,815	CA	$655,815	$655,188	$627
10	$622,801	$617,226	$623,182	CA	$623,182	$622,801	$381
Average	**$626,557**	**$619,520**	**$625,288**	**50% CTA**	**$626,873**	**$626,557**	**Total EVPI = $315**

WTP, willingness-to-pay = $50,000/QALY; CTA, multidetector computed tomographic angiography; No test, no imaging test; CA, conventional catheter coronary angiography; sample best, the strategy with the highest net benefit of the sample; baseline max, the outcome of the strategy with the overall optimal outcome; sample max, the outcome of the strategy with the highest net benefit of the sample.

We built a tornado diagram of all estimated parameters, with the exception of parameters that are correlated, such as Dirichlet distributions.[92] We used net monetary benefit (NMB) as the expected outcome, combining the outcomes cost and effect: NMB = effect * WTP – cost.5 Figure 1A presents a tornado diagram for the ten estimated parameters with the largest impact on the expected outcome (i.e., NMB) at the decision node, given a WTP of $50,000/QALY. The black mark in the top bar demonstrates that only this parameter – the utility of non-specific chest pain – causes a change in optimal strategy at the lower end of its range. Three other parameters changed the optimal strategy within their range. These are presented in the tornado diagram of Figure 1B.

3.2 Probabilistic sensitivity analysis

PSA or second-order Monte Carlo analysis, evaluates the joint effect of uncertainty about all estimated parameter values in the model.[16, 17] The uncertainty about the parameter values is represented by probability distributions and propagated in the model resulting in a probability distribution of the expected outcome for each strategy. The probability distributions of the parameters are often obtained using traditional parametric statistical methods. Briggs et al. give guidance on selecting the appropriate distributions.[80] Alternatively, bootstrapping has the advantage that the analyst does not have to make parametrical assumptions about the parameter distribution.[93, 94] However, standard bootstrapping methods may lead to misleading inferences, for example, when cost data exhibit highly skewed distributions.[95] Correlations between parameters can be modeled by drawing values from

joint distributions of the correlated parameters. Bayesian Markov chain Monte Carlo methods are increasingly used to model correlations, but require dedicated software such as WinBUGS. [7, 27-30]

We performed PSA by randomly drawing a value for each parameter from its probability distribution. This set of values is commonly referred to as a "sample". The model was then recalculated for this sample using a Markov cohort analysis. We repeated this for 10,000 samples. Table 2 (column 1-4) presents ten samples of the PSA. In column 5, for each sample we identified the optimal alternative as the strategy with the highest net benefit. The probability that CTA is the "true" optimal strategy is the percentage of samples in which it has the highest net benefit. Table 2 shows that CTA is optimal in 50% of the ten samples. Out of 10,000 samples, CTA was optimal in 54%.

Acceptability curves (Figure 2) [26, 96-99] demonstrate that the probability that CTA is cost-effective is less than 55% for any reasonable value of the WTP. Note also that the probability that CA is cost-effective is smaller than the probability that CTA is cost-effective for any WTP, even when the expected net benefit of CA exceeds the *expected* net benefit of CTA (i.e., for

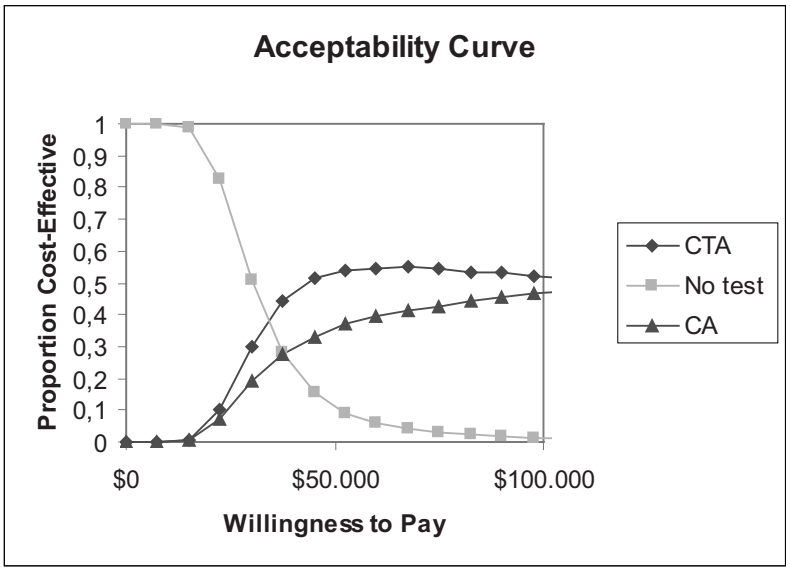

Figure 2 Cost-effectiveness acceptability curves. The probability is presented that each strategy is the "true" optimal strategy across a range of values for the willingness-to-pay. For each value of the willingness-to-pay the total probability adds to 1.
CTA, computed tomographic angiography; No test, no imaging test; CA, coronary angiography.

BOX 1: SECOND-ORDER MONTE CARLO SIMULATION (PSA) IN PATIENT-LEVEL MODEL.

1. Draw a sample from all parameter distributions.
2. Plug in the sampled values and perform a first-order Monte Carlo simulation with N trials.
3. Calculate the expected outcome of the N trials for each strategy.
4. Repeat steps 1 to 3 M times.
5. The M expected outcomes obtained at step 3 represent the distribution of the expected outcome characterizing parameter uncertainty.

WTP>$85,000/QALY, see Table 1). The cost-effectiveness acceptability frontier (CEAF) has been introduced within acceptability curves to indicate the intervention with the highest *expected* net benefit.[100]

Next, we estimate the 95% uncertainty interval for the incremental net benefit of CTA versus CA from the results of the PSA. For each sample we calculate the difference in net benefit between the two strategies. The 95% uncertainty interval for the incremental net benefit of CTA versus CA is –$430 to +$516, with an expected difference of $64. The equivalent values in quality-adjusted life days are -3 to +4 days, with an expected difference of half a day. These values are small, but typical for incremental benefits of diagnostic tests and screening programs.

BOX 2: TOTAL EVPI

1. Perform a second-order Monte Carlo simulation, sampling all estimated parameters.
2. For each sample of the PSA, calculate the opportunity loss as the difference between the maximum expected benefit of the sample and the sample's expected benefit of the baseline optimal strategy.
3. The total EVPI is the average opportunity loss.

The use of PSA also has a more technical justification. A deterministic analysis is valid only if the expected outcome of the model (f) equals the model outcome when evaluated in the expected values of the parameters (x): $E[f(x)] = f(E[x])$. Correlated parameters in which the model is linear do not require sampling from their probability distribution to obtain unbiased expected outcomes. A linear function of parameters $x_1, x_2, \ldots x_n$ is defined as a function of the form $a_1*x_1 + a_2*x_2 + \ldots + a_n*x_n$ for certain constants $a_1, a_2, \ldots a_n$. Uncorrelated parameters in which the model is multilinear also do not require sampling from their probability distribution to obtain unbiased expected outcomes. A function is multilinear if it is a linear function of each parameter when the other parameters are given fixed values.

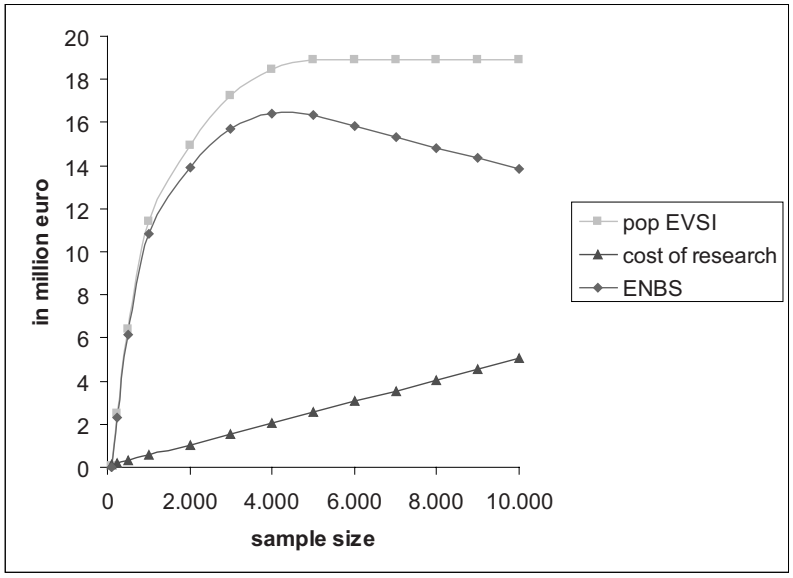

Figure 3 The population expected value of sample information (population EVSI), the research cost, and the expected net benefit of sampling information (ENBS) as functions of the sample size of a future observational study gathering data on the utilities of the angina states. The maximum ENBS corresponds with an optimal sample size of about 4,000 patients for each strategy.

BOX 3: PARTIAL EVPI – ONE-LEVEL ALGORITHM

1. Perform a second-order Monte Carlo simulation, but sample only the parameters of interest. The parameters not-of-interest remain fixed at their mean values.
2. For each sample, calculate the opportunity loss as the difference between the maximum expected benefit of the sample and the sample's expected benefit of the baseline optimal strategy.
3. The partial EVPI is the average opportunity loss.

Nonlinearity is the rule rather than exception: in Markov models the transition probabilities are multiplied by themselves repeatedly in the calculation of expected outcomes. The same issue arises in section 3.4 when we estimate the partial expected value of perfect information. In appendix 1 we present a numerical example. Although the expected outcomes of the Markov cohort analysis in section 1 are biased in theory, we could not detect this bias: a PSA of 100,000 samples found expected outcomes similar to the deterministic analyses.

A PSA becomes more complicated in patient-level models (see section 2) that do not allow for Markov cohort analyses. Instead, the recalculation for each sample requires an entire first-order Monte Carlo simulation (e.g., 1,000 trials). The PSA represents an outer loop of

BOX 4: PARTIAL EVPI – TWO-LEVEL ALGORITHM

1. Draw a sample from the distributions of the parameters of interest.
2. Draw a sample from the distributions of those parameters not-of-interest,
   - in which the model is nonlinear,
   - in which the model is multilinear and that are correlated with other parameters not of interest,
   - that are correlated with parameters of interest and for which an analytic expression for the conditional mean value is not available.

   All other parameters not-of-interest may be fixed at their mean values.
3. Recalculate the model performing a cohort analysis for the values of step 1 and 2.
4. Repeat step 2 and 3 K times. – inner loop
   Calculate the expected benefit of each strategy.
5. Calculate the opportunity loss as the difference between the maximum expected benefit in step 4 and the expected benefit in step 4 of the baseline optimal strategy.
6. Repeat step 1 to 5 J times. – outer loop
7. The partial EVPI is the average opportunity loss.

M samples; the first-order Monte Carlo simulation represents an inner loop of N trials. This is sometimes called the "M by N problem" and it is a major obstacle to performing PSA because of the time-consuming calculations required.[19, 20, 90, 101] Confusion often exists about performing a PSA of a patient-level model. Box 1 presents a step-by-step algorithm.

3.3 Total EVPI

Information obtained in future quantitative research – for example, a randomized controlled trial – can reduce parameter uncertainty. A decrease in parameter uncertainty may avoid reimbursement of suboptimal interventions, and consequently is expected to benefit patients and/or reduce costs. Value of information analyses explicitly estimate the expected benefit of collecting information in future research. The value of information is not the

BOX 5: ALGORITHM PARTIAL EVSI – PARAMETERS OF INTEREST WITH A BETA DISTRIBUTION

1. Draw a sample θ_j^I from the prior distributions Beta(a,b) of the parameters of interest.
2. Select a sample size n for the proposed data collection.
3. Draw a sample r_j from the binomial distribution (θ_j^I, n).
4. Obtain the posterior distribution of each parameter of interest: Beta($a+ r_j$,$b+n - r_j$).
5. Recalculate the model using the posterior distributions of the parameters of interest.
   All nonlinear and correlated multilinear parameters should be sampled in a nested second-order Monte Carlo simulation. – inner loop
6. Calculate the opportunity loss.
7. Repeat steps 1-6 J times. – outer loop
8. The partial EVSI is the average opportunity loss.

Table 3 Individual model parameters ranked by partial expected value of perfect information (pEVPI) as a measure for the importance of uncertainty. The Tornado rank in the third column demonstrates that it is not a valid proxy for the importance of uncertainty. All parameters not in this table had a zero pEVPI.

parameter	p**EPVI**	Tornado rank
utility of nonspecific angina	42	1
cost CA	32	15
cost CTA	25	34
TPR CTA	14	39
utility of mild angina	0.001	2
annual costs of mild angina pat with normal LVEF	0	6
cost CABG	0	9
cost PCI	0	10
reduction in mortality after CABG in LMD	0	8
RR of dying with 1 or 2 VD (compared to no CAD)	0	3
RR of dying with 3 VD (compared to no CAD)	0	7
RR of dying with LMD (compared to no CAD)	0	5
utility of severe angina	0	4

pEVPI, partial expected value of perfect information; CA, coronary angiography; CTA, computed tomographic angiography; TPR, true positive rate; LVEF, left ventricular ejection fraction; CABG, coronary arterial bypass graft; PCI, percutaneous coronary intervention; LMD, left main disease; RR, relative risk; VD, vessel disease; CAD, coronary artery disease.

actual value of future research – which we will only learn after performing future research – but the "expected" value of future research, *ex ante*. It is expressed in the same units as the model outcome, typically net monetary benefit. Value of information analysis was introduced by Grundy[29] in the late fifties and developed by Raiffa and Schlaifer.[102, 103] Since the late eighties it has received increasing attention in the risk analysis literature and more recently in health care.[14, 28, 81, 104]

The total EVPI is the expected benefit per patient of a hypothetical study with an infinite sample size that would eliminate all parameter uncertainty. It is estimated by the average opportunity loss of the samples of the PSA (Table 2).[105] The opportunity loss of a sample is defined as the difference between the maximum expected benefit of that sample (sample max) and the sample's expected benefit of the baseline optimal strategy (baseline max).[106] For example, the maximum expected benefit of the second sample in Table 2 is the expected benefit of CA: $608,743. CTA is the baseline optimal strategy; the expected benefit of CTA in sample 2 is $608,478. The opportunity loss of sample 2 is the difference between these values: $608,743 – $608,478 = $265. The final column of Table 2 presents the opportunity loss for each sample. The average opportunity loss of all samples is $315 per patient and is an estimate of the total EVPI per patient. See Box 2 for a step-by-step

algorithm. Using 10,000 samples instead of the 10 samples of Table 2, we obtained a more precise estimate of the total EVPI: $294 per patient. This means that after eliminating uncertainty we can expect an improvement in net monetary benefit of $294 per patient. The probability that the actual value is zero is 54%, identical to the current probability that CTA is the optimal strategy. This result also implies that the current expected harm due to uncertainty is $294 per patient, with a health equivalent of 2 quality-adjusted life days (WTP=$50,000/QALY).

More research to decrease uncertainty is justified if the expected benefit to future patients exceeds the cost of research. The population EVPI represents the expected benefit to all future patients. It is estimated as the product of the total EVPI per patient and the population that is expected to benefit from future research, discounting expected benefit in future years. The population EVPI is a ceiling level for the expected return on investment of research. If research is more expensive than the population EVPI, it is a bad investment: the uncertainty is not important enough to be resolved. The annual population to benefit is usually ambiguous, because it is not obvious whether we should consider the local setting, one country, or all patients worldwide. Moreover, the period that patients will benefit from the proposed data collection is typically uncertain because of future improvements, novel interventions, or new insights. These ambiguities, however, are not drawbacks of value of information analysis in itself, but inherent to the problem of allocating resources wisely.

We estimated the annual population to benefit (males aged 50-59 years with atypical chest pain) for the US at 44,000 patients (i.e., 11% of 400,000).[84, 85] Assuming a period of 5 years and a discount rate of 3%, we found a total EVPI for the population to benefit of $61 million.

3.4 Partial EVPI

Instead of estimating the total EVPI of all parameters, we can estimate the partial EVPI of one or more parameters (i.e., the parameters of interest). First, we demonstrate the so-called one-level algorithm.[82, 105] Analogous to the estimation of total EVPI, a second-order Monte Carlo simulation is performed, but only the parameters of interest are sampled from their distributions. The parameters *not*-of-interest (i.e., all other parameters) remain fixed at their mean values. For each sample the model is recalculated and the opportunity loss is estimated analogous to the estimation of total EVPI. The average opportunity loss of many samples is an estimate of the partial EVPI of the parameters of interest. See Box 3 for a step-by-step algorithm of the one-level algorithm.

Keeping the parameters *not*-of-interest fixed at their mean value may result in the same bias that we discussed in section 3.2 (technical justification of PSA). In a two-level algo-

rithm, a second-order Monte Carlo simulation is performed for each sample of the parameters of interest, to avoid this bias. All parameters of interest are sampled in the outer loop second-order Monte Carlo simulation; selected parameters *not*-of-interest are sampled in the inner loop second-order Monte Carlo simulation. The criteria to decide if a parameter *not*-of-interest requires sampling in the inner loop are the same as the criteria of section 3.2 to avoid a bias when obtaining expected outcomes of a model. If correlations exist between parameters of interest and a parameter *not*-of-interest, the latter should also be sampled in the inner loop when an analytic expression for the conditional mean value is not available. See Box 4 for a step-by-step algorithm of the two-level algorithm.

The outer loop – sampling the parameters of interest – determines the precision of the estimate of the partial EVPI. More samples in the inner loop – recalculating the model by sampling the parameters *not*-of-interest – yields less biased results. Brennan et al. recommend a 1 to 5 ratio of samples of the outer versus the inner loop.[82] The correct order of magnitude was found with a minimum of 100 samples in the outer loop.

In a tornado diagram we identified four parameters that have the capability to change the optimal strategy somewhere along their range of likely values. We first estimated the partial EVPI of these parameters because they should exceed zero. We divided the remaining parameters into several groups. Most groups have a zero partial EVPI, and therefore all constituent parameters have a zero partial EVPI. The tornado diagram identified four out of five parameters with a nonzero partial EVPI. The ranking in the tornado diagram, however, is not a good predictor of the importance of uncertainty as reflected by the partial EVPI. Table 3 presents the ranked results for partial EVPIs with their rankings in the tornado diagram. Note that the sum of the partial EVPIs of individual parameters generally does not equal the total EVPI.[106, 107]

To assess whether more research is justified we should estimate the partial EVPI of a set of parameters that could be measured in a specific study. Parameters with a partial EVPI of zero may seem useless to consider in future research. A combination of such parameters, however, may jointly have a nonzero partial EVPI. We analyzed the partial EVPI of eight study designs. Six of these were observational studies, measuring, respectively: test characteristics of CTA, complications of coronary angiography, utilities of chest pain states, costs of interventions, complications of PCI, and complications of CABG. The other two study designs were randomized trials: medical treatment versus PCI in one- or two-vessel disease and medical treatment versus CABG in three-vessel disease or left main disease. For the observational study measuring utilities we found the highest partial EVPI: $91 per patient, with a population EVPI of $19 million. The cost study had a partial EVPI of $48 per patient, and the diagnostic study for test characteristics of CTA had a partial EVPI of $31 per pa-

tient. All other study designs, including the two RCTs had a (near) zero partial EVPI.
The (partial) expected value of sample information (EVSI) estimates the expected value of obtaining information for finite sample sizes.[47-50] With increasing sample size, the partial EVSI will reach a ceiling: the partial EVPI, representing an infinite sample size. At the same time, the cost of research increases with increasing sample size. The expected net benefit of sampling (ENBS) is defined as the difference between the EVSI and the cost of research. The maximum ENBS is associated with the optimal sample size of a proposed study design. We refer to Ades et al. for an extensive coverage of EVSI.[32]

4. MODEL STRUCTURE UNCERTAINTY

The structure of a model typically depends on assumptions about the natural course of a disease and how medical interventions may influence this course. The amount and form of available data will also determine the model structure. Moreover, analysts can opt for more or less model complexity: for example, a simple decision tree versus a Markov model. This may depend on the research question, available time, and the required validity and precision of the results. Consequently, different teams of experts may come up with different models to represent the same decision problem. Campbell et al. demonstrated how this can result in large variation in the outcomes.[45] Evaluating only the plausible single best model may result in an underestimation of uncertainty.[78] Ideally, analysts should build a model for every imaginable set of assumptions regarding a decision. In real life most analysts settle for using their single best model. Claxton et al. pointed out that more research is needed regarding the trade-off between the realism of the model and the available time.[19] If alternative modeling assumptions may affect the decision, more evidence to justify the use of one assumption instead of the others should be found. If this is impossible, the results of each model can be presented, and the policy maker can decide. Sometimes model structure uncertainty can be dealt with using parameterization.[108] Finally, some sources of uncertainty are not covered by parameter or structural uncertainty: for example, uncertainty about the appropriate evidence sources, and uncertainty about the selection of interventions.

We performed a structural sensitivity analysis by assuming an additive mortality function. The CTA strategy remained the optimal strategy at a WTP of $50,000/QALY, with an ICER of $28,000 compared to no imaging test. The ICER differed about 9% from the ICER using the multiplicative model.

5. PATIENT HETEROGENEITY

Patient heterogeneity is usually analyzed to identify differences in the optimal strategy for subgroups of patients.[109] Clinical guidelines and reimbursement decisions reflect these differences between subgroups. Moreover, a strategy could have a high incremental cost-effectiveness ratio for the total population, but a low ratio for a certain subgroup; the mean value may obscure the cost-effectiveness of a strategy for a subgroup. Baseline patient characteristics can influence each estimated parameter in the model: for example, we can distinguish heterogeneity in treatment effects, heterogeneity in costs, and heterogeneity in utilities. In practice it is often difficult to determine whether a difference between subgroups is genuine or simply reflects noise in the data. Criteria are being developed to decide when it is justified to model heterogeneity in a parameter.[110]

Sensitivity analysis may evaluate the optimal strategy for various subgroups and involves repeated analysis of a model for, for example, different age groups. If patient heterogeneity is modeled using a continuous or ordinal variable, sensitivity analysis can calculate the expected outcomes over a range of values for the patient characteristic, analogous to deterministic sensitivity analysis of parameters representing parameter uncertainty.

Differences in setting - as opposed to differences in patient characteristics – can also cause heterogeneity in parameter values. This type of heterogeneity arises, for example, when a model developed based on data from one country is used to make inferences for another country. For example, the cost of an appendectomy or the sensitivity of ultrasound for appendicitis can differ across countries or hospitals. A policy maker should be able to assess what the model implies for a situation that may be identical to the base-case assumptions of the model except for a limited set of parameter values.

In our example model, gender, age, and type of chest pain are the most relevant patient characteristics. We assessed 30 subgroups: 5 age groups, both genders, and 3 chest pain groups. The higher the prior probability of CHD the more likely it is that CA is cost-effective. The prior probability of CHD is increased by advanced age, male gender, and typical instead of nonspecific or atypical chest pain. CTA is cost-effective for intermediate-risk patients, no imaging for low-risk patients.

The analysis of patient heterogeneity is also required when uniform decisions are considered for rather heterogeneous populations. For example, policy makers may want to know the overall ICER of a colon cancer screening program for everyone over 50 years. For population-level decisions it is important that the heterogeneity of the target population is reflected in the model. Patient heterogeneity can be represented by distributions for

each patient characteristic or by bootstrapping "real" subjects from a study population. The expected outcome for the heterogeneous population is the average outcome of many randomly drawn patients performing a random walk in the model. Typically this requires a patient-level first-order Monte Carlo simulation. Nijhuis et al. used this approach to model a heterogeneous population using data from a large study population. [111] Parameter uncertainty in such models adds an additional level of complexity.[112]

DISCUSSION

We demonstrated various methods to analyze uncertainty and patient heterogeneity in decision models. The analyses resulted in outcomes serving various purposes and stakeholders. Policy makers (e.g., NICE) determine the optimal strategy by combining these results with (typically) unmodeled considerations such as ethical viewpoints or the transition costs of a new intervention.8 More recently, research-funding agencies can use the results of value of information analyses to guide future research. Patients and doctors also combine model results with their preferences regarding health states and risk attitude, as well as specific patient characteristics that are often not accounted for in the model.

The analysis of uncertainty and patient heterogeneity faces several challenges. Increasing model complexity impedes probabilistic sensitivity analysis and value of information analysis. Nested analyses are required in patient-level models, and for partial value of information analyses in nonlinear models. The required calculation time in decision analytic software using personal computers is prohibitive. Linear cohort models (i.e., no Markov nodes, no patient-level simulation) avoid nested simulations, but could be an unrealistic representation of the decision problem. Accounting for correlations between estimated parameters is another challenge. Correlations are influential on the value of information, but often no data is available to model this. Thirdly, it is not feasible to consider the characteristics and preferences of each individual patient in a decision model. However, doctors rarely see patients that match the base case analysis of a cost-effectiveness study. Instead, doctors care for a variety of patients each with their own unique set of risks and preferences. Finally, model structure uncertainty remains problematic, because more than one structure may be reasonable. We often lack time to build various models, and placing a weight on each model is an arbitrary choice.

Despite these challenges, decisions regarding current implementation and future research have to be made. We hope this paper will stimulate and help analysts and policy makers evaluating uncertainty and patient heterogeneity to inform these decisions.

4

The combined analysis of uncertainty and patient heterogeneity in medical decision models

Bas Groot Koerkamp
Theo Stijnen
Milton C. Weinstein
M.G. Myriam Hunink

Submitted

ABSTRACT

The analysis of both parameter uncertainty and patient heterogeneity in decision models is increasingly recommended to improve decision making. In addition, the complexity of current medical decision models commonly requires simulating individual subjects, which introduces stochastic uncertainty. The combined analysis of uncertainty and heterogeneity often involves complex nested Monte Carlo simulations to obtain the model outcomes of interest.

In this paper we distinguish eight model types, each dealing with a different combination of patient heterogeneity, parameter uncertainty, and stochastic uncertainty. The model outcomes of interest include the expected outcome, the distribution of the expected outcome reflecting lack of perfect knowledge, the distribution of the expected outcome reflecting patient heterogeneity, and the distribution of the individual outcome. The analyses that are required to obtain these model outcomes are expressed in equations, explained in stepwise algorithms, and demonstrated in examples. Patient heterogeneity is represented by frequency distributions and analyzed with Monte Carlo simulation. Parameter uncertainty is represented by probability distributions and analyzed with second-order Monte Carlo simulation (i.e., probabilistic sensitivity analysis). Stochastic uncertainty is analyzed with first-order Monte Carlo simulation (i.e., trials or random walks). This article can be used as a reference for analyzing complex models with more than one type of uncertainty and with patient heterogeneity.

INTRODUCTION

Policy making in health care can benefit from the evaluation of uncertainty and patient heterogeneity in decision models. Organizations and journals therefore increasingly demand a probabilistic analysis of parameter uncertainty.[18] Modeling patient heterogeneity is also recommended, because the optimal intervention may depend on patient characteristics.[110] In addition, analysts are forced to deal with stochastic uncertainty in models that require the simulation of individual subjects.

Nevertheless, many published models consider neither uncertainty nor patient heterogeneity, and hardly any consider all 3 types of uncertainty and heterogeneity simulteneously.[89] When more than one type of uncertainty and/or patient heterogeneity is modeled, the correct algorithm to obtain the model outcomes of interest can be complicated. Nested Monte Carlo simulations, for example, may be necessary. The purpose of this paper is to demonstrate the combined analysis of uncertainty and patient heterogeneity in medical decision models.

In section 1 we discuss the terminology for uncertainty and patient heterogeneity, as well as various decision model outcomes of interest. In sections 2 and 3 we demonstrate how the outcomes of interest are estimated in eight models, each exhibiting a different combination of parameter uncertainty, stochastic uncertainty, and patient heterogeneity. We distinguish models in which the analysis of stochastic uncertainty is not required (section 2 – "macrosimulations") from models in which the analysis of stochastic uncertainty is required (section 3 – "microsimulations"). The analyses are expressed in equations, and explained in stepwise algorithms. In section 4 we demonstrate some of the analyses with an example. We adopt a Bayesian approach, treating model parameters as random variables, and with probability distributions representing uncertainty. The focus is on Monte Carlo methods to perform the analyses. We assume some familiarity with Markov models and Monte Carlo simulation.[38, 83, 113]

1. TERMINOLOGY

Typically a cohort analysis of a deterministic model is performed to obtain the mean value of the outcome, or *expected outcome*. Introducing stochastic uncertainty, parameter uncertainty and/or patient heterogeneity into a model complicates the estimation of the expected outcome. At the same time, it allows the estimation of other model outcomes in addition to the expected outcome:

- stochastic uncertainty: distribution of the individual outcome in a cohort or population, reflecting randomness,
- parameter uncertainty: distribution of the expected outcome, reflecting lack of perfect knowledge, and
- patient heterogeneity: distribution of the expected outcome, reflecting patient heterogeneity.

Stochastic uncertainty

Stochastic uncertainty is the uncertainty about the outcome of an individual due to chance. This uncertainty is also referred to as first-order uncertainty.[88, 114] The analysis of stochastic uncertainty results in a *distribution of the individual outcome* in a cohort of population. This distribution may be of interest, in addition to the expected outcome, if a decision maker is concerned about minimizing the risk of an unfavorable outcome due to risk aversion. Moreover, the individual-level uncertainty might be of interest to patients for purely prognostic reasons.

The analysis of stochastic uncertainty, however, is applied primarily to estimate the expected outcome in so-called microsimulation (i.e., patient-level) models.[20] Chance occurrences within these models determine the value of downstream parameters. For example, after experiencing a myocardial infarction, a patient's annual probability of mortality will increase. Because the model is required to "memorize" such previous events in order to set parameter values, subjects must be simulated one at a time. Adding more Markov states to deterministic models to retain memory of previous events may result in prohibitively large models.[101] Therefore, microsimulation models may be used because of computational necessity, rather than because of an interest in patient-level uncertainty.

A disadvantage of microsimulation is the noise introduced by stochastic uncertainty in the estimate of the expected outcome. This noise can only be resolved by simulating an infinitely large number of subjects. The standard *error* of the mean outcome can be used to ascertain that sufficient simulations were performed. In contrast, the standard deviation of the outcome is a measure of uncertainty at the patient-level.

Parameter uncertainty

Parameter uncertainty arises from lack of perfect knowledge. The "true" value of an input parameter (e.g., a relative risk, a treatment efficacy, or a utility) will remain unknown, unless it was estimated in an infinitely large sample. This uncertainty is also referred to as second-order uncertainty and is typically represented by a probability distribution for each parameter.[88] Parameter uncertainty can be propagated through the model, resulting in a *distribution of the expected outcome reflecting lack of perfect knowledge*.

The probabilistic analysis of parameter uncertainty has several objectives: it quantifies the confidence in the expected outcome, it enables the synthesis of model results with risk attitude, and it is required for value of information analysis – analysis of the value of obtaining evidence that can reduce the uncertainty and guide future decisions.[14, 32, 115] Moreover, in models where the outcome is a nonlinear function of parameters (e.g., probabilities in a Markov model), a probabilistic analysis is required to obtain unbiased estimates of the expected outcome, even if the uncertainty is not of interest in its own right.[115]

Patient heterogeneity

Patient heterogeneity (also known as variability) concerns patient characteristics that can influence the expected outcome of a decision model. Because of patient heterogeneity, the optimal intervention may vary across individuals. Unlike parameter uncertainty, patient heterogeneity cannot be reduced with better data.[116] Heterogeneity can be modeled using discrete subgroups (e.g. age ranges), within which all individuals are assumed identical. The analyst can assign subgroup-specific parameter values, resulting in subgroup-specific expected outcomes. When using discrete subgroups, patient heterogeneity is represented by a (discrete) *distribution of the expected outcome*.

Infinitely many subgroups arise when patient heterogeneity is modeled using continuous covariates.[111] The frequency distribution of a covariate can represent patient heterogeneity. For example, the annual probability of developing a stroke depends on cholesterol level, and a frequency distribution can represent how the cholesterol level varies across individuals. These frequency distributions should be distinguished from the probability distributions representing parameter uncertainty. The importance of the covariate(s) can be illustrated by a (continuous) *distribution of the expected outcome* reflecting its dependence on the covariate(s).

The analysis of patient heterogeneity is needed in order to identify the optimal intervention for different subgroups. For example, annual mammography screening could be recommended for women aged 40 to 79, but not for other ages. However, sometimes a policy maker has to make a uniform recommendation for a heterogeneous population, as in the

choice of contrast media for imaging tests. In that case, the expected outcome averaged over the heterogeneous population is required in order to evaluate the optimal course of action for the population as a whole.

2. MACROSIMULATIONS

In this section we discuss models for which the cohort is the fundamental unit of the analyses. We use the term macrosimulations for these models, as opposed to microsimulations that require simulation of individual subjects (Table 1). In macrosimulation models, Markov cohort analysis can be applied to calculate the expected outcome for a set of parameter values and patient characteristics. The analysis of stochastic uncertainty will be demonstrated later, in the microsimulation section. We first introduce some notation.

Table 1 Model complexity

Model	Parameter uncertainty	Patient heterogeneity
Macrosimulation		
D	-	-
H	-	+
P	+	-
PH	+	+
Microsimulation		
S	-	-
HS	-	+
PS	+	-
PHS	+	+

D: deterministic
H: patient heterogeneity
P: parameter uncertainty
S: stochastic uncertainty

Let M be the expected outcome of a model, for example, the life expectancy. M is a function of patient characteristics (X) and model parameters (B). X is a vector of covariates and B a vector of parameters. The patient characteristics vary within the population of interest according to some joint probability distribution, $f(x)$, reflecting patient heterogeneity. Parameter uncertainty is also reflected by a joint probability distribution, $g(b)$. We denote X_i as a subgroup of patients and b_j as a possible set of values for the parameters.

If patient heterogeneity is not modeled, then patients are assumed to be identical, and $f(x)$ resolves at a single point, which we will refer to as x^*. If parameter uncertainty is not modeled, then parameter values are assumed to be certain and $g(b)$ resolves at a single point, which we will refer to as b^*. Thus, we can distinguish four macrosimulation models (Table 1): deterministic models that consider neither patient heterogeneity nor parameter uncertainty (model D), models that consider patient heterogeneity only (model H), models that consider parameter uncertainty only (model P), and models that consider both patient heterogeneity and parameter uncertainty (model PH). We now demonstrate how to evaluate the appropriate outcomes of interest for each of these models.

Model D: deterministic model

The simplest model is a deterministic model:

$$M = h(x^*, b^*)$$

This model type is probably the most prevalent model type in the literature. The expected outcome, M, is typically calculated using a single (Markov) cohort analysis. Patient heterogeneity can still be evaluated by analyzing the model for discrete subgroups, and parameter uncertainty can be considered by using deterministic sensitivity analysis.

Model H: patient heterogeneity

In this model, the expected outcome is subject to patient heterogeneity only:

$$M = h(X, b^*).$$

The expected outcome, M, in this model varies across the possible patient characteristics, X, which are distributed according to a frequency distribution, $f(X)$. M_i is the expected outcome given the patient characteristics x_i.

Cohort analysis can be applied to calculate the expected outcome for any subgroup x_i: $M_i = h(x_i, b^*)$. With second-order Monte Carlo simulation we can obtain the distribution (P_X) of the expected outcome, reflecting patient heterogeneity: $P_X[M]$. For each randomly drawn set of values, x_i, a cohort analysis is performed to calculate $M_i = h(x_i, b^*)$. A distribution is then made of the results of the cohort analyses.

If we want to make a decision for a heterogeneous population, our primary interest is in the expected outcome of the aforementioned distribution: $E_X[M]$. The expected outcome (for a heterogeneous population) is estimated as the average outcome of the Monte Carlo simulation that resulted in $P_X[M]$. When the patient characteristics can be quantified numeri-

cally (such as age in years, or diastolic blood pressure in mmHg), this average outcome may not equal the expected outcome for patients whose characteristics are at the mean of their distributions. That is, $E_X[M] = E_X[h\ (X, b^*)]$ may differ from $h(E[X], b^*)$. This is true in particular when the outcome is a nonlinear function of the characteristic, X. Therefore, avoiding Monte Carlo simulation by using the mean values of the patient characteristics may result in a biased expected outcome.

Model P: parameter uncertainty

When parameter uncertainty is modeled and patient heterogeneity is not, the expected outcome is subject to parameter uncertainty only:

$$M=h(x^*, B).$$

M_j is the expected outcome if b_j represents the "true" set of parameter values. Second-order Monte Carlo simulation (i.e., probabilistic sensitivity analysis) is typically applied to obtain the distribution of the expected outcome (reflecting lack of perfect knowledge): $P_B[M]$.[16, 17, 80] For each randomly drawn set of values, b_j, a cohort analysis is performed to obtain $M_j=h(x^*, b_j)$. The distribution of the expected outcome is then built from the results of the cohort analyses. The expected value of this distribution is $E_B[M]$. In nonlinear models, this expected outcome may not equal the expected outcome that would result if all parameters were fixed at their mean values: $E_B[M] = E_B[h\ (x^*, B)]$ is unequal to $h(x^*, E[B])$. Therefore, avoiding Monte Carlo simulation by using the mean values for the uncertain model parameters – as practiced in deterministic models – may result in a biased expected outcome.

Model PH: parameter uncertainty and patient heterogeneity

The analyses become more complicated when the expected outcome is subject to both parameter uncertainty and patient heterogeneity:

$$M=h(X, B).$$

In this model, the outcome M varies across both patient characteristics, X, and the parameter space, B. For each set of values (x_i, b_j) we can calculate the expected outcome,

$$M_{ij}=h(x_i, b_j).$$

Expected outcome: heterogeneous population

If we want to make a decision for a heterogeneous population, our main interest is in the expected outcome given both parameter uncertainty and patient heterogeneity:

$$E_{X,B}=[M].$$

Second-order Monte Carlo simulation can be applied to estimate this expected outcome. For each set of values (x_i, b_j) – randomly drawn from the distributions of X and B – a cohort analysis is performed to obtain $M_{ij}=h(x_i, b_j)$. The expected outcome of the model is estimated by the average outcome of many cohort analyses. The step-by-step algorithm is:

1. Draw a value for (x_i, b_j).
2. Calculate $M_{ij}=h(x_i, b_j)$, for example, using cohort analysis.
3. Repeat steps 1 and 2 many times.
4. Calculate the average of the expected outcomes, M_{ij}, obtained at step 2.

Distribution of the expected outcome, reflecting lack of perfect knowledge

In addition, we may be interested in the distribution of the expected outcome reflecting lack of perfect knowledge, for a heterogeneous population:

$$P_B[E_X\,[M]]. \qquad \text{[equation 1]}$$

This is the same distribution as in model P, but now for a heterogeneous population. The analysis requires a nested Monte Carlo simulation: for each parameter value, b_j, we have to average over X. For each b_j the expected outcome is a function of X: $M_j=h(X, b_j)$. The algorithm to obtain this distribution consists of the following steps:

1. Draw a value b_j.
2. Draw a value x_i.
3. Calculate $M_{ij}=h(x_i, b_j)$, for example, using cohort analysis.
4. Repeat steps 2 and 3 many times, for the same value b_j.
5. Calculate the average of the expected outcomes, M_{ij}, obtained at step 3.
6. Repeat steps 1 to 5 many times.
7. Make a probability distribution of the averages at step 5.

The mean of this distribution is $E_B[E_X\,[M]] = E_{X,B}\,[M]$: if we are only interested in the expected outcome, the nested simulation is not necessary, and the method described in the previous section ("Expected outcome: heterogeneous population") would apply.

Distribution of the expected outcome, reflecting the impact of patient heterogeneity

To obtain the expected outcome for a subgroup, a second-order Monte Carlo simulation is performed to average over the parameter space B for the subgroup x_i:E_B $[M|X = x_i]$. For each, the expected outcome is a function of B:$M_i = h(x_i,B)$. The distribution of the expected outcome reflects the impact of patient heterogeneity:

$$P_X(E_B[M]). \qquad \text{[equation 2]}$$

This is the same distribution as in model H, but averaged over B for each x_i. The algorithm for P_X is identical to the algorithm for P_B in this model, except for a reversed order of steps one and two and correspondingly different interpretations of the result.

The mean of this distribution is $E_X[E_B\,[M]] = E_{X,B}\,[M]$: the same expected outcome for a heterogeneous population is obtained, independent of the order of the nested Monte Carlo simulation.

3. MICROSIMULATIONS

In this section we demonstrate models that require the simulation of individual subjects (Table 1). The setup of this section is analogous to the previous section on macrosimulations. The simulation of individual subjects, however, adds a layer of complexity. Instead of a cohort analysis, a first-order Monte Carlo simulation is now required to estimate the expected outcome given the patient characteristics and parameter values. In addition, the simulation of individual subjects enables the estimation of the distribution of the individual outcome. Whereas M was the expected model outcome in the previous section, in this section L is the model outcome of an individual. Like M, L is a function of X and B. For given values of X and B, however, L is uncertain at the patient level. We denote a realized value for Las l_k.

We can distinguish four microsimulation models, all of which are subject to stochastic uncertainty (Table 2): stochastic models that are subject to neither patient heterogeneity nor parameter uncertainty (model S), models that consider patient heterogeneity (model HS), models that consider parameter uncertainty (model PS), and models that consider both patient heterogeneity and parameter uncertainty (model PHS).

Model S: stochastic uncertainty

When parameter uncertainty and patient heterogeneity are not modeled, the outcome of an individual is subject only to stochastic uncertainty:

$$L=h(x^*, b^*).$$

The distribution of the individual outcome $P[L]$ can be obtained with first-order Monte Carlo simulation.[83, 113] The Monte Carlo simulation involves many random walks, each resulting in a realized value (l_k) for L. The expected outcome, $E[L]$, is estimated as the average individual outcome across many random walks.

Model HS: patient heterogeneity and stochastic uncertainty

In this model, the outcome for an individual is subject to patient heterogeneity and stochastic uncertainty:

$$L=h(X, b^*)$$

For any patient with characteristics x_i, $L_i=h(x_i, b^*)$ represents the range of values that have a defined probability of being realized.

Distribution of the individual outcome

The distribution of the individual outcome is $P_{X,L}[L]$. The algorithm to obtain this distribution is:

1. Draw a value x_i.
2. Draw a value l_k (i.e., perform a random walk) given x_i.
3. Repeat steps 1 and 2 many times.
4. Make a probability distribution of the individual outcomes, l_k, obtained at step 2.

Expected outcome (heterogeneous population)

If we are interested in the expected outcome in a heterogeneous population we can estimate $E_{X,L}[L]$. The expected outcome is estimated as the average of the previous distribution.

Distribution of the expected outcome, reflecting patient heterogeneity

The expected outcome of a subgroup x_i is: $E_L\,[L|X = x_i]$. This can be estimated with a first-order Monte Carlo simulation. The patient heterogeneity in the expected outcome is presented as a probability distribution:

$$P_X\,[E_L[L]].$$

This distribution can be estimated with a nested Monte Carlo simulation. The algorithm is:

1. Draw a value x_i.
2. Draw a value l_k (i.e., perform a random walk) given x_i.
3. Repeat step 2 many times.
4. Calculate the average of the random walks, l_k, obtained at step 2.
5. Repeat steps 1 to 4 many times.
6. Make a probability distribution of the averages at step 4.

Model PS: parameter uncertainty and stochastic uncertainty

In this model, the outcome of an individual is subject to both parameter uncertainty and stochastic uncertainty:

$$L=h(x^*, B).$$

For any set of parameter values, b_j $L_j=h(x^*, b_j)$ represents the range of values that have a defined probability of being realized.

Distribution of the individual outcome

The distribution of the individual outcome is $P_{B,L}[L]$. This distribution can be obtained by performing first-order and second-order Monte Carlo simulation simultaneously. The algorithm is:

1. Draw a value b_j.
2. Draw a value l_k (i.e., perform a random walk) given b_j.
3. Repeat steps 1 and 2 many times.
4. Make a probability distribution of the individual outcomes at step 3.

This distribution reflects uncertainty about the individual outcome due to both randomness (stochastic uncertainty) and the lack of perfect knowledge (parameter uncertainty).

Distribution of the expected outcome, reflecting lack of perfect knowledge

The expected value of the previous distribution is the expected outcome of the model: $E_{B,L}(L)$. A nested Monte Carlo simulation is needed in order to obtain a probability distribution for the expected outcome:

$$P_B\,[E_L[L]].$$

The algorithm is:
1. Draw a value b_j.
2. Draw a value l_k (i.e., perform a random walk) given b_j.
3. Repeat step 2 many times.
4. Calculate the average of the random walks at step 2.
5. Repeat steps 1 to 4 many times.
6. Make a probability distribution of the averages at step 4.

Model PHS: parameter uncertainty, patient heterogeneity, and stochastic uncertainty

In this model, the outcome of an individual is subject to patient heterogeneity, parameter uncertainty, and stochastic uncertainty:

$$L=h(X, B).$$

Distribution of the individual outcome

The distribution for the outcome of an individual with unknown patient characteristics is:

$$P_{X,B,L}[L].$$

This distribution can be estimated with a combined first-order and second-order Monte Carlo simulation. The algorithm is:
1. Draw values for (x_i, b_j).
2. Draw a value l_k (i.e., perform a random walk), given (x_i, b_j), drawn in step 1.
3. Repeat steps 1 and 2 many times.
4. Make a distribution of the outcomes found at step 2.

This distribution reflects uncertainty about the individual outcome due to randomness (stochastic uncertainty), the lack of perfect knowledge (parameter uncertainty), and unknown patient characteristics (patient heterogeneity).

Distribution of the expected outcome, reflecting lack of perfect knowledge

If we want to make a decision for a heterogeneous population, we can estimate the expected outcome of the heterogeneous population as the average of the previous distribution:

$$E_{X,B,L}[L].$$

The probability distribution for the expected outcome reflecting lack of perfect knowledge is:

$$P_B[E_{X,L}[L]].$$

We can estimate this distribution using both first-order and second-order Monte Carlo simulation. The expected outcome is estimated for $B = b_j$, for , $j = 1,\ldots, N$ by averaging over X and L. The step-by-step algorithm is:

1. Draw a value b_j.
2. Draw a value x_i.
3. Draw a value l_k (i.e., perform a random walk), given the values b_j and x_i drawn in steps 1 and 2.
4. Repeat steps 2 and 3 many times, for the same value of b_j.
5. Calculate the average found at step 3.
6. Repeat steps 1 to 5 many times.
7. Make a probability distribution of the averages found at step 5.

Distribution of the expected outcome, reflecting the impact of patient heterogeneity
To obtain the expected outcome for a subgroup, we can estimate:

$$E_{B,L}[L|X = x_i].$$

The algorithm to estimate this expectation using Monte Carlo simulation is:

1. Draw a value b_j.
2. Draw a value l_k (i.e., perform a random walk), given b_j from step 1 and x_i.
3. Repeat steps 1 and 2 many times.
4. Calculate the average of the outcomes found at step 2.

The distribution of the expected outcome that reflects the impact of patient heterogeneity is:

$$P_X[E_{B,L}[L]].$$

The algorithm is identical to the algorithm for equation 3, but with steps 1 and 2 switched.

4. EXAMPLE

A multivariable risk function to predict the risk of cardiovascular disease (CVD) is a key component of many CVD decision models. D'Agostino used data from the Framingham Heart Study to build such a function to predict the 10-year risk of developing a first CVD event.117 Cox proportional hazards regression resulted in an equation for the 10-year risk (p):

$$p = 1 - S_o(10)^{\exp\left(\sum_{i=1}^{p} \beta_i \left(X_i - \overline{X_i}\right)\right)},$$

where $S_o(10)$ is the baseline 10-year survival, β_i is the ith estimated regression coefficient (log hazard ratio), X_i is the (log-transformed) ith covariate, and $\overline{X_i}$ is the (log-transformed)

Table 2 Multivariable risk function to predict the 10-year risk of cardiovascular disease in men.

MODEL

Parameter	-mean	beta-sem	mean X
Age (log)	3.061	0.214	3.856
Total cholesterol (log)	1.124	0.207	5.342
HDL (log)	-0.933	0.142	3.769
SBP if untreated (log)*	1.933	0.290	4.354
SBP if treated (log)#	1.999	0.286	0.502
Smoking	0.655	0.078	0.352
Diabetes	0.574	0.110	0.065

SUMMARY STATISTICS FRAMINGHAM COHORT

Covariate	mean	Sd	proportion yes
Age	48.5	10.8	-
Total cholesterol	212.5	39.3	-
HDL	44.9	12.2	-
SBP (log)	129.7	17.6	-
Antihypertensive treatment	-	-	402/3969=0.10
Smoking: 0=no 1=yes	-	-	1398/3969=0.35
Diabetes: 0=no 1=yes	-	-	258/3969=0.18

SBP: systolic blood pressure
sem: standard error of the mean
sd: standard deviation
Total cholesterol and HDL in mg/dL

* The covariate associated with the regression coefficient "SBP if untreated" is 0 if an individual has antihypertensive treatment, and it is log (SBP) if an individual has no antihypertensive treatment.
‡ The covariate associated with the regression coefficient "SBP if treated" is log (SBP) if an individual has antihypertensive treatment, and it is 0 if an individual has no antihypertensive treatment.

mean value of the ith covariate. The regression coefficients β_i can be represented by distributions, reflecting parameter uncertainty. The covariates X_i reflect patient heterogeneity. Table 2 presents the covariates and coefficients; appendix 4 demonstrates the calculation.

We now demonstrate the analyses described in section 2 in models whose patient heterogeneity consists only of variations in the covariates of this risk function. Both a spreadsheet and dedicated decision analysis software are suitable for the analyses.[118, 119] We performed 1,000,000 runs for single Monte Carlo simulations and 10,000*10,000 runs for nested Monte Carlo simulations.

In a deterministic model (model D) we only need the mean estimates of the regression coefficients. Next, a base case analysis is usually performed involving a patient with a typical set of covariates. For our example (see appendix 4), we considered a non-smoker, non-diabetic patient with the mean values of the other covariates in Table 2. The 10-year risk for the base case is 22%. In addition to the base case analysis we can estimate the 10-year risk for any set of covariates. Sensitivity analyses are typically performed to evaluate the importance of uncertainty about the regression coefficients. Sensitivity analyses can also assess the impact of a covariate on the outcome. For example, a sensitivity analysis across the frequency distribution of the total cholesterol level (keeping all other covariates fixed at their base-case values) resulted in a range of the ten-year risk of 14% to 30%.

Patient heterogeneity is introduced by using distributions instead of fixed values for the covariates, resulting in model H. A second-order Monte Carlo simulation resulted in the expected outcome of a heterogeneous population or the expected outcome of a patient with random values for the covariates: 20%. The 95% uncertainty interval of the simulation outcomes was [2%;61%]. This is the uncertainty interval of the distribution of the expected outcome reflecting the impact of patient heterogeneity. Such wide intervals reflecting patient heterogeneity mandate separate analyses for subgroups, which may result in subgroup-specific guidelines.

Parameter uncertainty is introduced into a deterministic model by using distributions instead of mean values for the regression coefficients, resulting in model P. Standard errors of the individual coefficient estimates – or ideally the joint variance-covariance matrix of the coefficient estimates – can be used to generate these parameter distributions. A second-order Monte Carlo simulation (probabilistic sensitivity analysis) resulted in the same 10-year risk of 22% as was found using model D. Note that only if the outcome is a linear function of the regression coefficients (which is approximately the case in this example) will the mean of the outcome of the probabilistic sensitivity analysis be identical to the expected outcome of the deterministic model. The 95% uncertainty interval of the simulation outcomes was

[20%;24%]. This is the uncertainty interval of the distribution of the expected outcome reflecting lack of perfect knowledge.

Finally, we introduced distributions for both regression coefficients and covariates, resulting in model PH. A second-order Monte Carlo simulation that sampled both the regression coefficients and the covariates resulted in the expected outcome of a patient with random values for the covariates: 20%. The expected outcome was identical to the expected outcome in model H because the outcome is approximately a linear function of the regression coefficients. To determine the distribution of the expected outcome reflecting lack of perfect knowledge, for a heterogeneous population, we performed a two-level simulation. In the outer loop the regression coefficients were sampled, and in the inner loop the covariates were sampled (as in equation 1 in the section on model PH). The 95% uncertainty interval of the simulation was [18%;21%]. To determine the distribution of the expected outcome reflecting the impact of patient heterogeneity, taking parameter uncertainty into account, we performed a similar two-level simulation. However, this time in the outer loop the covariates were sampled, and in the inner loop the regression coefficients (as in equation 2 in the section on model PH). We found a 95% uncertainty interval of [2%;61%]. The interval is identical to the interval of model H, because in our example the model outcome is approximately a linear function of the regression coefficients.

DISCUSSION

We have reviewed the methods for the combined analyses of uncertainty and patient heterogeneity in medical decision models. To the best of our knowledge a comprehensive overview of decision models that are subject to more than one type of uncertainty or patient heterogeneity does not exist in the literature. However, several applications of the analysis of parameter uncertainty in microsimulation models are available in the literature.[20, 101] A debate in this journal concerning the correct combined analysis of parameter uncertainty and stochastic uncertainty exemplified that such combined analyses are not straightforward.[101, 120, 121]

The combined analysis of patient heterogeneity and uncertainty has drawn increasing attention in the risk analysis literature since the early nineties.[122-125] Nested (or two-dimensional) simulations were suggested for quantifying both patient heterogeneity and uncertainty in the assessment of risks.[123, 126] In addition, a decomposition of the total variance into components attributable to parameter uncertainty and patient heterogeneity was proposed in the risk analysis literature, and more recently in the statistical literature.[112, 127]

Decision models are built primarily to inform policy makers about the expected outcomes of competing strategies.[90] We demonstrated that the expected outcome and the distribution of the individual outcome can always be obtained in a single Monte Carlo simulation (i.e., without using nested Monte Carlo simulations). Nested Monte Carlo simulations are inevitable if we are interested in the distribution of the expected outcome (reflecting lack of perfect knowledge) for a heterogeneous population or or in microsimulation models. In addition, nested analyses should be performed to obtain the distribution of the expected outcome reflecting patient heterogeneity, both in microsimulation models and models with parameter uncertainty. These nested Monte Carlo simulations are computer-intensive. More time-efficient methods, such as Gaussian process modeling, have been developed, but few applications have been published.[91] The importance of such time-efficient methods will depend on the balance between increasing model complexity and increasing computer performance.

5

Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis

Bas Groot koerkamp
M.G. Myriam Hunink
Theo Stijnen
James K. Hammitt
Karen M. Kuntz
Milton C. Weinstein

Med Decis Making. 2007 Mar-Apr;27(2):101-11.

Presented at the Annual Meeting of the Society for Medical Decision Making in 2004, and awarded a Lee B. Lusted Student Prize for outstanding presentation.

ABSTRACT

Clinical journals increasingly illustrate uncertainty about the cost and effect of health care interventions using cost-effectiveness acceptability curves (CEACs). CEACs present the probability that each competing alternative is optimal for a range of values of the cost-effectiveness threshold. The objective of this manuscript is to demonstrate the limitations of CEACs for presenting uncertainty in cost-effectiveness analyses.

These limitations arise because the CEAC is unable to distinguish dramatically different joint distributions of incremental cost and effect. A CEAC is not sensitive to any change of the incremental joint distribution in the upper-left and lower right quadrant of the cost-effectiveness plane, neither is it sensitive to radial shift of the incremental joint distribution in the upper-right and lower-left quadrants. As a result, CEACs are ambiguous to risk-averse policy makers, inhibit integration with risk-attitude, hamper synthesis with other evidence or opinions, and are unhelpful to assess the need for more research. Moreover, CEACs may mislead policy makers and can incorrectly suggest medical importance. Both for guiding immediate decisions and for prioritizing future research, these considerable drawbacks of CEACs should make us rethink their use in communicating uncertainty.

As opposed to CEACs, confidence and credible intervals do not conflate magnitude and precision of the net benefit of health care interventions. Therefore, they allow (in)formal synthesis of study results with risk attitude and other evidence or opinions. Presenting the value of information in addition to these intervals allows policy makers to evaluate the need for more empirical research.

INTRODUCTION

Clinical journals increasingly present the uncertainty about the costs and effects of health care interventions using cost-effectiveness acceptability curves (CEACs). The British Medical Journal, for example, has published 16 cost-effectiveness analyses – decision models as well as economic trials – that presented a CEAC.[128-143] These CEACs present the probability that each evaluated intervention is optimal, for varying values of the willingness-to-pay (WTP) per quality-adjusted life year (QALY). Decision making is assumed to improve when CEACs are presented in addition to point estimates of the incremental cost-effectiveness ratios (ICERs).

The ICER is typically considered the most important outcome of economic trials and decision models. It allows for comparison of the "value for money" of the interventions under evaluation with other unrelated health care interventions, whether preventive, diagnostic, or therapeutic. Organizations and journals increasingly recommend that uncertainty about the cost-effectiveness of interventions be evaluated. Because inference for the ICER is associated with statistical problems,[5] CEACs were introduced in 1994 to assess the uncertainty surrounding the ICER.[26] Unfortunately, CEACs have several noteworthy limitations that have received little attention in the literature.

The objective of this manuscript is to demonstrate the limitations of CEACs for presenting uncertainty in cost-effectiveness analyses. In the next section, we first review the relevant background, including the cost-effectiveness plane, the WTP, the net benefit framework, uncertainty about costs and effects, and the construction of CEACs. Then we discuss why uncertainty is of interest to policy makers because it may not be obvious that the evaluation of uncertainty results in better decisions. The limitations of CEACs for presenting uncertainty are then examined. Finally, we will discuss alternative presentations of uncertainty.

BACKGROUND

Cost-effectiveness plane

Cost-effectiveness analyses in health care typically result in 2 outcome measures for each competing alternative: the costs per patient and an effect measure, for example, QALYs or symptom-free days per patient. The cost-effectiveness plane (CE-plane) was introduced to illustrate the resulting 2-dimensional "health policy space".[22] Figure 1 presents a CE-plane with the expected costs and effects of alternatives A to F each compared to a common comparator strategy. None of the alternatives is dominant – that is, more effective and less costly than all other alternatives. A policy maker will therefore need to be explicit about how she or he values costs compared to effects. Because novel health care interventions are often both more effective and more costly, a trade-off between costs and effects has to be made.

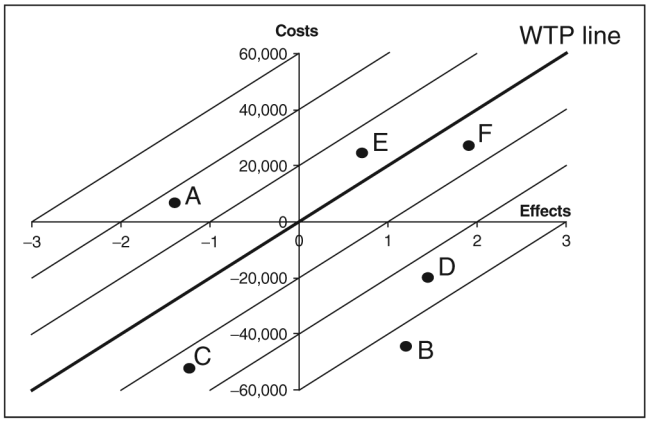

Figure 1 Cost-effectiveness plane with the expected costs (C) and expected effects (E) of six alternatives each compared to a common comparator strategy. The WTP line is indicated through the origin for a WTP of $20,000 per QALY. The parallel lines are indifference curves, connecting points with the same net benefit. The NHB of the indifference curves corresponds to the value of the x-axis where the indifference curves intersect the x-axis. A to F represent the cost and effects of 6 alternatives: B is the preferred alternative.

Threshold WTP and the Net benefit framework

The threshold WTP has been introduced as the substitution rate at which society is indifferent to "trade" costs for effects or vice-versa. It is typically assumed that the WTP is independent of the value of costs or effects. Using the substitution rate (i.e., the WTP), we can express the cost of each alternative in a health equivalent. Cost and effects can then be combined in a single outcome measure, net health benefit (NHB):[5]

$$NHB = effect - cost/WTP.$$

Analogously, we can express the effect of each alternative in a monetary equivalent, resulting in the net monetary benefit (NMB):

$$\text{NMB} = \text{effect}*\text{WTP} - \text{cost}.$$

Identifying the preferred alternative is now simplified to selecting the strategy with the maximum net benefit (NHB or NMB).

We can also present this graphically. The slope of the straight line through the origin of the CE-plane in Figure 1 equals the WTP per unit of the effect measure (e.g., $20,000 per QALY). Rotating the WTP line around the origin varies the WTP. As we rotate the WTP line counterclockwise from horizontal to vertical, the WTP increases from 0 to infinity. Alternatives to the lower right of the WTP line are preferred over alternatives to the upper left of the WTP line.

Alternatives with the same net benefit lie on the same line parallel to the WTP line. Because society is assumed to be indifferent between alternatives with the same net benefit, these parallel lines are indifference curves.[87] The indifference curves create a preference gradient, from large negative net benefits in the upper left corner of the CE-plane to large positive net benefits in the lower right corner. The NHB of the indifference curves corresponds to the x-value, where the curves intersect the x-axis: the NMB of the indifference curves corresponds to the y-value, where the curves intersect the y-axis. The WTP line through the origin represents a net benefit of 0. The indifference curves illustrate that alternative B is the most preferred, given a WTP of $20,000 per QALY.

Uncertainty about costs and effects

Estimates of the mean costs and effects are uncertain, because trials have limited sample sizes, and model parameters are not known with absolute certainty. This uncertainty is the uncertainty about the mean outcomes (cf. standard error of the mean), and is different from the uncertainty about individual outcomes (cf. standard deviation). Although uncertainty about individual outcomes may affect decision making through consideration of equity, the focus of this article is on uncertainty about mean outcomes.

In an economic trial, the costs and effects are measured for each patient. By drawing bootstrap replicates from the trial data, we can derive the joint distribution of the mean costs and effects of each alternative to represent uncertainty.[24] These joint distributions take correlations between costs and effects into account.

In a decision model, probabilistic sensitivity analysis (also known as second-order Monte Carlo simulation) can be applied to obtain the joint distribution of the mean costs and effects.[17, 38, 80] This involves 3 steps: first, many samples are drawn from the joint probability

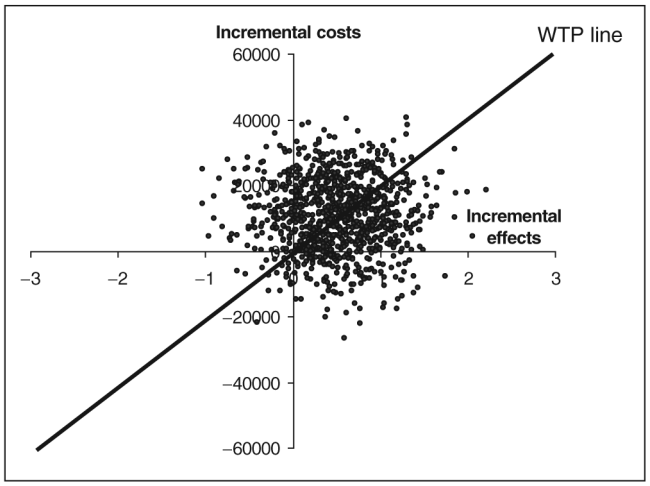

Figure 2 Cost-effectiveness plane with the joint distribution of the incremental expected costs and incremental expected effects of two competing alternatives. The word "incremental" refers to the difference between two alternatives (e.g., alternative B minus alternative A).

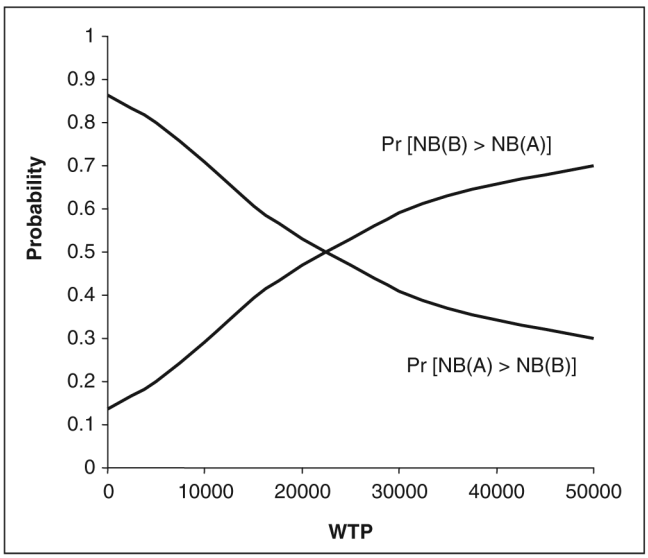

Figure 3 Cost-effectiveness acceptability curves for both alternatives. The ascending curve presents the probability that the net benefit (NB) of alternative B exceeds the net benefit of alternative A: Pr [NB(B) > NB(A)]. The descending curve presents the probability that the net benefit (NB) of alternative A exceeds the net benefit of alternative B: Pr [NB(A) > NB(B)]. Often only the ascending curve is presented.

distribution of the input parameters (e.g., transition probabilities, utilities, and costs). Next, using the model, the expected costs and effects are calculated for each alternative, for each sample. Finally, the expected costs and effects for each alternative and each sample are plotted in the CE-plane. The expected costs and effects can be correlated within and between alternatives: for example, a transition probability in the model can impact several alternatives.

Table 1 The NHB (in QALYs) for ten samples are presented for the alternatives A, B, and C. Alternative B is the preferred alternative with the maximum average NHB (i.e., 0.06). Alternative B is preferred in 70% of the samples, alternative C in 20%, and alternative A in 10%. Estimating these proportions over a range of the WTP will result in the CEACs. More samples will result in more precise estimates of the probabilities.

Sample	A	B	C	preferred
1	-0.14	0.06	-0.02	B
2	-0.27	0.07	0.05	B
3	-0.08	0.10	0.05	B
4	-0.12	0.05	0.02	B
5	-0.07	0.06	0.04	B
6	0.04	0.18	0.19	C
7	-0.15	0.04	0.02	B
8	-0.36	-0.05	-0.01	C
9	0.06	-0.02	-0.02	A
10	-0.06	0.07	0.06	B
Average	-0.11	0.06	0.04	B

If only 2 alternatives are considered, the joint distribution of the incremental costs and effects can be presented. In Figure 2, for each sample of a probabilistic sensitivity analysis, the difference in expected costs and effects is plotted. Note that the axes of Figure 2 present incremental costs and effects, whereas the axes of Figure 1 present the costs and effects of each alternative compared to a common comparator strategy. The incremental distribution reflects correlations between the outcomes of the 2 alternatives. The outcomes of alternatives in economic trials, however, are not correlated – with the exception of cross-over trials. To create an incremental joint distribution of the cost and effects of an economic trial, it is typically assumed that no correlation exists between the alternatives.

Constructing CEACs for 2 alternatives: graphically

CEACs present uncertainty as the probability that each alternative has the greatest net benefit as a function of the WTP. CEACs were introduced to present uncertainty about a choice between 2 alternatives (e.g., alternative A representing standard care and alternative B representing a new treatment).[26] The probability that one alternative (B) is preferred over the other (A) is represented graphically in Figure 2 as the proportion of the incremental joint distribution to the lower-right of the WTP line. We can estimate this proportion repeatedly while rotating the WTP line counterclockwise from horizontal (i.e., WTP = 0) to vertical (i.e., WTP = infinity). The ascending curve in Figure 3 presents the probability that alternative B is preferred over alternative A for a range of values for the WTP. The probability

that alternative A is preferred over alternative B is 1 minus the probability that alternative B is preferred over alternative A (descending curve in Figure 3). When only 2 alternatives are compared, the CEAC typically presents only the ascending curve.

Constructing CEACs for 2 or more alternatives: numerically

The net benefit framework enables the construction of CEACs for any number of alternatives.[5, 96] Table 1 presents the hypothetical results in NHB of a probabilistic sensitivity analysis of a model with 3 alternatives. Each row represents a sample from the joint distribution of the input parameters with the resulting net benefit of the model for each alternative. Each row is equally likely to be "true". For each row, we can determine the alternative with the maximum NHB. The probability that each alternative is preferred is the fraction of rows in which it is the alternative with the maximum NHB: (A;B;C) = (10%;70%;20%). The estimated fractions become more precise with more replications than the 10 samples in Table 1. Repeating this process for a range of values for the WTP yields the CEACs for each of the alternatives. The probabilities that each alternative is optimal – for a given WTP – sum to 1.

WHY POLICY MAKERS CARE ABOUT UNCERTAINTY

Uncertainty about the outcome (e.g., net health benefit) has been considered irrelevant, assuming that the preferred alternative is simply the alternative with the highest expected outcome.[27, 144] The analysis and presentation of uncertainty may, therefore, seem pointless. A study, however, showed that reimbursement decisions of the National Institute of Clinical Excellence (NICE) in the UK were considerably influenced by uncertainty about the expected outcome.[61] We will discuss 3 rationales for presenting uncertainty about the outcomes of economic randomized controlled trials (RCTs) and decision models. Two additional rationales to evaluate uncertainty are especially relevant to decision models.

Integrate uncertainty with risk attitude

Analysts typically assume that policy makers are risk neutral. In reality, however, policy makers (as representatives of society at large) tend to be risk averse rather than simply expected outcome maximizers. They are especially worried about new interventions with a nonnegligible probability of being very harmful. For example, an established drug with a small but certain benefit may be preferred over a new drug with a higher – but uncertain –benefit. When the cost of a new intervention is particularly uncertain, policy makers may worry about budgetary problems. If considerable uncertainty exists about the cost-effectiveness of the alternative with the most favorable cost-effectiveness ratio, future evidence

may show that another alternative is actually preferred. Policy makers may be reluctant to reimburse a new alternative with uncertain cost-effectiveness because changing policy and practice costs time and effort and commonly meets resistance from those involved in delivering and receiving care. For example, repeatedly starting and discontinuing a breast cancer screening program whenever the evidence tips the balance may be unacceptable to both providers and the public.

In evaluating uncertainty, risk-averse policy makers consider 2 aspects of uncertainty: the probability of not selecting the "true" preferred alternative and the possible consequences of not selecting the "true" preferred alternative. No matter how precisely estimated the net benefits of the alternatives are, the preferred alternative - based on available evidence - has a non-zero probability of not being the "true" preferred alternative. The difference in net benefit between the "true" preferred alternative and the selected alternative is the consequence or harm of not selecting the "true" preferred alternative. A policy maker can only appraise uncertainty when both the probability and the consequences of not selecting the "true" preferred alternative are presented. A decision with a 4% probability of being wrong, for example, may be perceived as very uncertain if the consequences could be devastating (e.g., outbreak of an epidemic). On the other hand, for a 50% probability of being wrong, little uncertainty may be perceived when the consequences of being wrong are negligible (e.g., 2 drugs with precise and nearly equal benefit). Because of the "flat maximum principle" the latter scenario is not entirely hypothetical in current day Western medical care: differences between competing alternatives tend to be small.[145]

An unambiguous presentation of uncertainty enables policy makers to combine the results of a cost-effectiveness analysis implicitly with their attitude towards risk. If desired, we could elicit a utility function for costs and effects for each policy maker and integrate uncertainty and risk attitude explicitly.[87] The decision is then simplified to selecting the alternative with the maximum expected utility.

Integrate uncertainty with other evidence and opinions

A quantitative analysis is rarely the only or final word on a decision.[115, 146] This is especially true for a single economic trial. Policy makers will wish to combine the results of the analysis with other (future) evidence and opinions. A quantitative statement of the uncertainty about the results will enable policy makers to weigh the results implicitly in combining them with other evidence and opinions. This synthesis could take place in a more formal Bayesian framework, if desired.[73]

Uncertainty and the need for more research

In addition to selecting the preferred alternative, policy makers have to decide whether to perform more empirical research (e.g., a trial) regarding the decision. The decision to perform more research can be considered as an independent decision from the selection of the preferred alternative.[14] Regarding the comparison of the current treatment and a new drug, for example, a policy maker can decide to:

1. implement the new drug and obtain more research,
2. implement the new drug and not obtain more research,
3. not implement the new drug and obtain more research, or
4. not implement the new drug and not obtain more research.

In the 2 previous subsections, we discussed how uncertainty influences policy makers in selecting the preferred alternative. Uncertainty, however, is also responsible for the policy maker's perception of the need for more research. Policy makers should compare the expected benefit of research with the expected research costs because collecting primary data is typically expensive. Value of information analysis is a method to explicitly estimate the expected benefit of research.[14, 27, 28, 32, 81, 105, 115, 146, 147] It considers both the probability and the consequences of not selecting the "true" preferred alternative.

Moral obligation

Scientists may feel a moral obligation to be clear about the lack of absolute certainty in their results.[115] The analysis of uncertainty serves this accountability by quantifying the confidence in the outcomes. Moreover, if uncertainty is not explicitly modeled, it is possible that investigators or institutions might (subconsciously) select values from the ranges of uncertainty that fit their preferences best.

Non-linearity

This rationale for assessing uncertainty about model parameter values in decision models has a mathematical nature and is often unappreciated. Ignoring uncertainty may result in a biased estimate of the expected net benefits of alternatives, when a nonlinear relation exists between the expected net benefit and the uncertain parameters. For example, a Markov model is nonlinear in the annual constant mortality rate. More generally we can state that a model outcome, f(b), is a nonlinear function of an uncertain model parameter, b, if the expected value of f(b) is not equal to the function of the expected value of b:

$$E[f(b)] \neq f(E[b]).$$

To account for such nonlinearities, nonlinear model parameters should ideally be modeled with probability distributions instead of point estimates.[17]

LIMITATIONS OF CEACS

In this section we discuss how CEACs perform in addressing the first 3 rationales – discussed in section 2 – for presenting uncertainty. In addition, we draw attention to 2 situations in which CEACs are easily misinterpreted and, as a result, could mislead policy makers. First, we demonstrate how dramatically different incremental joint distributions of 2 alternatives may result in the exact same CEAC. The limitations of CEACs arise because of this inability to distinguish completely different incremental joint distributions.

CEACs don't distinguish dramatically different incremental joint distributions

CEACs are identical if the proportion of the incremental joint distribution to the lower-right of the WTP line is the same for all values of the WTP. This allows points of the incremental joint distribution in the upper left or lower right quadrant to move anywhere else within the same quadrant, without changing the CEAC. The CEAC remains unchanged, because the proportion of the incremental joint distribution to the lower right of the WTP line for any value of the WTP remains unchanged. Consequently, if 1 alternative could be more costly and less effective, CEACs ignore whether the adverse health effect and budgetary impact could be tiny or giant. Moreover, CEACs are not affected when points of the incremental

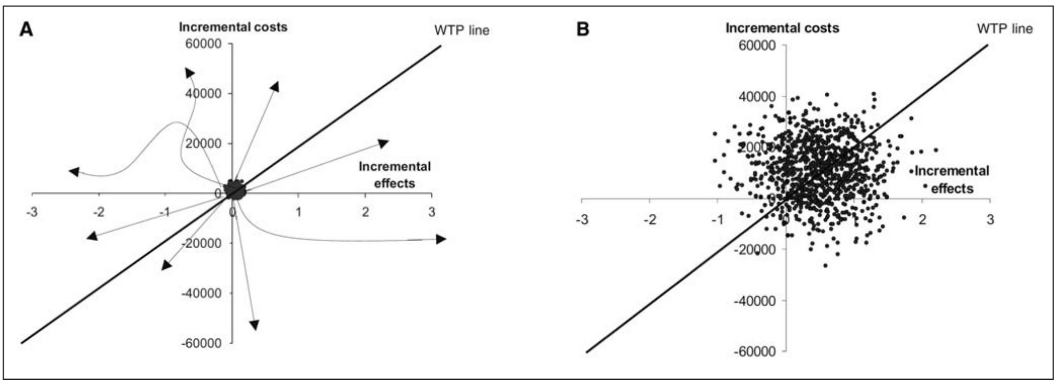

Figure 4 The CE-plane in Figure 4A reflects small expected incremental costs and effects, and some uncertainty. The arrows represent (hypothetical) changes in the distribution that would not modify the CEAC of this CE-plane. Points in the upper-left or lower-right quadrant can move anywhere else within the same quadrant, without changing the CEAC. Points in the two other quadrants can move radially towards or away from the origin, again without changing the CEAC. In Figure 4B the points have moved, resulting in a different distribution with a large expected incremental costs and effects and more uncertainty. However, the CEACs of Figure 4A and B are identical.

joint distribution in the upper right and lower left quadrants move radially toward or away from the origin. The CEAC remains unchanged because the proportion of the incremental joint distribution to the lower right of the WTP line for any value of the WTP remains unchanged.

Figure 4 illustrates 2 remarkably different incremental joint distributions with identical CEAC, presented in Figure 3. The distributions represent 2 different (hypothetical) datasets regarding the same competing health care interventions. The distribution in Figure 4A has small incremental expected costs and effects (1100 euro and 0.05 QALYs), some uncertainty (standard errors of 1100 euro and 0.05 QALYs), and no correlations. The arrows represent hypothetical changes in the incremental distribution that would not modify the CEAC. Regarding the upper left and lower right quadrants the arrows represent shift of points of the incremental distribution to anywhere within the same quadrant. In the upper-right and lower-left quadrants, they represent radial shift within the same quadrant. Figure 4B (identical to Figure 2) presents such a change in the distribution, resulting in large incremental expected costs and effects (11,000 euro and 0.5 QALYs), increased uncertainty (standard errors of 11,000 euro and 0.5 QALYs), and no correlations. We will use these 2 sets of hypothetical results to illustrate the drawbacks of CEACs and some advantages of others methods.

CEACs are useless to risk-neutral and ambiguous to risk-averse policy makers

Risk-neutral policy makers don't need CEACs. They will prefer the alternative with the maximum expected net benefit, no matter what the probability is that this alternative is optimal. To a risk-averse policy maker, the probability of not selecting the "true" preferred alternative – as presented by the CEAC – is ambiguous for assessing the importance of uncertainty without knowing the consequences (as explained earlier). As a result, any cut-off value for this probability is an ambiguous criterion to risk-averse policy makers for making decisions. Figure 4 illustrates that CEACs fail to capture the importance of uncertainty. Uncertainty is clearly more important in B than in A: possible consequences of making the wrong decision (in retrospect) are several times larger in Figure 4B, compared to Figure 4A. Therefore, a risk-averse decision maker would be more worried about the uncertainty in Figure 4B. However, the incremental distributions of Figure 4A and 4B have identical CEACs.

The risk-attitude of a policy maker - as reflected by a utility function - applies to the uncertain net benefits of the alternatives. We can formally integrate the risk attitude of a policy maker with the uncertain outcome of each alternative, to identify the alternative with the maximum expected utility.[87] Because CEACs only present the probability of not selecting the "true" preferred alternative – and not the consequences– they imply that policy makers have a binary utility function: a value of 1 is assigned if the preferred alternative is the "true" preferred alternative and a value of 0 if the preferred alternative is not the "true" preferred alternative. Policy makers are unlikely to care only about the probability that an alternative is best with no consideration of how much better or worse it may be.

CEACs hamper synthesis of study results with other evidence and opinions

To integrate a CEAC with existing evidence and opinions, we need to choose an appropriate weight for the evidence presented by the CEAC. The weight depends on the precision and validity of the evidence. Unfortunately, we cannot tell whether results presented as a CEAC are precise. CEACs conflate the precision of the net benefit estimates of each alternative with the magnitude of the estimates. For the same CEAC, the precision and magnitude may both be small as in Figure 4A, or large as in Figure 4B. Because a CEAC does not reveal the precision of estimates, we cannot assign a meaningful weight to the evidence contained in the cost-effectiveness study.

The conflation of magnitude and precision has an analogy in epidemiology. For decades, statisticians have tried to convince the medical research field that estimation (i.e., reporting confidence intervals) is better than hypothesis testing (i.e., reporting p-values).[9, 148] Hypothesis testing only provides an indication of statistical significance, which is not the same as medical importance. A confidence interval conveys the (im)precision of an estimate in addition to its magnitude.

CEACs are unhelpful in value of information analysis

CEACs provide a measure of uncertainty. It has therefore been argued that creating a CEAC is an appropriate first step towards the evaluation of the need for more research.[96] However, CEACs do not reflect the importance of uncertainty (i.e., the value of information), because they ignore the consequences of not selecting the "true" preferred alternative. For a probability of 65% that an alternative is the "true" preferred alternative, the value of information could be negligible or enormous. The same holds for a probability of 99%. The incremental distributions in Figure 4 share the same CEAC, but have an expected value of perfect information (EVPI) of 0.03 QALYs and 0.32 QALYs respectively (WTP=20,000). CEACs are therefore unreliable to evaluate the need for future research. Because a probability of 65% may suggest the need for more research, and 99% may suggest minimal value from more research, CEACs can even be considered potentially misleading regarding the need for more research.

CEACs may mislead policy makers regarding the preferred alternative

The CEACs do not report which alternative has the maximum expected benefit given the WTP. Policy makers may be easily misled to believe that the alternative with the maximum probability of having the maximum benefit is the alternative with the maximum expected benefit. However, outcomes from decision models typically have asymmetric distributions. If the distribution of the incremental net benefit is asymmetric, the alternative with the

maximum probability of having the maximum benefit may not have the maximum expected benefit. To overcome this interpretation problem, the CEAC frontier was recently introduced: the alternative with the maximum expected benefit is highlighted for all values of the WTP.[96] This "fix", however, may be confusing for the policy maker unfamiliar with these concepts.

CEACs may incorrectly suggest medical urgency or importance

When a CEAC presents a high probability (e.g., 99%) that a new drug is the "true" preferred alternative, a need for immediate action may be perceived for this seemingly convincing breakthrough drug. The high statistical significance is easily equated to medical importance. The additional benefit of the new drug compared to current standards, however, cannot be determined from the CEAC or the ICER. The mean expected benefit may in fact be negligible and not even justify the implementation costs.

OTHER REPRESENTATIONS OF UNCERTAINTY

Cost-effectiveness plane

The joint distribution of costs and effects for each alternative illustrates the comprehensive results of a cost-effectiveness analysis. Policy makers interested in cost or effects, in particular, can get an impression of the expected costs and effects, as well as their uncertainty intervals for each alternative. If only 2 alternatives are considered, the incremental joint distribution can be presented. The results can be combined qualitatively with risk-attitude, other evidence, and implementation costs. Moreover, the incremental joint distribution gives an unbiased impression of the value of more information.

If more than 2 alternatives are presented, however, the CE-plane is ambiguous about the value of information and therefore about the importance of uncertainty. Correlations between more than 2 alternatives cannot be presented in the CE-plane, but have a large impact on the value of information. The uncertain outcomes of alternatives in a decision model are often correlated because the same uncertain parameter may influence the outcome of several alternatives. When 2 alternatives are compared in a cross-over RCT or a decision model, the incremental joint distribution in the CE-plane captures the correlation between the alternatives. When more than 2 correlated alternatives are compared, the CE-plane can present the joint distributions of each alternative. Overlapping joint distributions may then be incorrectly interpreted as a reflection of important decision uncertainty. Correlations between overlapping alternatives, however, may cause one alternative to always dominate the others. This misperceived uncertainty, however, does not deter the selection

of the preferred alternative with the maximum expected outcome or expected utility for risk-averse policy makers.

The CE-plane has been recommended as a visual presentation of the results of a cost-effectiveness analysis. Summary measures may be useful in addition to the CE-plane.

Intervals for the ICER

Confidence or credible intervals for the ICER are ambiguous when the incremental joint distributions are not confined to either the upper right or the lower left quadrant. This is because the interpretation of the ICER varies across the quadrants. Unfortunately virtually all decisions worth formal consideration extend into more than 1 quadrant: uncertainty about the sign of incremental cost and effects is the rule rather than the exception. Interventions in the upper right quadrant of the CE-plane are more effective and more costly: a low ICER is preferred. Interventions in the lower left quadrant are less effective and cost saving: a high ICER is preferred.[149, 150] ICERs in the other 2 quadrants lack an inherent hierarchy, because they are a ratio of something good (i.e., positive effect or negative cost) and something bad (i.e., positive cost or negative effect). If incremental effects are positive and equal, we prefer an ICER of -$10,000 over an ICER of -$5,000, because the former has a higher negative incremental cost. If incremental costs are negative and equal, however, we do not prefer an ICER of -$10,000 over an ICER of -$5,000, because the former has a smaller incremental effect.

The question now arises, whether a valid ICER interval helps policy makers in dealing with uncertainty. The interval limits do not seem particularly informative to the risk-averse policy maker, because the incremental net benefit, the incremental effect, and the incremental costs per patient remain indeterminate from an ICER. The same upper limit of an ICER interval (e.g., $100,000 per QALY) could represent an outcome with a small change in incremental net benefit, or a huge change. It is therefore impossible to predict how bad – considering both health and costs – it would be if the ICER would resolve at the upper limit. Moreover, ICER intervals are not informative for the weight the current study should get for synthesis with other evidence, or for the value of more research.

In both Figure 4A and B about 13% of the distribution falls in the upper left, and about 13% in the lower right quadrant. As explained earlier, a hierarchy for the ICER in these quadrants does not exist.

Intervals for the net benefit

A confidence or credible interval for the net benefit of each alternative provides policy makers with information about the precision of the outcomes. Intervals enable risk-averse policy makers to weigh the precision with their risk attitude implicitly. If desired, they can

explicitly integrate their single-attribute utility function with the distribution of the net benefit for each alternative to identify the preferred alternative. The separate presentation of the magnitude and the precision of the outcomes also allows policy makers to weigh the results with other evidence and opinions. Moreover, when 2 alternatives are compared, the interval for the incremental net benefit provides an impression of the value of more information. If considerable uncertainty exists about the WTP, the intervals for net benefit can be presented across a range of values for the WTP. The intervals of the net benefit can be presented in addition to separate intervals for the cost and effects of each alternative. Policy makers may be more risk averse for harmful effects or high costs than is reflected in the net benefit framework.

This presentation has several additional advantages. The benefit of each alternative can be easily extrapolated to the population level by multiplying the benefit per individual with the appropriate number of patients. The population benefit enables comparison with implementation or transition costs. Analogously, the interval limits can be multiplied with the appropriate number of patients. Moreover, policy makers and clinicians are used to confidence/credible intervals to present uncertainty: results of RCTs and meta-analyses are typically presented as confidence intervals.

The 95% confidence intervals for the incremental net benefit of the joint distributions are –0.15 to 0.14 QALYs (Figure 4A) and –1.5 to 1.4 QALYs (Figure 4B).

Expected value of perfect information (EVPI)

When 2 alternatives are compared, the incremental distribution on the CE-plane and intervals for the INB provide only an impression of the EVPI. When more than 2 alternatives are compared, CE-planes and incremental intervals do not reflect correlations between the outcomes of alternatives. Such correlations often arise in decision models and are influential on the EVPI. The EVPI is therefore valuable to policy makers, in addition to other presentations of uncertainty, to evaluate the need for more research.

DISCUSSION

We have demonstrated some of the limitations of cost-effectiveness acceptability curves (CEACs), which are increasingly applied to represent uncertainty about the choice between competing alternatives. Policy makers are interested in uncertainty in order to assimilate study results with their risk attitude and other evidence and opinions in choosing the preferred alternative. In addition, they evaluate uncertainty to decide whether more quantita-

tive research regarding the decision uncertainty is justified. We have shown that CEACs do not allow integration with risk attitude or other evidence and opinions. Moreover, they are ambiguous regarding the need for more research.

It is not clear why CEACs are so popular in the light of these limitations. The analogy of CEACs with hypothesis testing - which is still widely used to present uncertainty in the health care literature instead of confidence (or credible) intervals - may explain its popularity. It has also been argued that CEACs are preferred over the probability distribution of net benefits, because they do not require the selection of a value for the WTP. However, the net benefit can be presented over a range of values for the WTP, analogous to CEACs. Other research fields that apply decision analysis seem to prefer the net benefit framework to present uncertainty over CEACs. For example, we are not aware of a single example of the use of CEACs in the environmental risk analysis and cost benefit literature.

We have discussed several alternative presentations of uncertainty. The cost-effectiveness plane is useful to visualize the joint distributions of the competing alternatives. In addition, summary measures of the CE-plane can help policy makers. Confidence and credible intervals for the net benefit allow integration with risk attitude as well as other evidence and opinions. They clearly separate the magnitude of outcomes from the precision of outcomes. Intervals for costs and effects can be presented separately, if the preferences of the policy maker are not represented by a constant WTP. Presenting the EVPI enables policy makers to evaluate the need for more quantitative research.

6

Identifying key parameters in cost-effectiveness analysis using value of information: a comparison of methods

Bas Groot Koerkamp
M.G. Myriam Hunink
Theo Stijnen
Milton C. Weinstein

Health Econ. 2006 Apr;15(4):383-92

ABSTRACT

Decisions in health care must be made, despite uncertainty about benefits, risks, and costs. Value of information analysis is a theoretically sound method to estimate the expected value of future quantitative research pertaining to an uncertain decision. If the expected value of future research does not exceed the cost of research, additional research is not justified, and decisions should be based on current evidence, despite the uncertainty.

To assess the importance of individual parameters relevant to a decision, different value of information methods have been suggested. The generally recommended method assumes that the expected value of perfect knowledge concerning a parameter is estimated as the reduction in expected opportunity loss. This method, however, results in biased expected values and incorrect importance ranking of parameters. The objective of this manuscript is to set out the correct methods to estimate the partial expected value of perfect information and to demonstrate why the generally recommended method is incorrect conceptually and mathematically.

INTRODUCTION

Decision makers in health care face two separate decisions when confronted with results from economic trials or decision models.[14] First, they have to decide which strategy to adopt. Next, the question arises whether more clinical research regarding the decision is desirable. More research is expected to have a benefit, since it usually decreases decision uncertainty and therefore the probability and harm of making the wrong decision. At the same time, research also has a financial cost and may result in harm because of forgone benefits from delaying adoption of beneficial interventions. More research is only justified if the expected benefit exceeds the cost.

Different methods are used to communicate (decision) uncertainty, for example, confidence intervals, Bayesian credible intervals, or acceptability curves. It is increasingly acknowledged that, when assessing uncertainty, value of information (VOI) analysis is the only method that explicitly estimates the expected benefit of future research.[151] The total expected value of perfect information (total EVPI) estimates the value of simultaneously eliminating all uncertainty of all uncertain parameters related to the decision. Estimation of partial expected value of perfect information (partial EVPI) can identify the key parameters, whose uncertainty drives the decision uncertainty. Other importance measures for parameters, such as rank correlation, n-way sensitivity analysis, elasticity or standardized regression coefficient, do not directly consider the expected net benefit per patient of further information.[151, 152] Since increasing expected net benefit per patient is the main objective of health care, VOI analysis has a sound theoretical basis.

Brennan et al. noticed that two different approaches for estimating partial EVPI using simulation have been described and applied in the literature.[153] The generally recommended approach assumes that a parameter is important if the reduction in expected opportunity loss as a result of obtaining its "true" value is high. [107, 144, 151, 154-156] The correct approach assumes that a parameter is important if the expected value of obtaining its "true" value is high. These approaches can result in large differences in outcome, and a different importance ranking. The objective of this manuscript is to set out the correct methods to estimate partial EVPI and to demonstrate why the generally recommended method is incorrect conceptually and mathematically. Before focusing on partial EVPI, we will first define some notation in a net benefit framework and start with total EVPI. This notation was recently applied by Ades and Claxton in their work about the more complicated expected value of sample information (EVSI).[32]

Net benefit framework

The net benefit (B) is a function of the strategy (a) and the uncertain parameters (θ): $B(a,\theta)$. It has two components, both functions of a and θ: an effectiveness component (Q in QALY) and a cost component (C in Dollars). In order to express effectiveness and costs in the same units, we express costs in quality adjusted life years (QALY) by dividing the costs by the willingness-to-pay for a QALY (λ) to obtain the net health benefit:[5]

$$B(a,\theta) = Q(a,\theta) - C(a,\theta)/\lambda.$$

Since θ is uncertain, the benefit of a strategy is uncertain. However, for each strategy we can calculate the expected benefit over the joint probability distribution of the parameters θ (E for expectation):

$$E_\theta B(a,\theta). \qquad (1)$$

This expectation is typically estimated using Monte Carlo simulation. The first step in Monte Carlo simulation is to randomly draw a value of θ (i.e., a vector with one value for each parameter) from the joint distribution of θ. Next, we calculate the benefit of strategy a, given the value of θ. The expected benefit of a strategy is estimated as the average benefit over many randomly drawn values of θ. If we then maximize over a, we obtain the expected benefit of the optimal strategy, without additional information:

$$\max_a E_\theta B(a,\theta). \quad (2)$$

The optimal strategy without additional information is also referred to as the optimal baseline decision or a^*. If $B(a,\theta)$ is a linear function of θ and the parameters θ are not correlated, we obtain the correct expected benefit by setting all parameters at their mean value: $\max_a B(a,E(\theta))$. This linearity is usually assumed implicitly in decision modeling where the analysis is called deterministic. Likewise, when the integral of $B(a,\theta)$ over the joint distribution of θ can be solved analytically, simulation is not necessary.

Total expected value of perfect information (EVPI)

The total EVPI is the expected value of obtaining perfect knowledge of the "true" values of all parameters (θ).[106] With perfect information there is no uncertainty, so we would know the "true" value of θ and make the correct decision. The expected benefit if we would know the "true" value of θ is obtained by integrating over θ:

$$E_\theta \max_a B(a,\theta). \quad (3)$$

Notice that the only difference between the equations (2) and (3) is the order of expectation and maximization. We can estimate equation 3 using Monte Carlo simulation. For each randomly drawn value of θ, we can imagine it is the "true" set of parameter values, and calculate the benefit of the optimal strategy given these values: $\max_a B(a,\theta)$. The average benefit over many randomly drawn values of θ is an estimate of the expected benefit *with* perfect information. The total expected value *of* perfect information (total EVPI) is the difference between the expected value *with* perfect information and the expected value without information:

$$\text{total EVPI} = E_\theta \max_a B(a,\theta) - \max_a E_\theta B(a,\theta). \qquad (4)$$

In words, the total EVPI is the difference between the expected value of the optimal decision we would make if we knew the "true" parameter values (equation 3), and the expected value of the strategy we would choose if we had to choose under conditions of uncertainty about the parameters (equation 2). If $B(a,\theta)$ is normally distributed, the total EVPI can be calculated analytically.[157]

Opportunity loss

An equally valid way of estimating and interpreting value of information is from the opportunity loss perspective.[158] In this section we demonstrate that the expected opportunity loss is equal to the expected value of information.

The opportunity loss is the difference between the benefit of the strategy that is optimal given the "true" value of θ and the benefit of the strategy that was optimal at baseline (a*). The expected opportunity loss over the joint probability distribution of the parameters θ is:

$$\begin{aligned} E_\theta\,[\text{opportunity loss}] &= E_\theta\,[\max_a B(a,\theta) - B(a^*,\theta)] \\ &= E_\theta \max_a B(a,\theta) - E_\theta B(a^*,\theta) \\ &= E_\theta \max_a B(a,\theta) - \max_a E_\theta B(a,\theta). \end{aligned}$$

Note that the final step results in the total EVPI as in equation 4. With Monte Carlo simulation we can estimate the expected opportunity loss. For each randomly drawn value of θ we can imagine it is the "true" value. We then calculate the difference in benefit between the strategy that is optimal given this value of θ and the strategy that was optimal at baseline (a*). The expected opportunity loss is estimated as the average benefit over many randomly drawn values of θ.

An example illustrates the equivalence of these definitions of EVPI. Table 1 presents the results of a Monte Carlo simulation to estimate the total EVPI of a decision between strate-

Table 1 Estimating total EVPI. Values are net benefits in QALYs.

θ	strategy A	strategy B	sample max	opportunity loss
1	4	0	4	4-4=0
2	1	2	2	2-1=1
3	3	1	3	3-3=0
4	4	5	5	5-4=1
…	…	…	…	…
N			max(a,b)	max(a,b)-a
mean	3	2	3.5	0.5

gies A and B. The bottom row, which gives the expected net benefit expressed in QALYs, shows that strategy A is optimal without additional information, resulting in an expected net benefit of 3 QALYs, compared to 2 QALYs for strategy B. The expected net benefit if we know the "true" value of θ is estimated as the average of the maximum net benefit across many (N) randomly drawn values of θ. In our example, this average is 3.5: the mean of the fourth column in Table 1. The total EVPI therefore is 3.5-3=0.5 QALYs, analogous to equation 4. The fifth column calculates the opportunity loss for each value of θ, analogous to the equation for the opportunity loss. The average opportunity loss is also 0.5 QALYs.

In the following sections we will explain several methods to estimate partial EVPI. We begin with the correct methods, based on the definition of partial EVPI. Then we will clarify why the reduction in expected opportunity loss approach is incorrect.

Partial EVPI as the increase in expected value

If the total EVPI suggests that more research is justified, the analysis of partial EVPI can identify the key parameters. For example, in an economic evaluation of alternative treatments, it is possible that only a few parameters – such as the waning of treatment benefits after the end of a clinical trial or the costs of components of resource use – account for virtually all decision uncertainty. Analogously, in a decision model comparing diagnostic tests, partial EVPI could indicate that future research is most valuable on health state utilities, rather than test characteristics.

Corresponding to the definition of the total EVPI, the partial EVPI is defined as the expected value of obtaining perfect knowledge about the "true" values of one or more parameters. Future research aims to increase the expected value of a decision. The higher this increase in expected value is, the more important the parameter is. To calculate partial EVPI we will explain a generally valid method (two-level method) and a short-cut method that is valid under certain conditions (one-level method).

Two-level method

For a subset of one or more parameters, θ_I (the parameters of interest), we estimate the expected value of learning their "true" values. The complementary set of parameters we refer to as θ_C. The expected benefit of the optimal strategy given values of θ_I is:

$$\max_a E_{\theta_C|\theta_I} B(a,\theta).$$

Here the expectation is over the conditional distribution of θ_C, given the values of θ_I. The expected benefit with perfect information on θ_I is the expectation of this quantity over the distribution of θ_I:

$$E_{\theta_I} \max_a E_{\theta_C|\theta_I} B(a,\theta).$$

BOX 1: TWO-LEVEL MONTE CARLO SIMULATION.

1. Draw a value for θ_I.
2. Draw a value for θ_C (draw from the conditional distribution given θ_I, if θ_I and θ_C are correlated).
3. Calculate the benefit for each strategy.
4. Repeat steps 2-3 many times (i.e.,θ_I remains fixed at value drawn at step 1).
5. Calculate, for each strategy, the average of the benefits calculated at step 3, and identify the optimal strategy.
6. Repeat steps 1-5 many times.
7. Calculate the average of the benefits of the optimal strategies calculated at step 5.
8. The partial EVPI is the difference between the average at step 7 and the benefit of the optimal baseline strategy.

The partial expected value of perfect information on θ_I is the difference between the expected benefit with perfect information on θ_I and the expected benefit without information:

$$p\text{EVPI}(\theta_I) = E_{\theta_I} \max_a E_{\theta_C|\theta_I} B(a,\theta) - \max_a E_\theta B(a,\theta). \quad [81, 159] \quad (5)$$

Again, we can estimate this expected benefit using Monte Carlo simulation. Because of the two expectations, we need a two-level Monte Carlo simulation. Box 1 presents a stepwise algorithm.

Based on the reasoning in this section, equation 5 is the correct definition of partial EVPI. Therefore, any definition of partial EVPI that is not mathematically equivalent to equation 5 is incorrect.

One-level method

Because of the double expectation in equation 5, the Monte Carlo simulation is computer-intensive. However, if $B(a,\theta)$ is a linear function of θ_C, or a multilinear function of θ_C and the parameters θ_C are not correlated, a one-level Monte Carlo simulation is mathematically equivalent (linearity assumption). A utility is an example of a parameter in which B is linear, a transition probability in a Markov decision model is an example of a parameter in which B is not linear.

Similar to the two-level method, we want to know the expected value of obtaining the "true" values of one or more parameters (θ_I). Under the above-mentioned conditions, it is valid to move the inner expectation of equation 5 within the net benefit formula:

$$E_{\theta_I} \max_a B(a, \theta_I, E(\theta_C|\theta_I)).$$

If θ_I and θ_C are independent this can be simplified as:

$$E_{\theta_I} \max_a B(a, \theta_I, E(\theta_C)).$$

BOX 2: ONE-LEVEL MONTE CARLO SIMULATION.

1. Draw a value for θ_I.
2. Calculate the mean for θ_C (use the conditional mean given θ_I, if θ_I and θ_C are correlated).
3. Calculate the benefit for each strategy, and identify the optimal strategy.
4. Repeat steps 1-3 many times.
5. Calculate the average of the benefits of the optimal strategies calculated at step 3.
6. The partial EVPI is the difference between the average at step 5 and the benefit of the optimal baseline strategy.

The resulting partial EVPI is mathematically equivalent to equation 5 if the linearity assumption holds. To distinguish this definition from equation 5, we use the suffix "short-cut".

$$p\text{EVPI}(\theta_I)_{\text{short-cut}} = E_{\theta_I} \max_a B(a,\theta_I,E(\theta_C|\theta_I)) - \max_a E_\theta B(a,\theta).^{81,\ 159,\ 160} \qquad (6)$$

Box 2 presents a stepwise algorithm for the simulation.

Partial EVPI as the reduction in expected opportunity loss

Reduction in expected opportunity loss approach

In the literature, many authors assume that the objective of future research is to reduce the expected opportunity loss, instead of increasing the expected value.[107, 144, 151, 154-156] Therefore, partial EVPI has been defined as a reduction in expected opportunity loss, which is equivalent to a reduction in EVPI as we showed in the section on opportunity loss. Defining total EVPI as a reduction in expected opportunity loss results in the same total EVPI as equation 4. Regarding partial EVPI, however, this approach results in a bias as we will demonstrate. To distinguish this alternative definition of partial EVPI from the correct definition we use the suffix "red" (for "reduction"):

$$\begin{aligned} p\text{EVPI}(\theta_I)_{\text{red}} &= \text{reduction in expected opportunity loss} \\ &= \text{reduction in EVPI} \\ &= \text{total EVPI} - [\text{EVPI} \mid \theta_I = E(\theta_I)] \\ &= \text{total EVPI} - [\text{EVPI of } \theta_C \mid \theta_I = E(\theta_I)]. \end{aligned} \quad (7)$$

The total EVPI, the first term on the right hand side in equation 7, is the same as in equation 4. The second (bracketed) term is the total EVPI, but keeping θ_I fixed at its mean value during the entire analysis. But since θ_I is fixed at its mean, this term is the same as what we have referred to as $p\text{EVPI}(\theta_C)$short-cut in equation 6:

$$\begin{aligned} p\text{EVPI}(\theta_I)_{\text{red,short-cut}} &= \text{total EVPI} - [\text{EVPI of } \theta_C \mid \theta_I = E(\theta_I)] \\ &= \text{total EVPI} - p\text{EVPI}(\theta_C)_{\text{short-cut}} \\ &= [E_\theta \max_a B(a,\theta) - \max_a E_\theta B(a,\theta)] - [E_{\theta_C} \max_a B(a,E(\theta_I),\theta_C) - \max_a E_\theta B(a,\theta)] \\ &= E_\theta \max_a B(a,\theta) - E_{\theta_C} \max_a B(a,E(\theta_I),\theta_C). \end{aligned} \quad (8)$$

This estimation of $p\text{EVPI}(\theta_C)_{\text{short-cut}}$ would be valid as an estimate of $p\text{EVPI}(\theta_C)$ if the linearity assumption holds. If this condition is not met, we should estimate $p\text{EVPI}(\theta_C)$ instead, as in equation 5:

$$\begin{aligned} p\text{EVPI}(\theta_I)_{\text{red}} &= \text{total EVPI} - p\text{EVPI}(\theta_C) \\ &= E_\theta \max_a B(a,\theta) - E_{\theta_C} \max_a E_{\theta_C|\theta_I} B(a,\theta). \end{aligned} \quad (9)$$

Equation 8 has been described and applied by many investigators.[144, 151, 153-156] However, neither equation 8 nor equation 9 correctly estimates the partial EVPI of θ_I. The reason, which we will demonstrate, is that they depend on the incorrect premise that partial EVPIs add up to the total EVPI.[107]

The bias explained

In this subsection we clarify why the reduction methods represented by equation 8 and 9 result in biased estimates of the partial EVPI as defined in their correct analogues, equations 6 and 5, respectively. Equation 6, and not equation 8, is correct if the linearity assumption holds. Equation 5, and not equation 9, is correct in the general case.

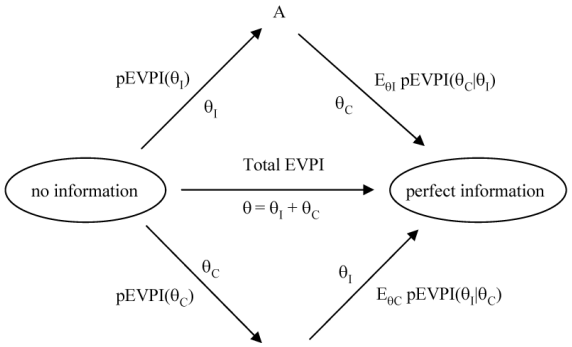

Figure 1 Partial EVPI: the order of gathering information matters.

Figure 1 shows different paths, represented by the arrows, from no additional information to total perfect information. For the total EVPI we go directly from no information to perfect information, learning the "true" values of all parameters θ. The quantity $p\text{EVPI}(\theta_I)$ in Figure 1 is the partial EVPI of θ_I as estimated correctly in equation 5, or equation 6 if the linearity assumption holds. It is the expected value of perfect information if we start out with no information and obtain perfect information on θ_I. If from no information we would first gather perfect information on θ_C, then $p\text{EVPI}(\theta_C)$ would be the expected value of perfect information on θ_C.

With perfect information on θ_I we arrive at A in Figure 1. From A, we can proceed to total perfect information by gathering perfect information on θ_C. The expected value of information when moving from A to total perfect information is conditional on the values for θ_I, and therefore not the same as the $p\text{EVPI}(\theta_C)$. Consequently, the sum of $p\text{EVPI}(\theta_I)$ and $p\text{EVPI}(\theta_C)$ is usually not equal to the total EVPI. If we know the "true" values for θ_I, we can directly estimate the expected value of perfect information on θ_C, conditional on the "true" values for θ_I: $p\text{EVPI}(\theta_C|\theta_I)$. If we have not yet gathered additional information on θ_I, we can use the joint distribution of θ_I to estimate the expected $p\text{EVPI}(\theta_C|\theta_I)$, which is E_{θ_I} $p\text{EVPI}(\theta_C|\theta_I)$. Analogously, $E_{\theta_C} p\text{EVPI}(\theta_I|\theta_C)$ is the expected value of perfect information on θ_I, given perfect information on θ_C. Intuitively, the sum of the value of information going from no information to total perfect information is independent of the path taken (see Appendix 5 for a mathematical proof):

$$\begin{aligned} \text{total EVPI} &= p\text{EVPI}(\theta_I) + E_{\theta_I}\, p\text{EVPI}(\theta_C|\theta_I) \\ &= p\text{EVPI}(\theta_C) + E_{\theta_C}\, p\text{EVPI}(\theta_I|\theta_C). \end{aligned} \qquad (10)$$

From equation 10,

$$p\text{EVPI}(\theta_I) = \text{total EVPI} - E_{\theta_I}\, p\text{EVPI}(\theta_C|\theta_I),$$

but from equation 9,

$$p\text{EVPI}(\theta_I)_{\text{red}} = \text{total EVPI} - p\text{EVPI}(\theta_C).$$

Therefore, $p\text{EVPI}(\theta_I)$ and $p\text{EVPI}(\theta_I)_{\text{red}}$ are equal if and only if:

$$p\text{EVPI}(\theta_C) = \text{E}_{\theta_I}\, p\text{EVPI}(\theta_C|\theta_I). \qquad (11)$$

In general, this equality will not hold. The left-hand side is the partial EVPI of θ_C starting with no information on θ_I, whereas the right-hand side is the partial EVPI on θ_C starting with perfect information on θ_I. In Figure 1, the left-hand side of equation 11 is the left portion of the bottom path, whereas the right-hand side of equation 11 is the right portion of the top path. These quantities need not be equal, and we give an intuitive counterexample below.

There is a way to fix the reduction in expected opportunity loss approach. We can define the reduction in expected opportunity loss as:

$$p\text{EVPI}(\theta_I)_{\text{red}} = \text{total EVPI} - E_{\theta_I}\, p\text{EVPI}(\theta_C|\theta_I). \qquad (12)$$

For this fix, we defined the second term on the right-hand side of equation 12 as the EVPI conditional on perfect information on θ_I. If interpreted this way, the reduction in expected opportunity loss equals the expected value of information. However, this is not the method of defining reduction in expected opportunity loss that is usually applied. Furthermore, equation 12 is clearly not the most direct path to estimate $p\text{EVPI}(\theta_I)$.

A simplified example

This example will demonstrate and help understanding why the generally recommended approach to calculating partial EVPI is incorrect, and likely to result in substantially biased outcomes and conclusions.

Figure 2 shows the decision tree for a decision between surgery and medication for a certain disease in newborns. If newborns receive medication, the expected net health benefit is 76 QALYs. With successful surgery, the expected net health benefit is extended to 80 QALYs. However, undergoing surgery will expose newborns to two independent risks. First, they face a risk of surgical mortality. If the newborn survives surgery, there is a risk of be-

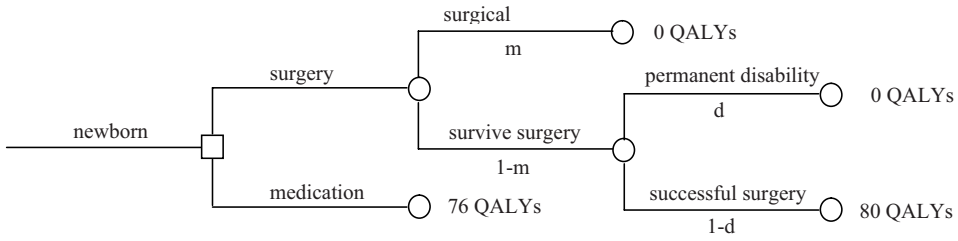

Figure 2 Decision tree for partial EVPI example.

coming so severely disabled that the utility equals zero. The "true" values of the probabilities of surgical mortality (m) and permanent disability (d) are uncertain. However, we know that they have identical and independent dichotomous distributions: 50% that the "true" value is 0, and 50% that the "true" value is 0.1. Since m and d are independent, and since the benefit is a linear function of each of these parameters, we can calculate the expected benefit of surgery using the mean values for m and d:

$$\begin{aligned}\text{Expected net health benefit of surgery} &= (1\text{-m})*(1\text{-d})*80 \\ &= (1\text{-}0.05)*(1\text{-}0.05)*80 = 72.2 \text{ QALYs.}\end{aligned}$$

Surgery seems too risky for the newborns. However, with value of information analysis we can assess whether more research is justified and if so, whether we should obtain more precise estimates of m, d, or both. Since the uncertain parameters in this example have dichotomous distributions, expected benefits can be calculated algebraically and simulation is not necessary. Moreover, since B is linear in both parameters, the short-cut method is valid (see Appendix B for the calculations). Using equation 4, we find a total EVPI of 1 QALY. This implies, that if future research could determine the "true" values for both m and d, the expected benefit of this knowledge would be 1 QALY, and the quality adjusted life expectancy would become 76+1=77 years. However, the partial EVPI of both m and d equals 0 (equation 5 or 6). This is because the optimal decision would be medication even if the uncertainty about either m or d – but not both – were resolved favorably. This means that obtaining a more precise estimate of only m or only d has no expected benefit.

The reduction in EVPI methods (equations 8 and 9) both result in partial EVPI's of 1 for m and d. These methods incorrectly recommend obtaining more information on m or d only. Notice also that the correct partial EVPIs (both equal to 0 QALY) don't add up to the total EVPI (equal to 1 QALY). However, $p\text{EVPI}(\theta_I)=0$ and $\text{E}_{\theta_I}\, p\text{EVPI}(\theta_C|\theta_I)=1$ do add up to the total EVPI, conforming to Figure 1 and equation 10. Using beta distributions for the uncertain probabilities instead of our artificial dichotomized distributions would result in similarly biased results for the reduction in EVPI approach.

DISCUSSION

In this manuscript we have explained and demonstrated that the generally recommended method to estimate partial EVPI is incorrect: it estimates partial EVPI as the reduction in expected opportunity loss instead of the increase in expected value. We have demonstrated the correct method to estimate partial EVPI, using a two-level Monte Carlo simulation. A computationally efficient one-level Monte Carlo simulation is mathematically equivalent if the outcome is a linear function of each parameter not of interest, or a multilinear function of each parameter not of interest, and the parameters not of interest are uncorrelated. Because many models are Markov models, the one-level method will typically result in a biased estimate of the partial EVPI. A numerical algorithm would be useful to determine whether the one-level method results in an acceptable approximation of the partial EVPI – when these conditions are not met. We are unaware of such algorithms in the literature, but hope that numerical mathematicians will develop them in coming years.

Some authors have suggested to empirically ascertain which one-level approach is better to estimate partial EVPI.[154] Based on this manuscript, we can conclude that there is no need to do such experiments. With analytic proof, we showed that methods to estimate partial EVPI as reduction in EVPI are conceptually and mathematically wrong. It has been suggested to use a one-level Monte Carlo simulation – when it fails to hold – as a rapid screening of all parameters.[151] The direction and magnitude of the bias, however, are uncertain.

Computation time of the two-level Monte Carlo simulation can be limited by selecting the appropriate number of runs for the inner and outer loop of the Monte Carlo simulation. Few runs in the inner loop results in a biased estimate of the partial EVPI, few runs in the outer loop in a lack of precision. Tappenden et al. built a practical method for calculating the necessary number of runs for the inner loop.[161] Confidence or credible intervals for any number of runs for the inner loop can be calculated once the simulation has been performed for a small number of runs. Nevertheless, computation time may still be impractical. Therefore, meta-models have been used to approximate partial EVPIs in complex decision models. A linear regression meta-model or a Gaussian process meta-model is a simplified version of the original model, which results in a substantial reduction of computation time. Applications of these meta-models and a tutorial of Gaussian processes have been published in the literature or are available on the internet.[161-163]

Estimating total EVPI and partial EVPI are only initial steps in guiding future research. Since perfect information is impossible, the next step assesses the expected value of sample information (EVSI).[157] Regarding EVSI, the same conclusion holds, that we should estimate the expected benefit of information, not the reduction in EVPI with information.

As the only method with a theoretically sound basis, we foresee an important role for value of information analysis in guiding future research by setting priorities amongst clinical research proposals, identifying key parameters, suggesting study designs, and estimating optimal sample sizes. Recently, Claxton et al. demonstrated the feasibility of value of information analysis to guide the research priority setting of the National Health Service in the United Kingdom. In a few months time they synthesized available evidence, built probabilistic models, and estimated the value of information to guide future research for several decision problems in health care.[31, 164] Synthesizing evidence and building probabilistic models is what many investigators have done. Value of information analysis is a logical and relatively simple additional step: the mathematics are straightforward, including only maximization and expectation. However, the concept needs time to gain familiarity and intuitive understanding.

7

Competing imaging tests for patients with chest pain: a value of information analysis to optimize study design

Bas Groot Koerkamp
Majanka H. Heijenbrok-Kal
Tessa S.S. Genders
Karen M. Kuntz
Kirsten E. Fleischmann
M.G. Myriam Hunink

In preparation.

Presented at the Annual Meeting of the Society for Medical Decision Making in 2004.

8

Value of information analysis of economic randomized controlled trails: the treatment of intermittent claudication

Bas Groot Koerkamp
Sandra Spronk
Theo Stijnen
M.G. Myriam Hunink

Value in Health. 2009: In press.

ABSTRACT

Objective: To design the optimal study comparing endovascular revascularization and supervised exercise training for patients with intermittent claudication, and 2) To demonstrate value of information (VOI) analysis of patient-level data from an economic randomized controlled trial to guide future research.

Methods: We applied a net benefit framework to patient-level data on costs and quality-of-life of a previous randomized controlled trial. VOI analyses were performed using Monte Carlo simulation. We estimated the total expected value of perfect information (total EVPI), the total expected value of sample information (total EVSI), the partial expected value of perfect information (partial EVPI), and the partial expected value of sample information (partial EVSI). These VOI analyses identified the key parameters and the optimal sample size of future study designs. Sensitivity analyses were performed to explore the robustness of our assumptions about the population to benefit, the willingness-to-pay threshold and the study costs. The VOI analyses are demonstrated in statistical software (R) and a spreadsheet (Excel) allowing other investigators to apply VOI analysis to their patient-level data.

Results: The optimal study design for the treatment of intermittent claudication involves a randomized controlled trial collecting data on the quality-adjusted life expectancy and additional admission costs for 525 patients per treatment arm. The optimal sample size remained between 400 and 600 patients for a willingness-to-pay threshold between €30,000 and €100,000 /QALY, for even extreme assumptions about the study costs, and for a range of 3 to 7 years that future patients will benefit from the results of the proposed study.

Conclusions: 1) The optimal study for patients with intermittent claudication collects data on 2 key parameters for 525 patients per trial arm, and 2) we have shown that value of information analysis provides an explicit framework to determine the optimal sample size and identify key parameters for the design of future clinical trials.

INTRODUCTION

The adoption of new medical interventions depends nowadays on evidence of cost-effectiveness in addition to evidence of effectiveness. Consequently, economic data are increasingly collected alongside clinical trials. We performed an economic trial to compare endovascular revascularization and supervised exercise training for patients with intermittent claudication.[174] Considerable uncertainty about the optimal cost-effective medical intervention remained after analysis of the trial. Given this uncertainty, policy makers should address two separate decisions: which intervention should be reimbursed, and is more research – for example, a larger economic trial – justified?[14] More quantitative research could be justified, because a decision based on a trial with a finite sample size can be wrong: that is, the intervention that is identified as optimal may not be the actual optimal intervention. A future study could justify a change in current care which may result in an improvement in quality-adjusted life expectancy of future patients and a decrease in health care costs. However, the actual benefit of a future study is uncertain. Before embarking on an expensive study, funding agencies like to know the expected study cost. Money spent on such a study cannot be spent otherwise, for example, to fund another study, or reimburse a new treatment. The cost of a study is usually specified in a grant proposal. The expected benefit of a study typically receives little formal consideration. The challenge of deciding whether more research is justified is to make the cost-benefit tradeoff of future clinical research prior to performing it. More research is justified only if the expected benefit exceeds the cost of a proposed study.

Value of Information (VOI) analysis provides a framework to guide the cost-benefit tradeoff of future cost-effectiveness research prior to performing it. VOI analysis estimates the expected benefit of a future study using available evidence (e.g., a previous randomized controlled trial) about a decision. VOI analysis can guide the design of a study that maximizes the difference between the expected benefit for future patients and the expected cost of the study. This study is characterized by its design (e.g., randomized or observational), the subset of sampled parameters (e.g., quality-of-life only, or a selection of cost parameters), the sample size, and the associated study costs. Claxton et al. have demonstrated the feasibility of VOI analysis to guide the research priority setting of the National Health Service in the United Kingdom.[31]

Most published VOI analyses involve decision models.[28] Economic trials, however, are attractive for VOI analysis because of their high internal validity. VOI analyses can be performed in addition to conventional analyses of economic trials.[6] They offer a sound alternative to significance testing when deciding if more research is needed. Moreover, VOI analysis provides a framework for sample size calculation. Based on patient-level data from

an economic trial VOI analysis can determine the optimal sample size of a future trial. The same methods can be used when patient-level data from a previous trial are not available, using elicited estimates with uncertainty intervals of the outcomes. Ideally, a VOI analysis is performed before and after a clinical trial.

The first objective of this article is to design the optimal study comparing endovascular revascularization and supervised exercise training for patients with intermittent claudication. The second objective is to demonstrate VOI analysis of patient-level data from an economic randomized controlled trial to guide future research. In the following section we briefly discuss the clinical problem, study design, and results of a previous trial of patients with intermittent claudication. Next, we explain the concepts and demonstrate the methods of the different VOI analyses, focusing on the application of VOI analyses to patient-level data from economic trials. The analyses are explained using mathematical notation (conform Ades)[32] and step-by-step algorithms specifically for VOI analysis of patient-level data of economic trials. We used Monte Carlo simulation to obtain the VOI estimates. In appendix 6 and 7 we present detailed instructions to perform VOI analyses on patient-level data using a spreadsheet such as Excel and statistical software such as R.[175, 176]

TREATMENT OF INTERMITTENT CLAUDICATION

Randomized controlled trial

Intermittent claudication is the mildest form of peripheral arterial disease. Patients suffer from a limited walking distance due to inadequate circulation of the legs. The treatment goal for intermittent claudication is to improve health-related quality-of-life. The general consensus is to treat these patients initially with exercise training.[177] Endovascular revascularization seems an attractive alternative with the advantage of immediate clinical success.[178] However, the drawbacks of endovascular revascularization include procedure related morbidity and mortality as well as increased costs.[179, 180]

Between September 2002 and September 2005, 150 patients with intermittent claudication were randomly allocated to endovascular revascularization or supervised exercise training.[174] During 12 months of follow-up all medical and non-medical costs (11 cost parameters) were assessed from the societal perspective and effects were measured with the EuroQol-5D questionnaire. We transformed the EuroQol-5D values into utilities using the Dutch scoring algorithm.[181] The improvement in quality-adjusted life years (QALYs) accumulated during the 12-month follow-up period was then used as effect measure in the cost-effectiveness analysis. We refer to the original article for more details on the study

design, analyses, and results.[174] The original article presented results with adjustment for age and gender. Here we used unadjusted data for the value of information analyses. The improvement in quality-adjusted life years (QALYs) was higher in the revascularization group than in the exercise group (mean difference 0.08; 95% CI 0.04, 0.12). The total mean cumulative cost per patient was also higher in the revascularization group than in the exercise group (mean difference €4254; 95% CI €1648, €7734).

Cost-effectiveness analysis

We adopted the net benefit approach to cost-effectiveness analysis.5 A decision between two strategies based on both costs and effect can only be made if a trade-off is made between cost and effect by putting a monetary value on health. We used a societal willingness-to-pay threshold (WTP) of 80,000 euro's per QALY, as has recently been recommended by a Dutch governmental institute.182 Costs and effect (in this case, quality-adjusted life expectancy) are combined into a single outcome called net (monetary) benefit:

$$\text{net benefit} = \text{WTP}*\text{effect} - \text{costs}$$

The net benefit is expressed in euro. The mean net benefit is denoted by B. The difference in mean net benefit between the two interventions is the incremental mean net benefit (IB). The uncertainty interval for the IB was estimated parametrically and non-parametrically.[5] The parametric intervals assume a normal distribution of the IB, justified by the central limit theorem. To check this assumption we also performed non-parametrical bootstrapping using 1 million bootstraps.[24]

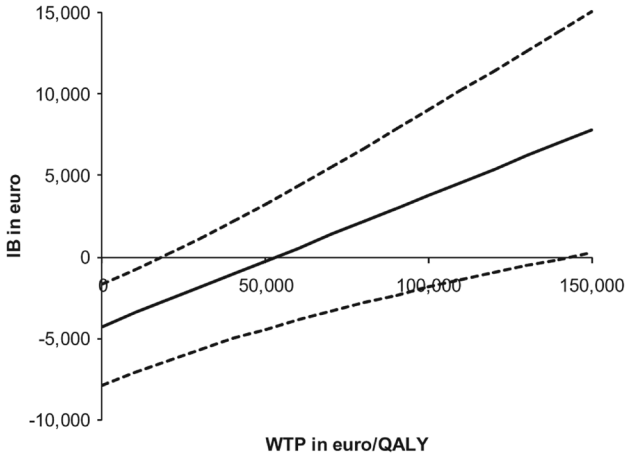

Figure 1 The incremental net benefit (IB) for the revascularization strategy in euro with 95% uncertainty boundaries based on 1 million bootstraps of patientlevel net benefits, across a range of values for the willingness-to-pay threshold (WTP) in euro/QALY.

The revascularization group had a higher net benefit (€4486 versus €2316 per patient), but the difference in net benefit was not significant (mean difference €2170; 95% CI €-2818, €6685). Based on these results we can conclude that revascularization seems cost-effective, but considerable uncertainty remains. Figure 1 presents the incremental net benefit (IB) in euro with 95% uncertainty interval across a range of values for the WTP in euro/QALY. The bootstrapping results showed that assuming normal distributions was justified.

TOTAL EVPI – ELIMINATING UNCERTAINTY

VOI analysis starts with estimating the total expected value of perfect information (total EVPI). It is the expected benefit per patient of a study with an infinite sample size, resulting in perfect information about all (total) uncertain cost and effect parameters. Such a study would eliminate uncertainty about the net benefit of each intervention, but is of course hypothetical. However, the total EVPI provides a ceiling level for the expected cost of a future study: studies with a finite sample size or studies that consider a subset of parameters all have a smaller expected benefit. Therefore, if the total EVPI does not exceed the fixed cost of research, more research is not justified. More research is potentially justified, if the total EVPI does exceed the fixed cost of research.

Total EVPI – equations and algorithm

The net benefit $B(a,\theta)$ of intervention a is a function of θ, where θ stands for the set of all unknown distributional cost and effect parameters involved. If all uncertainty about the parameters would be eliminated the actual net benefit of each intervention would be known. The cost of not knowing the actual net benefit of each intervention is the cost of uncertainty or the opportunity loss. It is defined as the difference between the maximum actual net benefit, and the actual net benefit of the supposedly optimal intervention (a*):

$$\text{opportunity loss} = \max_a B\,(a, \theta_{\text{actual}}) - B\,(a^*, \theta_{\text{actual}})$$

We cannot calculate the opportunity loss, because we don't know the actual parameter values θ_{actual}. However, the *expected* opportunity loss is the expectation over the distribution of the parameters of each intervention. The total EVPI equals the expected opportunity loss:

$$\text{total EVPI} = E_\theta\,[\max_a B(a, \theta) - B\,(a^*, \theta)]$$

To calculate the total EVPI, the parameters of each intervention must be characterized by a probability distribution based on the data of the initial trial. We can also use the probability

distribution of the mean net benefit $B(a,\theta)$ of each intervention. Although the net benefit may not be normally distributed in the population, the uncertainty about the *mean* net benefit typically is (central limit theorem). As the distribution of the mean net benefit we take $\mathrm{N}\ (\mu_0;\ \sigma_{\mathrm{pop}} / \sqrt{n_0})$, for each intervention; μ_0 is the estimated mean net benefit in the initial study, for each intervention; σ_{pop} is the estimated standard deviation in the initial study, for each intervention; and n_0 is the sample size in the initial study, for each intervention. For simplicity σ_{pop} is treated as a known parameter, based on the initial study [32] b_j is a random value of the distribution of the mean net benefit, for intervention a. The algorithm for the estimation of total EVPI involves the following steps:

1. raw a value b_j for the net benefit from $\mathrm{N}\ (\mu_0;\ \sigma_{\mathrm{pop}} / \sqrt{n_0})$, for each intervention a
2. calculate the opportunity loss: $\max_a b_j^a - b_j^{a^*}$
3. repeat step 1 and 2 N times
4. the total EVPI is estimated by averaging over the opportunity losses at step 2

The standard error of the mean opportunity loss ($\sigma_{\mathrm{opp\ los}} / \sqrt{N}$) reflects how precisely the total EVPI was estimated. The process of drawing a random value of each distribution (an iteration) is sometimes referred to as Monte Carlo simulation.

The example – total EVPI

Table 1 presents the results of the Monte Carlo simulation for 10 iterations. Each iteration consists of a random value of the distribution of the mean net benefit of each intervention. For each iteration the opportunity loss is calculated. For example, exercise was the optimal intervention in the first iteration, and revascularization was the optimal intervention based

Table 1 – total EVPI (in euro): 10 iterations of Monte Carlo simulation

Iteration	Revascularization	Exercise	opportunity loss
1	3348	4372	1024
2	7997	3198	0
3	3129	3311	182
4	2267	2991	724
5	-466	711	1177
6	3716	2276	0
7	3179	2220	0
8	5204	962	0
9	3543	2435	0
10	7679	4489	0
Mean			311

EVPI: expected value of perfect information

on the initial trial results. Therefore, the opportunity loss of the first iteration equals €4372 – €3348 = €1024. Based on these ten iterations only, the total EVPI is estimated by the mean opportunity loss of €311.

A total EVPI of €249 per patient was found with 10 million simulations in R. This means that after eliminating uncertainty we can expect an improvement in net monetary benefit of €249 per patient. Endovascular revascularization was the optimal intervention given the results of the initial trial, with an expected net benefit of 4486 euro per patient. With perfect information the expected net benefit of the optimal intervention (which could be endovascular revascularization or exercise training) is 4486 + 249 = €4735 per patient. The total EVPI per patient should be extrapolated to the entire population that will benefit from the study results, to allow for comparison with the fixed study cost.

Population EVPI and study cost

The expected benefit of a study should include the benefit of all future patients from some predetermined perspective: single hospital, health insurance agency, country, or worldwide. The number of years (T) that future patients are expected to benefit from the results of a proposed study is difficult to determine. It depends on improvement in technology and future evidence. A sensitivity analysis can illustrate the importance of this uncertainty. The expected benefit to future patients is discounted by a discount rate of typically 3% per year: that is, each year further ahead a smaller benefit is assigned on behalf of these patients.[157] The population EVPI equals:

$$\text{population EVPI} = \text{total EVPI} * \sum_{t=1}^{T} \frac{annualpopulation}{(1+discountrate)^t}$$

The study costs are typically estimated as fixed cost (e.g., salary of a PhD-student) and variable cost per patient in the study [15].

The example – population EVPI and study cost

Because the initial trial was funded by a national governmental agency, we used the national perspective for the annual population to benefit. This annual population was estimated at 10,000 patients for The Netherlands. We assumed patients would benefit from the results for 5 years and discounted these benefits at 3% per year. We found a discounted population to benefit of about 46,000 patients and a population EVPI of 11 million euro. We estimated the fixed cost of an additional clinical study at 200,000 euro, based on the cost of our previous study. Because the EVPI for the population exceeds the expected costs of an additional study, it is potentially justified to perform some sort of additional study.

TOTAL EVSI – REDUCING UNCERTAINTY

The total expected value of sample information (total EVSI) is an estimate of the expected benefit of studies with a finite sample size, collecting information on all cost and effect parameters. Instead of eliminating uncertainty, uncertainty about the mean net benefit of each intervention is only reduced. With increasing sample size, the total EVSI will reach a ceiling which equals the total EVPI, representing an infinite sample size.

Total EVSI – equations and algorithm

A proposed study provides data about all cost and effect parameters of n patients for each intervention. The study will improve the mean estimates of the parameter values, and consequently of the net benefit of each intervention. Once we have observed the actual study data D, the expected benefit of treatment a^* is $E_{\theta/D_{\text{actual}}} B\,(a^*, \theta)$. The best treatment then has expected benefit $\max_a E_{\theta/D_{\text{actual}}} B\,(a, \theta)$. Thus the current cost of uncertainty about the actual study data D is the opportunity loss:

$$\text{opportunity loss} = \max_a E_{\theta/D_{\text{actual}}} B\,(a, \theta) - E_{\theta/D_{\text{actual}}} B\,(a^*, \theta)\ .$$

Because B is linear in θ this simplifies to:

$$\text{opportunity loss} = \max_a B\,(a, E(\theta|D_{\text{actual}})) - B\,(a^*, E(\theta|D_{\text{actual}})).$$

We cannot calculate the opportunity loss, because we don't know the actual data before performing the study. However, the expected opportunity loss is the expectation over all possible values of the new data. The total EVSI equals the expected opportunity loss:

$$\text{total EVSI} = E_D \left[\max_a B\,(a, E(\theta|D)) - B\,(a^*, E(\theta|D))\right]$$

Again the analysis is simplified using the probability distribution of the mean net benefit $\text{N}\,(\mu_0; \sigma_{\text{pop}} / \sqrt{n_0})$, of each intervention a in the initial study. The data D of the proposed study is characterized by the sample size n_1, the sample mean net benefit μ_1, and the standard error $\sigma_{\text{pop}} / \sqrt{n_1}$ for each intervention.[32] μ_1 is unknown, but can be sampled from the distribution of the actual net benefit $\text{N}\,(\mu_{\text{actual}}; \sigma_{\text{pop}} / \sqrt{n_1}$ for each intervention. μ_{actual} is also unknown, but can be sampled from the distribution of the mean net benefit of the initial trial $\text{N}\,(\mu_0;$ $\sigma_{\text{pop}} / \sqrt{n_0}$ for each intervention. The algorithm for the estimation of the total EVSI involves the following steps:

1. choose a sample size n_1 per intervention of the proposed study
2. draw a value $\mu_{\text{actual},\,j}$ from $\text{N}\,(\mu_0; \sigma_{\text{pop}} / \sqrt{n_0})$, for each intervention

3. draw a value $\mu_{1,j}$ from N ($\mu_{\text{actual},j}$; $\sigma_{\text{pop}} / \sqrt{n_1}$) , for each intervention
4. calculate the posterior mean net benefit for each intervention a: $b_j^a = \frac{\mu_0 \cdot n_0 + \mu_{1,j} \cdot n_1}{n_0 + n_1}$
5. calculate the opportunity loss: $\max_a b_j^a - b_j^{a^*}$
6. repeat step 1 to 5 N times
7. the total EVSI is estimated by averaging over the opportunity losses at step 5

ENBS - the optimal sample size

The expected net benefit of sampling (ENBS) is defined as the difference between the total EVSI and the study cost. The optimal sample size is reached when the ENBS reaches a maximum. At this maximum the additional benefit of one more patient in the study equals the additional study costs of one more patient in the study.

In addition to the fixed cost and variable cost per patient, the cost of clinical trials should also include the forgone net benefit of each patient that is randomized to an intervention that is supposedly suboptimal. This amounts to the sample size of the inferior arm of the trial times the difference in net benefit based on the previous trial.

The example – total EVSI, study costs, and ENBS

For a clinical trial collecting data on all parameters (*total* EVSI) we estimated a fixed cost of 200,000 euro and a variable cost of €1000 per patient, both based on our previous (identical) study. The difference in net benefit between the revascularization and the exercise group was €2170 . The total cost in euro was therefore:

$$\text{total study costs} = 200{,}000 + 1{,}000*2*n^1 + 2170*n^1$$

Note that n^1 is the sample size per study arm and not the total sample size of the study. We used the same population to benefit of about 46,000 patients (see section on total EVPI). Figure 2 presents the study cost, the total EVSI, and the ENBS as a function of the sample size n^1 per study arm. A maximum ENBS of €7.3 million is reached for a sample size of about 475 patients per study arm. The study cost of this study would be 2.2 million, of which 1.2 million is accounted for by the forgone net benefit of 475 patients assigned to the supposedly suboptimal intervention (supervised exercise training).

The ENBS of the proposed study is not the actual benefit of the study – which we will only learn after analyzing the results of the study – but the *expected* benefit, prior to performing the study. The actual benefit to an individual patient is zero if the new study does not lead to a change in current care, because the patient's outcomes remain unchanged. The actual benefit to an individual patient is nonzero if the new study demonstrates that our initial decision was suboptimal and current care is changed accordingly. However, the

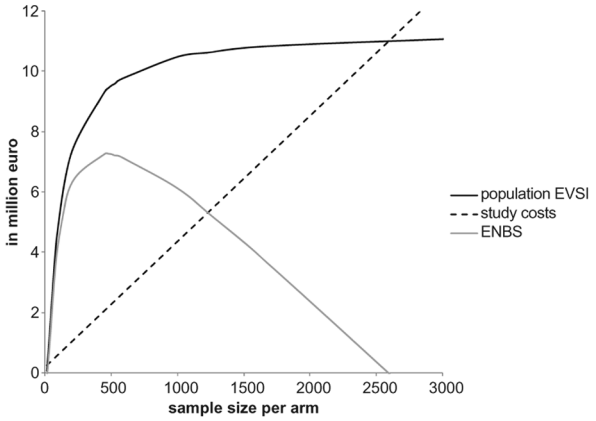

Figure 2 Population EVSI, study costs and ENBS in million Euro for different sample sizes per study arm.The proposed study collects data on all parameters that were considered in the initial trial. The optimal sample size is about 500 patients per study arm with an ENBS of €7.2 million, and total study costs of €2.3 million.Of the total study costs, 1.1 million is accounted for by the forgone net benefit of 500 patients assigned to the supposedly suboptimal intervention (supervised exercise training).

proposed study will almost always improve the precision of the outcomes and reduce the cost of uncertainty.

PARTIAL EVPI – ELIMINATING UNCERTAINTY OF SOME PARAMETERS

The partial EVPI estimates the expected benefit of eliminating uncertainty for individual parameters or subsets of parameters. Typically it is not justified to consider all parameters in a future study: a few key parameters are the source of most decision uncertainty. Eliminating uncertainty about the other parameters has no additional expected benefit or an expected benefit that does not justify the additional study costs.

Partial EVPI – equations and algorithm

For the estimation of partial EVPI, the parameters are divided into two groups: the parameters-of-interest θ^I are considered in a future study, the parameters-not-of-interest θ^C are not considered in a future study. If all uncertainty about the parameters-of-interest θ^I for each intervention would be eliminated, the actual values of these parameters would be known. The current cost of uncertainty of not knowing the actual values of the parameters-of-interest θ^I_{actual} of each intervention is called the opportunity loss. It is defined as the difference between the maximum expected net benefit given θ^I_{actual}, and the expected net benefit given θ^I_{actual} of the supposedly optimal intervention (a*). Because $B(a,\theta^C,\theta^I)$ is linear in θ^C and θ^I we get:

$$\text{opportunity loss} = \max_a B\,(a, \theta^I_{\text{actual}}, \mathrm{E}_{\theta^C}(\theta^C|\theta^I_{\text{actual}})) - B\,(a^*, \theta^I_{\text{actual}}, \mathrm{E}_{\theta^C}(\theta^C|\theta^I_{\text{actual}}))$$

We cannot calculate the opportunity loss, because we don't know the actual values of the parameters-of-interest θ^I. However, the expected opportunity loss is the expectation over all possible values of the parameters-of-interest θ^I of each intervention. The partial EVPI equals the expected opportunity loss:

$$\text{partial EVPI} = E_{\theta^I}\left[\max_a B\,(a, \theta^I, \mathrm{E}_{\theta^C}(\theta^C|\theta^I)) - B\,(a^*, \theta^I, \mathrm{E}_{\theta^C}(\theta^C|\theta^I))\right]$$

Again, the analysis can be simplified by avoiding the distributions of individual parameters. Instead, we created distributions for the mean net benefit of the parameters-of-interest $\mathrm{N}\,(\mu_0^I, \sigma_{\text{pop}}^I / \sqrt{n_0})$ and the mean net benefit of the parameters-not-of-interest $\mathrm{N}\,(\mu_0^I, \sigma_{\text{pop}}^I / \sqrt{n_0})$, with correlation ρ , for each intervention. The partial EVPI is estimated in the following algorithm. The conditional mean net benefit of the parameters-not-of-interest θ^C is calculated in step 2 of the algorithm using the general equation for the conditional mean value of a bivariate normal distribution. [183]

1. draw a value b_j^I from $\mathrm{N}\,(\mu_0^I, \sigma_{\text{pop}}^I / \sqrt{n_0})$, for each intervention
2. calculate the conditional mean net benefit of θ^C, for each intervention:

$$b_j^C = \mu_0^C + \frac{\sigma_{pop}^C * \rho * \left[b_j^I - \mu_0^I\right]}{\sigma_{pop}^I}$$

3. calculate the mean net benefit for each intervention: $b_j = b_j^I + b_j^C$
4. calculate the opportunity loss: $\max_a b_j^a - b_j^{a^*}$
5. repeat step 1 to 4 N times
6. the partial EVPI is estimated by averaging over the opportunity losses at step 4

Subsets of parameters

In theory we could estimate the partial EVPI for each subset of parameters. To evaluate whether more research regarding each subset is potentially justified, the partial EVPI is compared with the subset-specific fixed study costs. The number of required analyses, however, would explode for even a small number of parameters. In practice, investigators typically first estimate the partial EVPI for each individual parameter. Unfortunately, the partial EVPI of individual parameters doesn't simply sum up to the partial EVPI of a subset of parameters. Even a subset of parameters with individual partial EVPIs of zero, together may have a nonzero partial EVPI. The subset of parameters with a nonzero EVPI or substantial individual partial EVPI seems a reasonable subset to consider for partial EVPI estimation. Other relevant subsets are found by changing this subset. If the partial EVPI of this subset is close to the total EVPI, we can remove parameters with a small individual partial EVPI or a substantial associated increase in study costs (e.g., the quality-of-life parameter). If the partial EVPI is much smaller than the total EVPI, we can add parameters with small

Table 2 – Partial EVPIs

Individual parameters	partial EVPI in euro
QALYs	104
additional admission costs	65
additional treatment costs	6
additional imaging costs	0.007
Subsets of parameters	
all 12 parameters = total EVPI	249
all 4 with nonzero individual pEVPI	248
all 3 with individual pEVPI>1	248
all 2 with individual pEVPI>10	244
all 8 with zero individual pEVPI	0
all 11 cost parameters	69

EVPI, expected value of perfect information; QALYs, quality-adjusted life-years.

additional study costs (e.g., a cost parameter that can be collected from administrative records).

The example – partial EVPI

Table 2 presents the partial EVPI of each individual parameter with 10 million simulations in R. Of the 12 parameters, only 4 had a nonzero individual partial EVPI. Together, these 4 parameters had a partial EVPI of €248 per patient; almost the same as the total EVPI of 249. The partial EVPI of the 3 parameters with the highest individual EVPI was also €248 per patient. The two parameters with the highest individual partial EVPI together had a partial EVPI of 244. For the subsets with 2 and 3 parameters we identified the optimal sample size in the next section. The partial EVPI of the subset of 8 parameters with an individual partial EVPI of zero, was still zero. Obtaining more information on these parameters is not justified, even for very small additional study costs.

Five cost parameters had a similar expected value, ranging from 437 euro to 742 euro per patient: the cost of material, personnel, overhead, treatment, and productivity loss. The partial EVPI of treatment cost was 6 euro per patient. The partial EVPI of the other four parameters *together* was zero. These results demonstrate that simply selecting the parameters with the highest expected value, or the parameters that differed most between treatments, is not a good alternative for partial VOI analysis.

PARTIAL EVSI – REDUCING UNCERTAINTY OF SOME PARAMETERS

The partial expected value of sample information (partial EVSI) is an estimate of the expected benefit of studies with a finite sample size, collecting information on a subset of parameters. With increasing sample size, the partial EVSI will reach a ceiling: the partial EVPI, representing an infinite sample size.

Partial EVSI – equations and algorithm

A future study provides data about the parameters-of-interest θ^I of n patients for each intervention. The study will improve the mean estimates of the parameters-of-interest θ^I, and consequently of the net benefit of each intervention. The current cost of not knowing the actual study data D_{actual} is called the opportunity loss. It is calculated as the difference between the maximum expected net benefit given D_{actual}, and the expected net benefit given D_{actual}, of the supposedly optimal intervention (a*):

$$\text{opportunity loss} = \max_a B\,(a, \mathrm{E}_{\theta^I}(\theta^I|D_{\text{actual}}), \mathrm{E}_{\theta^C}(\theta^C|\theta^I|D_{\text{actual}}))) - B\,(a^*, \mathrm{E}_{\theta^I}(\theta^I|D_{\text{actual}}), \mathrm{E}_{\theta^C}(\theta^C|\theta^I|D_{\text{actual}})))$$

We cannot calculate the opportunity loss, because we don't know the actual data before performing the study. However, the expected opportunity loss is the expectation over all possible values of the new data. The partial EVSI equals the expected opportunity loss:

$$\text{partial EVSI} = E_D\,[\max_a B\,(a, \mathrm{E}_{\theta^I}(\theta^I|D), \mathrm{E}_{\theta^C}(\theta^C|\theta^I|D))) - B\,(a^*, \mathrm{E}_{\theta^I}(\theta^I|D), \mathrm{E}_{\theta^C}(\theta^C|\theta^I|D)))]$$

Analogous to estimating the partial EVPI, we used the observed distributions for the mean net benefit of the parameters-of-interest $\mathrm{N}\,(\mu^I_0, \sigma^I_{\text{pop}} / \sqrt{n_0})$ and the mean net benefit of the parameters-not-of-interest $\mathrm{N}\,(\mu^C_0, \sigma^C_{\text{pop}} / \sqrt{n_0})$, with correlation ρ, for each intervention. The data of the proposed study is characterized by the mean net benefit of the parameters-of-interest μ^I_1 and the standard error $\sigma^I_{\text{pop}} / \sqrt{n_1}$ for each intervention.[32] μ^I_1 is unknown, but can be sampled from the distribution of the actual net benefit of the parameters-of-interest $\mathrm{N}\,(\mu^I_{\text{actual}}; \sigma^I_{\text{pop}} / \sqrt{n_1})$, for each intervention. μ^I_{actual} is also unknown, but can be sampled from the distribution of the mean net benefit of the parameters-of-interest observed in the initial study $\mathrm{N}\,(\mu^I_0, \sigma^I_{\text{pop}} / \sqrt{n_0})$, for each intervention. The partial EVSI is estimated in the following algorithm.

1. choose a sample size n_1 per intervention of the proposed study
2. draw a value $\mu^I_{\text{actual, j}}$ from $\mathrm{N}\,(\mu^I_0, \sigma^I_{\text{pop}} / \sqrt{n_0})$, for each intervention

3. draw a value $\mu^I_{1,j}$ from N ($\mu^I_{\text{actual, j}}$; $\sigma^I_{\text{pop}} / \sqrt{n_1}$), for each intervention
4. calculate the posterior mean net benefit of θ^I, for each intervention:

$$b^I_j = \frac{\mu^I_0 * n_0 + \mu^I_{1,j} * n_1}{n_0 + n_1}$$

5. calculate the conditional mean net benefit of θ^C, for each intervention:[183]

$$b^C_j = \mu^C_0 + \frac{\sigma^C_{pop} * \rho * \left[b^I_j - \mu^I_0 \right]}{\sigma^I_{pop}}$$

6. calculate the mean net benefit for each intervention: $b_j = b^I_j + b^C_j$
7. calculate the opportunity loss: $\max_a b^a_j - b^{a^*}_j$
8. repeat step 1 to 7 N times
9. the partial EVSI is estimated by averaging over the opportunity losses at step 7

Although the proposed data collection sampled only the parameters-of-interest, the net benefit of each intervention is estimated using all available data of all parameters (step 6). The net benefit of the parameters-of-interest is estimated using data from both the initial study and the proposed data collection (step 4). The net benefit of the parameters-not-of-interest is estimated using data from the initial study only, unless a correlation (ρ) exists with the parameters-of-interest (step 5).

The example - partial EVSI, study costs, and ENBS

We estimated the partial EVSI, for various sample sizes, for two subsets of parameters that were identified with the analysis of partial EVPI. These subsets include respectively three and two parameters with the highest individual partial EVPI. We estimated that collecting information on the quality-adjusted life expectancy represents 50% of the fixed and variable study costs: 100,000 euro and 500 euro per patient. Also collecting hospital admission costs implies searching administrative records: we estimated an increase in study costs of 10,000 euro plus 50 euro per patient. Also collecting the additional treatment costs requires more resources: we estimated an increase in study costs of 20,000 euro plus 100 euro per patient. We used the same population to benefit of about 46,000 patients (see section on total EVPI).

For the subset of two parameters the optimal sample size was 525 patients per study arm and the ENBS was 7.6 million euro. For the subset of three parameters we found an optimal sample size of 500 patients per study arm and an ENBS of also 7.6 million euro. While the expected benefit of the studies is equal, we prefer the former study because of lower study costs. This optimal study design obtains data on the QALY and additional admission costs for 525 patients per study arm. The required study costs are 690,000 euro. Assign-

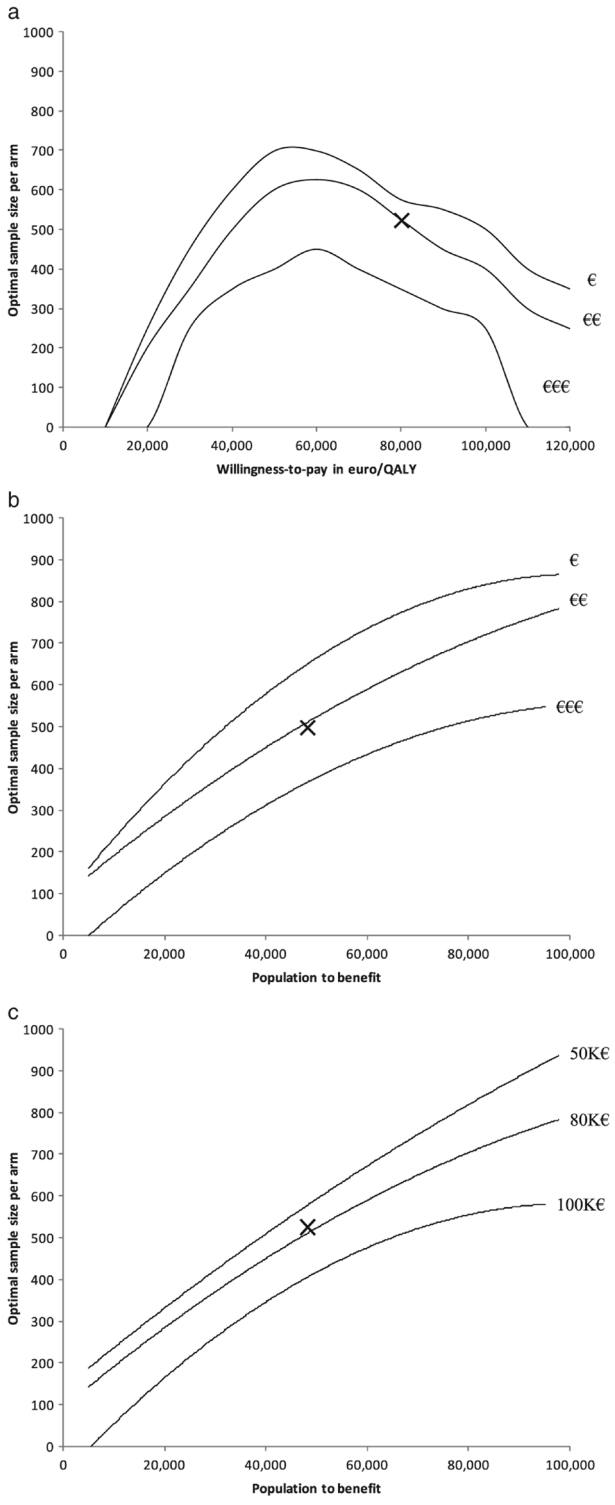

Figure 3 (a) Sensitivity analysis for the willingness-to-pay threshold and the study costs.The proposed study collects information on the quality-adjusted life expectancy (QALE) and the additional admission costs. The population to benefit is about 46,000 patients.The study costs estimates are: fixed 10,000 and variable 100 euro (€); fixed 110,000 and variable 550 (€€: base-case); and fixed 500,000 and variable 2000 (€€€).The X presents the base-case willingness-topay threshold and study costs. (b) Sensitivity analysis for the population to benefit and the study costs. The proposed study collects information on the QALE and the additional admission costs. The willingness-to-pay threshold is 80,000 euro/quality-adjusted life-years (QALY).The study costs estimates are: fixed 10,000 and variable 100 euro (€); fixed 110,000 and variable 550 (€€: base-case); and fixed 500,000 and variable 2000 (€€€). The X presents the base-case population to benefit of about 46,000 patients. (c) Sensitivity analysis for the population to benefit and the willingness-to-pay threshold. The proposed study collects information on the QALE and the additional admission costs.The study costs are €110,000 for fixed costs and €550 for the costs per patient.The X presents the base-case population to benefit.

ing 525 patients to the supposedly suboptimal arm has an associated cost of 1.1 million euro. To estimate study costs we assumed a randomized controlled trial as optimal design to reassure validity of the results. Figure 3 presents several sensitivity analyses for the WTP threshold, the population to benefit, and the study costs.

ALTERNATIVE METHODS

We performed VOI analysis assuming a normal distribution of the mean net benefit of each intervention, based on the central limit theorem. The total EVPI can also be estimated using bootstrapping techniques, which do not rely on the normality assumption.[24] With 1 million bootstraps in R we found a total EVPI of €264 per patient: very close to the €249 that we found assuming a normal distribution of the mean net benefits. Unfortunately, it is not obvious how bootstrapping should be implemented to estimate partial value of information and sample information. Other nonparametric methods for VOI analysis are being developed.[32]

We assessed the total EVPI using unadjusted estimates for the mean net benefit of each intervention. As another alternative, we used regression analysis in the net benefit framework to obtain estimates of the mean net benefit adjusted for potential imbalances of baseline characteristics between the treatment groups. The resulting total EVPI was €119 per patient. The adjusted outcomes probably underestimate the VOI; the unadjusted outcomes may overestimate the VOI. The regression analysis ambiguously decomposes the total variance into components attributable to patient heterogeneity and uncertainty. Both uncertainty and estimates of the VOI will decrease, when more variation is attributed to heterogeneity.

Both Claxton and Willan estimated the total EVPI and total EVSI using closed form (analytical) methods that do not rely on simulation.[15, 157] Using closed form solutions, Claxton also considered the VOI of trial designs assigning unequal sample sizes to each intervention.[172] Although closed form solutions for estimating the partial EVPI and partial EVSI are not available in the literature, it should be possible to derive such solutions. The advantages of closed form solutions are exact outcomes and negligible calculation time. Simulation, however, also has several advantages as compared to closed form solutions. The simulations that we applied to our example can be easily modified to accommodate more than two comparators, prior distributions other than normal, and nonlinear functions of the parameters. Moreover, simulation has educational appeal: by following the steps of the simulation, the reader understands how it works.

DISCUSSION

With value of information (VOI) analysis we found that more research is justified regarding the choice between endovascular revascularization and supervised exercise training for patients with intermittent claudication. The optimal study design for a future study involves a randomized controlled trial collecting data on the quality-adjusted life expectancy and additional admission costs for 525 patients per treatment arm. The outcome of this trial could justify a change in current care. As a result, future patients may benefit from an increase in quality-of-life, or cost savings may allow the health care system to reimburse other (unrelated) beneficial interventions. Although we don't know the actual benefit of the proposed study, the VOI analysis estimated an expected net benefit of sampling (ENBS) of 7.6 million euro for The Netherlands, using the results of a previous trial. The study costs of 690,000 euro were accounted for in this estimate. No other study design had a higher ENBS. Sensitivity analyses demonstrated that the optimal sample size was fairly stable. It remained between 400 and 600 patients for a willingness-to-pay threshold between €30,000 and €100,000 /QALY, for extreme assumptions about the study costs, and for a range of 3 to 7 years that future patients will benefit from the results of the proposed study. These results were used for a recent grant proposal. The €7.6 million of the proposed study can be compared with the ENBS of other (unrelated) study proposals. The ENBS can guide a funding agency to set priorities if the research budget is limited: study proposals with a higher ENBS should be reimbursed first. However, funding is justified for any study proposal with an ENBS exceeding zero.

An important assumption of VOI analysis is that health care costs and research costs are ultimately paid for from the same resources, which is largely true for The Netherlands. The appropriate perspective of the VOI analysis regarding the population to benefit from the results of a study proposal is not obvious. Because the initial trial of our example was funded by a national governmental agency, we used the national perspective. For a European Union or world-wide perspective the population to benefit would increase at least 20-fold, resulting in an optimal sample size of several thousand patients. The period that patients will benefit from the proposed data collection is also uncertain because of uncertainty about future technological improvements and evidence from future studies. These ambiguities, however, are not drawbacks of VOI analysis in itself, but inherent to setting research priorities.

We demonstrated VOI analysis using patient-level data from a single clinical trial. We may have overestimated uncertainty and the value of information, because we didn't consider all available evidence pertaining to the decision. A decision model can bring together evidence from various sources and also extrapolate costs and effects beyond the follow-up

period in the initial trial. However, decision models also have several drawbacks. Many assumptions are required when building decision models, to the extent that for the same research question different models report diverging results.[45] As a consequence the validity of models is often challenged. Moreover, building models is very time-consuming and requires expertise on both the clinical subject matter as well as the methodology of decision modelling. On the other hand, this investment of time may be worthwhile, because VOI analyses of such a decision model could conclude that more research is *not* justified. Finally, most decision models are nonlinear and require extremely computer-intensive nested VOI analyses.32 VOI analysis of patient-level data can avoid these nested analyses because the net benefit is a linear function of the cost and effect parameters. Clinical trials remain attractive because of their high internal validity and timeliness.[6] For many clinical decisions a trial constitutes the best available evidence, not only to decide what medical intervention should be adopted, but also to address the question whether more research is needed. VOI analysis of trials is aimed at replacing the use of significance testing to determine whether more research is justified; it is *not* aimed at replacing decision models to guide further research. Advocates for economic trials will emphasize their internal validity, while advocates for models stress their consideration of all available evidence. Whether a decision model is required to guide future research will depend on the methods and results of the initial study, as well as the importance of evidence from other sources.

A limitation of trials as compared to models is that clinical trials rarely have a lifetime follow-up of costs and effects. This is not only a drawback of VOI analysis of trials, but inherent to any analysis of a trial. Sometimes a model is used to extrapolate the trial data beyond the follow-up period.[37] The follow-up of our initial trial was only 12 months. However, both interventions – endovascular revascularization and exercise training – give only temporary relief of symptoms. Improvement of quality-of-life in our study was more immediate after revascularization, but at 12 months no difference was detected in quality-of-life. Like most interventions in surgery, the costs are largely incurred upfront. The base case analysis of a model based on our trial data would assume that no difference in costs and effects is anticipated beyond the follow-up of the trial.

We recommend a randomized controlled trial as the optimal design for the future study. Alternatively, an observational study could collect data on quality-of-life and admission costs. The drawback of a nonrandomized design is that it is more difficult to avoid that differences in (known and unknown) baseline patient characteristics (i.e., confounders) are responsible for differences between the outcomes of interventions. Moreover, a randomized design is not necessarily associated with additional study costs.

Few applications of VOI analysis to guide the design of clinical trials have been published. In 2005, a guidance document for designing and analyzing cost-effectiveness analyses conducted as part of clinical trials did not mention VOI analysis.[6] However, the philosophy of a formal cost-benefit trade-off prior to experimental studies is not new. VOI analysis was introduced by Grundy29 in the late fifties and developed by Raiffa and Schlaifer.[4] Howard noted in 1966 that: "Placing a value on the reduction of uncertainty is the first step in experimental design, for only when we know what it is worth to reduce uncertainty do we have a basis for allocating our resources in experimentation designed to reduce the uncertainty." In 1989 Detsky evaluated the effect of design choices made in the planning stages of a clinical trial on the costs and benefits derived from conducting the trial.[184, 185] Claxton introduced VOI analysis to the clinical audience of the Lancet.[14] Moreover, he demonstrated VOI analysis to estimate the optimal sample size of a trial using hypothetical data.[157] More recently, Willan applied VOI analysis to estimate the optimal sample size of a trial using patient-level data from a previous trial.[186] Both studies used closed form solutions that are not available in the literature for partial value of information analyses. Further research could find out closed form solutions for some partial value of information analyses.

Although the VOI analyses are relatively complex and technical, the fundamental ideas of VOI analysis is rather straightforward and may appear familiar to clinicians. A clear analogy exists between the Bayesian framework for VOI analysis and Bayesian diagnostic reasoning. A future trial can change the probability that an intervention is optimal, just as a diagnostic test can change the probability that a patient has a certain disease. Both a trial and a diagnostic test are costly and require a cost-benefit trade-off. The probability and the consequences of implementing a suboptimal intervention or misdiagnosing a patient determine whether more research or a diagnostic test is justified.

Clinical trials often show no statistically significant difference between the treatments compared. We demonstrated that VOI analysis allows for a formal comparison of the expected benefit and the cost of a proposed study, before concluding whether or not more research is justified. Two erroneous conclusions are common about the need for more research when no significant difference is found. Some authors conclude that the decision has been settled: the interventions are assumed to be equivalent and more research is not needed. It has long been demonstrated, however, that a difference is often not found because the study was underpowered to detect even a large difference.[11] In a famous quote, Altman warned that: "Absence of evidence is not evidence of absence".[12] Other authors conclude the exact opposite - more research is needed - when no significant difference is found. They seem to reason that there must be a difference and because it wasn't found with the current study, another (larger) study is needed. Phillips pointed out that the conclusion

"more research is needed" requires some assessment of the expected benefit for future patients that would come from more research in comparison to the cost of research (e.g., VOI analysis). He observed that studies in health care typically conclude that "more research is needed" without such an assessment.[13]

If more research is justified, the investigators should decide on a sample size. We demonstrated VOI analysis as an explicit framework to perform sample size calculations. Classical sample size calculations are based on arbitrary values for the minimal clinically significant difference in treatment effect, a type I error (typically α=0.05) and a type II error (typically β=1-power=0.2). In practice the equation is often back-solved after substituting in a sample size that primarily reflects feasibility and cost. VOI analysis considers the actual harm to future patients of making a type I error (i.e., rejecting the null-hypothesis when it is true) and a type II error (not rejecting the null-hypothesis when it is false). The optimal sample size is estimated by considering the marginal cost and benefits of sampling patients. Moreover, VOI analysis can conclude that more research is not justified.[187] We recommend performing VOI analyses before and after an economic trial.

9

Value of information analysis used to determine the necessity of additional research: MR imaging in acute knee trauma as an example

Bas Groot Koerkamp
Jeroen J. Nikken
Edwin H. Oei
Theo Stijnen
Abida Z. Ginai
M.G. Myriam Hunink

Radiology. 2008 Feb;246(2):420-5

Presented at the Annual Meeting of the Radiological Society of North America in 2004.

ABSTRACT

Objective: To guide future outcomes research regarding the use of magnetic resonance (MR) imaging for patients with acute knee trauma in the emergency room setting, with use of prospective data from a randomized clinical trial and value of information analysis.

Methods: A total of 189 patients (123 male, 66 female; mean age 33.4 years) were randomly assigned to undergo radiography alone (n=93) or radiography followed by MR imaging (n=96). Institutional review board approval and informed consent (parental consent for minors) were obtained. During 6-months of follow-up, data on quality of life and 39 cost parameters were collected. Value of Information analysis was used to estimate the expected benefit of future research to eliminate the decision uncertainty that remained after trial completion. In addition, the parameters that were responsible for most of the decision uncertainty were identified, the expected benefits of various study designs were evaluated, and the optimal sample size was estimated.

Results: Only three parameters were responsible for most of the decision uncertainty: number of quality-adjusted life years, cost of overnight hospital stay, and friction costs. A study in which data on these three parameters are gathered would have an optimal sample size of 3500 patients per arm and would result in a societal benefit of €5.6 million, or 70 quality-adjusted life years.

Conclusion: The optimal study design for use of MR imaging to evaluate acute knee trauma involves a trial in which there are 3500 patients per trial arm, and data on the number of quality-adjusted life years, cost of overnight hospital stay, and friction costs are collected.

INTRODUCTION

Patients who present to the emergency room with acute knee trauma may benefit from immediate magnetic resonance (MR) imaging. Moreover, the initial cost of the MR imaging may be offset by a reduction in the subsequent medical and societal costs because the patient may be able to return to work sooner than if MR imaging had not been performed. In a randomized controlled trial, we assessed the additional value of MR imaging in patients with acute knee trauma.[188] After a 6-month follow-up of 189 patients we found no statistically significant difference in costs and a small transient significant difference in patient outcome. These results made us question whether a second larger trial, in which we would attempt to identify a difference in costs and a durable difference in patient outcome that may have remained undetected in our initial study, would be justified.

If study results show no significant differences between the primary outcomes, the researchers invariably conclude that more clinical research is needed to reduce decision uncertainty.[13] Uncertainty could result in the adoption of suboptimal medical interventions, which could harm patients or result in inefficient allocation of limited health care funds. More research – for example, another clinical trial – is expected to decrease this uncertainty and benefit patients, save money, or both. However, research is costly and money spent on one research project cannot be spent on another. Furthermore, while further research is being performed, a potentially cost-effective intervention is withheld from patients. These problems raise the question whether more research regarding an uncertain decision is a good value for the money. More clinical research is justified only if the expected benefit of this research exceeds the expected costs. Value of Information analysis is a method that expands on cost-effectiveness analysis and can be used to determine if more research is justified regarding a medical decision. This method is used to estimate the expected benefit of a proposed study given the currently available evidence. In addition, Value of Information analysis can be used to identify the optimal study design and sample size. Use of Value of Information analysis has been embraced and recommended by the National Institute for Clinical Excellence (NICE) in the United Kingdom as a framework for setting research priorities in health care.[14]

The purpose of this study was to help guide future outcomes research regarding the use of MR imaging in patients with acute knee trauma in an emergency room setting, with use of prospective data from a randomized clinical trial and value of information analysis.

METHODS

Randomized controlled trial

In a previously published diagnostic randomized controlled trial[188], we (EHO, JJN) enrolled 189 consecutive patients (123 male [mean age, 31.9 years; age range, 12.6-74.6 years], 66 female [mean age, 36.1 years; age range, 16.6-72.9]) with a mean age of 33.4 years between August 1999 and May 2001. These patients had recent knee trauma and were referred to the radiology department for conventional radiography.[188] Patients were randomly assigned to undergo conventional radiography alone (n=93) or radiography followed by a short dedicated MR imaging examination (n=96). Institutional review board approval and informed consent (parental consent for minors) were obtained for the randomized study.

During 6 months of follow-up, quality of life was measured four times with a valuative device (EuroQol). All relevant societal costs were recorded during the follow-up period. These costs included medical and non-medical costs. Medical costs consisted of costs of diagnostic procedures and treatment both inside and outside the hospital and were estimated with 36 resource-use parameters for each strategy. Nonmedical costs were estimated with 3 parameters: patient travel costs, patient time cost, and friction costs. The latter was an estimate of societal production losses. In total, 40 parameters (39 cost parameters and quality of life) were sampled for each strategy. Mean values and 95% confidence intervals were calculated for costs and effects of both strategies. (See the original article[188] for more details on study design and analysis.)

Cost-effectiveness analysis

To perform cost-effectiveness analysis, we (BG, JJN) transformed the EuroQol values into utility values.[189] For each patient an author (BG) calculated the overall number of quality-adjusted life years during the study period as the effect parameter.[38]

A choice of one of the two strategies that is based on both cost and effect can be made only if a trade-off between cost and effect is made by placing a monetary value on health. We used a societal willingness-to-pay of €80,000 per quality-adjusted life year, as recently recommended by a Dutch governmental institute.[182] Subsequently we (BG, TS) combined cost and effect into one outcome, which we termed net (monetary) benefit.5 The net benefit was calculated by multiplying effect by willingness to pay and subtracting cost. The strategy with the maximum net benefit is the strategy that is preferred.

Value of Information analysis

We (BG, TS) applied Value of Information analysis as described in the literature.[27, 32, 81, 105, 153] First, we estimated the total expected value of perfect information (EVPI) per patient. This is the value of collecting data about the effect parameter and all cost parameters in an infinitely large study. In other words, it is the value of removing all uncertainty related to the decision problem.

Subsequently, we estimated the expected value for the entire patient population that can potentially benefit from more research (population EVPI). To calculate the population EVPI we (BG, MGMH) estimated the effective lifetime of the technology to be 10 years. Benefits to future patients were discounted at a rate of 3% per year.[38] For the Netherlands perspective we estimated the annual population that could potentially benefit from the results of a future study to be 20,000 patients. We performed an additional analysis for the European Union perspective. By extrapolating the annual population of 20,000 patients to the European Union, we determined that an annual population of 561,000 patients could benefit from more research. If the population EVPI is substantial, it is of interest to estimate the EVPI for individual parameters, or sets of parameters. We termed this the partial EVPI. Partial EVPI is used to identify the parameters that have the highest informational value regarding decision uncertainty.

If the total EVPI is substantial, we are interested to learn the expected benefit of reducing uncertainty by obtaining information from a future study with a finite sample size. This is referred to as the total expected value of sample information (EVSI). Moreover, we can assess the expected benefit of future studies with a finite sample size that collect information on a limited set of parameters. This is referred to as the partial EVPI. An author (BG) estimated the partial EVSI for several sets of parameters to assess various study designs. Comparing the EVSI with the costs of performing research enables us to determine whether an additional study is justified given the cost. Subtracting the cost of research from the EVSI results in the expected net benefit of sampling (ENBS). The optimal sample size is determined by calculating the sample size that maximizes the ENBS.

For a future multicenter trial with a 3-year duration (assuming the study requires 2 full-time equivalent junior researchers and a senior researcher with 0.4 full-time-equivalent responsibility) we (BG, MGMH) assumed a fixed cost of €500,000 and a variable cost of €500 per patient if all parameters in the initial trial were to be measured. If data on only 3 parameters (friction cost, overnight hospital stay, and quality-adjusted life years) were to be collected, we assumed a fixed cost of €250,000 and a variable cost of €250 per patient. These cost estimates were based on our current expenses for similar studies.

Technical details

To allow for Value of Information analysis, we (BG, TS) represented the joint uncertainty about the mean values of all parameters by using a multivariable normal distribution, with variances equal to the estimated squared standard errors of the mean and correlations between the different parameters calculated from the dataset. The central limit theorem justified the normality assumption.

An author (BG) performed 10 million simulations for each analysis, resulting in standard errors in our estimates of about 1%. Nested simulations were not required to estimate partial EVPI and EVSI because the relation between the net benefit and each parameter was linear and because the multivariable normal distribution allowed us to calculate conditional mean values.[32, 183] To estimate EVSI, an author (BG) derived posterior normal distributions for the sampled parameters by using Bayesian updating of the prior normal distributions.[190] We (BG, TS) assumed that the standard deviations and correlations between parameters in future research would be the same as those in the initial trial. All analyses were performed with R software (version 1.7.1; R Foundation for Statistical Computing, Vienna, Austria) that can be accessed at http://www.r-project.org.[175]

RESULTS

Cost-effectiveness analysis

The combination of radiography and MR imaging was more effective (that is, it resulted in more quality-adjusted life years during the study period), was less costly, and had a higher net benefit than radiography alone (Table). The differences in effect and net benefit were statistically significant but the difference in costs was not.

Table Mean cost, effect, and net (monetary) benefit.

	Strategy 1: Radiography only	**Strategy 2: Additional MRI**	**Difference: strategy 2 - 1**
mean cost Euro (95% CI)	2231 (1561;2901)	1815 (1277;2353)	-416 (-1275;443)
mean effect QALY (95% CI)	0.35 (0.34;0.36)	0.38 (0.36;0.39)	0.03 (0.01;0.04)
mean net benefit* Euro (95% CI)	25,848(24,696;27,000)	28,315(27,399;29,231)	2467 (515;4419)

CI: confidence interval
QALY: quality-adjusted life years
*Mean net benefit = effect x willingness-to-pay – cost, with willingness-to-pay = 80,000 Euro/QALY

Population EVPI: overall importance of uncertainty

We found a total EVPI of €2.1 per patient. The resulting population EVPI was €365,000 for the Netherlands and €10.2 million for the European Union. These values have an equivalent benefit of 5 quality-adjusted life years for the Netherlands and 128 quality-adjusted life years for the European Union. An effective lifetime of the technology of 5 years instead of 10 years would reduce these benefits by approximately half.

Partial EVPI: important parameters

In the initial study, only two of the 40 collected data parameters had a non-zero partial EVPI. The partial EVPI of the quality-adjusted life year was €1.0 per patient; the partial EVPI of the friction cost was €0.01 per patient. These two parameters had a synergistic effect, and together they had a partial EVPI of €1.9 euro per patient. This synergistic effect was augmented by considering the cost of an overnight hospital stay. A future study in which data on the number of quality-adjusted life years, the cost of an overnight hospital stay, and the friction costs per patient would be gathered, would have a partial EVPI of €2.0 per patient, which would be nearly equal to the total EVPI.

Total EVSI and ENBS: optimal sample size

We first considered the optimal sample size for a future study collecting data on all parameters. The population EVSI for a study from the perspective of the Netherlands did not exceed the study costs for any sample size. This meant that more research was not justified. The population EVSI for a study from the perspective of the European Union increased as

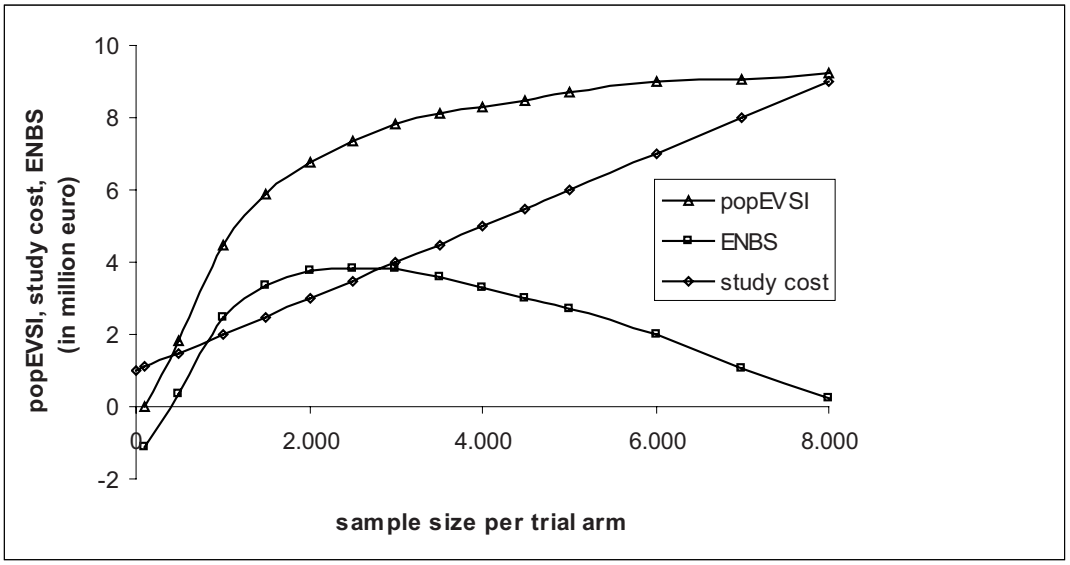

Figure 1 Population expected value of sample information for all parameters (popEVSI), study costs, and expected net benefit of sampling (ENBS). The ENBS curve reaches a maximum (equal to 3.8 million euro) at a sample size of 2500 patients per trial arm.

the sample size increased until a plateau was reached; this plateau was equivalent to the population EVPI (Figure 1). The study costs increased linearly as the sample size increased. The maximum ENBS of €3.8 million was reached at a sample size of 2500 patients per trial arm. One should note, however, that there was a decreasing marginal gain in the ENBS: A study with 1500 patients per trial arm was expected to reach a net benefit of €3.4 million.

Partial EVSI and ENBS: optimal sample size

To calculate partial EVSI, we considered a study that would enable us to collect data on only the three most important parameters: quality-adjusted life years, cost of an overnight hospital stay, and friction costs per patient. The population (partial) EVSI for a study from the perspective of the Netherlands did not exceed the study cost for any sample size. The same study from the perspective of the European Union had a slightly lower population EVSI and substantially lower study costs compared with the study in which data were collected for all parameters (Figure 2). Therefore, the ENBS was higher when only the three most important parameters were sampled. The optimal sample size was 3500 patients per trial arm, resulting in an ENBS of €5.6 million or 70 quality-adjusted life years. Again, because of the decreasing marginal gain, we found an ENBS of €5.1 million for a study with 2000 patients.

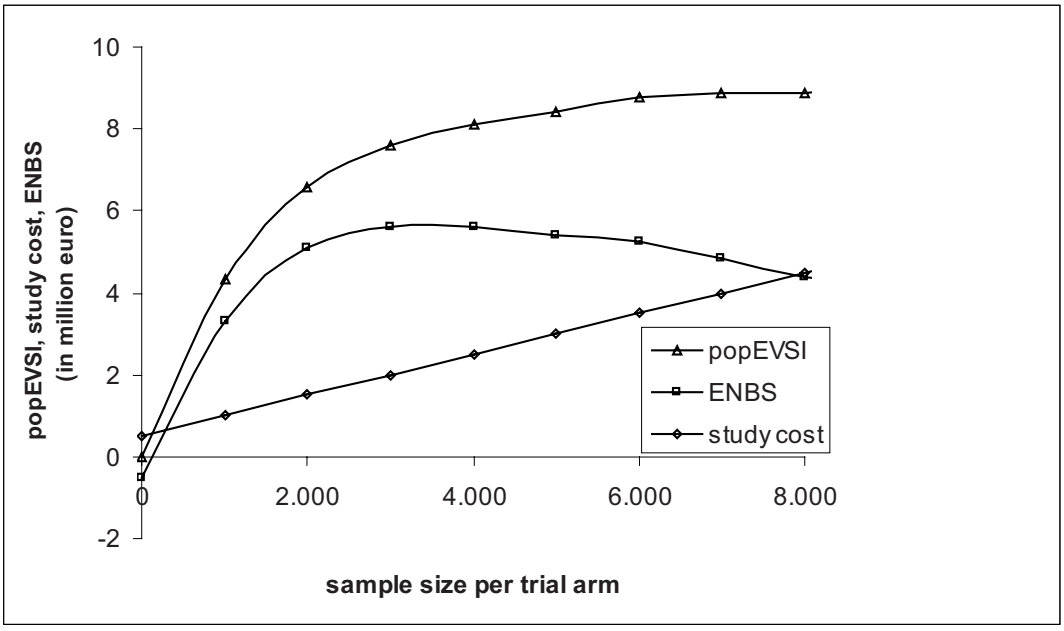

Figure 2 Population expected value of sample information (popEVSI) for three parameters (friction cost, overnight hospital stay, and quality-adjusted life time), study costs, and expected net benefit of sampling (ENBS). The ENBS curve reaches a maximum (of 5.6 million euro) at a sample size of 3500 patients in each trial arm.

DISCUSSION

We found a population EVPI of €10.2 million for a study from the perspective of the European Union regarding the decision of whether to add an MR imaging to the current initial work-up of patients with acute knee trauma. This indicates that if we would eliminate all uncertainty regarding this decision, we could expect a societal financial benefit of €10.2 million, which is equivalent to a societal health benefit of 128 quality-adjusted life years. Only three parameters were responsible for the decision uncertainty: the number of quality-adjusted life years, cost of an overnight hospital stay, and friction costs per patient. Collecting data on the other 37 cost parameters has almost no additional benefit. A study in which data on these three parameters were gathered would have an optimal sample size of 3500 patients per trial arm, and it would be expected to result in a societal benefit of €5.6 million or 70 quality-adjusted life years. From the perspective of The Netherlands, however, more research was not justified.

It is important to realize that the calculated societal benefit of €5.6 million euro or 70 quality-adjusted life years is an *expected* net benefit: It is a probability weighted average over all possible outcomes of a future study. We learn the actual benefit of a study only after we have initiated the study, collected the data, and analyzed the actual results. Often there is no actual benefit. The findings of the future study are more likely than not to confirm that the strategy that we believe to be optimal is indeed optimal. If a future study results in a change in the optimal strategy, the benefit may be a reduction in cost, an increase in quality-adjusted life years, or a combination of these benefits.

The expected societal benefit of €5.6 million should be compared with the expected societal benefit of other unrelated proposed clinical research projects to set research priorities. The decision uncertainty regarding imaging for patients with acute knee trauma turns out to be relatively small in comparison with other clinical problems that have been addressed in Value of Information analyses.[31] More research regarding MR imaging in patients with acute knee trauma is justified, but other clinical studies are expected to result in up to a 100-fold higher benefit. The prioritization of research studies will ultimately depend on the portfolio of potential studies submitted to a funding agency, their corresponding expected value of information, and the available research budget.

Our results were sensitive to the uncertain magnitude of the population expected to benefit from reducing decision uncertainty. This is, by definition, true for all Value of Information analyses. However, it is not a drawback but rather inherent to the assessment of the expected benefit of future research. Both the annual population that can potentially benefit from the research and the effective lifetime of the technology are influential and uncertain.

The annual population that can benefit from research depends on the perspective of the policy maker: For example, is it the perspective of the hospital, the state, the country, or something even larger? When research proposals are compared, they need to be judged and compared from one perspective. Furthermore, the effective lifetime of the technology is uncertain because we do not know when improvements in diagnosis and treatment will come about and how they will influence decision uncertainty.

A few limitations pertain specifically to our study. We applied Dutch medical and nonmedical costs to the entire European Union. This may have biased our results. Moreover, we assumed that medical care in the entire European Union was similar to that in the Netherlands. In addition, in our analyses we assumed that the intervention has no effect on the costs and effects after the 6-months follow-up period. Although these limitations may affect the precise figure that results from the calculations, they are unlikely to have a substantial effect on our conclusions.

Our results imply that a Dutch funding agency seeking to maximize the future health in the Netherlands should not fund more research regarding the value of MR imaging in patients with acute knee trauma. A European agency, however, should consider funding a multicenter trial with about 3500 patients per trial arm in which the friction costs, the cost of an overnight hospital stay, and the quality-adjusted life years are measured. However, other unrelated research proposals with a higher expected benefit should receive priority.

Value of Information analysis is an analytic tool that can help researchers decide whether more clinical research regarding an uncertain medical decision is justified. It is a logical initial step when clinical research is considered regarding a medical decision or when the results of a randomized clinical trial are inconclusive. Value of information analysis can be used to determine whether the decision uncertainty justifies the cost of the research. Decision uncertainty can be modeled by using all available evidence in the literature.[38] Alternatively, the results of a previous clinical study or meta-analysis can be used for value of information analysis, as in the current study. Ideally, the analysis should involve all competing strategies to include all decision uncertainty. If the expected benefit of more research is substantial, Value of Information analysis can be used to identify key parameters, evaluate various study designs, and estimate optimal sample sizes. Claxton et al. have demonstrated the feasibility of Value of Information analysis to help guide the research priority setting of the National Health Service in the United Kingdom.[31, 191] Although the mathematics are relatively simple, we realize that it takes time to understand the concepts of Value of Information analysis. To our knowledge, this is the only method with a theoretically sound basis; therefore, we foresee an important role for Value of Information analysis in guiding future research. The budget for clinical research is limited and money should be spent where the

expected benefits are greatest. Moreover, more clinical research is justified only if the expected benefit of more research exceeds the expected research costs.

10

Summary

Making decisions about the care of individual patients is fundamental to health care. In daily practice, most medical decisions are based on experience and judgment. Medical decision making was developed because of concerns about human judgment, practice variation, and the proliferation of diagnostic and treatment options. The aim of medical decision making is to perform a complete formal assessment of every aspect that is relevant for a decision. This assessment includes patient preferences, rare events, and health care costs, all of which are typically ignored within the evidence-based medicine framework. Considerable uncertainty about the optimal intervention typically remains after evaluating all available evidence. However, a decision between the interventions has to be made, regardless of the extent of uncertainty. Evaluation of uncertainty is particularly relevant to determine whether more quantitative research is justified. A future study could reduce decision uncertainty which is expected to benefit future patients, reduce health care costs, or both. Value of information (VOI) analysis was introduced to estimate the expected benefit of a future study. Moreover, VOI analysis can guide the design of a study by identifying key parameters as well as the optimal sample size. This dissertation concerns the analysis and presentation of uncertainty in medical decision making with a focus on VOI analysis.

In Chapter 2 we introduced cost-effectiveness analysis to a clinical audience of surgeons. New interventions in surgery typically provide small additional health benefits and are more expensive in comparison to current care. If a health care system cannot afford all beneficial new interventions, cost-effectiveness analysis can help set priorities. The application of cost-effectiveness analysis faces several challenges, including credibility, generalizability, and ethical implications. Additional challenges more specific to surgery include the learning curve for new surgical interventions and the gradual improvement of surgical technology. Adherence to guidelines for cost-effectiveness analyses could address some challenges; other challenges simply reflect the difficulty of making decisions under uncertainty. Despite these challenges, priorities have to be set.

Chapter 3 is a tutorial in which we demonstrated Monte Carlo simulation and VOI analysis to analyze stochastic uncertainty, parameter uncertainty, and patient heterogeneity in a decision model. First-order Monte Carlo simulation allows modeling of the influence of patient history on subsequent events. Second-order Monte Carlo simulation evaluates the joint effect of uncertainty about all estimated parameter values in the model. VOI analyses explore whether more research is justified, identify key parameters, and assess the optimal sample size of a proposed study design. Increasing model complexity, often requiring microsimulation, is a major challenge for second-order Monte Carlo simulation and VOI analyses. We provided step-by-step algorithms for the nested analyses that are required in patient-level models and in nonlinear models for partial VOI analyses.

Chapter 4 concerned the combined analysis of uncertainty and patient heterogeneity in medical decision models. When more than one type of uncertainty and heterogeneity is analyzed, the correct algorithm to obtain the model outcomes of interest can be complicated. We distinguished eight model types, each dealing with a different combination of heterogeneity, parameter uncertainty, and stochastic uncertainty. Decision models are primarily built to inform policy makers about the expected outcome of competing strategies. We demonstrated that the expected outcome and the distribution of the individual outcome can always be obtained in a single Monte Carlo simulation (i.e., without using nested Monte Carlo simulations). Nested Monte Carlo simulations are inevitable if we are interested in the uncertainty about the expected outcome, reflecting lack of perfect knowledge, for a heterogeneous population or in microsimulation models. In addition, nested simulations are required to obtain the distribution of the expected outcome reflecting patient heterogeneity, both in microsimulation models and models with parameter uncertainty. These nested Monte Carlo simulations require intensive computer processing capabilities.

In Chapter 5 we clarified the limitations of acceptability curves for presenting uncertainty in cost-effectiveness analyses. Clinical journals increasingly illustrate uncertainty about the cost and effect of health care interventions using cost-effectiveness acceptability curves (CEACs). CEACs present the probability that each competing alternative is optimal for a range of values of the cost-effectiveness threshold. The limitations of CEACs arise because a CEAC is not sensitive to any change of the incremental joint distribution in the upper-left and lower-right quadrant of the cost-effectiveness plane, neither is it sensitive to radial shift of the incremental joint distribution in the upper-right and lower-left quadrants. As a result, CEACs are ambiguous to risk-averse policy makers, inhibit integration with risk-attitude, hamper synthesis with other evidence or opinions, and are unhelpful to assess the need for more research. Moreover, CEACs may mislead policy makers and can incorrectly suggest medical importance. Both for guiding immediate decisions and for prioritizing future research, these considerable drawbacks of CEACs should make us rethink their use in communicating uncertainty.

In Chapter 6 we evaluated two different approaches for estimating the partial expected value of perfect information (partial EVPI) to identify key parameters in cost-effectiveness analysis. A method that is generally recommended assumes that the EVPI for a parameter is estimated as the reduction in expected opportunity loss instead of the increase in expected value. With analytic proof, we showed that this method is incorrect and results in biased partial EVPIs and incorrect importance ranking of parameters. We demonstrated the correct method to estimate partial EVPI, using a two-level Monte Carlo simulation. A computationally efficient one-level Monte Carlo simulation is mathematically equivalent if the outcome is a multilinear function of each parameter not of interest and — additionally

— the parameters not of interest are uncorrelated. Because Markov models are nonlinear, they require two-level simulations to obtain unbiased partial EVPIs.

VOI analysis was applied in Chapter 7 to a decision model comparing four imaging tests to diagnose coronary heart disease in patients with chest pain. We found that the optimal imaging test for patients with chest pain is uncertain for most subgroups of patients. The main objective was to design the optimal future study to reduce uncertainty regarding the optimal imaging test for patients with chest pain. We compared the partial EVPIs of eight study designs: four observational studies for test characteristics and quality-of-life weights, a cost study, and three clinical trials measuring treatment effects. The partial EVPI was highest for an observational study evaluating the quality-of-life weights for varying severity of chest pain. By comparing the expected net benefit of sampling (ENBS) for various sample sizes, we determined that the optimal sample size for this study would be 1500 patients for each severity-level of chest pain.

In Chapter 8 we demonstrated VOI analysis using data from an economic clinical trial. Most published applications of VOI analysis use decision models or hypothetical trial data. We explained how VOI analysis using patient-level data on costs and effects is simplified using several justifiable assumptions. Step-by-step instructions in statistical software (R) and a spreadsheet (Excel) allows other investigators to apply VOI analysis to their patient-level data. We designed the optimal study comparing endovascular revascularization and supervised exercise training for patients with intermittent claudication, based on data from a recent economic clinical trial. The optimal study should maximize the difference between the expected benefit of the anticipated study results for future patients and the expected cost of the study. The proposed study would be a randomized controlled trial collecting data on the quality-adjusted life expectancy and additional admission costs for 525 patients per treatment arm. Sensitivity analyses showed that the optimal sample size was fairly robust: it remained between 400 and 600 patients for a willingness-to-pay threshold between €30,000 and €100,000 per QALY, for extreme assumptions about the study costs, and for a period in which future patients are expected to benefit from the results of the proposed study ranging from 3 to 7 years.

In Chapter 9 we applied VOI analysis to guide future outcomes research regarding the use of MR imaging for patients with acute knee trauma in the emergency room setting. An economic clinical trial randomizing 189 patients with acute knee trauma to radiography only or radiography followed by MR imaging had found no significant difference in cost or effect. VOI analysis was performed to determine if more research was justified. From the perspective of only the Netherlands, the country that had funded the initial study, more research was not justified. From the perspective of the European Union, we found a population EVPI

of €10.2 million. VOI analysis further demonstrated that out of 40 parameters in the initial trial, only three parameters were responsible for most of the decision uncertainty: quality-adjusted life time, cost of an overnight hospital stay, and friction costs. A study that gathers data on these three parameters would have an optimal sample size of 3500 patients per arm and would be expected to result in a societal benefit of €5.6 million, or 70 quality-adjusted life years.

Some recommendations can be made based on the research presented in this dissertation:

- Most cost-effectiveness analyses (whether decision models or clinical trials) can be improved by adhering to guidelines.
- The combined analysis of uncertainty and patient heterogeneity in decision models requires careful selection of the correct algorithms to obtain the outcomes of interest.
- Uncertainty in cost-effectiveness analysis is best presented by uncertainty intervals for the incremental net benefit (when two interventions are compared) or the total EVPI.
- Methods that estimate the reduction in the partial EVPI are incorrect and should not be used to identify key parameters.
- The analyses of a probabilistic decision model should culminate in VOI analysis to find the key parameters and optimal sample size of a future study.
- VOI analysis should be the first step in designing a clinical study, as it assesses the importance of decision uncertainty, the key parameters, and the optimal sample size.

Many topics on uncertainty in medical decision making remain unexplored, especially regarding VOI analysis. Future research could evaluate the feasibility of requiring a VOI analysis whenever a clinical study is considered for funding. In chapters 8 and 9, simulation methods were used for partial VOI analyses of clinical trial data. In theory, closed form solutions should exist for these analyses. Most published VOI analyses consider both costs and health effects, while clinical research is mainly focused on health outcomes, such as mortality. Future research could explore the use of VOI analysis to prioritize clinical research with outcomes such as mortality.

Samenvatting

Beslissingen nemen met betrekking tot de zorg voor individuele patiënten is een essentieel onderdeel van de gezondheidszorg. In de dagelijkse praktijk zijn de meeste beslissingen gebaseerd op de persoonlijke ervaring en beoordeling van de behandelend arts. Het vakgebied medische besliskunde is ontstaan om bezorgdheid over de feilbaarheid van deze beoordeling, verschillen in beleid tussen artsen en het toenemend aantal diagnostische en therapeutische mogelijkheden. Het doel van medische besliskunde is een formele evaluatie van alle aspecten die relevant zijn voor een bepaalde beslissing. In tegenstelling tot het vakgebied evidence-based medicine, worden ook de voorkeur van de patiënt, zeldzame complicaties en de kosten van de zorg expliciet meegewogen. Nadat al het beschikbare bewijs is meegewogen blijft het vaak onzeker welke interventie (diagnostische onderzoek of behandeling) het beste is voor een bepaalde patiënt. Echter, de arts moet een keuze maken, ongeacht hoeveel onzekerheid er is. De evaluatie van onzekerheid is met name van belang om te beoordelen of meer klinisch onderzoek gerechtvaardigd is. Door een toekomstige studie kan de onzekerheid afnemen, wat kan resulteren in betere uitkomsten voor toekomstige patiënten en/of een kostenreductie. Value of information (VOI) analyse is geïntroduceerd om de verwachte baten van een toekomstige studie te schatten. Bovendien kan VOI analyse de opzet van een studie begeleiden bij het selecteren van de belangrijkste parameters en het kiezen van het optimaal aantal patiënten in de studie. Dit proefschrift gaat over de analyse en presentatie van onzekerheid in de medische besliskunde met een focus op VOI analyse.

Hoofdstuk 2 is een introductie van kosteneffectiviteits analyse voor chirurgen. Nieuwe behandelingen in de chirurgie hebben vaak een kleine toegevoegde waarde voor de patiënt en zijn aanmerkelijk duurder dan de huidige zorg. Kosteneffectiviteits analyse kan helpen bij het stellen van prioriteiten als de gezondheidszorg zich niet iedere nieuwe interventie kan veroorloven. Het gebruik van kosteneffectiviteits analyse gaat gepaard met enkele uitdagingen, waaronder geloofwaardigheid, de toepasbaarheid op een specifieke patiënt, en ethische implicaties. Ook zijn er uitdagingen specifiek voor de chirurgie, zoals de leercurve voor nieuwe chirurgische behandelingen en de geleidelijke verbetering van chirurgische technologie. Het volgen van richtlijnen voor kosteneffectiviteits analyse lost sommige uitdagingen ten dele op; andere uitdagingen weerspiegelen dat het moeilijk is beslissingen te nemen in onzekerheid. Ondanks deze uitdagingen moeten keuzes worden gemaakt.

Hoofdstuk 3 is een tutorial waarin Monte Carlo simulatie en VOI analyse worden gedemonstreerd om stochastische onzekerheid, parameter onzekerheid en patiënten heterogeniteit te analyseren in een besliskundig model. Eerste-orde Monte Carlo simulatie maakt het mogelijk om de invloed te modelleren van de voorgeschiedenis van een patiënt op toekomstige kansen en uitkomsten. Tweede-orde Monte Carlo simulatie evalueert het gemeenschappelijk effect van de onzekerheid van alle geschatte parameter waarden in het

besliskundige model. VOI analyses evalueren of meer klinisch onderzoek gerechtvaardigd is en stellen vast wat de belangrijkste parameters zijn evenals de optimale studie grootte. De toename van de complexiteit van besliskundige modellen, waardoor vaak microsimulatie nodig is, is een grote uitdaging voor tweede-orde Monte Carlo simulatie en VOI analyses. Algoritmes worden gepresenteerd voor de geneste analyses in besliskundige modellen waarin individuele patiënten worden gesimuleerd en in niet-lineaire modellen voor de partiële VOI analyses.

Hoofdstuk 4 behandelt de gecombineerde analyse van onzekerheid en patiënten heterogeniteit in medische besliskundige modellen. Als meer dan één type onzekerheid of patienten heterogeniteit wordt geanalyseerd, kan het moeilijk zijn vast te stellen wat het juiste algoritme is om de correcte uitkomsten van een model te verkrijgen. Onderscheid wordt gemaakt tussen acht model types die ieder een verschillende combinatie van patiënten heterogeniteit, stochastische onzekerheid en parameter onzekerheid beschouwen. Besliskundige modellen worden met name gemaakt om de optimale interventie te kiezen op basis van de verwachte uitkomsten. We hebben laten zien dat de verwachte uitkomst en de distributie van de individuele uitkomst altijd verkregen kunnen worden met een enkele Monte Carlo simulatie (d.w.z. zonder een geneste simulatie). Geneste Monte Carlo simulaties zijn onvermijdelijk als we geïnteresseerd zijn in de onzekerheid (door gebrek aan bewijs) over de verwachte uitkomst, zowel als het gaat om beslissingen voor een heterogene populatie als om microsimulatie modellen. Ook is een geneste simulatie nodig als het gaat om de verdeling (door patiënten heterogeniteit) van de verwachte uitkomst, zowel in microsimulatie modellen als in modellen met parameter onzekerheid. Deze geneste Monte Carlo simulaties maken intensief gebruik van computer processor capaciteit.

Hoofdstuk 5 beschouwt de beperkingen van de kosteneffectiviteits acceptability curve (KEAC) voor het weergeven van onzekerheid in kosteneffectiviteits analyses. Klinische tijdschriften illustreren steeds vaker onzekerheid over de kosten en effecten van interventies in de gezondheidszorg met KEACs. Deze curves geven de kans weer dat een interventie kosteneffectief is voor bepaalde waarden van de kosteneffectiviteits drempel. De beperkingen van KEACs ontstaan doordat KEACs niet gevoelig zijn voor veranderingen van de incrementele gezamelijke verdeling van kosten en effecten in de kwadranten linksboven en rechtsonder van het kosteneffectiviteits vlak. Hierdoor zijn KEACs niet eenduidig voor besluitvormers met een risico aversie, maken zij het onmogelijk ander bewijs en opinies mee te nemen in de besluitvorming, en zijn ze ongeschikt om te beoordelen of meer klinisch onderzoek gerechtvaardigd is.

Bovendien kunnen KEACs beleidsmakers misleiden door ten onrechte een groot medisch belang te suggereren. Gezien deze beperkingen voor het communiceren van onzekerheid

moet het gebruik van KEACs heroverwogen worden, zowel voor het besluit om een nieuwe interventie te implementeren, als voor de beoordeling of meer klinisch onderzoek gerechtvaardigd is.

Hoofdstuk 6 evalueert twee verschillende methoden voor het schatten van de partiële expected value of perfect information (EVPI) om de belangrijkste parameters te vinden in een kosteneffectiviteits analyse. Een veel gebruikte methode veronderstelt dat de EVPI van een parameter berekend kan worden met de reductie in de EVPI in plaats van met de toename in expected value. We bewezen analytisch dat deze methode onjuist is en leidt tot foutieve waarden voor de partiële EVPI en een foutieve rangschikking van de parameters. De juiste methode voor het berekenen van de partiële EVPI werd gedemonstreerd met behulp van geneste Monte Carlo simulaties. Een snellere methode met behulp van een enkele Monte Carlo simulatie leidt tot dezelfde partiële EVPI mits de parameters die niet in de voorgestelde studie worden gemeten ongecorreleerd zijn en bovendien de uitkomst van het model een multilineaire functie is van deze parameters. In Markov modellen is een geneste simulatie onvermijdelijk omdat deze modellen niet-lineair zijn.

In hoofdstuk 7 is VOI analyse toegepast op een besliskundig model waarin vier beeldvormende onderzoeken, voor het diagnosticeren van coronair lijden bij patiënten met pijn op de borst, worden vergeleken. Het meest kosteneffectieve beeldvormende onderzoek voor patiënten met pijn op de borst bleek onzeker voor de meest subgroepen van patiënten. Het doel in dit hoofdstuk was om de optimale klinische studie op te zetten ter vermindering van onzekerheid met betrekking tot de keuze van een beeldvormend onderzoek voor patiënten met pijn op de borst. We vergeleken de partiële EVPI van acht studie opzetten: vier observationele studies naar de testkarakteristieken en de kwaliteit van leven, een kostenstudie en drie klinische trials voor de behandeling van coronair lijden. De partiële EVPI was het hoogst voor een observationele studie naar de kwaliteit van leven voor patiënten zonder pijn op de borst, met milde en met ernstige pijn op de borst. De expected net benefit of sampling werd berekend voor verschillende waarden van de studie grootte. De optimale studie grootte was 1500 patiënten per patiëntengroep.

Hoofdstuk 8 demonstreert VOI analyse met data van een economische klinische studie. De meeste gepubliceerde toepassingen van VOI analyse gebruiken besliskundige modellen of hypothetische klinische data. Met behulp van enkele veronderstellingen is het mogelijk om VOI analyse toe te passen op patiënten data met kosten en kwaliteit van leven uitkomsten. Met onze stap-voor-stap instructies in statistische software (R) en in een spreadsheet (Excel) kunnen anderen VOI analyse toepassen op hun eigen patiënten data. Op basis van patiënten data van een recente economische klinische studie hebben we de optimale studie opgezet om endovasculaire revascularisatie te vergelijken met gesuperviseerde looptraining

voor patiënten met intermitterende claudicatie klachten. De optimale studie maximaliseert het verschil tussen de verwachte baten van de studieresultaten voor toekomstige patiënten en de kosten van de studie. De optimale studie was een gerandomiseerde studie naar de twee belangrijkste parameters, kwaliteit van leven en de kosten van ziekenhuisopname, voor 525 patiënten per studie arm. Gevoeligheids analyses lieten zien dat de optimale studie grootte vrij stabiel was: de studie grootte bleef tussen de 400 en 600 patiënten voor een willingness-to-pay drempel tussen de €30.000 en de €100.000 voor een kwaliteit gecorrigeerd levensjaar, voor extreme waarden voor de kosten van de studie, en voor een periode waarin toekomstige patiënten profiteren van de studieresultaten van drie tot zeven jaar.

Hoofdstuk 9 is een toepassing van VOI analyse om toekomstig onderzoek op te zetten naar de waarde van MRI voor patiënten op de spoed-eisende hulp met acuut knieletsel. Een economische klinische studie randomiseerde 189 patiënten met acuut knieletsel tussen alleen een röntgenfoto en een röntgenfoto gevolgd door een MRI. Deze studie vond geen significant verschil in kosten en effecten. VOI analyse werd toegepast om te evalueren of meer klinisch onderzoek gerechtvaardigd was.Vanuit het perspectief van de gezondheidszorg in Nederland, van waaruit de eerdere klinische studie was gefinancierd, was meer klinisch onderzoek niet gerechtvaardigd. Vanuit het perspectief van de Europese Unie werd een population EVPI gevonden van €10,2 miljoen. VOI analyse toonde ook aan dat van de veertig parameters in de eerdere studie slechts drie parameters verantwoordelijk waren voor het overgrote deel van de onzekerheid: kwaliteit van leven, de kosten van ziekenhuisopname, en de frictiekosten. Een toekomstige studie naar deze drie parameters heeft een optimale studie grootte van 3500 patiënten per studie arm en een verwachte opbrengst voor de maatschappij van €5,6 miljoen, of 70 voor kwaliteit gecorrigeerde levensjaren.

Op basis van het onderzoek in dit proefschrift kunnen enkele aanbevelingen worden gedaan:

- De meeste kosteneffectiviteits analyses (zowel besliskundige modellen als klinische studies) kunnen worden verbeterd door het strict volgen van richtlijnen.
- Voor de gecombineerde analyse van onzekerheid en patiënten heterogeniteit in besliskundige modellen is het van belang het juiste algoritme te kiezen om de correcte uitkomst te berekenen.
- Onzekerheid in kosteneffectiviteits analyse kan het best worden gepresenteerd met betrouwbaarheids intervallen voor de incrementele net benefit (als twee interventies worden vergeleken) of de totale EVPI.
- Methoden die de reductie in de partiële EVPI berekenen zijn onjuist en moeten niet worden gebruikt om vast te stellen welke parameters het belangrijkst zijn.

- VOI analyses om de belangrijkste parameters vast te stellen evenals de optimale studie grootte, zijn een essentieel onderdeel van de analyse van een probabilistisch besliskundig model.
- VOI analyses moeten de eerste stap zijn bij het opzetten van een klinische studie om de mate van onzekerheid met betrekking tot een keuze tussen interventies vast te stellen, evenals de belangrijkste parameters en de optimale studie grootte.

Veel onderwerpen over onzekerheid in de medische besliskunde verdienen meer aandacht, met name op het gebied van VOI analyse. Toekomstig onderzoek kan evalueren of het haalbaar is om een VOI analyse te verplichten voor iedere subsidie aanvraag voor een klinische studie. In de hoofdstukken 8 en 9 werd patiënten data gebruikt om partiële EVPIs te berekenen met simulatie methoden. Toekomstig onderzoek kan de formules vinden waarmee deze partiële EVPIs eenvoudiger zijn te berekenen. In de literatuur worden VOI analyses meestal toegepast in het kader van kosteneffectiviteits studies, terwijl de meeste klinische studies zich richten op uitkomsten zoals mortaliteit. Toekomstig onderzoek kan het gebruik van VOI analyses evalueren om prioriteiten te stellen voor onderzoek met dergelijke uitkomsten.

11

Epilogue

Knowing how little you know

Uncertainty is an essential aspect of medical decision making: the optimal intervention for a patient is typically uncertain to some extent. Knowing how little you know — that is, knowing the extent of uncertainty — has long been recognized as a sign of wisdom.[115] Socrates went so far as to say: "As for me, all I know is I know nothing".[207] Interestingly, the extent of uncertainty has no immediate relevance when deciding on the optimal intervention for a patient. A decision has to be made, no matter how little you know. Moreover, while doubt may be a sign of wisdom, indecisiveness can be detrimental, especially in medical emergencies.

Knowing how little you know is especially relevant when considering if more research is justified to decrease uncertainty about a decision. Less uncertainty is expected to benefit patients. An estimate of the extent or importance of uncertainty is required to set priorities among clinical studies competing for funding. In this dissertation value of information (VOI) analysis was demonstrated as a set of methods to determine the importance of uncertainty and guide future quantitative research. Through VOI analysis you can know how little you know.

The analysis of uncertainty

The analysis of uncertainty in cost-effectiveness analyses is the initial focus of this thesis. Organizations, journals, and experts increasingly recommend the analysis of parameter uncertainty and patient heterogeneity.[110, 208] However, most published cost-effectiveness analyses, for example in surgery, use deterministic models (Chapter 2). For stochastic uncertainty, experts disagree about the use of microsimulation in decision models: some consider it an avoidable obstacle to perform probabilistic sensitivity analyses; others point out that cohort analyses often fail to reflect real-life complexity.[90, 209, 210] Fortunately, increasing computer performance continues to facilitate nested analyses in microsimulation models. The analysis of patient heterogeneity is complicated because no unambiguous method exists to decompose the total variance of a parameter into components attributable to parameter uncertainty and to patient heterogeneity. The methodological complexity of the combined analysis of stochastic uncertainty, parameter uncertainty, and patient heterogeneity presents a barrier for their application. The appropriate algorithms and step-by-step instructions in Chapters 3 and 4 have addressed this difficulty.

The presentation of uncertainty

The debate about the optimal presentation of uncertainty in cost-effectiveness analysis (CEA) has not been settled. Uncertainty intervals for the incremental cost-effectiveness ratio and separate hypothesis testing for costs and effects have been discredited and are currently rarely presented. We demonstrated the serious limitations of cost-effectiveness

acceptability curves (CEACs; Chapter 5), which often appear in clinical journals such as the British Medical Journal. As Howard stated: "No theory that involves just the probabilities of outcomes without considering their consequences could possibly be adequate in describing the importance of uncertainty to a decision maker."[30] Nevertheless, CEACs continues to have their advocates.[211]

For the comparison of two interventions, we recommend the presentation of the incremental distribution on the cost-effectiveness plane, the uncertainty interval of the incremental net benefit, and the total expected value of perfect information (total EVPI). The CE-plane and uncertainty intervals of the net benefits overestimate the extent of uncertainty when more than two interventions are compared, because the outcomes are typically positively correlated in decision models. Therefore, the total EVPI is the only unambiguous measure of decision uncertainty for the comparison of more than two interventions. The total EVPI is a new concept for most clinicians and policy makers. However, the total EVPI is no harder to estimate or understand than a CEAC. Both the total EVPI and the CEAC are straightforward to calculate from the results of a probabilistic sensitivity analysis (Chapter 3).

VOI analysis – methods

Value of information (VOI) analysis is the second focus of this thesis. The framework of cost-effectiveness analysis requires that health care interventions are cost-effective; analogously, VOI analysis requires that quantitative health care research is cost-effective. Phillips observed that studies in health care typically conclude that "more research is needed" without some assessment of the expected benefit for future patients that would come from more research.[13] VOI analysis evaluates uncertainty resulting in a formal comparison of the expected benefit and the cost of a proposed study.

Our initial explorations of VOI analysis were held back because two different methods to estimate the partial EVPI were applied in the VOI literature. In Chapter 6 we demonstrated that methods to estimate the partial EVPI as reduction in EVPI are conceptually and mathematically incorrect. Brennan et al. confirmed our conclusion.[82] All VOI applications published in the last few years have applied the correct methodology for estimating partial EVPI.

VOI analysis – decision models

Yokota's comprehensive overview of VOI analyses in 2004 [28], as well as the more recent literature, shows that most published VOI studies are limited to estimating total (population) EVPI and sometimes partial EVPIs.[33, 34, 212, 213] Such analyses are relevant to evaluate the importance of overall uncertainty and identify key parameters responsible for decision uncertainty. However, they are insufficient to justify and guide further research. Estimation

of the partial expected value of sample information (partial EVSI) is required to compare a study's expected benefit with the expected study costs across a range of sample sizes, to find the optimal sample size. In Chapter 7, we compared the partial EVPI of various study designs. Next, we demonstrated the feasibility of finding the optimal sample size for a study design using VOI analysis in a decision model. For the decision between diagnostic tests for patients with chest pain, the quality-of-life weights for varying severity of chest pain were the most important parameters. The optimal study design was an observational study with 1500 patients for each level of chest pain severity.

VOI analysis – clinical trials

Most published applications of VOI analysis use decision models or hypothetical trial data.[28] Decision models are often preferred for cost-effectiveness, and VOI analyses because of their external validity and consideration of all available evidence. However, a randomized controlled trial constitutes the best available evidence for many clinical decisions, not only to decide what medical intervention should be adopted, but also to guide future research. Currently, a non-significant difference between the study outcomes typically results in a call for a larger trial.[13] In Chapters 8 and 9, we developed and demonstrated VOI methods to use patient-level data on costs and effects from randomized controlled trials. For the treatment of intermittent claudication and for the use of MR imaging for acute knee injury, we identified the key parameters and optimal sample size. Chapter 8 includes programming syntax and step-by-step instructions to allow other investigators to apply VOI analyses to their clinical trial data.

VOI analysis – assumptions and challenges

While we anticipate a bright future for VOI analysis, we acknowledge several assumptions and limitations. VOI analyses are typically performed in a cost-effectiveness analysis setting. As a consequence, VOI analysis has the same assumptions and limitations associated with cost-effectiveness analysis, such as the use of QALYs.[214] Moreover, VOI analysis determines that more research is justified if the expected benefit of a proposed study exceeds the study costs. This trade-off assumes that health care costs and research costs are ultimately paid from the same resources. The extent to which this assumption holds is different in different countries.

The populationVOI, which is compared with the study costs, depends greatly on the future population benefiting from the acquired information. The annual population to benefit depends on the appropriate perspective of the VOI analysis. Typically the appropriate perspective (e.g., national or worldwide) is not obvious. The period that patients will benefit from the proposed data collection is also uncertain because of uncertainty about future technological improvements and evidence from future studies. These ambiguities about

the population to benefit, however, are not drawbacks of VOI analysis in itself, but are inherent to identifying research priorities.

VOI analyses often result in recommending study designs with very large sample sizes, up to several thousand patients. Such sample sizes may seem unfeasible, especially for a single-center study. On the other hand, some multinational trials in cardiology have managed to randomize thousands of patients. Moreover, Altman has pointed out that the sample size of controlled trials is generally inadequate.[12] Using VOI analyses, funding agencies may select a few large multinational studies for funding, instead of many studies with each an inadequate sample size.

VOI analyses in decision models can be challenging for both the analyst and the computer. First, the analyst needs to build a probabilistic model. A probabilistic sensitivity analysis is then required to estimate the total EVPI. Moreover, a nested simulation is required to estimate the total EVPI for microsimulation models. In Chapter 6, we explained that a nested (two-level) simulation is also required to estimate the partial EVPI or EVSI, because typically parameters which are not of interest are not a multilinear function of the model outcome. Step-by-step algorithms are now available for these complicated analyses. Unfortunately, the nested simulations require a great deal of computer-processing capabilities. Obtaining a precise and unbiased estimate in complex models using dedicated decision making software may take weeks of simulation. More time-efficient methods, such as Gaussian process modeling, have been developed, but few applications have been published.[91] The importance of such time-efficient methods will depend on the balance between increasing model complexity and increasing computer performance.

VOI analysis – recommendations and future research

VOI analysis can help set clinical research priorities by comparing the expected benefit and the study costs of proposed study designs. Detsky developed methods to estimate the benefit of clinical trials after they were carried out, and he compared the benefit with the cost of these trials.[184] VOI analysis allows this comparison before the study is performed. Informally, some comparison of the expected benefit and the costs of proposed study designs is also made without a VOI analysis. Funding proposals typically include an estimate of the study costs and of the population that is expected to benefit from the results of the proposed study. The main novelty of VOI analysis is the estimation of the importance of uncertainty per patient, based on evidence available before the study is performed. VOI analysis can replace significance testing (e.g., p-values < 0.05) to evaluate whether more research is justified. Moreover, VOI analysis can determine the optimal sample size of a proposed study, replacing more arbitrary conventional sample size calculations.

Analysts increasingly incorporate parameter uncertainty into their decision models. VOI analysis is the main justification of building such a probabilistic model, since the analysis of uncertainty is especially relevant to guide future research.[27] Analysts should therefore proceed to identify the optimal study design of a future study, if the total EVPI exceeds the study costs. This optimal study is characterized by its design (e.g., randomized or observational), the subset of sampled parameters (e.g., test characteristics or treatment effects), the sample size, and the associated study costs.

We used simulation methods for partial VOI analyses of clinical trial data. In theory, it should be possible to derive closed form solutions for these analyses. We recommend finding these solutions facilitating VOI analysis of clinical trial data to guide future clinical research.

Most published VOI analyses consider both costs and health effects. However, most clinical research considers only health outcomes such as mortality. VOI analysis could be applied to set priorities among studies with mortality as the principle outcome. For example, the total EVPI could be estimated as a 2% expected reduction in mortality with perfect information. For a population-to-benefit of 10,000 patients, the associated expected population benefit of perfect information could be expressed as 200 averted deaths. Applying VOI analyses to studies with multiple health outcomes (e.g., mortality as well as quality-of-life or various adverse events) requires integration of the outcomes into a single outcome such as quality-adjusted life expectancy. We recommend exploring the feasibility of such VOI analyses to guide priority setting for clinical research with only health outcomes.

In general, we recommend evaluating the feasibility of requiring a VOI analysis whenever a clinical study is considered for funding. The VOI analysis could be performed within a decision model (Chapter 7) or using data from a previous clinical study (Chapters 8 and 9). Alternatively, VOI analyses can be performed using expert estimates of the uncertainty intervals of costs and effects of each intervention (e.g., using Delphi methods). The latter requires less time than building a decision model, but is more susceptible to bias. Currently, most study proposals include an estimate of the study costs and of the population which will benefit from the results of the study. Study proposals based on VOI analyses also include estimates of the expected population benefit given the optimal sample size. Proposed studies with a higher population benefit should receive priority. As Howard noted: "Placing a value on the reduction of uncertainty is the first step in experimental design, for only when we know what it is worth to reduce uncertainty do we have a basis for allocating our resources in experimentation designed to reduce the uncertainty."[30]

12

Appendix

APPENDIX 1 – CHAPTER 3

Notation

x : model parameter
$f(x)$: model outcome for a value of x
$E[f(x)]$: expected outcome of model
$E[x]$: expected value of x

Numerical example of nonlinearity

$f(x)=\frac{1}{x}$ is a nonlinear function of x.
If x has a discrete distribution, being 0.1 or 0.3 both with 50% probability, then:

$E\left[f(x)\right]=\frac{1}{2}*\frac{1}{0.1}+\frac{1}{2}*\frac{1}{0.3}=6.7$, which is not equal to:

$f\left(E[x]\right)=\frac{1}{E[x]}=\frac{1}{0.2}=5.$

APPENDIX 2 – CHAPTER 3

Notation

We used a notation similar to Ades et al.32 The notation is slightly modified to allow for stochastic uncertainty and patient heterogeneity.
Notation
θ : stochastic variable for model parameters
θ^I : parameters of interest in partial value of information analyses
θ^C : parameters not of interest in partial value of information analyses
X : stochastic variable for covariates
t : strategy
t^* : baseline optimal strategy; i.e., prior to collecting more data
$B(t,\theta,X)$: stochastic variable for patient-level model outcome (e.g., lifetime)
$\overline{B}(t,\theta,X)=E_B\left(B(t,\theta,X)\right)$: expected outcome of cohort analysis (e.g., life expectancy)
D : data collection on parameters of interest in proposed future study
E : expectation
$\max$: maximization
P : probability distribution

Stochastic uncertainty

Models with no parameter uncertainty and no patient heterogeneity.
The expected outcome of the optimal strategy is: $\max_t [E_B (B(t))]$,

estimated by: $\max_t \left[\frac{1}{N} \sum_{n=1}^{N} B_n(t) \right]$.

Parameter uncertainty

Models with parameter uncertainty, and no patient heterogeneity.
Markov cohort model
The expected outcome of the optimal strategy is: $\max_t \left[E_\theta \left(\overline{B}(t,\theta) \right) \right]$,

estimated by: $\max_t \left[\frac{1}{M} \sum_{m=1}^{M} \overline{B}(t,\theta_m) \right]$.

The distribution of the expected outcome of strategy a is: $P_\theta \left\{ \overline{B}(t,\theta) \right\}$.
Patient-level model
The expected outcome of the optimal strategy is: $\max_t [E_{B,\theta} (B(t,\theta))]$,

estimated by: $\max_t \left[\frac{1}{NM} \sum_{n=1,m=1}^{N,M} B_n(t,\theta_m) \right]$.

The distribution of the expected outcome of strategy a is: $P_\theta \{ E_B (B(t,\theta)) \}$,

estimated by: $P_\theta \left\{ \frac{1}{N} \sum_{n=1}^{N} B_n(t,\theta) \right\}$

TOTAL EVPI

Markov cohort model with parameter uncertainty, and no patient heterogeneity.

$$\text{total EVPI} = E_\theta \left(\textit{opportunity loss} \right) = E_\theta \left(\max_t \left(\overline{B}(t,\theta) \right) - \overline{B}\left(t^*, \theta \right) \right),$$

estimated by: $\frac{1}{M} \sum_{m=1}^{M} \left(\max_t \left(\overline{B}(t,\theta_m) \right) - \overline{B}\left(t^*, \theta_m \right) \right)$.

See Box 2 for a step-by-step algorithm. See Groot Koerkamp et al. for a proof that the following is mathematically equivalent:[105]

$$\text{total EVPI} = E_\theta \max_t \left(\overline{B}(t,\theta) \right) - \max_t E_\theta \left(\overline{B}(t,\theta) \right).$$

PARTIAL EVPI

Markov cohort model with parameter uncertainty, and no patient heterogeneity.

$$pEVPI(\theta^I) = E_{\theta^I}\left(\max_t E_{\theta^C|\theta^I}\left(\overline{B}(t,\theta)\right) - E_{\theta^C|\theta^I}\left(\overline{B}\left(t^*,\theta\right)\right)\right),$$

estimated by: $\frac{1}{J}\sum_{j=1}^{J}\left(\max_t \frac{1}{K}\sum_{k=1}^{K}\overline{B}\left(t,\theta_j^I,\theta_k^C\middle|\theta_j^I\right) - \frac{1}{K}\sum_{k=1}^{K}\overline{B}\left(t^*,\theta_j^I,\theta_k^C\middle|\theta_j^I\right)\right)$

See Box 4 for a step-by-step algorithm. If the model is a multilinear function of uncorrelated B^C, and a linear function of correlated B^C, the equations simplify to:

$$pEVPI(\theta^I) = E_{\theta^I}\left(\max_t\left(\overline{B}\left(t,\theta^I,E\left(\theta^C\middle|\theta^I\right)\right)\right) - \overline{B}\left(t^*,\theta^I,E\left(\theta^C\middle|\theta^I\right)\right)\right),$$

estimated by: $\frac{1}{J}\sum_{j=1}^{J}\left(\max_t\left(\overline{B}\left(t,\theta_j^I,E\left(\theta^C\middle|\theta_j^I\right)\right)\right) - \overline{B}\left(t^*,\theta_j^I,E\left(\theta^C\middle|\theta_j^I\right)\right)\right).$

See Box 3 for a step-by-step algorithm.

Variability

Markov cohort model with patient heterogeneity, and no parameter uncertainty.
The expected outcome of a heterogeneous population is: $E_X\left(\overline{B}(t,X)\right)$,

estimated by: $\frac{1}{P}\sum_{p=1}^{P}\overline{B}\left(t,X_p\right).$

APPENDIX 3 – CHAPTERS 3 AND 7

Table 1 includes parameter values related to exercise echocardiography and exercise SPECT that were not considered in the model of chapter 3.

Table 1A Model parameters – test characteristics and pre-test probability of CHD

nr.	parameter	mean (in %)	95% confidence limits		source
	angiography: short-term risks				
1	myocardial infarction if 0-VD‡	0.02	0.02	0.03	[192]
2	myocardial infarction if 1-VD	0.06	0.05	0.07	[192]
3	myocardial infarction if 2-VD	0.08	0.07	0.10	[192]
4	myocardial infarction if 3-VD	0.08	0.07	0.09	[192]
5	myocardial infarction if LMD§	0.17	0.15	0.18	[192]
6	die if 0-VD	0.02	0.01	0.02	[193]
7	die if 1-VD	0.05	0.04	0.06	[193]
8	die if 2-VD	0.07	0.06	0.08	[193]
9	die if 3-VD	0.12	0.11	0.13	[193]
10	die if LMD	0.55	0.52	0.58	[193]
	CTCA*				
11	Sensitivity	99	95	100	[194]
12	Specificity	88	59	100	[194]
13	percentage uninterpretable	1.69	0.04	6.16	[194]
14	percentage morbidity	0.05	0.04	0.06	expert opinion
	exercise echocardiography				
15	Sensitivity	85	83	87	[195]
16	Specificity	77	74	80	[195]
17	percentage uninterpretable	5	0	10	[196]
18	percentage morbidity	0.05	0.04	0.06	expert opinion
19	percentage mortality	0.005	0.004	0.006	expert opinion
	exercise SPECT				
20	Sensitivity	87	86	88	[195]
21	Specificity	64	60	68	[195]
22	percentage uninterpretable	2	0	4	expert opinion
23	percentage morbidity	0.05	0.04	0.06	expert opinion
24	percentage mortality	0.005	0.004	0.006	expert opinion
25	pre-test probability of CHD given age, gender, and chest pain type				see Table 3

‡ VD = vessel disease
§ LMD = left main disease
* CTCA = computed tomographic coronary angiography

Table 1B Model parameters – costs in US$.

nr.	Parameter	mean	credible interval ‡		source
	cost tests				
26	cost angiography	3186	2230	4141	Medicare (30% inpatient)
27	cost CTCA	705	494	917	Medicare
28	cost echocardiography	241	169	313	Medicare
29	cost SPECT	599	419	779	Medicare
	cost treatments				
30	cost CABG	23052	16136	29968	Medicare
31	cost PCI	11795	8257	15334	Medicare (70% stents)
	cost events				
32	cost myocardial infarction	6690	4683	8697	Medicare
	annual costs chest pain patient				
33	no, abnormal LVEF	3870	2709	5031	Medicare
34	mild, abnormal LVEF §	5689	3982	7395	Medicare
35	mild, normal LVEF	1818	1272	2363	Medicare
36	severe, abnormal LVEF	9148	6403	11892	Medicare
37	severe, normal LVEF	4105	2874	5337	Medicare

‡ For costs we assumed a ±30% credible interval on Medicare estimates.
§ LVEF = left ventricular ejection fraction

Table 1C Model parameters – utilities.

nr.	Parameter	mean	95% confidence limits		source
	utilities chest pain states				
38	No	0.87	0.80	0.92	197
39	Mild	0.81	0.76	0.86	197
40	Severe	0.67	0.56	0.77	197
	disutilities (in days lost)				
41	CABG	30	24	36	expert opinion
42	Angiography	1	0.8	1.2	expert opinion
43	myocardial infarction	10	8	12	expert opinion
44	PCI	2.5	2	3	expert opinion

Table 1D Model parameters – medical treatment.

nr.	parameter ‡	mean	95% confidence limits		source
	effects medical treatment only				
45	risk myocardial infarction with 0-VD	1.0	0.7	1.5	198
46	risk myocardial infarction with 1/2-VD	2.1	1.5	2.9	198
47	risk myocardial infarction with 3-VD / LMD	2.7	2.1	3.5	198
48	relative risk die with 1/2-VD *	2.3	1.9	2.8	199
49	relative risk die with 3-VD *	3.6	3.1	4.1	199
50	relative risk die with LMD *	9.6	6.2	14.3	199
51	% for annual transitions between chest pain states after medical treatment				see Table 2
52	risk future procedure 0-VD	0.5	0.2	1.0	200
53	risk future procedure 1-VD	1.0	0.5	1.7	200
54	risk future procedure 2-VD	4.2	3.0	5.5	200
55	risk future procedure 3-VD/LMD	7.5	6.0	9.2	200
56	% CABG of future procedures 0/1-VD	16	13.8	18.3	200
57	% CABG of future procedures 2-VD	58	55	61	200
58	% CABG of future procedures 3-VD/LMD	87	85	89	200
	LVEF (treatment independent) §				
59	risk abnormal LVEF with 0-VD	0.5	0.4	0.7	198
60	risk abnormal LVEF with 1-VD	2.0	1.7	2.3	198
61	risk abnormal LVEF with 2-VD	5.0	4.5	5.5	198
62	risk abnormal LVEF with 3-VD	6.0	5.5	6.5	198
63	risk abnormal LVEF with LMD	6.0	5.5	6.5	198

‡ All risks – except relative risks – are annual risks and expressed in %.
§ LVEF = left ventricular ejection fraction
* Relative risks of dying are in comparison to life table mortality.

Table 1E. Model parameters – PCI.

nr.	parameter ‡	mean	95% confidence limits		source
	short-term risks with PCI				
64	myocardial infarction if 1-VD	3.5	2.7	4.4	[201]
65	myocardial infarction if 2-VD	5.2	4.3	6.2	[201]
66	die if 1-VD	0.2	0.1	0.5	[202, 203]
67	die if 2-VD	0.9	0.5	1.4	[202, 203]
	effects PCI				
68	reduction in myocardial infarction	17	12	22	[204]
69	reduction in mortality if 1/2-VD	15	0	50	[200]
70	% for annual transitions between chest pain states after PCI				see Table 2
71	relative risk chest pain improvement of PCI vs. CABG	0.85	0.80	0.90	[200]
72	relative risk of reprocedure for stent vs. PTCA§	0.5	0.2	0.8	[205]
73	% eligible for PCI	76	61	91	[205]
74	annual risk of reprocedure after PCI	4	2	5	[200]
75	% CABG of reprocedures	22	19	25	[200]

‡ All values are expressed in %, except the relative risks
§ PTCA = percutaneous transluminal coronary angioplasty

Table 1F. Model parameters – CABG.

nr.	parameter ‡	mean	95% confidence limits		source
	short-term risks CABG				
76	myocardial infarction if 3-VD	7.0	5.9	8.2	[201]
77	myocardial infarction if LMD	7.0	5.9	8.2	[201]
78	die if 35yr male	2.3	0.2	6.6	[206]
79	die if 35yr female	2.9	0.1	10.9	[206]
80	die if 45yr male	2.7	1.6	4.2	[206]
81	die if 45yr female	3.5	1.6	6.1	[206]
82	die if 55yr male	3.2	2.3	4.2	[206]
83	die if 55yr female	4.1	2.6	5.8	[206]
84	die if 65yr male	4.4	3.5	5.4	[206]
85	die if 65yr female	5.7	4.1	7.5	[206]
86	die if 75yr male	6.3	5.1	7.6	[206]
87	die if 75yr female	8.0	6.0	10.3	[206]
	effects CABG				
88	reduction in myocardial infarction	42	29	55	[198]
89	reduction in mortality if 3-VD	48	29	64	[199]
90	reduction in mortality if LMD	67	40	84	[199]
91	% from mild to no angina	60	55	65	[198]
92	% from severe to mild angina	70	65	75	[198]
93	% from severe to no angina	10	7	13	[198]
94	% for annual transitions between chest pain states after CABG				see Table 2
95	annual risk of reprocedure after CABG	2	1	3	[200]
96	% CABG of reprocedures	7	5	9	[200]

‡ All values are expressed in %.

	sample size	none	mild	severe
Medical treatment, 1-VD				
none	164	136	26	2
mild	309	46	244	19
severe	20	5	7	8
CABG/PCI, 1-VD				
none	263	226	34	3
mild	188	43	134	11
severe	18	5	6	7
Medical treatment, 2-VD				
none	193	154	37	2
mild	449	49	359	40
severe	42	8	10	23
CABG/PCI, 2-VD				
none	428	376	47	4
mild	262	34	225	3
severe	14	3	4	8
Medical treatment, 3-VD/LMD				
none	202	135	65	2
mild	360	86	259	14
severe	31	3	3	25
CABG/PCI, 3-VD/LMD				
none	316	262	51	3
mild	208	31	164	12
severe	15	4	5	6

Table 2

Dirichlet distributions were used to represent uncertainty about the annual transition probabilities between chest pain severity states. These transitions depend on the extent of vessel disease and the treatment. The parameters of the dirichlet distributions are the number of patients that transit to each health state. For example, for patients with 1 vessel disease, medical treatment, and no chest pain, 26 out of 164 patients had mild chest pain at the end of the year. These numbers were derived and extrapolated from the CASS-registry.[85, 198]

subgroup	sample size	0-VD	1-VD	2-VD	3-VD	LMD
NON-SPECIFIC						
Men						
30-39	249	242	5	0	2	0
40-49	391	344	35	8	4	0
50-59	440	361	48	18	4	9
60-69	129	94	15	12	6	1
>70	22	8	2	7	5	0
Women						
30-39	135	130	3	1	0	1
40-49	425	408	13	4	0	0
50-59	585	550	23	12	0	0
60-69	215	194	15	2	2	2
>70	23	23	0	0	0	0
ATYPICAL						
Men						
30-39	171	92	32	31	12	3
40-49	568	244	131	80	85	28
50-59	919	267	221	221	147	64
60-69	434	95	91	104	100	43
>70	46	3	10	11	15	7
Women						
30-39	39	31	4	2	0	2
40-49	257	177	46	18	5	10
50-59	414	290	54	37	25	8
60-69	264	137	58	34	24	11
>70	34	15	5	5	8	1
TYPICAL						
Men						
30-39	66	11	25	15	10	5
40-49	402	48	109	105	96	44
50-59	840	42	168	244	269	118
60-69	539	27	86	129	199	97
>70	67	2	6	10	33	16
Women						
30-39	10	3	1	3	2	1
40-49	70	31	15	12	12	0
50-59	161	52	35	31	31	13
60-69	137	26	19	34	44	14
>70	23	1	5	2	13	2

Table 3

Dirichlet distributions representing uncertainty about pretest probability of CHD were based on observational study results.[85] The probabilities depend on gender, age, and type of chest pain. For example, 249 men with non-specific chest pain were observed of whom 5 had one vessel disease.

APPENDIX 4 – CHAPTER 4

Example calculation of 10-year risk for base case patient:
base case patient: men 60 yr, average total cholesterol (212.5), average HDL (44.9), average (treated) SBP (129.7), nonsmoker, nondiabetic

$$\sum_{i=1}^{p} \beta_i \left(X_i - \overline{X_i} \right) = 3.061*(\log(60)-3.856)+1.124*(\log(212.5)-5.342)-0.933*(\log(44.9)-3.769)+1.933(0-4.354)+1.999*(\log(129.7)-0.502)+0.655*(0-0.352)+0.574*(0-0.065)=0.7517$$

$$p = 1 - S_o\left(10\right)^{\exp\left(\sum_{i=1}^{p} \beta_i \left(X_i - \overline{X_i} \right)\right)} = 1 - 0.8894^{\exp(0.7517)} = 0.22 = 22\%$$

Note:
For the analysis of parameter uncertainty we assumed that the regression coefficients of "SBP if treated" and "SBP if untreated" had a correlation of 1. We used the standard eauation for the conditional random value of a bivariate normal distribution: $y = \overline{y} + \frac{\sigma_y}{\sigma_x} * \left[x - \overline{x} \right]$. x and y represent the correlated beta's.

APPENDIX 5 – CHAPTER 6

We will prove equation 10, showing that the total expected value of perfect information (total EVPI) is independent of the path in Figure 1. We want to show that:

$$\text{Total EVPI} = p\text{EVPI}(\theta_I) + E_{\theta I}\, p\text{EVPI}(\theta_C|\theta_I).$$

Substituting the definitions of total and partial EVPI (equations 4 and 5), we get:

$$\begin{aligned}&\mathrm{E}_\theta \max_a \mathrm{B}(a,\theta) - \max_a \mathrm{E}_\theta \mathrm{B}(a,\theta) = \\ &[\mathrm{E}_{\theta I} \max_a \mathrm{E}_{\theta C}|_{\theta I} \mathrm{B}(a,\theta) - \max_a \mathrm{E}_\theta \mathrm{B}(a,\theta)] + \\ &E_{\theta I}\, [E_{\theta C} \max_a E_{\theta I|\theta C} \mathrm{B}(a,\theta) - \max_a \mathrm{E}_\theta \mathrm{B}(a,\theta)].\end{aligned}$$

Collecting and canceling terms leads to:

$$\begin{aligned}&\mathrm{E}_\theta \max_a \mathrm{B}(a,\theta) = E_{\theta I} \max_a E_{\theta I|\theta C} \mathrm{B}(a,\theta) + \mathrm{E}_\theta \max_a \mathrm{B}(a,\theta) \\ &- E_{\theta I} \max_a E_{\theta I|\theta C} \mathrm{B}(a,\theta),\end{aligned}$$

which proves the equality we set out to prove.

APPENDIX 6 - CHAPTER 8

SYNTAX FOR EXCEL

The algorithms for the VOI analyses are described using a spreadsheet such as Excel. The analyses are easier and faster using visual basics for Excel. We didn't use visual basics, because many are not familiar with it. Moreover, if you're used to writing syntax, we recommend R because it is much faster than visual basics.

general preparation for VOI analyses

1. Create a separate worksheet for the data of each intervention.
2. Set up the data with patients in rows and parameters in columns. The first row is the header.
3. Set costs as negative values.
4. The last column of the datasheet should contain the monetarized values of the quality-adjusted life expectancy: choose a value for the WTP threshold (e.g., 80,000 euro/QALY), and multiply each quality-adjusted life expectancy with the WTP.
5. Calculate the net monetary benefit for each patient in the column on the right of the datasheet.

The syntax below assumes a total of 12 parameters in columns A to L, with patient-level data on 75 patients for each intervention. Column M of the datasheet contains the net monetary benefit. We perform the analyses for 10,000 simulations. Each algorithm is performed in a separate sheet. We refer to the cells in a bold capital and number followed by a colon. 'PTA data'! refers to the worksheet named PTA data.

preparation for total EVPI and total EVSI analyses

```
A1: =AVERAGE('PTA data'!M2:M76)              #  mean net benefit of PTA
B1: =AVERAGE('EX data'!M2:M76)               #  mean net benefit of EX
A2: =SQRT(VAR('PTA data'!M2:M76)/75)         #  standard error of the mean net
                                                benefit of PTA
B2: =SQRT(VAR('EX data'!M2:M76)/75)          #  standard error of the mean net
                                                benefit of EX
```

total EVPI – in Excel

```
A4: =NORMINV(RAND(),A$1,A$2)                 #  step 1
B4: =NORMINV(RAND(),B$1,B$2)
C4: =MAX(A6:B6)-IF(A$1>B$1,A6,B6)            #  step 2
copy cells A4, B4, and C4 to rows 5 to 10003 #  step 3
C1: =AVERAGE(C4:C10003)                      #step 4
```

total EVSI – in Excel

```
D1: =500                                                          # step 1
A4: =NORMINV(RAND(),A$1,A$2)                                      # step 2
B4: =NORMINV(RAND(),B$1,B$2)
C4: =NORMINV(RAND(),A4,SQRT(VAR('PTA data'!M$2:M$76)/D$1))        # step 3
D4: =NORMINV(RAND(),B4,SQRT(VAR('EX data'!M$2:M$76)/D$1))
E4: =(A$1*75+C4*D$1)/(75+D$1)                                     # step 4
F4: =(B$1*75+D4*D$1)/(75+D$1)                                     # step 4
G4: =MAX(E6:F6)-IF(A$1>B$1,E6,F6)                                 # step 5
copy cells A4, B4, C4, D4, E4, F4, and G4 to rows 5 to 10003      # step 6
C1: =AVERAGE(G4:G10003)                                           # step 7
```

preparation for partial EVPI and partial EVSI analysis

```
A1: =AVERAGE('PTA data'!M2:M76)                  # mean net benefit of PTA
B1: =AVERAGE('EX data'!M2:M76)                   # mean net benefit of EX
A4 to A15: "select value 1 or 0 for each cell"   # PTA: 1= θ^I, 0= θ^C
B4 to B15: "select value 1 or 0 for each cell"   # EX: 1= θ^I, 0= θ^C
C4: =IF(A4=1,0,1)                                # PTA: 0= θ^I, 1= θ^C
D4: =IF(B4=1,0,1)                                # EX: 0= θ^I, 1= θ^C
copy cells C4 and D4 to rows 5 to 15
E4: =MMULT('PTA data'!A2:L76,A4:A15)             # matrix multiplication
select cells E4 to E78, press F2, press ctrl-shift-enter   # net benefit per patient of PTA
                                                           for θ^I
E1: =AVERAGE(E4:E78)                             # mean NB of PTA for θ^I
E2: =sqrt(var(E4:E78)/75)                        # sem of mean NB of PTA for θ^I
F4: =MMULT('EX data'!A2:L76,B4:B15)              # matrix multiplication
select cells F4 to F78, press F2, press ctrl-shift-enter   # net benefit per patient of EX
                                                           for θ^I
F1: =AVERAGE(F4:F78)                             # mean NB of EX for θ^I
F2: =sqrt(var(F4:F78)/75)                        # sem of mean NB of EX for θ^I

G4: =MMULT('PTA data'!A2:L76,C4:C15)             # matrix multiplication
select cells G4 to G78, press F2, press ctrl-shift-enter   # net benefit per patient of PTA
                                                           for θ^C
G1: =AVERAGE(G4:G78)                             # mean NB of PTA for θ^C
G2: =sqrt(var(G4:G78)/75)                        # sem of mean NB of PTA for
                                                           θ^C
```

```
H4: =MMULT('EX data'!A2:L76,D4:D15)                    # matrix multiplication
select cells H4 to H78, press F2, press ctrl-shift-enter  # net benefit per patient of EX
                                                          for θ^C
H1: =AVERAGE(H4:H78)                                   # mean NB of EX for
H2: =sqrt(var(H4:H78)/75)                              # sem of mean NB of EX for θ^C
```

partial EVPI – in Excel

```
I4: =NORMINV(RAND(),E$1,E$2)                                  # step 1
J4: =NORMINV(RAND(),F$1,F$2)
K4: =G$1+((CORREL(E$4:E$78,G$4:G$78)*G$2*(I4-E$1))/E$2)       # step 2
L4: =H$1+((CORREL(F$4:F$78,H$4:H$78)*H$2*(J4-F$1))/F$2)
M4: =I4+K4                                                    # step 3
N4: = J4+L4
O4: =MAX(M4:N4)-IF(A$1>B$1,M4,N4)                             # step 4
copy cells I4, J4, K4, L4, M4, N4, and O4 to rows 5 to 10003  # step 5
O1: =AVERAGE(O4:O10003)                                       # step 6
```

partial EVSI – in Excel

```
D1: =500                                                      # step 1
I4: =NORMINV(RAND(),E$1,E$2)                                  # step 2
J4: =NORMINV(RAND(),F$1,F$2)
K4: =NORMINV(RAND(),I4,SQRT(VAR(E$4:E$78)/D$1))               # step 3
L4: =NORMINV(RAND(),J4,SQRT(VAR(F$4:F$78)/D$1))
M4: =(E$1*75+K4*D$1)/(75+D$1)                                 # step 4
N4: =(F$1*75+L4*D$1)/(75+D$1)
O4: =G$1+(CORREL(E$4:E$78,G$4:G$78)*G$2*(M4-E$1))/E$2         # step 5
P4: =H$1+(CORREL(F$4:F$78,H$4:H$78)*H$2*(N4-F$1))/F$2
Q4: =M4+O4                                                    # step 6
R4: =N4+P4
S4: =MAX(Q4:R4)-IF(A$1>B$1,Q4,R4)                             # step 7
copy cells I4, J4, K4, L4, M4, N4, O4, P4, Q4, R4, and S4 to rows 5 to 10003  # step 8
S1: =AVERAGE(S4:S10003)                                       # step 9
```

APPENDIX 7 – CHAPTER 8

SYNTAX FOR R

The R syntax is presented for each step of the algorithms. We used vectors to create n random draws in one step, avoiding computer-intensive loops.

preparation for VOI analyses

Set up (e.g., in Excel) the data with patients in rows and parameters in columns. The first row is the header. The last parameter is the quality-adjusted life expectancy. Create a separate file for each intervention. Save the file with extension txt (tab delimited). The syntax below assumes a total of 12 parameters, and 75 patients per treatment group.

general preparation

```
library("MASS", character.only=TRUE)              # open statistical package
pta<-read.delim("d:/r/PTA.txt",header=TRUE)       # import data from folder d:/r
ex<-read.delim("d:/r/Ex.txt",header=TRUE)
pta<-as.matrix(pta)                               # allow for matrix calculations
ex<-as.matrix(ex)
WTP<-80000                                        # select a WTP
pta[,12]<-pta[,12]*WTP                            # monetarize QALYs
ex[,12]<-ex[,12]*WTP
pta[,1:11]<- -pta[,1:11]                          # set costs as negative values
ex[,1:11]<- -ex[,1:11]
nbPTA<- pta[,12]+ apply(pta[,1:11],1,sum)         # create vector for net benefit
nbEX<- ex[,12]+ apply(ex[,1:11],1,sum)
N<-100000                                         # select number of simulations
```

preparation for partial EVPI and partial EVSI analysis

```
PTAi<- c(0,0,0,0,0,0,0,0,0,1,0,1)                 # vector to select θ^I of PTA with 1
EXi<- c(0,0,0,0,0,0,0,0,0,1,0,1)                  # vector to select θ^I of EX with 1
PTAc<-c(1,1,1,1,1,1,1,1,1,1,1,1)-PTAi             # vector to select θ^C of PTA with 1
EXc<-c(1,1,1,1,1,1,1,1,1,1,1,1)-EXi               # vector to select θ^C of EX with 1
nbPTAi<-pta%*%PTAi                                # total NB for θ^I of PTA
nbEXi<-ex%*%EXi                                   # total NB for θ^I of EX
nbPTAc<-pta%*%PTAc                                # total NB for θ^C of PTA
nbEXc<-ex%*%EXc                                   # total NB for θ^C of EX
```

The parameters of interest can be changed in the definitions of PTAi and EXi. In the current setting the parameters-of-interest are parameter number 10 and 12 of both interventions.

total EVPI – algorithm in R

```
rPTA<-rnorm(N,mean(nbPTA),sqrt(var(nbPTA)/75))                    # step 1
rEX<-rnorm(N,mean(nbEX),sqrt(var(nbEX)/75))
VPI<-apply(cbind(rPTA,rEX),1,max)- rPTA                           # step 2
EVPI<-mean(VPI)                                                   # step 4
```

Step 3 is skipped because at step 1 a vector of n random values is drawn. The last term of step 2 should be replaced with rEX if exercise is the optimal treatment given the WTP.

total EVSI – algorithm in R

```
nEVSI<-500                                                        # step 1
rPTA<-rnorm(N,mean(nbPTA),sqrt(var(nbPTA)/75))                    # step 2
rEX<-rnorm(N,mean(nbEX),sqrt(var(nbEX)/75))
rPTAs<-rnorm(N,rPTA,sqrt(var(nbPTA)/nEVSI))                       # step 3
rEXs<- rnorm(N,rEX,sqrt(var(nbEX)/nEVSI))
pPTA<-(mean(nbPTA)*75+ rPTAs*nEVSI)/(75+nEVSI)                    # step 4
pEX<-(mean(nbEX)*75+rEXs*nEVSI)/(75+nEVSI)
VSI<-apply(cbind(pPTA,pEX),1,max)-pPTA                            # step 5
EVSI<-mean(VSI)                                                   # step 7
```

Step 6 is skipped because at step 2 a vector of n random values is drawn. The last term of step 5 should be replaced with pEX if exercise is the optimal treatment given the WTP.

partial EVPI – algorithm in R

```
rPTAi<-rnorm(N,mean(nbPTAi),sqrt(var(nbPTAi)/75))                 # step 1
rEXi<-rnorm(N,mean(nbEXi),sqrt(var(nbEXi)/75))
rPTAc<-mean(nbPTAc)+cor(nbPTAi,nbPTAc)*sqrt(var(nbPTAc))*
(rPTAi-mean(nbPTAi))/sqrt(var(nbPTAi))                            # step 2
rEXc<-mean(nbEXc)+cor(nbEXi,nbEXc)*sqrt(var(nbEXc))*
(rEXi-mean(nbEXi))/sqrt(var(nbEXi))
rPTA<-rPTAi+rPTAc                                                 # step 3
rEX<-rEXi+rEXc
VPI<-apply(cbind(rPTA,rEX),1,max)-rPTA                            # step 4
pEVPI<-mean(VPI)                                                  # step 6
```

Step 5 is skipped because at step 1 a vector of n random values is drawn. The last term of step 4 should be replaced with rEX if exercise is the optimal treatment given the WTP.

partial EVSI – algorithm in R

```
nEVSI<-500                                                    # step 1
rPTAi<-rnorm(N,mean(nbPTAi),sqrt(var(nbPTAi)/75))             # step 2
rEXi<-rnorm(N,mean(nbEXi),sqrt(var(nbEXi)/75))
rPTAs<-rnorm(N,rPTAi,sqrt(var(nbPTAi)/nEVSI))                 # step 3
rEXs<-rnorm(N,rEXi,sqrt(var(nbEXi)/nEVSI))
pPTAi<-(mean(nbPTAi)*75+rPTAs*nEVSI)/(75+nEVSI)               # step 4
pEXi<-(mean(nbEXi)*75+rEXs*nEVSI)/(75+nEVSI)
pPTAc<-mean(nbPTAc)+cor(nbPTAi,nbPTAc)*sqrt(var(nbPTAc))*
(pPTAi-mean(nbPTAi))/sqrt(var(nbPTAi))                        # step 5
pEXc<-mean(nbEXc)+cor(nbEXi,nbEXc)*sqrt(var(nbEXc))*
(pEXi-mean(nbEXi))/sqrt(var(nbEXi))
pPTA<- pPTAi+pPTAc                                            # step 6
pEX<-pEXi+pEXc
VSI<-apply(cbind(pPTA,pEX),1,max)-pPTA                        # step 7
pEVSI<-mean(VSI)                                              # step 9
```

Step 8 is skipped because at step 2 a vector of n random values is drawn. The last term of step 7 should be replaced with pEX if exercise is the optimal treatment given the WTP.

13

References

1. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974;185(4157):1124-31.
2. Stiell IG, Greenberg GH, McKnight RD, et al. Decision rules for the use of radiography in acute ankle injuries. Refinement and prospective validation. JAMA 1993;269(9):1127-32.
3. Elstein AS. On the origins and development of evidence-based medicine and medical decision making. Inflamm Res 2004;53(Supplement 2):S184-S9.
4. Raiffa H, Schlaifer RO. Applied Statistical Decision Theory. Cambridge (MA): Division of Research, Graduate School of Business Administration, Harvard University; 1961.
5. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998;18(2 Suppl):S68-80.
6. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. Value Health 2005;8(5):521-33.
7. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ 2006;15:677-87.
8. Pearson SD, Rawlins MD. Quality, innovation, and value for money: NICE and the British National Health Service. JAMA 2005;294(20):2618-22.
9. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J (Clin Res Ed) 1986;292(6522):746-50.
10. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. Br Med J 1999;319:508-12.
11. Freiman JA, Chalmers TC, Smith HJ, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized controlled trial. Survey of 71 "negative" trials. N Engl J Med 1978;299(13):690-4.
12. Altman DG, Bland JM. Absence of evidence is not evidence of absence. Br Med J 1995;311:485.
13. Phillips CV. The economics of 'more research is needed'. Int J Epidemiol 2001;30(4):771-6.
14. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). Lancet 2002;360(9334):711-5.
15. Willan AR, Pinto EM. The value of information and optimal clinical trial design. Statist Med 2005;24:1791-806.
16. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. Comput Biomed Res 1986;19(3):254-65.
17. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Med Decis Making 1985;5(2):157-77.

18. Guide to the methods of health technology appraisal. London: National Institute for Clinical Excellence; 2004.
19. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14(4):339-47.
20. Cronin KA, Legler JM, Etzioni RD. Assessing uncertainty in microsimulation modelling with application to cancer screening interventions. Stat Med 1998;17:2509-23.
21. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med 2003;22:3687-709.
22. Black WC. The CE-plane: a graphic representation of cost-effectiveness. Med Decis Making 1990;10(3):212-4.
23. Briggs AH, O'Brien BJ. The death of cost-minimisation analysis? Health Econ 2001;10:179-84.
24. Efron B TR. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
25. Fieller EC. Some problems in interval estimation. J R Stat Soc, Ser B 1954;16:175-83.
26. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 1994;3(5):309-19.
27. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999;18(3):341-64.
28. Yokota F, Thompson KM. Value of information literature analysis: a review of applications in health risk management. Med Decis Making 2004;24(3):287-98.
29. Grundy PM, Healy MJR, Rees DH. Economic choice of the amount of experimentation. J R Stat Soc, Ser A 1956;18:32-48.
30. Howard RA. Information Value Theory. Trans Syst Sci Cybern 1966;SSC2 (1):22-6.
31. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. Health Technol Assess 2004;8(31):1-103, iii.
32. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Med Decis Making 2004;24(2):207-27.
33. Colbourn TE, Asseburg C, Bojke L, et al. Preventive strategies for group B streptococcal and other bacterial infections in early infancy: cost effectiveness and value of information analyses. Br Med J 2007;335(7621):655.
34. Hoomans T, Fenwick EA, Palmer S, Claxton K. Value of Information and Value of Implementation: Application of an Analytic Framework to Inform Resource Allocation Decisions in Metastatic Hormone-Refractory Prostate Cancer. Value Health 2008;Jul 24. [Epub ahead of print].
35. Eddy DM. Clinical decision making: from theory to practice. Cost-effectiveness analysis. A conversation with my father. JAMA 1992;267(12):1669-75.

36. Fishman A, Martinez F, Naunheim K, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. N Engl J Med 2003;348(21):2059-73.
37. Ramsey SD, Berry K, Etzioni R, Kaplan RM, Sullivan SD, Wood DE. Cost effectiveness of lung-volume-reduction surgery for patients with severe emphysema. N Engl J Med 2003;348(21):2092-102.
38. Hunink MGM, Glasziou P. Decision Making in Health and Medicine. Cambridge, UK: Cambridge University Press; 2001.
39. Zanocco K, Angelos P, Sturgeon C. Cost-effectiveness analysis of parathyroidectomy for asymptomatic primary hyperparathyroidism. Surgery 2006;140(6):874-81; discussion 81-2.
40. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med 1977;296(13):716-21.
41. Eddy DM. Oregon's methods. Did cost-effectiveness analysis fail? JAMA 1991;266(15):2135-41.
42. Choosing Interventions that are Cost Effective (WHO-CHOICE). (Accessed at http://www.who.int/choice/costs/CER_levels/en/index.html.)
43. CIA - The World Factbook. (Accessed at https://www.cia.gov/library/publications/the-world-factbook/geos/us.html.)
44. Drummond MF. Experimental versus observational data in the economic evaluation of pharmaceuticals. Med Decis Making 1998;18(2 Suppl):S12-8.
45. Campbell H, Briggs A, Buxton M, Kim L, Thompson S. The credibility of health economic models for health policy decision-making: the case of population screening for abdominal aortic aneurysm. J Health Serv Res Policy 2007;12(1):11-7.
46. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. Br Med J 1996;313(7052):275-83.
47. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276(16):1339-41.
48. Briggs AH, Claxton K, Sculpher MJ. Decision Modeling for Health Economic Evaluation: Oxford University Press; 2006.
49. Drummond MF. Economic evaluation of treatment strategies in gastroenterology. Am J Gastroenterol 2005;100(10):2143-5.
50. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials: Oxford University Press; 2007.
51. Eddy DM. Clinical decision making: from theory to practice. Applying cost-effectiveness analysis. The inside story. JAMA 1992;268(18):2575-82.

52. Eddy DM. Clinical decision making: from theory to practice. Cost-effectiveness analysis. Will it be accepted? JAMA 1992;268(1):132-6.
53. Eddy DM. Clinical decision making: from theory to practice. Cost-effectiveness analysis. Is it up to the task? JAMA 1992;267(24):3342-8.
54. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1997;277(19):1552-7.
55. O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA 1997;277(22):1802-6.
56. How to use an Article on Economic Analysis of Clinical Practice., 1999. (Accessed 2008, at http://www.cche.net/usersguides/economic.asp.)
57. Neumann PJ. Using cost-effectiveness analysis to improve health care: Opportunities and barriers.: Oxford University Press; 2005.
58. Kruper L, Kurichi JE, Sonnad SS. Methodologic quality of cost-effectiveness analyses of surgical procedures. Ann Surg 2007;245(1):147-51.
59. Blackmore CC, Smith WJ. Economic analyses of radiological procedures: a methodological evaluation of the medical literature. Eur J Radiol 1998;27(2):123-30.
60. Bell CM, Urbach DR, Ray JG, et al. Bias in published cost effectiveness studies: systematic review. Br Med J 2006;332(7543):699-703.
61. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ 2004;13(5):437-52.
62. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. N Engl J Med 2003;349(22):2117-27.
63. Finlayson SR, Birkmeyer JD. Cost-effectiveness analysis in surgery. Surgery 1998;123(2):151-6.
64. Gillick MR. Medicare coverage for technological innovations--time for new criteria? N Engl J Med 2004;350(21):2199-203.
65. Tunis SR. Why Medicare has not established criteria for coverage decisions. N Engl J Med 2004;350(21):2196-8.
66. Neumann PJ, Sullivan SD. Economic evaluation in the US: what is the missing link? Pharmacoeconomics 2006;24(11):1163-8.

67. Vidal-Trecan GM, Stahl JE, Eckman MH. Radioiodine or surgery for toxic thyroid adenoma: dissecting an important decision. A cost-effectiveness analysis. Thyroid 2004;14(11):933-45.
68. Clegg A, Colquitt J, Sidhu M, Royle P, Walker A. Clinical and cost effectiveness of surgery for morbid obesity: a systematic review and economic evaluation. Int J Obes Relat Metab Disord 2003;27(10):1167-77.
69. Michaels JA, Drury D, Thomas SM. Cost-effectiveness of endovascular abdominal aortic aneurysm repair. Br J Surg 2005;92(8):960-7.
70. Visser K, de Vries SO, Kitslaar PJ, van Engelshoven JM, Hunink MG. Cost-effectiveness of diagnostic imaging work-up and treatment for patients with intermittent claudication in The Netherlands. Eur J Vasc Endovasc Surg 2003;25(3):213-23.
71. Van Den Brink M, Van Den Hout WB, Stiggelbout AM, et al. Cost-utility analysis of preoperative radiotherapy in patients with rectal cancer undergoing total mesorectal excision: a study of the Dutch Colorectal Cancer Group. J Clin Oncol 2004;22(2):244-53.
72. Longworth L, Young T, Buxton MJ, et al. Midterm cost-effectiveness of the liver transplantation program of England and Wales for three disease groups. Liver Transpl 2003;9(12):1295-307.
73. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. Med Decis Making 2002;22(4):359-71.
74. Jaisson-Hot I, Flourie B, Descos L, Colin C. Management for severe Crohn's disease: a lifetime cost-utility analysis. Int J Technol Assess Health Care 2004;20(3):274-9.
75. Polsky D, Mandelblatt JS, Weeks JC, et al. Economic evaluation of breast cancer treatment: considering the value of patient choice. J Clin Oncol 2003;21(6):1139-46.
76. Mahadevia PJ, Fleisher LA, Frick KD, Eng J, Goodman SN, Powe NR. Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis. JAMA 2003;289(3):313-22.
77. Heijenbrok-Kal MH. Assessment of Diagnostic Imaging Technologies for Cardiovascular Disease. Rotterdam, The Netherlands: Erasmus Medical Center; 2004.
78. Draper D. Assessment and Propagation of Model Uncertainty. Journal of Royal Statistical Society B 1995;57(1):45-97.
79. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. Value in Health 2005;8(1):1-2.
80. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Med Decis Making 2002;22(4):290-308.

81. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. Med Decis Making 1998;18(1):95-109.
82. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. Med Decis Making 2007;27(4):448-70.
83. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-38.
84. Heart disease and stroke statistics., 2008. (Accessed 2008, at www.americanheart.org.)
85. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation 1981;64(2):360-7.
86. Kuntz KM, Fleischmann KE, Hunink MG, Douglas PS. Cost-effectiveness of diagnostic strategies for patients with chest pain. Ann Intern Med 1999;130(9):709-18.
87. Keeney RL, Raiffa H. Decisions with Multiple Objectives: Preferences and Value Trade-offs. New-York: Wiley; 1976.
88. Stinnett AA, Paltiel AD. Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. Med Decis Making 1997;17(4):483-9.
89. Coggon DI, Martyn CN. Time and chance: the stochastic nature of disease causation. Lancet 2005;365(9468):1434-7.
90. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic analysis and computationally expensive models: Necessary and required? Value Health 2006;9(4):244-52.
91. Stevenson MD, Oakley J, Chilcott JB. Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. Med Decis Making 2004;24(1):89-100.
92. Briggs AH, Ades AE, Price MJ. Probabilistic sensitivity analysis for decision trees with multiple branches: use of the Dirichlet distribution in a Bayesian framework. Med Decis Making 2003;23(4):341-50.
93. Efron B TR. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
94. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non- parametric approach to confidence interval estimation. Health Econ 1997;6(4):327-40.
95. O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? Health Econ 2003;12(1):33-49.
96. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ 2001;10(8):779-87.
97. Lothgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. Health Econ 2000;9(7):623-30.

98. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ 2004;13(5):405-15.
99. Groot Koerkamp B, Hunink MG, Stijnen T, Hammitt JK, Kuntz KM, Weinstein MC. Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. Med Decis Making 2007;27(2):101-11.
100. Barton GR, Briggs AH, Fenwick EA. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI). Value Health 2008;11(5):886-97.
101. Halpern EF, Weinstein MC, Hunink MG, Gazelle GS. Representing both first- and second-order uncertainties by Monte Carlo simulation for groups of patients. Med Decis Making 2000;20(3):314-22.
102. Raiffa H, Schlaifer R. Probability and statistics for business decisions. New York: McGraw-Hill; 1959.
103. Raiffa H, Schlaifer R. Applied statistical decision theory. Cambridge, MA: Harvard University; 1961.
104. Yokota F, Thompson KM. Value of information analysis in environmental health risk management decisions: past, present, and future. Risk Anal 2004;24(3):635-50.
105. Groot Koerkamp B, Myriam Hunink MG, Stijnen T, Weinstein MC. Identifying key parameters in cost-effectiveness analysis using value of information: a comparison of methods. Health Econ 2006;15(4):383-92.
106. Howard RA. Information value theory. IEEE Transact Systems Science and Cybernetics 1966;SCC-2:22-6.
107. Thompson KM, Graham JD. Going beyond the single number: using probabilistic risk assessment to improve risk management. Hum Ecol Risk Assess 1996;2:1008-34.
108. Bojke L, Claxton K, Palmer S, Sculpher M. Defining and characterising structural uncertainty in decision analytic models. In. York: Centre for Health Economics (CHE); 2006.
109. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. Med Decis Making 2007;27(2):112-27.
110. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics 2008;26(9):799-806.
111. Nijhuis RL, Stijnen T, Peeters A, Witteman JC, Hofman A, Hunink MG. Apparent and Internal Validity of a Monte Carlo-Markov Model for Cardiovascular Disease in a Cohort Follow-up Study. Med Decis Making 2006;26(2):134-44.
112. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med 2003;22(23):3687-709.

113. Beck JR, Pauker SG. The Markov process in medical prognosis. Med Decis Making 1983;3:419-58.
114. Brown RV. Assessment uncertainty technology for making and defending risky decisions. J Behav Dec Making 1990;3:213-28.
115. Henrion M. The value of knowing how little you know: The advantages of probabilistic treatment of uncertainty in policy analysis. Pittsburgh: Carnegy Mellon University; 1982.
116. Thompson KM. Variability and uncertainty meet risk management and risk communication. Risk Anal 2002;22(3):647-54.
117. D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117(6):743-53.
118. CEA Registry of Tufts Medical Center in Boston, MA (https://research.tufts-nemc.org/cear/default.aspx)
119. Data Professional. In. release 6 ed. Williamstown, MA: TreeAge Software; 2001.
120. Craig BA, Black MA, Sendi P. Uncertainty in decision models analyzing cost-effectiveness. Med Decis Making 2000;20(1):135-7.
121. Hunink MG, Bult JR, de Vries J, Weinstein MC. Uncertainty in decision models analyzing cost-effectiveness: the joint distribution of incremental costs and effectiveness evaluated with a nonparametric bootstrap method. Med Decis Making 1998;18(3):337-46.
122. Morgan MG, Henrion M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. New York: Cambridge University Press; 1990.
123. Hoffman FO, Hammonds JS. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. Risk Anal 1994;14(5):707-12.
124. Slob W. Uncertainty analysis in multiplicative models. Risk Anal 1994;14(5):571-6.
125. Nayak TK, Kundu S. Calculating and describing uncertainty in risk assessment: the Bayesian approach. Hum Ecol Risk Assess 2001;7(2):307-28.
126. Frey HC, Burmaster DE. Methods for characterizing variability and uncertainty: comparison of bootstrap simulation and likelihood-based approaches. Risk Anal 1999;19(1):109-30.
127. Rai SN, Krewski D. A general framework for the analysis of uncertainty and variability in risk assessment. Hum Ecol Risk Assess 1996;2:972-89.
128. Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes: UKPDS 40. UK Prospective Diabetes Study Group. Br Med J 1998;317(7160):720-6.
129. Multicentre aneurysm screening study (MASS): cost effectiveness analysis of screening for abdominal aortic aneurysms based on four year results from randomised controlled trial. Br Med J 2002;325(7373):1135.

130. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: cost effectiveness of physical treatments for back pain in primary care. Br Med J 2004;329(7479):1381.
131. Bower P, Byford S, Barber J, et al. Meta-analysis of data on costs from trials of counselling in primary care: using individual patient data to overcome sample size limitations in economic analyses. Br Med J 2003;326(7401):1247-50.
132. Korthals-de Bos IB, Hoving JL, van Tulder MW, et al. Cost effectiveness of physiotherapy, manual therapy, and general practitioner care for neck pain: economic evaluation alongside a randomised controlled trial. Br Med J 2003;326(7395):911.
133. Martikainen J, Valtonen H, Pirttila T. Potential cost-effectiveness of a family-based program in mild Alzheimer's disease patients. Eur J Health Econ 2004;5(2):136-42.
134. Moayyedi P, Soo S, Deeks J, et al. Systematic review and economic evaluation of Helicobacter pylori eradication treatment for non-ulcer dyspepsia. Dyspepsia Review Group. Bmj 2000;321(7262):659-64.
135. Plant PK, Owen JL, Parrott S, Elliott MW. Cost effectiveness of ward based non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease: economic analysis of randomised controlled trial. Bmj 2003;326(7396):956.
136. Chilcott J, McCabe C, Tappenden P, et al. Modelling the cost effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. Commentary: evaluating disease modifying treatments in multiple sclerosis. Bmj 2003;326(7388):522; discussion
137. Legood R, Gray A, Wolstenholme J, Moss S. Lifetime effects, costs, and cost effectiveness of testing for human papillomavirus to manage low grade cytological abnormalities: results of the NHS pilot studies. Bmj 2006;332(7533):79-85.
138. Iglesias C, Nixon J, Cranny G, et al. Pressure relieving support surfaces (PRESSURE) trial: cost effectiveness analysis. Bmj 2006.
139. Rivero-Arias O, Campbell H, Gray A, Fairbank J, Frost H, Wilson-MacDonald J. Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial. Bmj 2005;330(7502):1239.
140. Raftery JP, Yao GL, Murchie P, Campbell NC, Ritchie LD. Cost effectiveness of nurse led secondary prevention clinics for coronary heart disease in primary care: follow up of a randomised controlled trial. Bmj 2005;330(7493):707.
141. Coast J, Noble S, Noble A, et al. Economic evaluation of a general practitioner with special interests led dermatology service in primary care. Bmj 2005;331(7530):1444-9.
142. Cost effectiveness of treatment for alcohol problems: findings of the randomised UK alcohol treatment trial (UKATT). Bmj 2005;331(7516):544.

143. Sculpher M, Manca A, Abbott J, Fountain J, Mason S, Garry R. Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. Bmj 2004;328(7432):134.
144. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. Int J Technol Assess Health Care 2001;17(1):38-55.
145. von Winterfeldt D, Edwards W. Flat maxima in linear optimization models. In: Technical report Number 011-313-4-T University of Michigan, Engineering Psychology Laboratory; 1973; Ann Harbor, MI; 1973.
146. Morgan MG, Henrion M. Uncertainty: A guide to dealing with uncertainty in quantitative risk policy analysis. New York: Cambridge University Press; 1990.
147. Raiffa H. Decision analysis: Introductory lectures on choises under uncertainty.: Addison-Wesley; 1968.
148. Poole C. Beyond the confidence interval. AJPH 1987;77(2):195-9.
149. Heitjan DF, Moskowitz AJ, Whang W. Problems with interval estimates of the incremental cost-effectiveness ratio. Med Decis Making 1999;19(1):9-15.
150. Kent DM, Fendrick AM, Langa KM. New and dis-improved: on the evaluation and use of less effective, less expensive medical interventions. Med Decis Making 2004;24(3):281-6.
151. Coyle D, Buxton MJ, O'Brien BJ. Measures of importance for economic analysis based on decision modeling. J of Clinical Epidemiology 2003;56:989-97.
152. Felli JC, Hazen GB. Do sensitivity analyses really capture problem sensitivity? An empirical analysis based on information value. Working paper. Monterey, CA: Defense Resource Management Institute, Naval Postgraduate School; 1997.
153. Brennan A, Chilcott J, Kharroubi S, O'Hagan A. A two level monte carlo approach to calculating expected value of perfect information: resolution of the uncertainty in methods.[abstract]. Med Decis Making 2002;22(6):552.
154. Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P. The role of modelling in prioritising and planning clinical trials. Health Technology Assessment 2003;7(23).
155. Payne N, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening: a rapid and systematic review. Health Technol Assess 2000;4(18):1-73.
156. Karnon J. Planning the efficient allocation of research funds: an adapted application of a non-parametric Bayesian value of information analysis. Health Policy 2002;61(3):329-47.
157. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Econ 1996;5(6):513-24.
158. Raiffa H, Schlaifer R. Applied Statistical Decision Theory: Harvard University; 1961.
159. Felli JC, Hazen GB. Correction: Sensitivity analysis and the Expected Value of Perfect Information. Med Decis Making 2003;23(1):97.

160. Thompson KM, Graham JD. Going beyond the single number: using probabilistic risk assessment to improve risk management. Hum Ecol Risk Assess 1996;2(4):1008-34.
161. Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. Health Technol Assess 2004;8(27):iii, 1-78.
162. Stevenson MD, Oakley J, Chilcott JB. Gaussian Process Modeling in Conjunction with Individual Patient Simulation Modeling: A Case Study Describing the Calculation of Cost-Effectiveness Ratios for the Treatment of Established Osteoporosis. Med Decis Making 2004;24:89-100.
163. Oakley J. Value of information for complex cost-effectiveness models. Sheffield: University of Sheffield; 2003. Report No.: 533/02.
164. Claxton K, Eggington S, Ginnelly L, Griffin S, McCabe C. A pilot study of value of information analysis to support research recommendations for the National Institute for Clinical Excellence. In. York, Sheffield, Nottingham; 2004:1-185.
165. Mowatt G, Cook JA, Hillis GS, et al. 64-slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart 2008.
166. Vanhoenacker PK, Heijenbrok-Kal MH, Van Heste R, et al. Diagnostic performance of multidetector CT angiography for assessment of coronary artery disease: meta-analysis. Radiology 2007;244(2):419-28.
167. Fox K, Garcia MA, Ardissino D, et al. Guidelines on the management of stable angina pectoris: executive summary: the Task Force on the Management of Stable Angina Pectoris of the European Society of Cardiology. Eur Heart J 2006;27(11):1341-81.
168. Wright JC, Weinstein MC. Gains in life expectancy from medical interventions - standardizing data on outcomes. N Eng J Med 1998;339(6):380-6.
169. Raiffa H. Decision Analysis: Introductory Lectures on Choices under Uncertainty. Reading, MA: Addison-Wesley; 1968.
170. Coyle D, Buxton MJ, O'Brien BJ. Measures of importance for economic analysis based on decision modeling. J Clin Epidemiol 2003;56(10):989-97.
171. Payne N, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening: a rapid and systematic review. Health Technol Assess 2000;4(18):1-73.
172. Claxton K, Thompson KM. A dynamic programming approach to the efficient design of clinical trials. J Health Econ 2001;20(5):797-822.
173. Detsky AS. Are clinical trials a cost-effective investment? Jama 1989;262(13):1795-800.

174. Spronk S, Bosch J, den Hoed P, Veen H, Pattynama P, Hunink M. Cost-effectiveness of endovascular revascularization compared to supervised hospital-based exercise training in patients with intermittent claudication: A randomized controlled trial. J Vasc Surg 2008;Dec;48(6):1472-80.
175. R. In. 1.7.1 ed: The R Development Core Team; 2003.
176. Microsoft. Excel. In; 2002.
177. Inter-Society Consensus for the Management op PAD. 2007. (Accessed at www.tasc-2-pad.org.)
178. Spronk S, Bosch J, den Hoed P, Veen H, Pattynama P, Hunink M. Clinical effectiveness of endovascular revascularization versus supervised hospital-based exercise training for intermittent claudication: A Randomized, Controlled Trial. Radiology; 2009 Feb;250(2):586-95.
179. de Vries SO, Visser K, de Vries JA, Wong JB, Donaldson MC, Hunink MG. Intermittent claudication: cost-effectiveness of revascularization versus exercise therapy. Radiology 2002;222(1):25-36.
180. Murphy TP, Ariaratnam NS, Carney WI, Jr., et al. Aortoiliac insufficiency: long-term experience with stent placement for treatment. Radiology 2004;231(1):243-9.
181. Dutch scoring algorithm for EuroQol-5D. (Accessed 2006, at www.xs4all.nl/~jannetvb/busschbach/faq_qoL&euroqol.htm.)
182. Zinnige en duurzame zorg. Zoetermeer, The Netherlands: Raad voor de Volksgezondheid en Zorg; 2006.
183. Anderson TW. An introduction to multivariate statistical analysis. 2 ed; 1984.
184. Detsky AS. Are clinical trials a cost-effective investment? JAMA 1989;262(13):1795-800.
185. Detsky AS. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. Statist Med 1990;9:173-84.
186. Willan AR. Clinical decision making and the expected value of information. Clinical Trials 2007;4:279-85.
187. Groot Koerkamp B, Nikken JJ, Oei EH, Stijnen T, Ginai AZ, Hunink MG. Value of information analysis used to determine the necessity of additional research: MR imaging in acute knee trauma as an example. Radiology 2008;246(2):420-5.
188. Nikken JJ, Oei EH, Ginai AZ, et al. Acute peripheral joint injury: cost and effectiveness of low-field-strength MR imaging--results of randomized controlled trial. Radiology 2005;236(3):958-67.
189. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095-108.
190. Gelman A, Carlin JG, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman and Hall; 1995.

191. Claxton K, Eggington S, Ginnelly L, Griffin S, McCabe C. A pilot study of value of information analysis to support research recommendations for the National Institute for Clinical Excellence. In. York, Sheffield, Nottingham; 2004:185.
192. Johnson LW, Lozner EC, Johnson S, et al. Coronary arteriography 1984-1987: a report of the Registry of the Society for Cardiac Angiography and Interventions. I. Results and complications. Cathet Cardiovasc Diagn 1989;17(1):5-10.
193. Lozner EC, Johnson LW, Johnson S, et al. Coronary arteriography 1984-1987: a report of the Registry of the Society for Cardiac Angiography and Interventions. II. An analysis of 218 deaths related to coronary arteriography. Cathet Cardiovasc Diagn 1989;17(1):11-4.
194. Nieman K, Cademartiri F, Lemos PA, Raaijmakers R, Pattynama PM, de Feyter PJ. Reliable noninvasive coronary angiography with fast submillimeter multislice spiral computed tomography. Circulation 2002;106(16):2051-4.
195. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. Jama 1998;280(10):913-20.
196. Rumberger JA, Behrenbeck T, Breen JF, Sheedy PF, 2nd. Coronary calcification by electron beam computed tomography and obstructive coronary artery disease: a model for costs and effectiveness of diagnosis as compared with conventional cardiac testing methods. J Am Coll Cardiol 1999;33(2):453-62.
197. Nease RF, Jr., Kneeland T, O'Connor GT, et al. Variation in patient utilities for outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. Jama 1995;273(15):1185-90.
198. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Survival data. Circulation 1983;68(5):939-50.
199. Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet 1994;344(8922):563-70.
200. Pocock SJ, Henderson RA, Rickards AF, et al. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. Lancet 1995;346(8984):1184-9.
201. Holmes DR, Jr., Holubkov R, Vlietstra RE, et al. Comparison of complications during percutaneous transluminal coronary angioplasty from 1977 to 1981 and from 1985 to 1986: the National Heart, Lung, and Blood Institute Percutaneous Transluminal Coronary Angioplasty Registry. J Am Coll Cardiol 1988;12(5):1149-55.

202. Yusuf S, Zucker D, Chalmers TC. Ten-year results of the randomized control trials of coronary artery bypass graft surgery: tabular data compiled by the collaborative effort of the original trial investigators. Part 2 of 2. Online J Curr Clin Trials 1994;Doc No 144:[3987 words; 31 paragraphs].
203. Yusuf S, Zucker D, Chalmers TC. Ten-year results of the randomized control trials of coronary artery bypass graft surgery: tabular data compiled by the collaborative effort of the original trial investigators. Part 1 of 2. Online J Curr Clin Trials 1994;Doc No 145:[3987 words; 38 paragraphs].
204. Wong JB, Sonnenberg FA, Salem DN, Pauker SG. Myocardial revascularization for chronic stable angina. Analysis of the role of percutaneous transluminal coronary angioplasty based on data available in 1989. Ann Intern Med 1990;113(11):852-71.
205. Meads C. -. Health Technol Assess 2000;4:1-153.
206. Hannan EL, Kilburn H, Jr., O'Donnell JF, Lukacik G, Shields EP. Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates. Jama 1990;264(21):2768-74.
207. Plato. Phaedrus: Hackett Pub Co.
208. Guide to the Methods of Technology Assessment. London: National Institute of Health and Clinical Excellence (NICE). 2004.
209. Caro JJ, Getsios D, Moller J. Regarding probabilistic analysis and computationally expensive models: necessary and required? Value Health 2007;10(4):317-8.
210. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics 2006;24(11):1043-53.
211. Fenwick E, Briggs A. Cost-effectiveness acceptability curves in the dock: case not proven? Med Decis Making 2007;27(2):93-5.
212. Bojke L, Claxton K, Sculpher MJ, Palmer S. Identifying research priorities: the value of information associated with repeat screening for age-related macular degeneration. Med Decis Making 2008;28(1):33-43.
213. Galani C, Al M, Schneider H, Rutten FF. Uncertainty in decision-making: value of additional information in the cost-effectiveness of lifestyle intervention in overweight and obese people. Value Health 2008;11(3):424-34.
214. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health 2009;12 Suppl 1:S5-9.

14

Contributing authors

Prof. K.E. Fleischmann
Division of Cardiology
UCSF Medical Center
San Francisco, CA, USA

Dr. A.Z. Ginai
Department of Radiology
Erasmus MC
Rotterdam, the Netherlands

Drs. T.S.S. Genders
Program for the Assessment of Radiological Technology
Departments of Radiology and Epidemiology
Erasmus MC
Rotterdam, the Netherlands

Prof. J.K. Hammitt
Department of Health Policy and Management
Harvard School of Public Health
Boston, MA, USA

Dr. M.H. Heijenbrok-Kal
Rijndam Rehabilitation Centre
Rotterdam, the Netherlands

Prof. dr. M.G.M. Hunink
Program for the Assessment of Radiological Technology
Departments of Radiology and Epidemiology
Erasmus MC
Rotterdam, the Netherlands
And:
Department of Health Policy and Management
Harvard School of Public Health
Boston, MA, USA

Prof. K.M. Kuntz
Department of Health Policy and Management
School of Public Health
University of Minnesota
Minneapolis, MN, USA

Dr. J.J. Nikken
Department of Radiology
Ziekenhuis Gelderse Vallei
Ede, the Netherlands

Dr. E.H. Oei
Department of Radiology
Erasmus MC
Rotterdam, the Netherlands

Dr. S. Spronk
Department of Epidemiology
Erasmus MC
Rotterdam, the Netherlands

Prof. dr. T. Stijnen
Department of Medical Statistics and Bioinformatics
Leiden University Medical Center
Leiden, the Netherlands

Prof. Y.C. Wang
Department of Health Policy and Management
Columbia Mailman School of Public Health
New York, NY, USA

Prof. M.C. Weinstein
Program in Health Decision Science
Department of Health Policy and Management
Harvard School of Public Health
Boston, MA, USA

List of Publications

Value of information analyses of economic randomized controlled trials: the treatment of intermittent claudication. Groot Koerkamp B, Spronk S, Stijnen T, Hunink MGM. *Value in Health*. 2009: In press.

Uncertainty and patient heterogeneity in medical decision models. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MGM. *Med Decis Making*. *2009*: In press.

Cost-effectiveness analysis for surgeons. Groot Koerkamp B, Wang YC, Hunink MGM. *Surgery*. 2009; 145(6):616-22.

Value of information analysis used to determine the necessity of additional research: MR imaging in acute knee trauma as an example. Groot Koerkamp B, Nikken JJ, Oei EH, Stijnen T, Ginai AZ, Hunink MGM. *Radiology*. 2008; 246(2):420-5.

Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. Groot Koerkamp B, Hunink MGM, Stijnen T, Hammitt JK, Kuntz KM, Weinstein MC. *Med Decis Making*. 2007; 27(2):101-11.

Identifying key parameters in cost-effectiveness analysis using value of information: a comparison of methods. Groot Koerkamp B, Myriam Hunink MGM, Stijnen T, Weinstein MC. *Health Econ*. 2006; 15(4):383-92.

Genomic FHIT analysis in RER+ and RER- adenocarcinomas of the pancreas. Hilgers W, Groot Koerkamp B, Geradts J, Tang DJ, Yeo CJ, Hruban RH, Kern SE. *Genes Chromosomes Cancer*. 2000; 27(3):239-43.

Novel homozygous deletions of chromosomal band 18q22 in pancreatic adenocarcinoma identified by STS marker scanning. Hilgers W, Su GH, Groot Koerkamp B, Tang DJ, Shekher MC, Sugar AY, Yeo CJ, Hruban RH, Kern SE. *Genes Chromosomes Cancer*. 1999; 25(4):370-5.

PhD portfolio

Name PhD student:	Bas Groot Koerkamp
Erasmus MC Department:	Epidemiology and Radiology
Research School:	Netherlands Institute for Health Sciences (NIHES)
PhD period:	2001-2009
Promotores:	Prof. dr. M.G.M. Hunink
	Prof. M.C. Weinstein
	Prof. dr. T. Stijnen

1. PhD training

Research skills

- Master of Science program in Clinical Epidemiology, Netherlands Institute for Health Sciences (NIHES), Rotterdam, the Netherlands, 2001-2003. Various courses in research methodology, including: Principles of Research in Medicine and Epidemiology, Study Design, Clinical Decision Analysis, Advanced Medical Decision Analysis, Introduction to Data-analysis, Regression Analysis, Topics in Evidence-based Medicine, Meta-analysis, Advanced Course on Diagnostic Research.
- Advanced courses in Causal Inference and Decision Theory, at the Harvard School of Public Health, Boston, USA, 2004.

Presentations

- Monte Carlo models in medical decision making, Karolinska Institute, Department of Epidemiology, Stockholm, Sweden, 2002.
- Value of information analysis directing further research: assessing the use of MRI for acute knee trauma. Annual Meeting of the Society for Medical Decision Making, Chicago, USA, 2003.
- Value of information analysis of a diagnostic randomized controlled trial: more research is not needed. Society for Medical Decision Making, Chicago, USA, 2003.
- Heterogeneity bias: pitfalls of the KISS-principle. Annual Meeting of the Society for Medical Decision Making, Chicago, USA, 2003. – poster presentation
- Value of information analysis directing further research. Erasmus MC, Department of Epidemiology and Biostatistics, Rotterdam, the Netherlands, 2003.
- Novel imaging technologies: is more research needed? Erasmus MC, Department of Radiology, Rotterdam, the Netherlands, 2003.
- Value of information analysis directing further research: assessing the use of MRI for acute knee trauma. Meeting for Medical Decision Analysis, UMC Radboud, Nijmegen, the Netherlands, 2003.

- Value of Information analysis directing further research: assessing the use of MRI for acute knee trauma. Annual Meeting and Scientific Assembly of the Radiological Society of North America, Chicago, USA, 2004. – presented by Dr. J.J. Nikken
- Limitations of acceptability curves. Annual Meeting of the Society for Medical Decision Making, Atlanta, USA, 2004.
- Competing diagnostic tests for coronary artery disease: which parameter uncertainties matter? Annual Meeting of the Society for Medical Decision Making, Atlanta, USA, 2004.
- Priority setting and optimizing study design in clinical research. Children's Hospital, Department of Neonatology, Boston, USA. 2004.
- Guiding future clinical cost-effectiveness studies of noninvasive diagnostic imaging tests for peripheral arterial disease. Annual Meeting and Scientific Assembly of the Radiological Society of North America, Chicago, USA, 2006. – presented by Dr. R. Ouwendijk

Awards

- Honorary Mention for excellent scientific research paper. Master of Science program in Clinical Epidemiology, Netherlands Institute for Health Sciences, 2003.
- Lee B. Lusted Student Prize. Prize for outstanding presentations of research at the Annual Meeting of the Society of Medical Decision Making, 2004.

Grants

- A.A. van Beek Fonds, 2001
- ZonMw, 2003
- Vereniging Trustfonds Erasmus Universiteit Rotterdam, 2003 and 2004
- Stichting De Drie Lichten, 2004
- Prins Bernhard Cultuurfonds, 2004

2. Teaching activities

- Teaching Fellow. Introduction to Decision Making in Medicine, Netherlands Institute for Health Sciences (NIHES), Rotterdam, the Netherlands, 2003.
- Teaching Fellow. Summer Program in Clinical Effectiveness, Harvard Medical School and Harvard School of Public Health, Boston, USA, 2004.
- Instructor and Teaching Fellow. Methods for Decision Analysis in Public Health and Medicine, Harvard School of Public Health, Boston, USA, 2004

Acknowledgements/dankwoord

Many people have contributed to this dissertation. I have been privileged to work with bright and thoughtful colleagues in both Rotterdam and Boston. At home I was fortunate to have the support and love of family and friends. I would like to thank all of them, in particular:

Prof. dr. M.G.M. Hunink, beste Myriam, jij bent de ideale promotor door een uitzonderlijke combinatie van intellect, oprechtheid en toewijding. Ik ben je dankbaar voor je persoonlijke begeleiding, vertrouwen en de mogelijkheden die je voor me hebt gecreëerd.

Prof. M.C. Weinstein, dear Milt, it was an honor and pleasure to work with you. Your experience and judgment were indispensable for my research. I am grateful for the inspiring year at Harvard.

Prof. dr. T. Stijnen, beste Theo, als geen ander kun jij statistiek helder uitleggen, zelfs aan een dokter. Een gesprek van een half uur met jou behoedde mij van radeloosheid en zette de wiskundige basis van mijn onderzoek steeds weer op het goede spoor.

Prof. dr. E.W. Steyerberg. Beste Ewout, in 2004 zaten we nog tegelijkertijd in Boston. Mooi dat je zo snel hoogleraar bent geworden om zitting te kunnen nemen in mijn kleine commissie. **Prof. dr. W.B.F. Brouwer**. Dank dat u zitting wilt nemen in mijn kleine commissie. Soms vind je een expert niet aan de andere kant van de wereld, maar aan de andere kant van de stad. **Prof. K. Claxton**. Dear Karl, it is an honor to have the godfather of VOI analysis in health care on my dissertation defense committee. **Prof. dr. A. Hofman**. Dank dat u zitting wilt nemen in mijn grote commissie. Dit proefschrift begon eigenlijk met een gesprek op uw kamer (voorjaar 2001), waarna u mij introduceerde aan Myriam Hunink. Uw Nihes opleiding is de perfecte bagage voor zowel wetenschappelijk onderzoek als patiëntenzorg. **Prof. dr. J. Kievit**. Het is een eer dat u als boegbeeld van de klinische besliskunde binnen de chirurgie, zitting wilt nemen in mijn grote commissie. **Prof. dr. F.F.H. Rutten**. Dank dat u zitting wilt nemen in mijn grote commissie.

Mijn paranimfen, **Michiel van der Heijden** en **Volkert Wreesmann**, mijn medico-vriendjes. Eindelijk hoop ik dan ook te promoveren en ben trots jullie 1 december aan mijn zijde te hebben.

ART-groep. Allereerst natuurlijk **Karen** en **Majanka** die mij wegwijs maakten in de klinische epidemiologie en besliskundige modellen. **Jeroen**, het was een eer jouw paranimf te zijn en natuurlijk bedankt dat ik jouw patiënten gegevens mocht gebruiken. **Rody**, dankzij jou kon ik nog af en toe een witte jas aantrekken om patiënten op de vaatpoli in je studie te includeren. **Ylian**, Comm Av – Darthmouth is de mooiste plek in Back Bay. **Sandra** en

Edwin, bedankt voor jullie tips en tricks voor de laatste loodjes. Ook alle anderen van de ART groep (**Ankie**, **Galied**, **Ineke**, **Joke**, **Marc**, **Rogier** en **Tessa**) wil ik bedanken voor het inspirerende klimaat en hun collegialiteit.

HSPH. Prof. K.M. Kuntz. Dear Karen, it was a great experience to be your teaching assistant. **Prof. J.K. Hammitt**. Dear Jim, your Decision Theory class was mathematically challenging, but it changed my perspective on decision making and uncertainty. Dear **Julia**, thanks for showing me the lunch variety around Harvard from Subway to falafel. Your L.L. Bean Tote bags for Sophie and Lidewij have become indispensable. Dear **Claire**, scientist and party girl, we share great memories such as the Halloween-party and the Four Seasons whisky tasting. Thanks for inviting me for the review in Surgery. All the others (**Amy**, **Jane**, **Jennifer**, **Jeremy**, **Natasha**, **Pamela**, **Stephen**, **Sun-Young**, and **Tanya**) it was a pleasure to work with so many clever and energetic colleagues. I would like to thank **Christine Bell** for her secretarial help.

Beste **Linda**, bedankt voor het editen van mijn Engelse brouwsels. Het is bijzonder dat jij niet alleen de Engelse grammatica en vocabulair beheerst, maar ook inhoudelijk mijn tekst verbeterde. Beste **Ton** en **Karin**, bedankt voor het verzorgen van de lay out van mijn proefschrift. De secretaressen van de 21e dank ik voor de goede samenwerking, in het bijzonder: **Marti**, **Petra**, **Marjolijn**, **Marion** en **Erica**.

Chirurgen en collega-assistenten van het AMC en Apeldoorn. Bedankt voor de goede werksfeer en alles wat ik van jullie leer. **Dr. Bouma** en **Dr. Busch**. Beste Wim en Olivier, dank voor het prettige opleidingsklimaat en de ruimte die jullie mij gegeven hebben om mijn promotie-onderzoek af te ronden.

Dear **Stephanie**, the other Swiss-American, thanks to you we lived in Back Bay and had a social network in Boston. We miss the live jazz in Wally's. **Mannen van de Druif**. Al weer 16 jaar samen. Dank voor jullie vriendschap en humor. Nog vele paasdiners, lustrums en skivakanties met club en herd om naar uit te kijken. **Mannen van Goed Bezig**. Onze discussie avonden in de Schreierstoren zijn me dierbaar.

Dear **John and Geneviève**. For many weeks in Kembs I was the son-in-law who is finishing his dissertation upstairs. Thank you for the great care and love. Lieve **Marjolijn** en **Marnix**, wat een geluk dat ik jullie als broertje en zusje heb. Lieve **papa** en **mama**, dank voor jullie onvoorwaardelijke steun, liefde en vertrouwen.

Mijn liefste **Salomé**, ich ha di so gärn. **Sophie en Lidewij**, kleine wondertjes, iedere dag laten jullie me lachen.

About the author

Bas Groot Koerkamp was born on June 26, 1975 in Eindhoven, the Netherlands. He graduated from the Gymnasium Celeanum in Zwolle, the Netherlands, in 1993. In the same year, he started a Master's program in Applied Physics at the Delft University of Technology, completing the first year (propaedeuse). In 1994 Bas enrolled in the Erasmus Medical School in Rotterdam, the Netherlands, from which he graduated cum laude in 2002. He took a sabbatical from his studies in 1998 in order to explore pancreatic cancer genetics at Johns Hopkins University in Baltimore, MD, USA, under the supervision of Prof. S.E. Kern.

In 2001 Bas started research at the Department of Epidemiology of the Erasmus MC under supervision of Prof. dr. M.G.M. Hunink, which research resulted in this dissertation. He obtained a Master's degree in Clinical Epidemiology at the Netherlands Institute for Health Sciences (NIHES) in Rotterdam in 2003. Part of the research presented in this dissertation was performed under supervision of Prof. M.C. Weinstein at the Harvard School of Public Health in Boston, MA, USA, where Bas worked in 2004 as Research Fellow and Teaching Assistant.

In 2005 he resumed his medical training with a residency in General Surgery at the Academic Medical Center in Amsterdam, the Netherlands (program directors: Prof. dr. D.J. Gouma, Prof. dr. J.J. van Lanschot, and currently Dr. O.R. Busch) and the Gelre Ziekenhuizen in Apeldoorn, the Netherlands (program director: Dr. W.H. Bouma). Bas Groot Koerkamp is married to Salomé and has two daughters, Sophie and Lidewij.