

HETEROGENEITY IN COST-EFFECTIVENESS ANALYSIS:

METHODS TO EXPLORE THE VALUE OF SUBGROUPS
AND INDIVIDUALIZED CARE IN A COLLECTIVELY
FUNDED HEALTH SYSTEM

A thesis submitted in fulfilment of the
requirements for the degree of PhD
by

Manuel Antonio Espinoza, MD, MSc

University of York
Department of Economics and Related Studies

September 2012

ABSTRACT

Cost-effectiveness analysis is increasingly being used to support decisions about the allocation of resources in health systems. However, decisions based on cost-effectiveness are usually made under imperfect and incomplete information. This thesis examines the value of understanding and characterizing heterogeneity for decision-making in healthcare. It proposes a methodological framework for a systematic cost-effectiveness subgroup analysis, providing guidance for identification and selection of subgroups. It suggests that the value of heterogeneity should be examined considering two dimensions: the value of making different decisions in different subgroups with current information (here termed static value); and the value of resolving parameter uncertainty conditional to a particular level of heterogeneity (here termed dynamic value). Finally, it provides empirical demonstration of such a framework through its implementation with a real case study.

The study of heterogeneity for decision-making led to an examination of the policy agenda for individualization of care. This thesis also presents a conceptual framework to address two aspects of the implementation of individualized decisions. First, the implementation of unrestricted choices for treatment responds to a positive value judgement, which is based on the expected health loss (or gain) associated to it. This chapter presents a novel analytical approach to estimating this magnitude, based on the characterization of the joint distribution of potential outcomes. Second, it is acknowledged that implementing individualized decisions also requires a normative judgement. This refers to a broader question in economic evaluation in health, which is the type of values that society seeks to maximise. Three categories are proposed here (paternalist, altruistic and welfarist) that define alternative normative positions. The positive and normative elements of this framework are illustrated with a real case study.

In conclusion, this thesis is a contribution to the development of methods for economic evaluation in healthcare, in particular, for understanding, characterizing and accounting for heterogeneity between members of the population.

Table of Contents

ABSTRACT	2
Table of Contents	3
List of Figures	6
List of Tables.....	7
Acknowledgments.....	8
Dissemination of the thesis	10
Author’s Declaration.....	11
1. CHAPTER 1: INTRODUCTION	12
1.1. Introduction	12
1.2. Health technology assessment for healthcare decision-making.....	14
1.3. Economic evaluation in health technology assessment.....	15
1.4. Cost-effectiveness analysis: A cornerstone in the understanding of value to the healthcare system	18
1.5. Objectives and structure of this thesis.....	23
2. CHAPTER 2 LITERATURE REVIEW	27
2.1. Introduction	27
2.2. Standard approaches to assess heterogeneity in evaluation of healthcare technologies.....	30
2.3. Value of Heterogeneity	36
2.4. Preferences and choice as a source of heterogeneity.....	38
2.5. Discussion	43
3. CHAPTER 3 THE VALUE OF HETEROGENEITY	47
3.1. Background	47
3.2. Heterogeneity and decision-making in healthcare	48
3.3. Subgroup cost-effectiveness analysis under current information.....	51

3.4.	Decision uncertainty in cost-effectiveness subgroup analysis	58
3.5.	Presentation of subgroup analysis and choice of the optimal number of subgroups: the role of the cost function	67
3.6.	Brief illustrative example	70
3.7.	Discussion	73
4.	CHAPTER 4: AN APPLICATION OF THE VALUE OF HETEROGENEITY FOR COST-EFFECTIVENESS SUBGROUP ANALYSIS	79
4.1.	Introduction	79
4.2.	Motivational example: RITA-3	80
4.3.	Methods	83
4.4.	Results	87
4.5.	Discussion	102
4.6.	Conclusions	108
5.	CHAPTER 5 INDIVIDUAL DECISIONS AND SOCIAL VALUE:	110
5.1.	Introduction	110
5.2.	Individual decisions and social value	113
5.3.	Alternative approaches to decision-making: a descriptive framework	128
5.4.	Alternative approaches to decision-making: a case study	130
5.5.	Discussion	133
6.	CHAPTER 6: DISCUSSION AND CONCLUSIONS	138
6.1.	Summary of the thesis contribution.....	138
6.2.	Discussion and opportunities for further research.....	145
6.3.	Recommendations for analysts and decision-makers.....	148
6.4.	Final conclusion	150
7.	APPENDIX 1 The expected Value of Individualized Care (EVIC) and the Expected Value of Perfect Information (EVPI): A source of confusion	152
8.	APPENDIX 2: Clinical effectiveness and cost-effectiveness of RITA-3 trial ..	156

9. APPENDIX 3: Cost-effectiveness subgroup analysis of the RITA-3 trial: details per subgroup	162
10. APPENDIX 4: Estimation of the Expected Value of Individualized Care (EVIC) and parameter-specific EVIC.....	169
11. APPENDIX 5: Stylized example of the extended method.....	172
12. APPENDIX 6: Societal versus patient values.....	175
13. List of Abbreviations.....	177
14. REFERENCES.....	179

List of Figures

Figure 3.1. Efficiency Frontier for Subgroup Analysis.....	57
Figure 3.2. The value of information with and without subgroups.....	62
Figure 3.3. Static and dynamic value of heterogeneity.....	64
Figure 3.4. Dynamic value of heterogeneity and decrease in EVPI.	65
Figure 3.5. Dynamic value of heterogeneity and its impact on subgroup specification.	66
Figure 3.6. Proportional and non-proportional value of heterogeneity.....	67
Figure 3.7. Representation of the gaining in expected NHB with current and perfect information when heterogeneity is considered.	68
Figure 4.1. Structure of the RITA-3 decision model.	81
Figure 4.2. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The two subgroup case	92
Figure 4.3. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The four subgroup case	95
Figure 4.4. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The eight subgroup case	96
Figure 4.5. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The sixteen subgroup case.....	97
Figure 4.6. Efficiency Frontier for Subgroup Analysis. Specifications that maximise the expected Net Health Benefits for each level of disaggregation.	98
Figure 4.7. Guideline for the specification that combines six covariates	99
Figure 4.8. Expected Net Health Benefits under current and perfect information for different level of disaggregation.	100
Figure 5.1. Social versus individual conflict: the case of rejection of a technology	117
Figure 5.2. The societal versus individual conflict: A stylized example.	121
Figure 5.3. Estimated joint distribution of potential outcomes from RITA-3 trial..	131
Figure 5.4. Minimum cost of restriction to consider decentralized decisions as the best option.	132

List of Tables

Table 2.1. Terminology in the study of heterogeneity. Relationship between different terms and the field where it is used.	32
Table 3.1. Total Net Health Benefits estimation for three treatments and four subgroups.	54
Table 3.2. Definition of subgroups. Grouping the population according to different specifications for different levels of disaggregation.	56
Table 3.3. Value of heterogeneity. A stylized example with two subgroups.	71
Table 3.4. Value of heterogeneity. A stylized example with three subgroups.	72
Table 4.1. Incremental cost-effectiveness ratios (ICERs) of the RITA-3 cost-effectiveness study for subgroups.	81
Table 4.2. Average marginal effects of nine covariates on the probability that the invasive strategy is the most cost-effective at a threshold of \$20,000 per QALY. Results based on a multivariable logit model.	87
Table 4.3. Cost-effectiveness subgroup analysis of RITA-3 trial: The two subgroup case.	92
Table 4.4. Cost-effectiveness subgroup analysis of RITA-3 trial: The four subgroup case.	94
Table 4.5. Subgroup analysis for additional combinations.	96
Table 4.6. Subgroup analysis for the baseline risk score.	101
Table 4.7. Akaike Information Criterion (AIC) of equation models with and without interaction with treatment effect.	102
Table 5.1. Sub-areas of the joint distribution of potential outcomes as societal gains or losses under alternative decisions.	118
Table 5.2. Summary of six approaches for decision-making in healthcare.	129

Acknowledgments

This thesis was carried out during the years 2009-2012 at the Centre for Health Economics of the University of York, at the Team for Economic Evaluation and Health Technology Assessment. This work was supported by the Scholarship “Beca Presidente de la República”, Ministry of Education of Chile. I would like to express my gratitude to this program. I also thank the Faculty of Medicine at Pontificia Universidad Católica de Chile and Hospital de Urgencias Asistencia Pública for their complementary financial support during these years.

This thesis is the result of my personal work during three years. However, during this period I have received important assistance and support from many people that I would like to acknowledge:

To my wife Báltica and our daughter Antonia, who have been my personal support during these years. I owe my deepest gratitude to them, for their love, patience and understanding.

To my supervisors, Dr. Andrea Manca and Professor Mark Sculpher, for the time dedicated to my development, the advice and support during these three years.

To my colleague Pedro Saramago who became one of my closest friends. I would like to thank him for his friendship, support and professional help during the whole PhD process.

To the members of the Centre for Health Economics, and especially to the members of the Team for Economic Evaluations and Health Technology Assessment. In particular I would like to thank Marta Soares, Susan Griffin and my PhD colleagues Ronan Mahon and Thomas Patton, who have generously shared their time, providing me with advice and support in different stages of my work.

To the former and current administration of the “Hospital de Urgencias Asistencia Pública in Santiago, Chile”. I am very grateful for their support during these four years in the UK. In particular, I would like to thank Mrs. Sonia Palominos, member

of staff at the Department of Emergency at Hospital de Urgencias Asistencia Publica, who has been a constant support.

To Professor Dr. Ignacio Sanchez, former dean of the Faculty of Medicine and current Rector of the Pontificia Universidad Católica de Chile, who believed in me and my professional development in the field of Health Economics.

Finally, I would like to make a very special mention to Professor Karl Claxton, member of my Thesis Advisory Group. I owe a great debt of gratitude to him, for his generosity, continuous encouragement and support in the most difficult periods of the PhD process.

Dissemination of the thesis

The work developed for this thesis has been communicated in the last two conferences of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Both presentations have been awarded with the prize “Best Student Podium Presentation”.

The abstracts have been published in the corresponding meeting proceeding, which is a special issue of the journal *Value in Health*. In addition, the presentations are available in the ISPOR website.

Espinoza, M.A., Manca, A., Claxton, K. & Sculpher, M. J. 2011. The Value of Heterogeneity for Cost-Effectiveness Subgroup Analysis: Theoretical Framework and Application. *Value in Health*, 14, A241-A241.

Presented at the ISPOR 14th Annual European Congress, 5-8 November 2011. Madrid, Spain. Available in: <http://www.ispor.org/awards/14euro/HG4-THE-VALUE-OF-HETEROGENEITY-FOR-COST-EFFECTIVENESS-SUBGROUP-ANALYSIS.pdf>

Espinoza, M. A., Manca, A., Sculpher, M. J. & Claxton, K. 2012. Individual Decisions and Social Value: A Conceptual Framework to Explore Alternative Decision-making Approaches and the Value of Heterogeneity in the Era of Individualized Care. *Value in Health*, 15, A2-A2.

Presented at the ISPOR 17th Annual International Meeting, June 2-6 2012. Washington DC, USA. Available in: <http://www.ispor.org/awards/17Meet/CO1-Individual-Decisions-and-Social-Value-ME.pdf>

Author's Declaration

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

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

1. CHAPTER 1:

INTRODUCTION

The Value of Health Technology Assessment and Economic Evaluation of Healthcare Interventions: An overview of the motivation of this thesis

1.1. Introduction

“Better health is of course the raison d’être of a health system, and unquestionably its primary or defining goal: if health systems did nothing to protect or improve health there would be no reason for them.”

The World Health Report (2000)

Health is the primary goal of a healthcare system. The aim of social planners in the health system is to make decisions such that limited resources can be allocated consistently with the maximization of health. However, this is not an easy task, primarily because the health market has some particularities that impose barriers for an optimal performance requiring governmental regulation (Dolan and Olsen, 2002). These barriers include demand and supply side market failures and most of the efforts undertaken in the field of health economics are an attempt to solve them. From a consequentialist point of view, any better understanding of these failures and methods to address them represents additional value to the system because they can be expressed in terms of potential health gains. The research undertaken in this thesis is focused on some particular elements of the demand side for healthcare, the further understanding of which might impact positively on the goals of the healthcare system.

In principle, people demand health from the healthcare system. However, people cannot demand health directly but only the inputs to produce health; this is what we call healthcare (Grossman, 1972). The demand for inputs requires knowledge about

the disease process and the alternative courses of action available for particular conditions - something that people usually do not have at the moment they need healthcare. In other words, people demand healthcare in conditions of uncertainty. This lack of information is a key characteristic of the functioning of a healthcare system (Arrow, 1963). Patients do not have comparative information about diagnostic tools and treatments, which force them to trust the health professional as a main source of information in order to make their decisions. Thus, a better functioning of the health market assumes, on one side, a fully informed health professional and, on the other side, a perfect principal-agency relationship between patient and doctor. The latter avoids any asymmetry of information so that patients can make decisions fully informed (Dolan and Olsen, 2002). This type of uncertainty can be called *ex-post* because it occurs when patients have a disease. Another type of uncertainty in the healthcare market is the *ex-ante uncertainty*. This is present when people are at risk of having a disease but they are not ill yet. It arises from the need of health systems to implement programmes for financial protection of the population. Where several insurance plans are available, people need to make choices from them with limited information. In the case of a single insurance system, information is even more important because it will define the inputs that should be included in a coverage set. Thus, the interaction of the individuals and the healthcare system is surrounded by uncertainty, whatever the point in time. This implies that the demand for healthcare is intimately associated with demand for information about healthcare.

Health technology assessment (HTA), of which economic evaluation of healthcare is part, is a scientific process created in response to the need for better information for decision-making in healthcare. This chapter presents an overview of HTA and economic evaluation in their role of supporting decision-making in health care. The objective is to offer a broad context of the role of HTA and, in particular, of economic evaluations in their capacity for exploring value in terms of health outcomes. The reader will find in this chapter some basic definitions that will be referred to in the subsequent chapters in this thesis. Finally, this overview contextualises the motivation of this thesis in pursuing a better understanding of the value of heterogeneity for healthcare decision-making and implications for the individualization of care.

1.2. Health technology assessment for healthcare decision-making

It has been argued that if the objective of the healthcare system is to produce the maximum amount of health with the limited resources defined exogenously, the aggregation of the demand of the healthcare system is a rational policy (Culyer, 1990). Three elements need to be considered to respond this policy. First, health systems should pursue efficiency as a goal. Second, decision-makers need more information about health needs and health outcomes. And third, financial protection schemes (i.e. insurance) should be centralized. In this section, it is argued that a HTA process, including economic evaluation, is consistent with the idea of a collective expression of demand through the specification and regulation of the needs of the population.

HTA has been defined as a “multidisciplinary field of policy analysis. It studies the medical, social, ethical, and economic implications of development, diffusion, and use of health technology” (INAHTA, 2012). Its main role is to provide decision-makers with high quality information resulting from a systematic and comprehensive process of evaluation of the available evidence. The implementation of a centralized HTA process, for example a national agency, is consistent with an adequate characterization of the demand for healthcare of the population (EunetHTA Work Package 8.). Throughout this process, a central authority is able to define the needs of the population and the alternative courses of action for each situation that maximises the health of the population. In this respect, Culyer (1990) says: “*I contend that all of the major ideological strengths of the NHS relate to the characteristics of demand. The job of the supply side is simply (!) to be cost-effective at meeting whatever demands are placed upon it by the demand side....The supply side is not judged by ideological but by practical criteria.*” (page 12).

A centralized process of HTA is efficient because it uses the best capacity of a jurisdiction¹ to undertake the evaluation instead of leaving it to doctors and patients.

¹ The term jurisdiction has been defined as the geographical region governed by a health authority able to make independent decisions about adoption or rejection of health technologies as well as their implementation (Drummond et al., 2009).

Doctors are not expected to assess the evidence as a centralized HTA process², but even if they were able, the process would be repeated several times, becoming inefficient. Neither would it take into account costs falling on others and benefits received by others. In addition, if the health system has monopsony power, HTA offers an opportunity for national (or jurisdictional) negotiation of price and alternative coverage schemes, since the assessment process allow analysts to identify the price that the system is willing to pay (Walker et al., 2012). If no monopsony power exists, HTA has value in informing the inputs of healthcare that should be included in a plan of financial protection. Finally, a less acknowledged effect of HTA is on innovation. The HTA process can identify need for innovation and provide the incentives for the development of technologies aligned with the needs of the population.

An indisputable element of the HTA process is the evaluation of effectiveness and safety. The analysis of these elements has followed the principles of evidence based medicine which are grounded on the epidemiological paradigm (Straus et al., 2011). This approach responds to the judgement as to whether a particular estimate of treatment effect can be considered a causal effect. As a consequence, the approach adopts a rather conservative position with respect to uncertainty from the viewpoint of a decision-maker. Conclusions such as *more evidence is needed* or *the results are insufficient to conclude in favour of one of the treatments* are useless for policy makers, who need to make a decision under conditions of uncertainty. Even when the decision is to postpone the decision, social planners need to justify this in terms of the opportunity costs that such delay might impose on society (Phillips, 2001, Griffin et al., 2011). In this context, economic evaluation becomes a central part of the HTA process. The fundamental and main methods for healthcare economic evaluation are presented in the next section.

1.3. Economic evaluation in health technology assessment

Neoclassical theory in welfare economics indicates that where markets work well, individuals will express their preferences making choices that lead to an optimal

² A central agency not only should be able to conduct a high quality assessment of the evidence but also should have access to more and more detailed information than single individuals (patients and individual doctors).

allocation of resources. In contrast, when markets fail, central authorities must intervene regulating private choice (Boardman et al., 2006). Economic evaluation is a tool used to support decision-makers in the allocation of resources on the basis of an efficiency criterion. It has been defined as the “*comparative analysis of alternative courses of action in terms of both their costs and benefits*” (Drummond et al., 2005) (page 9) and their different approaches are used across different areas of the economy.

The earliest version of economic evaluation is cost-benefit analysis (CBA). CBA is founded on the grounds of neoclassical welfare economics, where social welfare is understood as the sum of individual welfares measured as individual utilities³. Further, CBA provides results consistent with the Pareto principle, i.e. social welfare increases only if the utility of at least one individual increases, and no one else’s falls. Given the restrictions of meeting the Pareto criterion in practice, the potential Pareto criterion has been suggested as more feasible. This is based on the Kaldor-Hicks criterion, which states that an intervention increases welfare, if those who gain benefits can fully compensate those who lose and still increase their welfare (Boardman et al., 2006). Classical descriptions of CBA emphasize the use of monetized individual outcomes to undertake this analysis. It has been argued that where markets do not exist, such as in healthcare, the estimation of the outcomes in terms of both the maximum willingness to pay (WTP) in the case of winners and willingness to accept (WTA) in the case of losers provide an appropriate valuation of utilities (Donaldson, 1999, Klose, 1999). A positive net benefit in CBA analysis is consistent with a potential Pareto criterion. Some of the criticisms of this approach are that values are affected by ability to pay, it ignores the initial health endowment, and the comparisons are restricted to uncompensated gains or losses (Brouwer et al., 2008, Dolan and Olsen, 2002). Although defenders of CBA suggest that economic evaluation of healthcare should not be different from other sectors (Pauly, 1995), several authors have criticized the adoption of a welfarist approach, not only to

³ The utility derived from one good or state is a representation of satisfaction. In addition, neoclassical welfare economics assumes individual sovereignty (i.e. individuals are the best judges to define a ranking of preferences), welfarism (i.e. welfare derived from a good or state can only be judged on the basis of utility) and that utility is derived exclusively from outcomes (consequentialism) (Hurley, 2000).

healthcare evaluation⁴ (Claxton et al., 2007), but also as a general approach to the evaluation of policy programs for welfare improvement (Sen, 1980).

Cost-effectiveness analysis (CEA) comes as an alternative to CBA for its use in healthcare decision-making. While CBA pursues recommendations about choices to improve social welfare, CEA is motivated by a less ambitious objective, the maximization of health subject to budget constraint (Brouwer and Koopmanschap, 2000), where health is one part of the social welfare function. CEA responds to the need of a legitimate health authority that is not able to make an explicit formulation of a social welfare function but only make recommendations about alternative allocations of an exogenous health budget focused on the main objective of the healthcare system, i.e. health (Claxton et al., 2011b). CEA has been criticized because it does not follow a well formulated and accepted theory as does neoclassical welfare economics (Tsuchiya and Williams, 2001). In response, CEA has been framed in an alternative view called extra-welfarism. This theory transcends the welfarist view in that relevant outcomes to measure welfare can be other than utilities. The classical extra-welfarist view in healthcare is that the only relevant argument is health (Culyer, 1989). Thus, the economic analysis should be understood as a tool to produce evidence that helps decision-makers to allocate resources of a fixed health budget. In addition, outcomes can be valued by the affected individuals but also by a legitimate authority, for example, through a normative set of representative values of the population, where the population provides legitimacy (Brouwer et al., 2008). Another more pragmatic view suggests that any person considered with ethical authority, for example as a result of a democratic legitimate process, can define the objective of any system (e.g. healthcare system). The subsequent analysis should produce evidence to help this legitimate authority, called the “decision-maker”, making decisions about allocation of a fixed budget in response to that (those) objectives. This defines the decision-maker approach (Sugden and Williams, 1978). Cost-effectiveness analysis has become a standard method for economic evaluation of healthcare technologies (Drummond,

⁴ Most of the arguments in healthcare economic evaluation focus on the difficulties of valuing health outcomes in monetary terms, for example, the robustness of hypothetical exercises in contingent valuation methods or the fact that willingness to pay is affected by ability to pay (Gold et al., 1996, Drummond et al., 2005), but also on the criticism of the fundamentals of the welfarist approach, i.e. individuals focused on individual interests and utilities, as a legitimate and independent part of the social welfare function (Brouwer et al., 2008).

2012). Many countries have adopted this approach to provide recommendations for pricing and reimbursement of health technologies (ISPOR, 2012). More recently, some jurisdictions have used economic analysis derived from standard CEA to make decisions about the additional research needed to resolve current uncertainty (Claxton and Sculpher, 2006). In general, any methodological development in this field responds to the need of producing better information for decision-makers. The next section presents the main elements of CEA emphasizing the areas where economic analysis offer additional value to the HTA process.

1.4. Cost-effectiveness analysis: A cornerstone in the understanding of value to the healthcare system

The comparative analysis in CEA is based on costs measured in monetary terms and benefits measured in a unit of health. Although health can be described and measured by some clinical specific parameters (e.g. weight, blood pressure), CEA has adopted a broader conceptualization of health in order to provide information for a comparative analysis among different activities developed within the health system, which is consistent with an extra-welfarist view. In this context, a health metric should incorporate one element of quantity (e.g. length of life) and quality related to health (Torrance, 2006). In addition, it has been argued that health related quality of life (HRQoL) is in part due to the objective consequences of the disease (impairments, disabilities and handicaps) and in part to the social participation of individuals. The fact that social participation depends on the preferences of individuals has become an important reason to construct a health metric that incorporates preferences (Brazier et al., 2007). One of the metrics that meet these conditions is the Quality Adjusted Life Year (QALY)⁵, which has become one of the most used units of health in CEA.

Costs and benefits (QALYs) are usually estimated from individual patient data or aggregated data or both using decision models (Sculpher et al., 2000). The modelling

⁵ QALYs are constructed on the basis of specific weights that reflect the relative value of a health state compared to the best imaginable health state (full health). These values respond to the preferences of individuals for alternative health states. Health states are descriptive states based on health dimensions that can include impairment, disabilities and social participation. When health states are described considering all those dimensions, for example based on the instruments EQ-5D or SF-6D, QALYs can be considered an appropriate metric of health and can be consistently used in CEA.

process includes some normative considerations, for example, perspective and discounting, but also some technical issues such as the construction of a feasible model or the methods to synthesize evidence and estimation of parameters (Briggs and Sculpher, 1998). This stage of the process ends with the estimation of expected (mean) costs and benefits for each alternative in the evaluation. Important developments in this area have been achieved in the last twenty years, and it is still an area where additional developments represent value for healthcare decision-making (Briggs et al., 2006).

The cost-effectiveness of a particular healthcare intervention is assessed comparing the incremental cost per unit of additional health (cost per QALY) and the opportunity cost of the healthcare budget ($1/\lambda$), i.e. the health forgone elsewhere in the health system due to the new allocation of resources. This can be represented as a decision rule where one treatment can be considered cost-effective compared to another insofar as the incremental cost-effectiveness ratio (ICER) is lower than the cost-effectiveness threshold (λ)⁶.

$$\frac{(C_1 - C_0)}{(B_1 - B_0)} < \lambda \quad (1.1)$$

where C_1 and C_0 are the costs and B_1 and B_0 are the benefits of the new (1) and old (0) treatments respectively. An alternative decision rule can be presented as a linear transformation of the previous expression (Stinnett and Mullahy, 1998). Thus, net monetary benefits can be expressed as a function of the cost-effectiveness threshold such that:

$$\lambda(B_1 - B_0) - (C_1 - C_0) > 0 \quad (1.2)$$

The same transformation can be done to express net benefits in health terms:

$$(B_1 - B_0) - \frac{(C_1 - C_0)}{\lambda} > 0 \quad (1.3)$$

⁶ This decision rule assumes that all costs fall in the healthcare budget, therefore, the cost-effectiveness threshold represents the opportunity cost of the alternative use of resources of the healthcare system and it does not include private consumption of healthcare.

As mentioned earlier in this section, costs and benefits correspond to expected values. The idea of adopting a decision rule based on expected values might be seen as opposed to the classical decision rules in statistical inference, as adopted by clinical epidemiology and evidence based medicine. However, it has been argued that for decision-making under uncertainty, and especially for decisions affecting social interests, a decision rule based on the expected utility theory is consistent with the objective of a social planner (Arrow and Lind, 1970). In the context of CEA for healthcare, in particular, Claxton (1999b) has argued that “*the rules of classical statistical inference and its Bayesian counterpart are arbitrary and inconsistent with the objectives of any coherent healthcare system and impose unnecessary costs*” (page 347). In other words, decisions regarding the cost-effectiveness of new healthcare technologies should be based on the mean net (health) benefits⁷.

According to Claxton, the fact that the adoption of a new technology is based on the mean net benefit (and not its confidence interval) should not lead us to consider that uncertainty is not relevant for the decision-making process (Claxton and Posnett, 1996, Claxton, 1999b). Indeed, uncertainty is relevant for a second and simultaneous question, whether to conduct further research to solve the current uncertainty represents a good use of limited resources. Moreover, decision-makers should consider whether the expected benefits of immediate adoption of new technologies (which is a disincentive of further research) is greater than the value that new research might have provided for future patients (Griffin et al., 2008). This might lead to postpone early adoption until the uncertainty is resolved. Likewise, even when technologies are expected to be cost-effective, the presence of significant unrecoverable costs⁸ might lead decision-makers to approve those interventions conditional to further research, for example, “approval with research” or “only in research” (Walker et al., 2012, Claxton et al., 2011a).

⁷ Expected Net (Health) Benefit corresponds to the integral of the density function of the net health benefits. Consequently, it can be interpreted as a weighted average of the all possible consequences of alternative decisions based on net health benefits and their probabilities of occurrence.

⁸ Irrecoverable costs refer to those costs which cannot be recovered at any time after the implementation of the guidance. They include capital expenditure of equipment or healthcare facilities, implementation of the guidance such as dissemination or training the staff in the utilization of the new technology.

The study of uncertainty in CEA has concentrated many efforts in the last ten years. Uncertainty can be categorized into three types, namely, methodological, structural and parameter or stochastic uncertainty (Briggs, 2001). Methodological uncertainty refers to the alternative methods adopted by different analysts to conduct an economic evaluation in healthcare. This source of uncertainty has been largely addressed through the harmonization of methods and generation of methodological guidelines (ISPOR, 2012). A second source is structural uncertainty, which corresponds to the set of scientific judgments and simplifications that analysts must make alongside the process whereby the cost-effectiveness results are estimated and interpreted⁹ (Bojke et al., 2009b). Finally, parameter uncertainty refers to the impossibility of collecting an infinite sample, which implies that any parameter estimation cannot be totally precise and, as a consequence, uncertain. A final, and often less acknowledged type of uncertainty, is unrevealed heterogeneity (O'Brien and Sculpher, 2000). Cost-effectiveness estimates are usually presented as average estimates without regard to the differences between individuals in terms of their costs and consequences. This lack of information responds to the intrinsic variability that cannot be solved with additional samples.

As a consequence of the previous explanations, uncertainty in CEA can be interpreted as an opportunity for additional value rather than a barrier for decision-making. Important developments have been made in the last few years in this field. For example, methods to estimate the value of information (VoI) have been applied in healthcare to assess the value of further research conducted to resolve parameter uncertainty¹⁰. Analysts can estimate the Expected Value of Perfect Information (EVPI) from the distribution of net (health) benefits. This metric expresses the value of collecting an infinite sample in order to obtain the true mean parameter (Ginnelly et al., 2005). It also represents an upper boundary and, therefore, a necessary condition for decisions about further research. In other words, the costs of any research proposal should not exceed the EVPI. However, because one of the goals of

⁹ The sources of structural uncertainty have been categorized in four types: (a) selection of relevant comparators, (b) selection of relevant events, (c) alternative statistical methods and (d) clinical uncertainty that affect the structure of the decision model. For a detailed review of these elements and alternative remedial actions see (Jackson et al., 2011) and (Bojke et al., 2009b).

¹⁰ This analysis requires the characterization of parameter uncertainty through a probabilistic sensitivity analysis in order to estimate the probability density function of net (health) benefits. For details see (O'Hagan et al., 2005) and (Fenwick et al., 2001).

the health system is efficiency, health resources should be spent in an optimal manner, something that it is also relevant for the part of the budget allocated to research. Consequently, decision-makers should be interested in two elements. First, which parameters explain most of the uncertainty so that specific research can be prioritized and conducted to resolve the uncertainty around those specific parameters, and second, the optimal sample size considering the additional value that each sample unit adds to the decrease of decision uncertainty.

The Expected Value of Perfect Information for Parameters (EVPPI) can be estimated for single parameters that have been considered in the decision model, for example, HRQoL values or treatment effect (Griffin et al., 2010). As a result, they can be ranked and compared in terms of the expected health if research is conducted. A second stage of the research design process is the definition of a sufficient condition to undertake further research, which is achieved when the optimal sample size for the new study is defined, something that can be found at the point where the marginal benefit equals the marginal cost of sampling. In order to do this, analysts can estimate the Expected Value of Sampling Information (EVSI_n) for a sample of size *n*, i.e. the marginal benefits of an additional unit of observation measured by the reduction in expected opportunity loss (Claxton and Posnett, 1996, Ades et al., 2004). Then, the difference between EVSI_n (when *n* takes alternative values of sample size) and the cost of research proposal of size *n*, corresponds to the Expected Net Benefit of Sampling (ENBS_n). All these metrics are function of *n*. The optimal sample size is found at the point of *n* where ENBS_n is the maximum.

Although the value of conducting CEA for decision-making is mainly related to decisions about reimbursement of new technologies, here its value is also presented in informing future research. However, there are two other important aspects of value that deserve attention. The first is the value of information about issues related to implementation. Because the decision concerning reimbursement does not guarantee perfect implementation, there is value in exploring whether investing in implementation strategies it is worthwhile from the health system point of view (Fenwick et al., 2008, Hoomans et al., 2009). A second issue, and more important for this thesis, is the value of understanding the variability between individuals. Several authors have raised the importance of considering heterogeneity in decision-making

based on cost-effectiveness (Coyle et al., 2003, Sculpher, 2008a, Basu, 2009). Although there is a vast body of literature in the field of comparative effectiveness research, mostly motivated by subgroup analysis (Stallones, 1987, Yusuf et al., 1991, Oxman and Guyatt, 1992, Feinstein, 1998, Parker and Naylor, 2000, Assmann et al., 2000, Sleight, 2000, Brookes et al., 2001, Cui et al., 2002b, Pocock et al., 2002, Rothwell, 2005, Lagakos, 2006, Wang et al., 2007, Sevdalis and Jacklin, 2008, Gabler et al., 2009, Sun et al., 2010), only few contributions have been carried out in the field of cost-effectiveness analysis (Nease and Owens, 1994, Sculpher, 1998, Coyle et al., 2003, Basu and Meltzer, 2007, Basu, 2011, van Gestel et al., 2012). It is not the aim of this chapter to present these contributions in detail but only to highlight that methodological developments in this field are recent, offering an opportunity to explore additional issues related to heterogeneity in cost-effectiveness, which is the main motivation of this thesis. The reader of this doctoral thesis will find in Chapter 2 a more exhaustive review of the methodological aspects of heterogeneity in CEA, their strengths and limitations as well as the gaps that need further research.

1.5. Objectives and structure of this thesis

The general objective of this thesis is to contribute to the understanding of the value of heterogeneity in cost-effectiveness analysis for healthcare decision-making. This objective is addressed through four specific objectives that provide the structure of this doctoral work. They include the examination of the current knowledge of the standard approaches to explore and understand the value of understanding heterogeneity for healthcare decision-making, the construction of a framework for cost-effectiveness subgroup analysis and its application, and the examination of the elements that determine the implementation of individualized decisions in a collectively funded healthcare system. A final chapter provides a summary of the contribution made by this thesis, their strengths and limitations as well as their relevance for a systematic consideration of heterogeneity in different types of healthcare systems. This thesis adopts an extra-welfarist conceptualization of the economic analysis for healthcare, which is consistent throughout the subsequent chapters.

1.5.1. Objective 1. Current knowledge of the value of heterogeneity for healthcare decision-making

The first objective is to review, describe and discuss the standard approaches to assessing heterogeneity for healthcare and clinical decision-making as well as the methods developed to realize its value. As a result, it is expected to identify areas requiring further development that can be addressed in this thesis.

This objective is addressed in Chapter 2, which presents the motivation for studying heterogeneity for decision-making in healthcare. The main concepts of variability and heterogeneity are defined, as well as the types of heterogeneity that can be assessed in healthcare decision problems. This aims to clarify some confusion generated by the wide use of these terms in the general literature of healthcare. Methods to express the value of exploring and characterizing heterogeneity are reviewed subsequently, introducing individual preferences and choices as central characteristics of the analysis. These concepts connect the study of heterogeneity with the agenda of individualization in healthcare, something that was identified at this point of the work. This chapter was based on an extensive review of the literature, allowing the author to identify the gaps in the current knowledge. A subsequent judgement about the feasibility of addressing these gaps was the starting point of the original pieces of work of this thesis. The remaining chapters are a consequence of this process.

1.5.2. Objective 2. Constructing a framework for cost-effectiveness subgroup analysis

The second objective of this thesis is to examine the fundamental elements that give support to cost-effectiveness subgroup analysis in order to construct an analytical framework for a systematic approach.

One of the main approaches to assess heterogeneity in healthcare evaluations is the examination of subsets of the studied population. In Chapter 3 the arguments are examined based on which cost-effectiveness subgroup analysis for healthcare decision-making can be considered the most appropriate approach. This analysis provides the elements to construct an analytical framework to explore, characterize, present and interpret results of these analyses. Thus, the methods presented in this

chapter expect to contribute to the decisions that central or local authorities have to make in terms of adoption and reimbursement of new technologies.

One of the most relevant elements of this chapter is the formal conceptualization of the value of heterogeneity as the sum of two parts, the value of making different decisions for different subgroups based on current information (called the static value of heterogeneity) and the value of resolving uncertainty collecting further information conditional to some level of heterogeneity (dynamic value of heterogeneity).

1.5.3. Objective 3. Application of a novel analytical approach to cost-effectiveness subgroup analysis

The third objective is to apply the analytical approach developed in Chapter 3 in order to explore the strengths and weaknesses of the technique, as well as contextualizing it with previous contributions in this field.

This objective is addressed in Chapter 4 of this thesis. The implementation of the conceptual basis developed in Chapter 3 demands the development of specific technical issues such as the prediction of individualized cost-effectiveness analysis and the characterization of the uncertainty under such estimations. Further, new concepts coined and formalized in this thesis (Chapter 3) are empirically illustrated and discussed contrasting with previous contributions.

1.5.4. Objective 4. Examination of the conflict between individual decisions and social values

The fourth objective of this thesis is to examine the elements that lead to the definition of a particular approach to the implementation of individualized care in a collectively funded healthcare system.

It follows from the arguments developed in Chapter 3 that decisions at individual level are consistent with a goal of efficiency, insofar as costs and benefits are taken into account. However, the only feasible way of implementing this in a collectively funded healthcare system is to allow patients and doctors to make choices without restriction. Social planners can take different perspectives on the problem of

individualization of care, bringing to bear social values that can ultimately serve as justification of different approaches to the decision-making process. Chapter 5 presents an analysis of these elements, distinguishing which of them corresponds to a pure normative judgement from those which need some consideration of value in terms of the consequences for the healthcare system (positive statement). This analysis provides the basis of a framework whereby the value of choice can be examined through a quantitative exercise based on a real case study.

More specifically this chapter examines, from a social planner point of view (i.e. central or local health authority), the effects of leaving doctors and patients making choices versus a centralized selection of treatments, under alternative normative judgements about health state values. Thus, the chapter contributes to the understanding of the conflict between society and the individual, the social value of choice for healthcare and the normative judgements towards individualization of care.

2. CHAPTER 2

LITERATURE REVIEW

Understanding heterogeneity for healthcare decision-making about health interventions

2.1. Introduction

The constant development of new health technologies (e.g. drugs or medical devices) responds to several factors: greater life expectancy of the population, leading to new healthcare needs; changes in environmental conditions creating new epidemiological profiles; and scientific development¹¹. This scenario has a huge impact on health systems, which are forced to increase their expenditure on healthcare to respond to the increasing demand for the new technologies. This situation is a problem even in the most liberal health system, where ultimately the individual citizen has to sacrifice his/her own consumption to allocate his/her individual budget to healthcare. Most countries, however, have a financial protection system through private or public insurance to protect patients from the financial burden of poor health, or at least part of it. Regardless of the health system structure, the problem of dealing with increasing expenditure on healthcare is unavoidable and forces planners to adopt a clear approach to allocating limited resources. Thus, the main objective of most European healthcare systems and certainly the National Health Service (NHS) in the UK, is to maximise some comprehensive measure of health outcome subject to the budget constraints of the healthcare system. This view is consistent with the extra-welfarist framework as explained in Chapter 1.

The allocation of resources requires, at the very least, information about which interventions work and the value of such technologies measured in terms of the health displaced elsewhere in the healthcare system. As introduced in Chapter 1,

¹¹ The scientific development refers to the ability of understanding the process of health and disease as well as the capacity to produce health technologies able to improve health of the population.

HTA has emerged as a multidisciplinary field of systematic analysis of evidence in order to provide decision-makers with necessary inputs for their decisions (Eddy, 2009). Most jurisdictions consider evidence about effectiveness and safety an essential component of the decision. In other countries, such as the UK, cost-effectiveness is also given formal consideration because it assesses the value of the additional gains for the health system. Both, effectiveness and cost-effectiveness, are usually considered as average estimates of the target population. This approach has been largely justified by the fact that it is impossible to observe the effect of alternative treatments in the same individual at the same time (Goodman, 1947), which is also referred to as the “fundamental problem of causal inference” (Holland, 1986). Instead, average treatment effects are unbiased estimates when two (or more) groups being compared are, on average, totally similar, so that the differences in the outcomes are only explained by the treatment. Typically, this is achieved through randomization of individuals to alternative courses of action such as diagnostic tools, treatments or both.

Although the arguments given for using average estimates in healthcare research have been widely accepted, they are essentially pragmatic and should not be considered a gold standard from a theoretical point of view. Rather, they are the best approach available hitherto, and should motivate further methodological development on the estimation of individual treatment effects. This has been recognized as an important concern for both comparative effectiveness research (CER) and CEA (Basu, 2009). While decisions based on the average might adopt one particular intervention for the whole population, decisions that consider heterogeneity in the population (for example at subgroup level) can deny the intervention in some patients where it is not (cost-)effective. As a result, resources can be re-allocated to other more (cost-)effective alternatives, leading to an improvement in the overall population health. At the limit, when decisions are made at individual level, the gains are the maximum possible, as long as the decision rule applied at that level is founded on the same principles of rationality that provide grounds for decisions at a more aggregated level.

The interest in heterogeneity for decision-making has been described in different ways. From a biomedical perspective, reflecting heterogeneity in decisions has been

promoted as achieving personalized medicine (PM), which aims to provide “the right treatment to the right patient at the right dose at the right time” (Hamburg and Collins, 2010). The idea of reaching PM requires the identification of measurable parameters (e.g. molecular biomarkers) that allow doctors to prescribe treatments according to specific individual characteristics. This idea has led to an increasing development of diagnostic tools based on molecular biology methods, including pharmaco-genetics (Evans and Relling, 2004, Conti et al., 2010). In nursing, on the other hand, this has been promoted as the implementation of a person-centred care model or individualized care model, where decisions take into account as much individual information as possible, including PM but also patient choices (Suhonen et al., 2002, Charalambous et al., 2010).

The idea that patient choices should play a central role in decision-making has been at the centre of political debate for a long time, in particular in countries with liberal health systems such as the USA. However, this is also part of the current policy agenda in countries such as the UK, where the right to choose healthcare at the individual level has been given formal consideration. In principle, this included the right to choose among healthcare providers (Department of Health, 2009). More recently, the British health authority has announced the introduction of choice for diagnostic testing and treatments (Department of Health, 2010).

The aim of this chapter is to review the key elements of the discussion about how heterogeneity should be examined, exploited and analysed for the purposes of decision-making about healthcare interventions. In terms of the methods for economic analysis, this review focuses on the role of heterogeneity as a source of value to achieve greater health, which is consistent with an extra-welfarist perspective. The chapter is structured in four sections. The first section seeks to review the standard approaches to assess heterogeneity. Second, there is an exploration of the methods developed to represent the value of considering heterogeneity in healthcare decision-making. Third, the role of patient choices and patient preferences is examined as a source of heterogeneity. The final section identifies some research needs and priorities.

The literature review was conducted in November 2010. Relevant methodological published papers were searched in Pubmed, Embase, Econlit, NHS Economic Evaluation Database. The following terms from the Pubmed MeSH database were considered for this search: “heterogeneity”, “cost effectiveness”, “cost benefit”, “comparative effectiveness research”, “individualized medicine” and “decision-making” (adapted to MeSH terms of each database). In addition, the following string terms were also considered “potential outcomes”, “subgroup analysis” and “methods”. There was no pre-defined limit of year or language. A total of 442 papers were potentially relevant according to several combinations of those terms. From these, only 42 papers were considered relevant for this review, because they corresponded to original methodological contributions. A further hand search through the reference list of those papers was conducted, which included official governmental documents (n=7). This literature was considered for a narrative review. An update of this literature search was conducted in April 2012. Seven further methodological papers were found and included in the chapter.

2.2. Standard approaches to assess heterogeneity in evaluation of healthcare technologies

The term variability has been used to express the differences in outcomes between individuals, which can be explained by observed and unobserved characteristics. Heterogeneity has been defined as the proportion of the variability that can be explained by a set of observed (*known*) characteristics at the time of analysis (Briggs et al. 2006). In this chapter, total variability is understood as equivalent to total heterogeneity, which includes observed and unobserved heterogeneity. In the unobserved heterogeneity, there is a proportion of variability that can be revealed through further research, which corresponds to those *knowable* characteristics that, once revealed, will explain the remaining observable heterogeneity. In this group we find the *known unknown* and the *unknown knowable unknown* characteristics. The first corresponds to the set of known covariates for which further research will provide specific individual information. The second refers to the unknown covariates able to further explain any residual observable heterogeneity and that can be identified only through additional exploratory research. Finally, the proportion of the variability that cannot be explained (the *unknowable* characteristics) corresponds to

random error¹². The unobserved part of the variability (unobserved heterogeneity) has also been called stochastic uncertainty or first-order uncertainty (Briggs et al. 2012). Complete information refers to the knowledge of the set of covariates that are able to explain differences of outcomes between all individuals in the population (total variability or total heterogeneity)¹³. From a decision-maker point of view, the main challenge is to take into account as much information about individual level characteristics as possible. The aim for health researchers is therefore to achieve a proper characterization of the total heterogeneity, i.e. not only to convert the knowable characteristics into observed measurable variables but also making some prediction of the expected individual outcomes considering unobservable heterogeneity.

The literature in different areas provides alternative nomenclatures to refer to the study of heterogeneity. For example, in the context of the evaluation problem in econometrics, unobserved heterogeneity has been termed non-essential heterogeneity when the selection of treatment does not depend on these unobserved characteristics. Instead, when treatment selection depends on the unobserved expected gains, this is called essential heterogeneity (Heckman et al., 2006). Epidemiology and biostatistics emphasize the importance of distinguishing between moderators, mediators or non-specific predictors of treatment outcomes (Kraemer et al., 2002). Variables considered moderators inform for whom and under which conditions the treatment works. Mediators, in contrast, inform about potential mechanisms that explain the causal effect. Non-specific predictors are variables that show an effect on the outcome without interacting with the treatment (Baron and Kenny, 1986, Kraemer et al., 2001). This distinction is relevant in understanding the underlying causal model of the health problem. More generally, terms such as observable or measurable heterogeneity are broadly used across the sciences. Figure 2.1 synthesises these

¹² These terminology, which in their former presentation included the “known-known”, “known-unknown” and “unknown-unknown”, was coined by Donald Rumsfeld in (2002) and lately supported by journalists, economists and philosophers. An additional category, the unknown-known, has been suggested by Zizek (2006) to refer the things that we know but we refuse to acknowledge.

¹³ Complete information should be distinguished from perfect information. Complete information is reached when all the covariates needed to explain differences between individuals are revealed. Perfect information instead refers to the knowledge of the true *mean* effect of such covariate (and its correlation with others) on the health outcome. Likewise, perfect information also refers to the knowledge of the true value of a particular covariate in one individual (e.g. presence of a genetic characteristic with 100% accuracy).

terms, making a parallel correspondence between them. For example, observable heterogeneity includes, on one side, mediators, moderators and non-specific predictors. On the other hand, it includes known and knowable heterogeneity. Unobserved heterogeneity, also called first order uncertainty or stochastic uncertainty, includes part of the observable heterogeneity (the part that has not been revealed yet) and the unobservable (or unknowable) heterogeneity. Essential and non-essential heterogeneity, as explained above, corresponds to the proportion of heterogeneity that has not been observed.

Table 2.1. Terminology in the study of heterogeneity. Relationship between different terms and the field where it is used.

TERMINOLOGY				AREA OF USE
Non-specific predictors				Epidemiology and biostatistics
Mediators			Moderators	
Observable (measurable) heterogeneity			Unobservable (un-measurable) heterogeneity	Econometrics
Observed heterogeneity	Unobserved heterogeneity, First order uncertainty or stochastic uncertainty			
Essential heterogeneity		Non-Essential heterogeneity		
Known heterogeneity	Knowable heterogeneity		Unknowable heterogeneity	Generally in social sciences and philosophy
Known known	Known Unknown	Unknown knowable Unknown	Unknown unknowable Unknown	
TOTAL HETEROGENEITY OR TOTAL VARIABILITY				

Exploration of heterogeneity has classically been driven by subgroup analysis. Usually, the dimensions explored correspond to baseline risk and treatment effect heterogeneity. Baseline risk refers to the set of characteristics that determine a particular *a priori* probability of presenting the health outcome. It is important to emphasize that this probability might or might not influence the treatment effect. However, even in the case where the treatment effect is the same across individuals, the absolute value of the health outcome varies across patients with different baseline risk profiles. Treatment effect heterogeneity, on the other hand, corresponds to the variation of the treatment effect among different patients. In statistical terms, it corresponds to the interaction between the treatment and the covariate that defines

the membership of an individual in one particular category¹⁴. Two elements can be identified affecting this type of effect: responsiveness and vulnerability (Kravitz et al., 2004). Responsiveness refers to a set of unobserved genetic, physiological or psychological factors that can ultimately explain differences in this effect between patients¹⁵. On the other hand, vulnerability corresponds to the patient's tolerance and subsequent adherence to a particular treatment. It is important to stress that both baseline risk and treatment effect heterogeneity are defined on the basis of one or more observable characteristics at baseline, assessed on the basis of health outcome(s)¹⁶. Other sources of heterogeneity include costs and preferences. Heterogeneity in costs can be observed as differences either in unit costs or use of resources. For example, in-hospital diabetic patients with a myocardial infarction require periodic control of glycemia and eventual use of insulin, leading to a systematic additional use of resources in this subgroup. On the other hand, in healthcare systems with a decentralized process of expenditure, where different healthcare trusts or clinics make different price arrangements with their providers, unit costs might also vary. Heterogeneity in preferences will be discussed in detail later.

Variation between contexts has also been a matter for attention in the evaluation of effectiveness and cost-effectiveness. Geographical differences have been explored mostly in the context of countries, though they could also be important in terms of different jurisdictions within a country with specific characteristics that affect, for example, the incidence or prevalence of a particular condition. These differences can be explained by several elements of the health system, clinicians, patients or wider socio-economic factors (Sculpher et al., 2004). For example, costs are very likely to vary across jurisdictions as well as the specific opportunity cost for that healthcare system (reflected as different cost-effectiveness thresholds). It has also been reported

¹⁴ Treatment effect heterogeneity can be categorized as quantitative interaction (differences between subgroups are in the same direction but they vary in terms of their magnitude), effect concentration (treatment effect is only seen in one subgroup) and qualitative interaction (treatment effect varies not only in magnitude but also in direction between subgroups) (Gail and Simon, 1985, Brookes et al., 2004)

¹⁵ For example, in most cases the treatment effect of drugs is related to the plasmatic levels and the metabolism rate of the drug by the liver. Some genetic characteristics determine that this metabolism rate is faster in some subset of patients with the subsequent lower plasmatic levels and smaller effect (Meyer, 1994).

¹⁶ This statement is founded on the conceptual grounds of the Rubin's model for causal inference (Holland, 1986)

that teaching and specialized hospitals incur higher expenditure than general hospitals, differences that can be observed within the same jurisdiction (Iezzoni et al., 1990, Mechanic et al., 1998). Further, more trained health professionals might be associated with better results and fewer costs, as a result of more efficient care (e.g. quicker diagnostics, lower complication rates in surgical procedures). This variation has been demonstrated by empirical work using specific statistical techniques which have been proposed to improve estimation of effectiveness and cost-effectiveness in particular locations (e.g. multilevel modelling and shrinkage estimation) (Manca et al., 2005).

Despite the growing interest in considering heterogeneity in decision-making in healthcare, researchers face some constraints that hinder its examination. One of these is the adherence to classical standards of statistical inference that mainly affect the study of heterogeneity in baseline risk and treatment effect. Because most clinical trials are designed to find significant average treatment effects and their sample size is calculated consistently with that goal, any attempt to make inference on subsets of the sample faces the problem of loss of power (increase of error type 2) due to a decrease in sample size (Grouin et al., 2005)¹⁷. Further, when additional testing is performed on the same data, there is a higher probability of finding differences between groups explained by chance (false positives or increase of error type 1) (Brookes et al., 2004, Lord et al., 2004, Rothwell, 2005, Ioannidis, 2005)^{18,19}. A third problem relates to the requirement of an interaction test to prove treatment effect heterogeneity in clinical studies (Wang et al., 2007, Oxman and Guyatt, 1992). If treatment effect is proved to be significant, authors usually report both baseline and treatment effect heterogeneity. If treatment effect is not significant, information about (significant) baseline risk heterogeneity might be omitted. Although from a clinical point of view, treatment effect heterogeneity might be considered the only

¹⁷ This concern is a statistical fact; however, there is also evidence in the opposite direction. It has been demonstrated that using pre-specified (baseline) covariates in a regression framework increases the statistical power (Hernandez et al., 2006, Hernandez et al., 2004, Hauck et al., 1998, Steyerberg, 2009). This increase can be explained by the magnitude of the prognostic effect of the covariate on the outcome (Lingsma et al., 2010).

¹⁸ This has led experts to recommend adjustments for multiple comparisons, a procedure that ultimately increases the p-value cut-off point to reject the null hypothesis (Aickin and Gensler, 1996, Bender and Lange, 2001, Bland and Altman, 1995)

¹⁹ The need of conducting multiple testing adjustments has also received critics. For a broader discussion of these arguments see (Rothman, 1990).

relevant attribute to report heterogeneity, variations in baseline risk are an important source of heterogeneity from a decision-maker's point of view. In addition, although these tests respond to a genuine interest in achieving good estimates of treatment effect differences between subgroups, they have been demonstrated to have low power and a high rate of false negatives (Brookes et al., 2004). Finally, loss of balance between arms of a trial has also been raised as a concern in comparing subgroups²⁰.

All these concerns are relevant for clinical studies and they do not necessarily apply in a similar way to cost-effectiveness analysis. Decision rules in CEA are grounded on the principles of expected utility theory, which not only considers the probability of a wrong decision (equivalent to a p-value or a confidence interval) but also its consequences (Claxton, 1999b). However, even in the case of decisions based on cost-effectiveness, there are some constraints on the study of heterogeneity. For example, characteristics used to explain differences between individuals should not transgress ethical or equity relevant principles for particular societies. Thus, guidelines of the National Institute for Health and Clinical Excellence (NICE) in the UK, for instance, states that subgroup analysis based purely on differences in treatment costs is not relevant (NICE, 2008). Furthermore, transaction costs involved in the operationalization of decisions at an individual level or in different subgroups are not usually considered and might hinder the implementation of these types of decisions²¹. Although heterogeneity in CEA has been indicated as one of the elements that require more discussion and methodological development (Sculpher, 2008b), there are no clear guidelines to systematically address, analyze and understand these different dimensions of heterogeneity.

²⁰ Stratified randomization has been suggested as a tool to keep balance on pre-specified subgroups (Torgerson and Torgerson, 2008).

²¹ The relevance of transaction costs is further developed in the next chapter

2.3. Value of Heterogeneity

As previously mentioned in the introduction, heterogeneity has value for the healthcare system simply because greater population health can be achieved by conditioning treatment decisions on those factors responsible for such between-patient heterogeneity. Subgroup analysis has been the most common approach to explore heterogeneity in HTA (Brookes et al., 2001). Coyle et al. (2003) have represented the value of considering subgroups in terms of Incremental Net Benefits (INBs) that can be gained from a stratified analysis for the simplest case where two strategies are compared. The authors explain that if policy makers restrict the adoption of technologies to those subgroups with positive INBs, then the gains derived from making different decisions for different subgroups is the difference between the total INBs (including positive and negative INBs) and the sum of the positive INBs. In other words, it is the absolute value of the sum of the INBs in those subgroups where the INB is negative. The approach suggested by Coyle et al. (2003) is restricted to express gains based on current information for a case where two strategies are compared²².

Nease and Owens (1994) introduced the idea of estimating individualized expected health benefits to realise the value of a guideline that considers individual preferences. Using a Markov decision model for mild hypertension, they showed that decisions guided on the basis of individualized utility assessment should be considered cost-effective compared to average utility estimates. Later, other studies have applied similar methods (Sculpher, 1998, Dowie, 1998). Basu and Meltzer (2007) move forward from this point proposing an estimation of the value of eliciting information at patient-level to make individualized decisions using the Net Benefits (NBs) framework (Stinnett and Mullahy, 1998). They introduced the Expected Value of Individualized Care (EVIC), a metric that reflects the population NBs forgone because of the ignorance of heterogeneity in preferences²³ when decisions are made based on the average estimates. EVIC is calculated as the difference between the average of the maximum NBs in each patient (individual Net

²² This is an important starting point of the conceptual framework for cost-effectiveness subgroup analysis developed in Chapter 3 and will be analyzed in detail.

²³ Heterogeneity in preferences refers to the variability among health state values and will be explained in detail in the next section.

Benefits or iNBs) and the maximum of the average NBs of the alternative treatments across patients. The authors point out that although EVIC was initially estimated for patient preferences, it might be estimated for any other (set of) parameter(s) of interest in the decision model. Indeed, a total EVIC captures all parameters of interest and should be interpreted as the expected gains that could be attained if individual information about every patient is considered when estimating the outcome of interest.

Basu and Meltzer (2007) distinguish two formalizations of EVIC depending on which decision rule is applied. When the objective is to maximize individual expected **net** benefits, EVIC is called “with cost-internalization”. On the other hand, if the objective is to maximize expected health benefits but without accounting for costs, EVIC is denominated “without cost-internalization”. In the example provided by the authors, it is demonstrated how the value of individualized information can be affected by the decision rule applied. While the EVIC with cost-internalization is greater than US\$70 million, this value falls to US\$0.9 million without cost-internalization, suggesting that the effort of eliciting individualized information is much more valuable if doctors (and patients) internalize costs when make their decisions.

Basu and Meltzer (2007) also presented the parameter-specific EVIC ($EVIC_{\theta_i}$), analogous to the expected value of perfect information for parameters (EVPPPI). An advantage of this metric is that by ranking parameters according to $EVIC_{\theta_i}$, the most valuable information for individualized decisions can be identified. In a more recent study, van Gestel et al. (2012) tested the feasibility and role of EVIC for decision-making in a cost-effectiveness study of glaucoma using discrete event simulation. They illustrated the interpretation of total EVIC as well as parameter-specific EVIC (with cost-internalization) on supporting partial decisions for subgroups (van Gestel et al., 2012).

These recent methodological approaches described above allow researchers to provide an adequate representation of the potential health that can be gained if heterogeneity is taken into account in healthcare decision-making. For example, Basu and Meltzer (2007) estimated the value of eliciting preferences directly from

patients (using time trade-off). In this particular case, the true value of the patient-health status is available and no remaining uncertainty needs to be resolved, at least for preferences. Likewise, in the example provided by van Gestel et al. (2012), total EVIC represent the value of patient specific decisions according to individual characteristics, assuming perfect information. An important issue that needs to be addressed is the role of decision uncertainty when heterogeneity is taken into account. It has been demonstrated that decision-makers not only face the question about adoption and rejection of interventions but also whether resolving the remaining uncertainty should be considered a good use of resources (Griffin et al., 2011). Therefore, these two dimensions should be analysed in a more integrated framework.

2.4. Preferences and choice as a source of heterogeneity

Preferences and choices are concepts with important implications for the study of heterogeneity across individuals and will be examined in this section. Although in principle they seem similar concepts, they encompass and express different meanings in the context of healthcare decision-making. In this section, the main concepts are defined in order to examine their implications further.

2.4.1. Preferences as a source of heterogeneity

Preferences arise from the conceptual grounds of economic analysis in healthcare. In particular, for CEA, where the primary objective is to maximise health subject to budget constraint, utility theory has been employed to measure the value of health outcomes (for example in the QALY model) generated by alternative programs (Torrance, 2006). Further, the tools used to measure the value of health outcomes (e.g. standard gamble) are entirely based on expected utility theory under a set of some restrictive assumptions (Pliskin et al., 1980)²⁴. These methods estimate a relative value for descriptive health states, which are a representation of a particular level of Health Related Quality of Life (HRQoL)²⁵. Under this framework, preferences refer to the rational element of judgment that guides individuals to rank

²⁴ These assumptions include utility independence, constant proportional trade-offs and risk neutrality (Pliskin et al., 1980).

²⁵ Health states are defined as a description resulting from combining dimensions of HRQoL. Therefore, by construction each health state has attached a unique absolute level of quality.

health states in a particular order to reveal the relative value of such states. These elicited values, used as weights of the observed quantity of life (e.g. years of survival), express the ultimate maximand of the CEA objective function, which can be denominated the *ultimate health effect*.

Preferences are usually elicited from a sample of the population. The sample can be representative of the general community (public preferences) or particular groups of patients (patient or private preferences)²⁶. Regardless of the source, these values are usually aggregated and presented as averages in a normative “tariff” to be subsequently used in economic evaluations (Kind, 2005). In practice, most of the time these values are taken as perfect information without even considering the uncertainty around them (Gray et al., 2012). Likewise, although the original surveys have shown the variability around those mean values (Zarate et al., 2011, Dolan et al., 1996, Zarate et al., 2008), heterogeneity is usually not considered in CEA.

Given the close link between health states and health state values, an important distinction must be made between heterogeneity in HRQoL and heterogeneity in preferences. HRQoL corresponds to the outcome upon which heterogeneity of different sources are expressed. They include baseline risk, treatment effects, costs or preferences²⁷. Therefore, heterogeneity is observed in the distribution of HRQoL but is explained by, for example, preferences. Thus, the real interest is to examine heterogeneity of different sources, such as heterogeneity in preferences, on the relevant health outcome, for example, HRQoL or QALY.

Few studies have addressed the idea of considering heterogeneity in preferences. Nease and Owens (1994) provided one of the earliest examples of how preferences from patients, i.e. private preferences (revealed by time trade-off), can be incorporated in formal CEA in order to take into account their heterogeneity. They showed that decisions based on individualized utility assessment can be considered cost-effective. Later, Sculpher (1998) compared different preference-based approaches to treatment allocation (expected individual health, expected individual

²⁶ Public and private preferences are elements of intense debate and will be addressed more in detail in chapter 5.

²⁷ As mentioned earlier in this chapter, other variables not directly related to the health problem might also explain heterogeneity.

cost-effectiveness and treatment selection²⁸) to evaluate the cost-effectiveness of making decisions considering specific patient-preferences. It was concluded that a decision rule based on individual net QALYs provided greater population health. In addition, it revealed that decisions based on expected individual QALYs and net QALYs are not well correlated with treatment selection. However, in theory, they should be correlated if patients are rational and have the same information from which QALYs have been estimated. This suggests that there is some information contained in the process of making choices that is not being considered in the whole process of evaluation or, alternatively, patients are not making decisions according to the expected utility theory and QALY maximization.

In the context of public preferences, Sculpher and Gafni (2001) suggested that there are subgroups of people in the population whose preferences are sufficiently different from others and that the incorporation of these differences in cost-effectiveness studies is consistent with the main objective of CEA. However, the implementation of a systematic evaluation of subgroup public preferences seems less feasible than patient preferences and even less likely to be good predictors of treatment selection.

2.4.2. Choice as a source of heterogeneity

An optimal (*treatment*) *choice* is a decision that allows the individual (e.g. the decision-maker or the patient) to maximise his/her welfare, utility or health depending on the elements in his/her objective function. In the context of healthcare we distinguish *ex-ante choices or treatment preferences* as the decisions made by patients before being subjected to critical scrutiny whereby other relevant information and point of views are taken into account. On the other hand, *treatment selection or revealed choices* is the individual decision resulting from the interaction of the patient with health professionals, relatives and other sources of information

²⁸ Treatment selection refers to individual choice for treatment (defined in the next section).

that are relevant for the decision. This can be operationalized in the context of a shared decision-making model of care²⁹ (Charles et al., 1997).

One of the central elements of the current policy agenda is the idea that patients' preferences and choices should be taken into consideration in the decision-making process. The autonomy of individuals is argued as an ethical imperative that justifies any effort in this direction. NICE in the UK recognizes this argument as part of its social value statement but it also highlights the importance of making adequate judgments to ensure good use of the limited resources (NICE, 2005). Consideration of patient preferences in HTA is consistent with both concerns. However, less clear is the extent to which treatment choices can be consistent with the health maximization objective. In other words, do patient choices provide some unobserved information about the expected potential gains due to a particular treatment?

In the clinical trials literature it has been reported that when patients are allocated to their preferred treatments, their outcomes are affected positively without effect on attrition rates (Preference Collaborative Review Group, 2008). This might indicate that the treatment works better in patients who prefer it, irrespective of the causes that explain loss in the follow-up³⁰. If ex-ante choices can be used to predict outcomes, then they could be used to help select treatment as a form of subgroup analysis. However, findings indicating that ex-ante choice are not good predictors have also been reported (Moffett et al., 1999). While the role of ex-ante choices as predictors is not clear, this might not be the case for ex-post choices. Given the process needed to reveal those choices, they are likely to be more informative of the health outcomes than ex-ante choices.

Because individuals do not necessarily behave according to the rules of rational choice theory to make their decisions, revealed choices are not necessarily the same as those predicted by a decision rule such as the maximization of expected QALYs according to expected utility theory. This suggests that patients bring additional

²⁹ Shared decision-making is a process whereby patients and health professionals share information about alternative diagnostic and treatment options as well as outcomes preferences with the aim of making the best choice among the alternative courses of action.

³⁰ This argument has led to the suggestion that open and adaptive trials should be more often considered to characterize this effect on treatment effect, especially in the context of the estimation of individual treatment effects (see below) (Basu, 2011).

elements of value to their decisions. There is evidence in the econometrics field (Heckman and Vytlačil, 2007) with some applications to healthcare indicating that treatment selection has value as predictor of health outcomes (Basu et al., 2007). In fact, choice for treatment might correlate very strongly with many other covariates that explain variability in health outcomes. Thus, by using adequate statistical techniques, the individual treatment effects could be estimated and characterized, producing a better understanding of the joint distribution of potential health outcomes³¹ (Basu, 2011).

The current debate in the UK has raised the desire to implement a healthcare system where people are given greater opportunity to choose the healthcare they receive without restrictions (Department of Health, 2010) in the context of a shared decision-making model (Gulland, 2011). In contrast, the actual approach determines that some patients sharing one particular condition or disease (will) face limited access to some very expensive technologies because of the resource constraints of the health budget. This political discourse responds – at least in part - to some empirical evidence that shows that more than 95% of British citizens think that there should be opportunity to choose between alternative treatments (British Social Attitudes Survey, 2009). Therefore, choice is valued in the current agenda for its own sake, which can be ethically justified on the principle of autonomy. However, choices are not only relevant because of a patient's autonomy, which is one element, but also because they (revealed choices) might provide information about individual treatment effects. As already mentioned, the selection of treatment might be informed partially by information publicly available but also by some information that cannot be observed by the social planner. If this unobserved information reveals some expected (individual) gains, treatment selection can be used to estimate individualized treatment effects. Consequently, a research agenda for understanding heterogeneity should also include new approaches to reveal individual choices and their role in explaining variability in health outcomes. This should address alternative study designs, as well as more sophisticated analytical techniques.

³¹ The joint distribution of potential outcomes refers to the Rubin's causality model. The potential outcomes are defined as the observed consequences (Y) of alternative treatments ($t=0,1$) in one particular individual (i), i.e. the outcome observed *de facto* and the counterfactual (unobservable). Thus, the joint distribution is defined by $G(Y_{0i}, Y_{1i})$ (Holland, 1986).

Therefore, the problem of considering individual choices of treatment in healthcare decision-making can be addressed using two approaches. First, taking the perspective of a centralized decision-maker, heterogeneity in preferences (health states values) can be reflected more systematically in HTA. Whether public or patient preferences should be used is a question that ought to be revisited in this context. Second, if patient choices are allowed, they should correspond to revealed choices because of their potential ability to generate greater health outcomes. In addition, further research should try to shed light on the extent to which these ex-post choices explain heterogeneity and whether they are consistent with efficiency goals of the healthcare system.

2.5. Discussion

The relevance of heterogeneity in decision-making has grown in the last few years. Despite the significant progress in many areas of health research, most of the evidence used for decision-making only provides average estimates, and the underlying variability is not acknowledged under a strong assumption of homogeneity (e.g. treatment effect homogeneity). However, the need for revealing more about heterogeneity is increasing not only because of theoretical arguments but also because of political motivation towards individualized decision-making. In this chapter, the main concepts and different approaches for the analysis of heterogeneity have been reviewed. They have been examined in the context of different sources of heterogeneity, as well as the relevance of realizing its value. In this section, the key elements presented in this chapter are summarized in order to discuss the gaps that should be considered in a research agenda for health research.

Healthcare systems face the problem of making decision about new health interventions, which are usually more expensive, within a limited budget. HTA tools, including CEA, provide a set of coherent methods to support decisions in this context. Further, growing concern about considering heterogeneity in decision-making is consistent with efficiency goals of a collectively funded health system and several methods have been developed to realize its value. Different sources of heterogeneity, such as baseline risk heterogeneity, treatment effect heterogeneity, contextual factors, health preferences and patient choices have been defined and explained throughout this chapter. In addition, the usual constraints faced by

researchers when they attempt to explore heterogeneity have been examined. In terms of the definition of heterogeneity, this chapter has contributed to the improvement of the understanding of its meaning. A comprehensive description of alternative terms to refer to heterogeneity has been developed. From this, there are two elements that worth to be highlighted. First, total heterogeneity can be considered the same as total variability. This is because the differences in outcomes can ultimately be explained by some individual characteristic. Thus, what has been called stochastic uncertainty corresponds to unobservable heterogeneity. Second, from the point of view of the healthcare decision-making process, the lack of knowledge about heterogeneity can be considered as another type of uncertainty. This dearth of information forces analysts to assume that the differences in outcomes observed between individuals are due to random variability, which has been termed first order uncertainty.

The alternative approaches to characterize the value of making decisions conditional to (some) heterogeneity include subgroups (stratification) and individualized decisions. Methods to express this value in terms of the opportunity cost of a health system have been developed in recent years. They focus on only one dimension of value, which is the expected health gains based on different decisions among individuals. However, most of the information is uncertain (as is revealed from samples of the target population), and this should be taken into account in the assessment of the value of heterogeneity. Thus, a following step should bring these two dimensions of the problem together in a more integrative framework. These elements represent the main motivation for chapters 3 and 4 of this thesis.

Another important element examined in this chapter is the potential role of preferences and treatment choices in health. As mentioned, the consideration of heterogeneity in preferences is consistent with the objectives of a public health planner. More recently, knowledge has progressed in the understanding of the potential role of revealed choices in the estimation of a joint distribution of potential outcomes. However, it still remains unclear in which cases we can rely on those choices as good predictors of health outcomes and in which cases they have been the consequences of other factors (out of the health construct).

Regardless of the potential role of revealed choices in explaining health outcomes, the act of making choices is presented in the public agenda as an essential value that

should be considered as another objective of a health system, especially in more libertarian health systems. Nevertheless, the idea of unrestricted choices could be introduced, consistent with the aim of a collectively funded health system (such as the National Health Service in the UK) under two considerations. First, by identifying subsets where the new alternative is cost-effective, patients who are members of such subgroups can make free choices between all alternatives (Brazier et al., 2005). Although the selection of less effective treatments could lead to health losses, this could, in part, be offset by transferring those resources to other efficient activities of the healthcare system. In addition, it remains unclear whether the true expected health outcome for that specific patient is actually lower than the expectation we assume on the basis of the current HTA process. A better understanding of patient preferences and the joint distribution of potential outcomes might reveal that in some cases patient choices maximize health. Second, unrestricted choices can also be consistent with the aim of a collectively funded NHS if the expected health forgone derived from their implementation is lower than the transaction costs due to restriction of choices. This latter idea is further developed in chapter 5.

It has been described in this chapter, that heterogeneity for decision-making entails several dimensions, each of them affected by constraints on their realization, further exploration and implementation. Different health systems face this problem in different ways. For example, transaction costs for implementation and equity constraints to reveal heterogeneity might be addressed very differently by private health insurance companies and public national health systems. For example, while private insurance adapts the co-payment or coverage schemes for a particular new technology in order to minimize moral hazard, the NHS assesses whether a policy based on subgroups can achieve more population health when compared to its opportunity cost. Likewise, while equity constraints might be important from the NHS point of view, they might be less relevant for a less regulated healthcare system.

In conclusion, heterogeneity in healthcare decision-making is occupying an important place in the health research agenda, not only because it is an intrinsic value for individualization of care but also because it is consistent with the objectives of maximising health under limited budgets. Important contributions have

been developed in the last few years; however, there are still several gaps that need more research. Future investigation should examine the need to produce a more systematic approach to exploring heterogeneity (e.g. through subgroup analysis), the incorporation of parameter uncertainty in a more integrative framework with heterogeneity and the exploration of the role of patient choices in the consideration of heterogeneity of health outcomes.

3. CHAPTER 3

THE VALUE OF HETEROGENEITY

A Framework for cost-effectiveness subgroup analysis

3.1. *Background*

As introduced in Chapter 1, a prominent objective of healthcare systems internationally is to deliver maximum health improvement to the population under its care given limited available and exogenously allocated resources. By comparing average net health benefits (net of costs) –typically expressed in terms of QALYs gained– associated with alternative uses of the healthcare budget, CEA generates the information needed to support resource allocation decision problems consistently with this objective (Brouwer and Koopmanschap, 2000).

It is increasingly recognised, however, that treatment decisions based on *average* measures of cost-effectiveness risk incorrect decisions for specific subsets of the population (Stevens and Normand, 2004), because they fail to reflect the fact that a treatment which may be cost-effective for one type or patient may not be so for another. This type of *heterogeneity* can be ascribed to both individual and contextual-level factors (Sculpher, 2008a) which, when taken into account in analysis and decisions, is consistent with an efficient allocation of resources and thus the realisation of greater population health benefits. Despite its importance in supporting decisions regarding the funding and provision of new medical technologies, there is relatively little in the literature on the methodology of characterising and understanding heterogeneity in cost-effectiveness analysis³². Key papers have explored the potential value of understanding and acting upon heterogeneity (Sculpher, 1998, Coyle et al., 2003, Basu and Meltzer, 2007, Sculpher, 2010) as well as its sources and the practicalities of analysis (Sculpher, 2008a).

³² This literature has been introduced in Chapter 2.

However, none of them provides a guide for systematic subgroup analysis, including identification, selection and presentation of cost-effectiveness results, and none of the issues regarding decision uncertainty have been addressed either.

Given the current policy debate and government agenda relating to the personalisation of health and social care services in the UK (Department of Health, 2010) and elsewhere (US Public Law Sec 6301, 2010, Hamburg and Collins, 2010, National Health and Hospitals Reform Commission, 2009) the need for a clear and coherent framework to support and guide decisions made by local and central health authorities in different groups of patients within the population is no longer an academic exercise but is rapidly becoming a key policy question.

The first section of the chapter summarises the debate on heterogeneity and subgroup analysis and the issues relevant to economic evaluation in supporting healthcare decision-making. Next, it shows how the existing CEA methods toolbox can be extended to address the identification and selection of subgroups. The framework is further expanded to consider how heterogeneity should be reflected and accounted for when assessing the potential costs associated with decision uncertainty, introducing the concepts of *static* and *dynamic value of heterogeneity*. Next, the role of transaction costs associated to the implementation of guidelines based on subgroups is examined. A simple stylized example is used to illustrate the concepts introduced in the chapter. The final section summarises the main contributions of the manuscript and provides a discussion of some of the issues to be considered in future research.

3.2. *Heterogeneity and decision-making in healthcare*

While the concept of subgroup analysis has existed for quite some time (Stallones, 1987, Oxman and Guyatt, 1992), its adoption has been met with caution in certain areas of healthcare decision-making, possibly due to concerns related to issues such as low statistical power and the risk of multiplicity in the analysis³³ (Cui et al., 2002a, Cook et al., 2004, Grouin et al., 2005). Many authors have indicated their preference for using an average measure of treatment effect (Oxman and Guyatt,

³³ This view has been fuelled by a classical frequentist interpretation of the statistical decision rules, which has been dominated for many years by a strict inferential paradigm (see Chapter 2 for details).

1992, Sun et al., 2009, Sun et al., 2010) and that findings from subgroup analyses should be considered exploratory in nature (Sun et al., 2009, Wang et al., 2007). More recently, economists and some statisticians have argued against the use of strict inferential rules regarding the effects of interventions in healthcare decision-making (Goodman, 1999b, Goodman, 1999a, McClosey and Ziliak, 1996, Claxton, 1999b). Specifically, the importance has been recognised of considering evidence on subgroups of patients when making probabilistic statements about the (cost-) effectiveness of a given treatment strategy (Sculpher, 2008a, Coyle et al., 2003).

This more encompassing attitude to sub-group analysis has been reflected in the position of agencies with responsibility for decisions about the funding of new medical technologies. For example, the possibility of considering subgroup analyses in its deliberations are now reflected in the methods guidance for technology appraisal issued by NICE for England and Wales (NICE, 2008). No specific guidance, however, is offered on how to explore and reflect heterogeneity when conducting subgroup cost-effectiveness analyses to inform the Institute's recommendations.

To date, most of the contributions to the literature have focused on specific practical issues. For instance, several authors have shown how baseline risk and treatment effect heterogeneity can be incorporated into decision models (Kuntz and Goldie, 2002, Groot Koerkamp et al., 2010). Others have discussed methods to estimate the cost-effectiveness of allocating patients to different treatments based on their preferences (Nease and Owens, 1994, Sculpher, 1998). Others have used covariates to assess cost-effectiveness of interventions for specific subgroups in the context of trial-based data and to characterise uncertainty as cost-effectiveness acceptability curves per subgroup (Nixon and Thompson, 2005). However, only a few contributors have explored the value of introducing these concerns in terms of population health.

As introduced in Chapter 2, the work by Coyle *et al.* (2003) represents one of the first attempts to produce a framework within which the value of subgroup CEA could be assessed. They show how consideration of treatment effect heterogeneity in a stratified CEA may lead to different recommendations being made for different

patients' subgroups, yielding higher societal payoffs compared to un-stratified cases. Basu and Meltzer (2007) have extended this framework to consider the notion of individualised care and develop the concept of EVIC, which represents the potential net health benefits attainable when research to explain the variability at individual level has been undertaken (Basu and Meltzer, 2010). Nevertheless, none of these approaches have taken into account decision uncertainty as part of an integrative framework.

The initial point of this work recognises that the analysis of heterogeneity has two dimensions of value, one related to the study of between-patient variability and the other to the analysis of parameter uncertainty. First, there is value in considering current information about the heterogeneity of estimated costs and benefits within a patient population. At the maximum, when all relevant individual level information is revealed and used to make decisions, the value of considering such information in decision-making is represented by EVIC³⁴. Second, in the context of estimating population 'average' costs and benefits, the analysis is limited to studying the value of collecting additional sample data in order to reveal the true *mean* value of a parameter. This is driven by applying the value of information methods to estimate EVPI and EVPPI. Although these metrics have been used for average estimates, the value they express is entirely relevant for the interpretation of those analyses that have included (to some extent) information about heterogeneity. Thus, average and individual levels define two extremes of a continuum that contains partial degrees of revealed heterogeneity but also partial degrees of unresolved uncertainty. Further, this dual conceptualization of heterogeneity for decision-making is consistent with the aim of the health system, since the use of finite resources to maximise health outcomes includes investment in additional research (Griffin et al., 2011). This chapter proposes a methodological framework that studies the value of heterogeneity, emphasizing the relevance of these two dimensions and their relationship. In addition, it argues that there are practical constraints to achieving individual-level decision-making which inevitably require the policy maker to make decisions at some point in the continuum between the average and individualized

³⁴ In practice, the value estimated by EVIC is conditional to the set of observed characteristics that have been used in the decision model to represent heterogeneity.

levels. As a consequence, the framework can be consistently operationalized using subgroup analysis.

The chapter presents the value of characterising heterogeneity and using it in decision-making as a two dimensional concept determined by (i) the benefits achieved in stratified versus average decisions and (ii) the potential benefits of resolving the parameter uncertainty at the disaggregation point where the transaction costs are not higher than the benefits. The approach seeks to locate the optimal level of disaggregation somewhere between the two extremes of population average and individual level. Ideally, this will be as far as possible from the average but is unlikely to be at the point of individual decisions because there are transaction costs that make this unfeasible. The framework proposed here is constructed to estimate these two parts. First, it extends the work of Coyle *et al.* (2003) to estimate the value of further interrogating the available evidence (e.g. secondary data analysis) to explore issues of heterogeneity. Since this activity does not involve extra data collection it is called the *static value of heterogeneity*. Second, Value of Information analysis is used to address the value of funding additional data collection. This is called the *dynamic value of heterogeneity*, since it involves further data collection conditional to specific levels of disaggregation.

3.3. Subgroup cost-effectiveness analysis under current information

This section examines subgroup CEA under current information on two dimensions: first, the comparison between different numbers of subgroups facilitated by the *disaggregation* of the population; and, second, the comparison between different specifications (i.e. the covariates used to define one particular subgroup) for a given number of subgroups (an analysis undertaken to compare alternative ways of *grouping* the population). It begins extending the framework previously developed by Coyle *et al.* (2003) to accommodate more than two treatment strategies.

3.3.1. Net Benefits for Subgroup Cost-Effectiveness Analysis

Classical decision rules in CEA (Briggs *et al.*, 2006) state that, under current information, the optimal strategy among j mutually exclusive alternatives, given a set of θ uncertain parameters, can be expressed as follows:

$$\max_j E_\theta NB(j, \theta) \quad (3.1)$$

That is, the optimal strategy is that with the maximum expected net benefit (NB). If we define the total (present and future) patient population expected to benefit from the intervention as

$$\sum_{t=1}^T \frac{I_t}{(1+r)^t} \quad (3.2)$$

where I represents the disease incidence for each period t , T indicates the period over which the technology is assumed to be relevant to clinical practice, and r is an appropriate discount rate, the population expected NB can be estimated as

$$\max_j E_\theta NB(j, \theta) \sum_{t=1}^T \frac{I_t}{(1+r)^t} \quad (3.3)$$

Coyle et al. (2003) showed that, by considering the differences in treatment effect heterogeneity (treatment effect moderators) within this framework, different recommendations can be made for different subgroups. This results in a greater expected NB compared to decisions based on the average across the patient population as a whole.

Using alternative notation, Total Incremental Net Benefit ($TINB$) (with $INB = E_\theta NB(j, \theta) - E_\theta NB(j^*, \theta)$, where $j \neq j^*$) across S subgroups can be written as

$$TINB = \sum_{s=1}^S INB_s w_s \quad (3.4)$$

where $w_s \in (0,1)$ is a weight indicating the proportion of the total population represented by subgroup s and $\sum_{s=1}^S w_s = 1$. The $TINB$ is, therefore, the weighted average of the INB across subgroups. Since some of the subgroup-specific INB may be negative, Coyle et al. (2003) state that the $TINB$, when the intervention is restricted to those subgroups where the new treatment is cost-effective, is

$$TINB_S = \sum_{s=1}^S INB_s w_s, \quad \forall_s \text{ where } INB_s > 0 \quad (3.5)$$

and the *INB* gained from considering patients' heterogeneity in the decisions - what Coyle et al. (2003) termed 'stratification' - can be written as the negative sum of the population weighted *INB* in those subgroups where *INB* is negative. This expresses the static value of heterogeneity.

$$\Delta_S TINB = TINB_S - TINB = - \sum_{s=1}^S INB_s w_s, \quad \forall_s \text{ where } INB_s < 0 \quad (3.6)$$

This estimation is generalized in this chapter, using absolute NB values³⁵ and illustrated with an example presented in Table 3.1. Let us define four subgroups (A, B, C and D) each representing 25% of the population. The objective is to assess the cost-effectiveness of three alternative strategies ($j=1,2,3$). Using a cost-effectiveness threshold of £20,000 per QALY, the maximum NB, expressed in terms of health, for the entire population is 20.75 QALYs per patient net of costs or net-QALYs. This corresponds to treatment 3. Allowing for the possibility of making different decisions for different sub-groups of patients, yields a *TINB* of 22.08 net-QALYs, which results in 1.33 net-QALYs gained from stratification (expressed as $\Delta_S TINB$) compared to the un-stratified case. Notice that $\Delta_S TINB$ is equivalent to the expected opportunity loss of a policy whereby all patients receive the new treatment on the basis of average estimates.

³⁵ The estimation of the total net benefits (*TNB_S*) based on absolute *NB* values can be expressed more generally as

$$TNB_S = \sum_{s=1}^S w_s (\max_j NB_s) \quad (3.7)$$

with the $\Delta_S TINB$ being

$$\Delta_S TINB = \sum_{s=1}^S w_s (\max_j NB_s) - \max_j NB \quad (3.8)$$

which can be easily generalized to the case with j multiple strategies and S subgroups.

Table 3.1. Total Net Health Benefits estimation for three treatments and four subgroups.

	B1	B2	B3	C1	C2	C3	NHB1	NHB2	NHB3
Average	21	18	22	20,000	10,000	25,000	20	17.5	20.75
A (25%)	23	16	21	20,000	7,000	24,000	22	15.65	19.8
B (25%)	19	21	22	20,000	8,000	30,000	18	20.6	20.5
C (25%)	18	17	24	20,000	10,000	25,000	17	16.5	22.75
D (25%)	24	18	21	20,000	15,000	21,000	23	17.25	19.95
$\text{TNHB}_s = [(22 \times 0.25) + (20.6 \times 0.25) + (22.75 \times 0.25) + (23 \times 0.25)]$ $\text{TNHB}_s = 22.08$ $\text{TNHB} = 20.75$ $\Delta_s \text{TNHB} = \text{NHB}_s - \text{NHB} = 22.0875 - 20.75 = 1.3375$									

Note: The estimation assumes a cost-effectiveness threshold of £20,000 per QALY. C: Costs; B: Benefits; QALYs: Quality Adjusted Life Years; NHB: Net Health Benefits; TNHB: Total Net Health Benefits on average; TNHB_s : Total Net Health Benefits considering subgroups

3.3.2. Defining Subgroups

Cost-effectiveness analysis includes more sources of heterogeneity than standard clinical studies (Sculpher, 2008a, Kravitz et al., 2004). The analysis of clinical trials is generally focused on inferences about treatment effects for the patient population defined by the study's inclusion criteria. Any interest in heterogeneity is generally confined to treatment effect moderators, and this is generally cautiously applied (Sun et al., 2010). In addition, there is often a clinical interest in heterogeneity in the underlying risk of adverse clinical events associated with a disease (Sculpher, 2010). This can lead to sub-group differences in the absolute benefit conferred by a treatment offering a fixed proportionate risk reduction across all types of patient. There may also be situations where baseline risk is correlated to relative treatment effect (Rothwell, 2005). These sources of heterogeneity relating to the intervention and the disease are also important in CEA. In addition, resource use (and hence costs) may systematically vary between individuals based on their characteristics and the geographical location of their treatment (Drummond et al., 2005). Finally, heterogeneity in preferences (Kravitz et al., 2004, Basu and Meltzer, 2007, Brazier et al., 2005) is increasingly being recognised as another key source of heterogeneity in cost-effectiveness.

There are, however, some potential constraints on sub-group analysis in CEA. One is the need to consider the costs of implementing sub-group-specific guidance made by the health system. These could include the costs of clinicians acquiring all relevant data on individual patients in order for their treatment decisions to adhere to the guidance. Costs can also be imposed on the health system if it wishes to monitor whether clinicians are implementing their guidance appropriately. There are also potential ethical and equity constraints in reflecting sub-group CEA in decisions. For example, NICE considers unethical to reflect age as a source of heterogeneity in its decisions unless it impacts directly on the efficacy of an intervention. These elements should be made explicit when subgroups are being defined.

After the identification of the subgroups that will be explored, decision-makers need to address the choice between *subgroup specifications* (f); that is, the covariate that defines the membership of individuals in one particular subset (Kent and Lindenauer, 2010). They should be at least biologically plausible and operationalizable in practice (Sculpher, 2008a). For example, patients at risk of cardiovascular events can be grouped on the basis of whether or not they have diabetes ($f=1$), hypertension ($f=2$) or a combination these variables³⁶. The question is how to select between alternative specifications. To help with this task, a criterion based on efficiency (measured in terms of expected NB) is proposed here.

Let us assume we wish to evaluate the cost-effectiveness of two different treatments for Acute Coronary Disease (ACD) and that there are F possible subgroup specifications that could be defined. Each subgroup (s) may be obtained using f alternative partitions³⁷. Clearly, further combinations of these partitions may generate more mutually exclusive subgroups, though there is a trade-off between the number of subgroups and the practicality of implementing these (e.g. data availability, our knowledge of the clinical problem).

³⁶ A more complicated situation is faced with a complex risk prediction model based on several baseline characteristics of the patient and the question is how to disaggregate a continuous risk variable into categories of risk that clearly and unequivocally reflect different subgroups. For instance the analyst may choose to report the results of the CEA by subgroups defined in terms of quartiles of the predicted risk distribution.

³⁷ For example, the sample can be divided in two ($S=2$) based on whether the individual has diabetes ($f=1$), i.e. diabetic patients ($s=1$) and non-diabetic patients ($s=2$) or subdivided on the presence of high blood pressure ($f=2$), i.e. hypertension ($s=1$) and non-hypertension ($s=2$).

The goal is to identify relevant subgroups and associated specifications which produce the highest expected NB , resolving the following maximisation problem

$$\max_{j,f} E_{\theta} TNB_{S,f}(j, \theta) \quad S = 1, 2, \dots, n; \quad f = 1, 2, \dots, F \quad (3.9)$$

A simple example is shown in Table 3.2. Four different subgroups are defined using different specifications in a population of patients with non-ST elevation acute coronary syndrome (NST-ACS). When only two subgroups are considered ($S=2$), the population can be subdivided using three alternative specifications, namely, based on the presence of diabetes ($f=1$), high versus low baseline TIMI risk-score (Antman et al., 2000), a well-known baseline risk score ($f=2$), or based on the presence of a highly sensitive and specific biomarker (Troponin) ($f=3$). The same concepts can be used to define alternative specifications for three and four subgroups. For each subgroup, total expected NBs can be estimated, under current information. Equation (3.9) suggests that the specification with the greatest total expected NBs should be preferred, *given current information*. These points would lead to an efficiency frontier similar to that depicted in Figure 3.1, which represents the range of possible expected NBs which could be achieved using alternative subgroup specifications for each given number of subgroups.

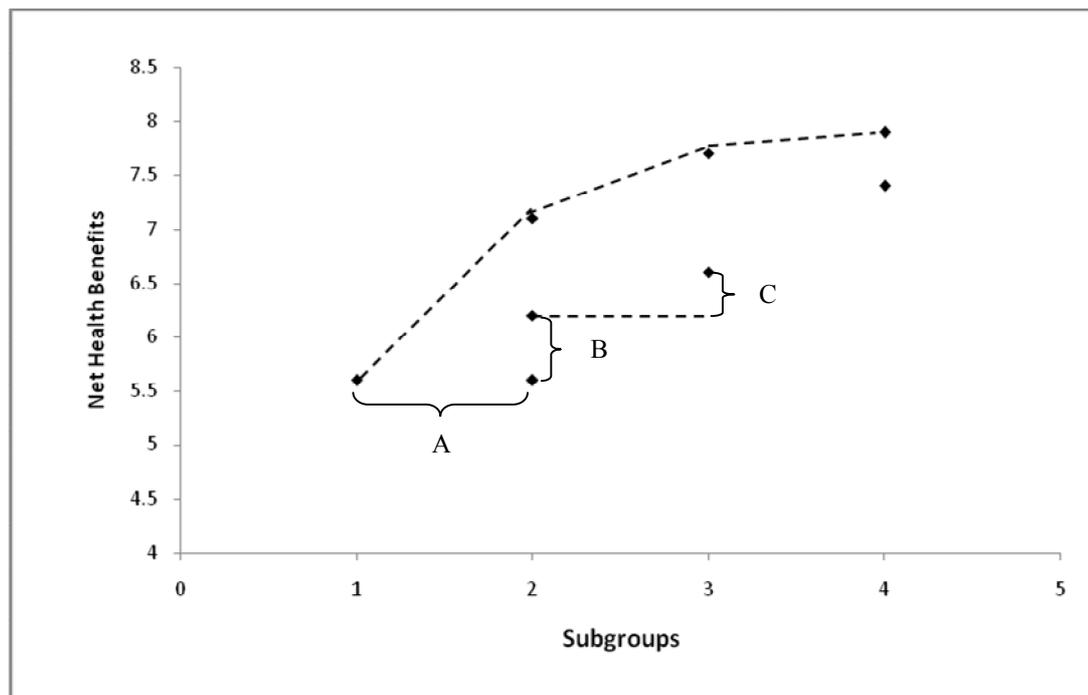
Table 3.2. Definition of subgroups. Grouping the population according to different specifications for different levels of disaggregation.

	$S=1$	$S=2$			$S=3$		$S=4$	
		$f=1$	$f=2$	$f=3$	$f=1$	$f=2$	$f=1$	$f=2$
	<i>All patients</i>	<i>No Diabetic</i>	<i>TIMI score <4</i>	<i>Normal Troponin</i>	<i>Non- Diabetic</i>	<i>TIMI score 1-2</i>	<i>TIMI <4 & non diabetic</i>	<i>TIMI <4 & Normal Troponin</i>
		<i>Diabetic</i>	<i>TIMI score ≥4</i>	<i>High Troponin</i>	<i>Compensated Diabetic</i>	<i>TIMI score 3-4</i>	<i>TIMI <4 & non diabetic</i>	<i>TIMI ≥4 & Normal Troponin</i>
					<i>Non compensated diabetic</i>	<i>TIMI Score 5-7</i>	<i>TIMI ≥4 & non diabetic</i>	<i>TIMI <4 & High Troponin I</i>
							<i>TIMI ≥4 & diabetic</i>	<i>TIMI ≥4 & High Troponin I</i>
Total NHB	5.6	5.6	6.2	7.1	6.6	7.7	7.4	7.9

Note; The matrix shows seven different specifications for subgroup analysis, three specifications to define two subgroups and two specifications to define three and four subgroups

The dotted line represents the set of the most efficient specifications for each partition. In addition, Figure 3.1 introduces two further important elements. First, there are instances where consideration of subgroups does not add any further societal benefit (in terms of expected NB) compared to what is achieved by providing treatment to the whole patient population. This might happen because the same treatment decision is appropriate for all subgroups (A). Second, in other cases further exploration of subgroups offers an additional societal benefit, even given current information. This could be exclusively explained by the effect of a different specification for the same number of subgroups (B) or because additional numbers of subgroups have been taken into account (C).

Figure 3.1. Efficiency Frontier for Subgroup Analysis.



Note: The dotted line joins the potential best specifications for each number of subgroups (S). The segment (A) shows that there is no value of heterogeneity between the points. Segment (B) illustrates the value of a specification used to define subgroups and (C) the value of considering additional subgroups.

This example also illustrates the problem of a limited analysis of heterogeneity. For instance, a treatment strategy might be cost-effective for patients with high troponin and not cost-effective for those with low troponin. Likewise, the treatment can provide value for money in diabetic patients but not in non-diabetic patients. What about diabetic patients with low troponin? While combining specifications might

resolve the problem, increasing the number of subgroups can become problematic as this is associated with transaction costs that will be discussed further in this chapter.

The elements presented in this section can be summarized in the following steps: (a) definition of subgroups must have biological, ethical and practical support; (b) to make explicit consideration about the sources of heterogeneity that will be explored, while baseline risk or treatment effect heterogeneity should always be included, the inclusion or omission of the analysis of costs and preferences should be properly justified³⁸; (c) graphical presentation of the efficiency frontier for subgroup analysis. This approach is consistent with the objective of CEA and highlights the value of exploring heterogeneity using existing data to identify potential best specifications.

3.4. Decision uncertainty in cost-effectiveness subgroup analysis

As decision-makers' interest moves away from population average recommendations towards more individualised decisions, it becomes important to be able to compare the uncertainty associated with each subgroup-specific decision against the potential benefits of resolving this uncertainty conditional on the new acquired insight into heterogeneity (Sculpher, 2008, Sculpher, 2010).

As explained in Chapter 2, uncertainty related to CEA can be classified in three categories, namely, *methodological*, *structural* and *parameter uncertainty* (Briggs, 2000). The latter is the focus of the present section. Parameter uncertainty is relevant since a wrong decision can be made because of the lack of precise (perfect) mean estimates. As a consequence, analysts should address the question of whether undertaking further research to resolve parameter uncertainty is worthwhile, which can be examined using *Value of Information (VoI)* analysis (Claxton and Posnett, 1996, Claxton, 1999a, Claxton, 1999b, Ginnelly et al., 2005). The adoption of this framework requires an estimation of the value of making decisions with an infinite sample (*perfect information*). In this circumstance, the decision-maker would be able

³⁸ Heterogeneity purely based on costs might be questioned, since new (more effective) technologies can be considered not cost-effective in patients with greater expected costs, other things being equal. NICE for, example, has recommended not to conduct subgroup analysis based purely on costs. The analysis of preferences is mainly limited by data, especially, when societal preferences are used.

to select the intervention that maximises NBs at the true value of the vector of parameters θ , that is

$$\max_j NB(j, \theta) \quad (3.10)$$

Since the true value of θ is unknown, we can only estimate an *expected* value of this quantity by averaging the maximum NBs over the joint distribution of θ ,

$$E_\theta \max_j NB(j, \theta) \quad (3.11)$$

The *expected value of perfect information* ($EVPI$) is the difference between the NBs derived from a decision made with perfect information (equation 3.11) and the NBs derived from decisions under current information (equation 3.1). This represents the expected gain (in NBs), for a single patient, from collecting further information and resolving the existing uncertainty.

$$EVPI = E_\theta \max_j NB(j, \theta) - \max_j E_\theta NB(j, \theta) \quad (3.12)$$

Notice that the population $EVPI$ can be derived by multiplying equation (3.12) and equation (3.2).

Claxton (1999b) shows that $EVPI$ represents an upper boundary for a new research proposal aimed at resolving the current levels of uncertainty. As long as obtaining new information is less costly than the population $EVPI$ there may be a positive potential payoff (Griffin et al., 2011). Thus, a necessary condition is achieved when this payoff is positive, otherwise investing in further research does not represent a good use of available resources.

3.4.1. Value of additional research and subgroup analysis

As we have seen in the previous sections when mutually exclusive subgroups are considered, different decisions can be made for different subgroups. Thus, under current information the decision-maker will need to choose for each subgroup s the strategy with the maximum NB . Equation (3.1) can therefore be re-expressed as

$$\max_j E_\theta NB_s(j, \theta) \quad (3.13)$$

with the expected value of the decision for subgroup s under perfect information being

$$E_\theta \max_j NB_s(j, \theta) \quad (3.14)$$

and the *EVPI* for the subgroup s is given by

$$EVPI_s = E_\theta \max_j NB_s(j, \theta) - \max_j E_\theta NB_s(j, \theta) \quad (3.15)$$

The $EVPI_s$ represents an upper boundary for further research on the target population considering that different decisions can be made for different subgroups. This expression considers the overall uncertainty of the population, which includes the uncertainty given by exchangeable and non-exchangeable parameters³⁹. Generalising equation (3.15), the total expected value of perfect information when considering S subgroups is the weighted average of each subgroup-specific *EVPI* weighted by the proportion of each subgroup in the population.

$$EVPI_{(S)} = \sum_{s=1}^S EVPI_s w_s \quad (3.16)$$

The population *EVPI* can be estimated by multiplying equation (3.16) by the future population of patients expected to benefit from the new information, which for s subgroups is given by

$$P_{(S)} = \sum_{s=1}^S \sum_{t=1}^{T_s} \frac{I_{s,t}}{(1+r)^t} \quad (3.17)$$

where T_s is the period over which the information that could be collected in the future about the actual decision problem is useful in the subgroup s (Philips et al.,

³⁹Exchangeability means the extent to which the information used to estimate $\theta_{s=1}$ can be used (or is exchangeable) to estimate $\theta_{s=2}$ (where s indicates subgroup 1 or 2). This point will be addressed with detail in the discussion section.

2008) and $I_{s,t}$ is the incidence over the period t . It follows that the population EVPI with S subgroups is

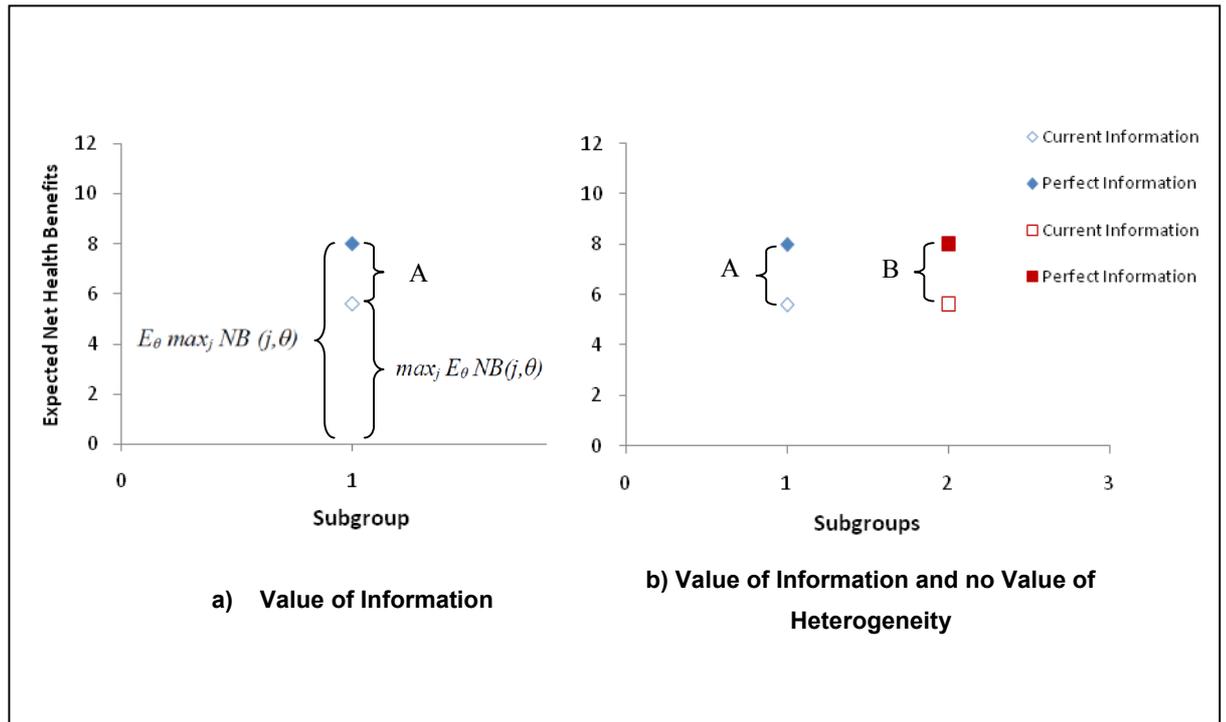
$$popEVPI_S = EVPI_S P_S \quad (3.18)$$

which represents the maximum amount of resources that the health system should be willing to pay for further research given a particular cost-effectiveness threshold.

This framework establishes a direct link between current decision uncertainty and the value of future research. Rather than considering uncertainty as a constraint for decision-making, VoI highlights its value as a potential source of health gain (Phelps and Mushlin, 1988, Claxton and Posnett, 1996, Claxton, 1999a). This chapter extends this concept to encompass the value of resolving heterogeneity and coin the term *Value of Heterogeneity* (VoH), to indicate that the additional health gains given by understanding heterogeneity involve two different sources. The first of these is the result of additional exploration of existing datasets in order to identify, characterise and quantify heterogeneity, which is a *static* assessment of the VoH. The second source involves the collection of new evidence to estimate parameters conditional to subgroups. This is the *dynamic* assessment of the VoH.

Several hypothetical scenarios are now presented to illustrate how uncertainty and heterogeneity converge in the decision problem. It starts in Figure 3.2 with the graphical illustration of the VoI concept. The vertical axis measures the expected NHB (assuming a particular value for the cost-effectiveness threshold) and the horizontal axis represents the possible alternative subgroup definitions (disaggregation) of the population, where ‘one subgroup’ expresses the population average NHB.

Figure 3.2. The value of information with and without subgroups.



Note: (a) The empty diamond represents the maximum expected NHBs (current information) and the filled diamond the expected maximum NHBs (perfect information). (b) The base reference case in (a) is also illustrated for two subgroups. A corresponds to the EVPI for the whole population and B the EVPI, when two subgroups have been considered ($s=2$). No value of heterogeneity is shown in this graph because the NHBs are the same for one or two subgroups either with current or perfect information.

Figure 3.2a illustrates the concept of the expected value of perfect information. Here, the empty diamond marker represents the expected NHB under current information, while the solid diamond marker indicates the maximum expected NHB achievable under perfect information. The difference between these two corresponds to the population-average EVPI.

In Figure 3.2b, the patient population is now split into two subgroups. It illustrates the situation where there is economic value in carrying out further research to resolve decision uncertainty. However, this value is not associated with investigating and reflecting heterogeneity in the population for the purposes of subgroup CEA. In fact, the maximum expected NHB obtained under current information for the whole population (represented by the empty diamond marker) is the same as what would be achievable by considering two subgroups (indicated by the empty square marker). The scenario reflects the fact that the same treatment decision is made for both

subgroups, and there is no further value associated with investigating heterogeneity, which is indicated by the fact that the EVPI for the entire population (the vertical distance A) is the same as the EVPI with two subgroups (the vertical distance B).

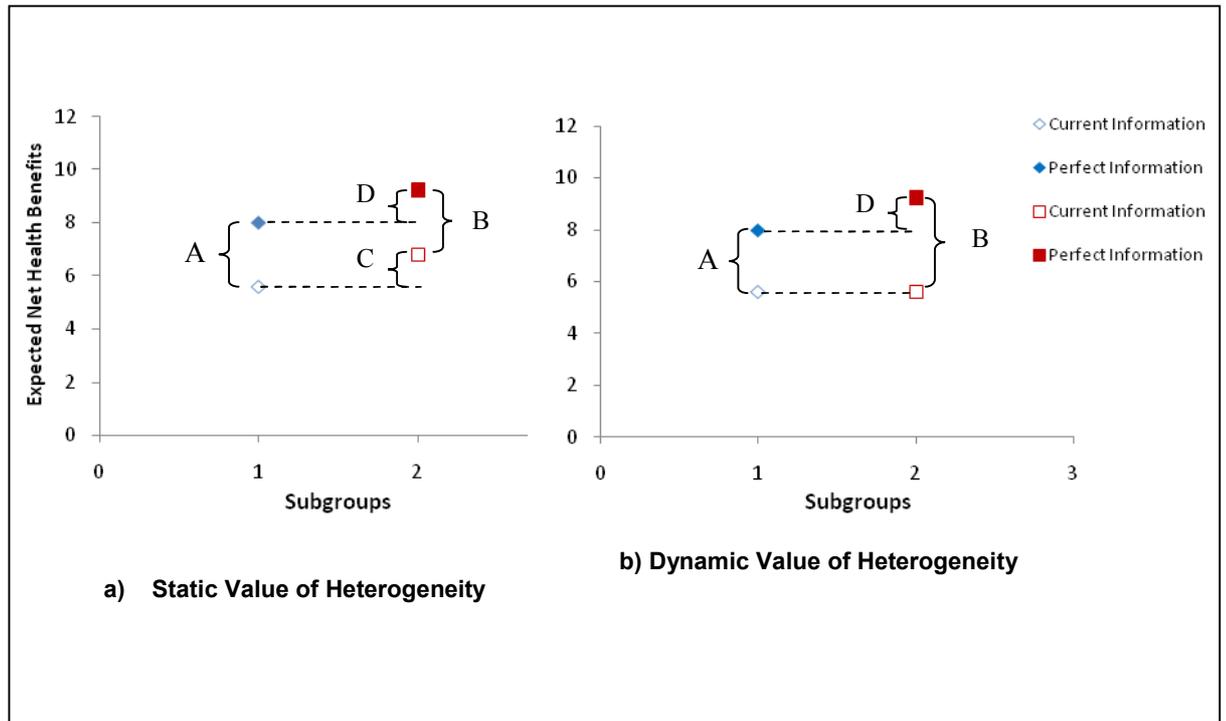
3.4.2. Static and Dynamic Value of Heterogeneity

A more likely scenario occurs when there is value in identifying and reflecting between-patient heterogeneity using existing datasets, over and above the value associated with undertaking further research. This case is presented in Figure 3.3a. Here, the total expected NHB with current information (represented by the empty markers) is greater when subgroups are considered. The difference between the total NHB for the entire population and the total NHB when considering subgroups (represented by the vertical distance C) is what is termed the *static value of heterogeneity*, which derives from analyses of existing data with a view to identifying relevant subgroups and increasing population total NHBs by facilitating different decision for different subgroups, but without any further primary research through additional data collection.

In addition, notice that the scenario depicted in Figure 3.3a indicates that even if we were able to collect additional data through new research and resolve any *decision uncertainty* for the subgroups, the expected NHB to be gained would be similar to what could be derived from the population-average case. This is indicated by the fact that the vertical distance B is equal to the vertical distance A. In this scenario, a policy maker might be interested in making different decisions for different subgroups, according to the evidence available, and further research is still worthwhile to resolve uncertainties not associated with heterogeneity. Figure 3.3b shows another situation. Here, the same decision would be made for both subgroups under current information. This would yield the same total expected NHB as for the whole population. However, the estimate under perfect information obtained when considering subgroups (represented by the filled squared) is greater than the NHB under perfect information derived from considering the population as a whole (represented by filled diamond). The difference between these two points (indicated by the vertical distance D) is the *dynamic value of heterogeneity*, since to benefit from this the policy maker has to fund new additional research. This additional value

is not only explained by greater uncertainty around the estimation of conditional parameters, but also from the fact that different decisions can be made for different subgroups.

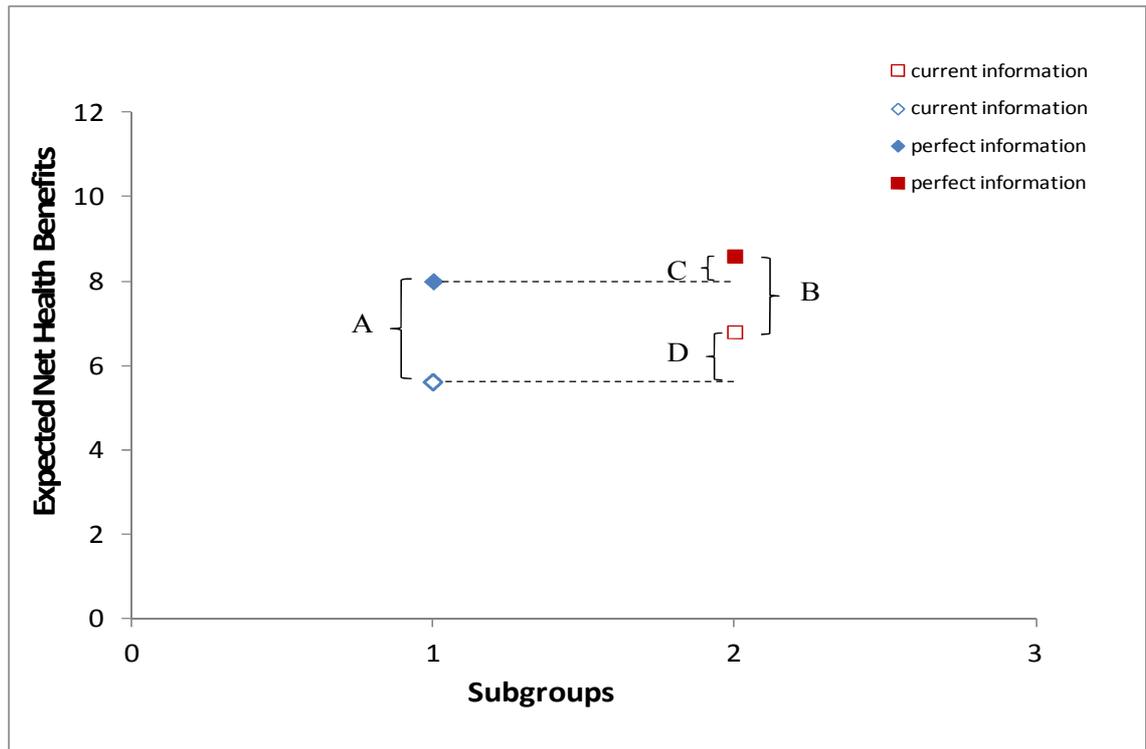
Figure 3.3. Static and dynamic value of heterogeneity.



Note: (a) A represents the EVPI for the average population, B is the EVPI for the population considering two subgroups, C is the increment in the expected maximum NHB and D is the increment in the maximum expected NHB. In this illustration C represents the static value of heterogeneity and no additional dynamic value in undertaking further research conditional to subgroups is observed since $A=B$ and $C=D$. The Dynamic Value of Heterogeneity is represented by D. (b) In this case there is only dynamic value of heterogeneity $B>A$, $D>0$ and $C=0$ (zero static value).

Figure 3.4 illustrates a situation in which there are both static and dynamic values of heterogeneity ($C>0$ and $D>0$). A particular feature of this example is that the EVPI when considering subgroups (distance B) is less than the average EVPI (distance A). This is expected to occur when the specification used to define subgroups is informative about heterogeneity. In this situation, the effect is not only observed under perfect information but also under current information (positive static value). Hence, the difference between current and perfect information is lower than the average.

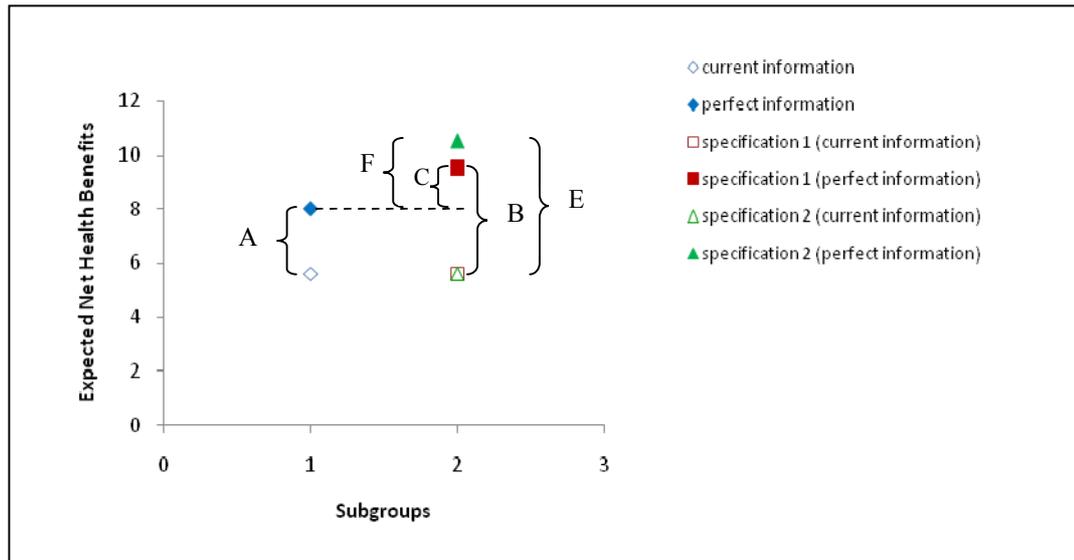
Figure 3.4. Dynamic value of heterogeneity and decrease in EVPI.



Note: A represents the average EVPI, B is the two-subgroups EVPI, C is the increment in the expected maximum NHB and D is the increment in the maximum expected NHB. In this case, $A > B$ and $C < D$.

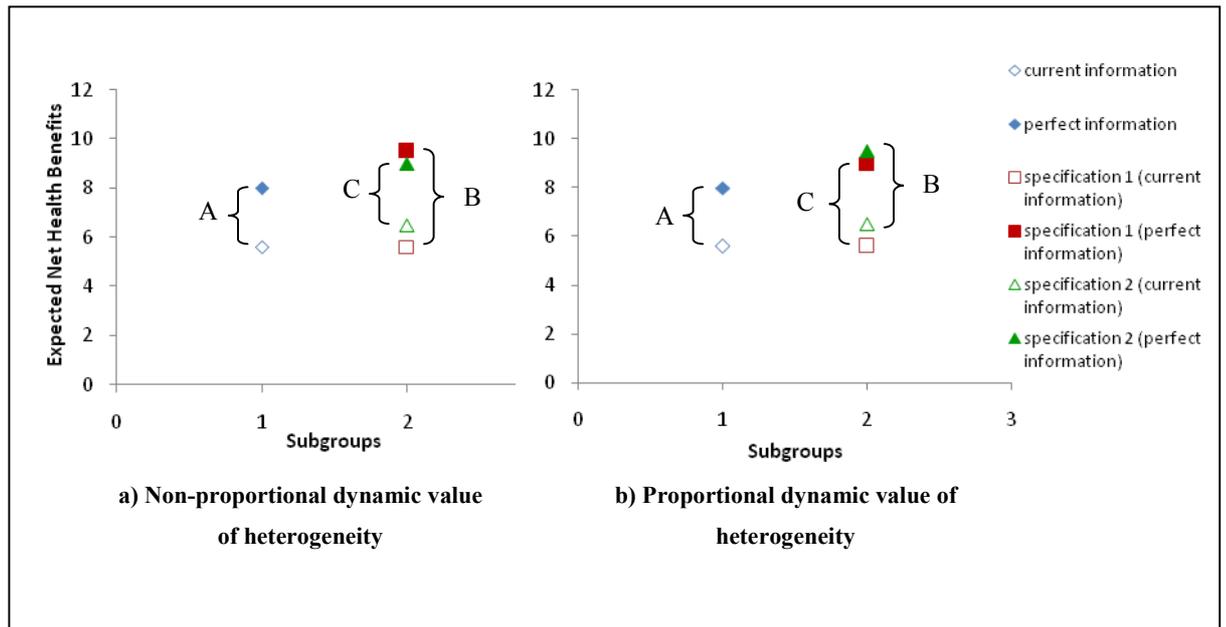
Figure 3.5 shows a further scenario in which an alternative subgroup specification has been introduced (indicated by the empty triangle). Similar to the scenario presented in Figure 3.3b, both specifications have the same expected *NHBs* under current information (i.e. no static value of heterogeneity). However, if further research were to be carried out, different treatment decisions could be made for different subgroups (regardless of the specification used), yielding much greater *NHBs* than those derived from a decision applied to the population as a whole. This is indicated by the fact that the vertical distance B is greater than A. Furthermore, it may be conceivable for this dynamic value of heterogeneity to differ for different specifications used to define the subgroup (i.e. $F > C$).

Figure 3.5. Dynamic value of heterogeneity and its impact on subgroup specification.



Finally, the impact of considering heterogeneity on the selection of subgroup specifications is presented in Figure 3.6 where we return to the scenario in which different specifications for two subgroups are available. The expected NHBs (with current information) estimated for one specification (empty triangle) is assumed to be greater than those obtained from the other (empty square), which itself is assumed to be equal to the NHBs derived for the population as a whole (empty diamond). Under current information, a *static* value of heterogeneity is observed for one specification but not for the other, which would guide the decision-maker's recommendation based on the specification with the greater expected NHBs. However, it is entirely conceivable that further research could reveal the specification with the lower static value of heterogeneity to be the one which could produce the greater expected NHBs under perfect information ($B > C = A$ in Figure 3.6a). This can be expressed as *non-proportional dynamic value of heterogeneity*, meaning that there is a dynamic value of heterogeneity for either specifications, but the potential NHBs obtained under perfect information for the specification with less static value is greater than the one with higher static value. One could also conceive a scenario (Figure 3.6b) where there is a *proportional dynamic value of heterogeneity* where the expected benefits estimated under perfect information (expected maximum NHB) increase in the same proportion as the static value (e.g. $C = B$).

Figure 3.6. Proportional and non-proportional value of heterogeneity.

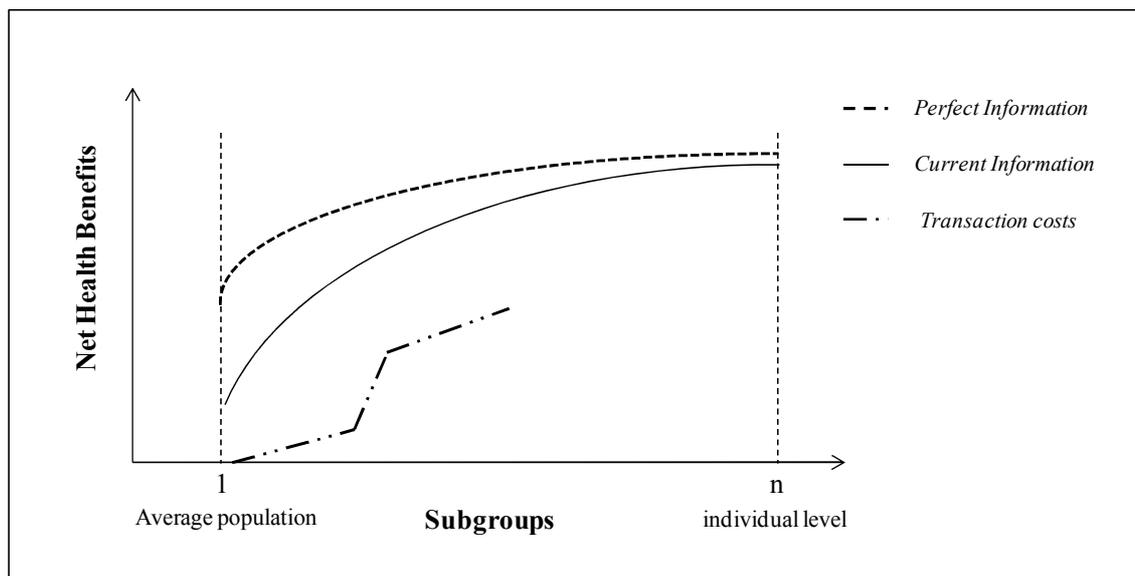


Note: (a) A represents the EVPI estimated from the whole sample, B is the EVPI considering two subgroups defined according to specification 1 and C is the EVPI considering two subgroups and specification 2. In this case $B > C$ and $C = A$. (b) A proportional dynamic value is presented where $B > A$, $C > A$ and $C = B$.

3.5. Presentation of subgroup analysis and choice of the optimal number of subgroups: the role of the cost function

By considering the static and dynamic dimensions at the same time, the trend of the expected NHBs as a function of the number of subgroups can be shown graphically for both current and perfect information. Since several alternative specifications are available for each number of subgroups, ranging from no subgroups (indicated by the vertical bar on the left in Figure 3.7) to individualized decisions (indicated by the vertical bar on the right in Figure 3.7) the analyst is required to plot -at least- the most efficient specifications (i.e. those with the highest NHB for each number of subgroups). In addition, those specifications that have lower NHB under current information but are expected to produce higher NHB with perfect information might also be included in the graph.

Figure 3.7. Representation of the gaining in expected NHB with current and perfect information when heterogeneity is considered.



Note: The transaction cost function shows two kink points illustrating a segment where the additional costs are higher than the additional gains. This represents the case where the optimal number of subgroups is the lower between those two points.

In principle, if there is no residual unexplained heterogeneity (i.e. there is complete knowledge of the individual characteristics that determine variability), the decision-maker has the best possible information to allocate resources efficiently. At the other extreme, the maximum value of exploring heterogeneity is observed when no subgroups have been taken into account (which can be represented by EVIC). In terms of uncertainty, it is not the case that the maximum decision uncertainty is at the average level. Indeed, more granular analysis could increase or decrease it, depending on the informative capacity of the specification used and the reduction of the sample size due to disaggregation⁴⁰. However, it is expected that if heterogeneity is progressively explained with informative covariates, then the remaining uncertainty should decrease with higher levels of disaggregation. In addition, although decision uncertainty will decrease as more heterogeneity is revealed, it is

⁴⁰ Total uncertainty depends on the standard error of the parameters. One particular specification can be informative in the sense that decreases the standard deviation of the key parameters leading to a lower standard error. However, these estimations are made on the basis of lower sample size which increases the standard error. Because the total uncertainty is the weighted average across subgroups, the subgroup with higher sample size will add more than the smaller subgroup. Hence, the final quantity might be affected more by the increment in precision rather than the loss in sample.

never going to be completely resolved. This is because the true value of the individual treatment effect can never be measured, since the counterfactual can never be observed. A graphical representation of these concepts is presented in Figure 3.7.

Under this framework, the optimal number of subgroups tends towards n ; however, because there are significant constraints to fully characterise heterogeneity the optimal number of subgroups will never be equal to n . One such constraint is the lack of information to characterize heterogeneity, for example, the data available usually do not have all the covariates in which we are interested (known unknown). A second element is the dearth of evidence to understand this heterogeneity, which leads to the fact that new research should be undertaken in order to find new factors able to explain some part of the variability (unknown knowable unknown). A third factor is related to the implementation of recommendations based on subgroups. Depending on the healthcare system, different strategies can be employed to monitor, enforce and ensure the fulfilment of healthcare recommendations. In general, this is expected to be more expensive as more heterogeneity is considered. Therefore, incorporating heterogeneity in health decision-making entails several costs not only limited to the required research but also other transactions that must be considered.

The formalization of a cost function which includes the components of research and other transaction costs is needed to answer the question about how many subgroups should be considered in a particular recommendation. A general form can be expressed in terms of three key elements:

$$\Delta C_t = f(\Delta X, \Delta Y, \Delta Z)$$

where ΔC_t defines the incremental transaction costs between two levels of disaggregation, ΔX represents the additional effort of implementing a further level of disaggregation, which may include additional health professionals' time, the cost of additional diagnostic tests required to categorize the patient and the cost of dissemination and implementation strategies. Next, ΔY corresponds to the effort required to enforce and monitor the compliance with guidelines. If the effort is paid in financial terms ΔY becomes the opportunity cost of an alternative allocation of

those resources. On the other hand, if the authority does not make any effort, ΔY becomes the health forgone due to partial implementation. Finally, ΔZ is the additional cost of future collection of information to resolve uncertainty of conditional parameters at the extra level of disaggregation. This source of cost is relevant when decisions are implemented conditional to further research funded by the health system. Assuming that all transaction costs fall on the health budget, the function can be expressed in NHB using the cost-effectiveness threshold (Figure 3.7).

The analysis of transaction costs and the efficiency frontier provides a rational element to make decisions about the optimal number of subgroups. It is argued that the optimal number of subgroups that should be implemented is the maximum level of disaggregation where the expected additional benefits are greater than the additional costs. This might be expressed in terms of the ratio between the incremental benefits and incremental costs of two adjacent levels of disaggregation (e.g. 1 and 2). If the ratio is lower than one, then the next relevant comparison is the levels 1 and 3. As a consequence of this analysis, the analyst can present the efficiency frontier net of transactions costs.

3.6. *Brief illustrative example*

A numeric example to illustrate the static and dynamic values of heterogeneity is provided in Table 3.3 for a simple hypothetical decision problem comparing two mutually exclusive treatments (T1 versus T2). For simplicity, we consider here two subgroups, each representing 50% of the population and only the first five initial iterations of a probabilistic sensitivity analysis are reported in the first column of Table 3.3. The maximum NHB at each iteration is reported (in the columns headed *Max*) and the expected NHB is calculated for each alternative (at the bottom of each). Table 3.3 shows a set of possible NHB values, for the population-average case and for each subgroup. Under current information, if decisions were taken at population-average, T2 would be the optimal choice since it generates a total of 12.8 net-QALYs. On the other hand, specific estimates for both subgroups suggest that T1 should be adopted for subgroup 1 (12.8 net-QALYs) and T2 for subgroup 2 (13 net-QALYs). Further, the total weighted average NHB $[(12.8 \times 0.5) + (13 \times 0.5)] = 12.9$

net-QALYs] is greater than the population average, indicating a positive static value (12.9-12.8=0.1 QALYs). Under perfect information, the expected maximum NHB for the population on average are 13.6 net-QALYs whereas for subgroups the expected maximum NHB would be 15.1 net-QALYs, indicating a positive dynamic value (15.1-13.6=1.5 net-QALYs). In terms of potential gains due to future research, the example shows that the EVPI increases with subgroups (2.2 versus 0.8 net-QALYs). In this case, a decision-maker should recommend T1 for subgroup 1 and T2 for subgroup 2, based on static value. In terms of future research, the positive static value associated with an increase of the EVPI suggests that conducting future research considering heterogeneity is worthwhile.

A second example is presented in Table 3.4. It shows the expected NHBs for a specification that defines three subgroups, i.e. it adds a subgroup to the previous example. Subgroups 1 and 3 each represent 25% of the population and subgroup 2 represents 50%. After 5 hypothetical iterations, T1 would be recommended for subgroup 1 and 3 and T2 for subgroup 2 on the basis of expected NHB given current information.

Table 3.3. Value of heterogeneity. A stylized example with two subgroups.

iteration	Population			Subgroup 1			Subgroup 2		
	T1	T2	Max	T1	T2	Max	T1	T2	Max
1	15	15	15	12	18	18	18	12	18
2	10	15	15	8	14	14	12	16	16
3	12	11	12	16	10	16	8	12	12
4	14	11	14	19	9	19	9	13	13
5	11	12	12	9	12	12	13	12	13
Average	12.4	12.8	13.6	12.8	12.6	15.8	12	13	14.4
E(NHB) with current information = 12.8 E(NHB) with perfect information = 13.6 EVPI = (13.6 – 12.8) = 0.8 E(NHB) _s with current information = (12.8 x 0.5) + (13x 0.5) = 12.9 E(NHB) _s with perfect information = (15.8 x 0.5) + (14.4 x 0.5) = 15.1 EVPI _s = (15.1 - 12.9) = 2.2						Static Value (with current information): E(NHB) _s – E(NHB) = 12.9-12.8 = 0.1 Dynamic Value (with perfect information): E(NHB) _s – E(NHB) = 15.1 – 13.6 = 1.5			

Interestingly, no additional static value of heterogeneity is observed compared with the previous case with two subgroups, since in both cases the expected NHB with

current information is 12.9 net-QALYs. Furthermore, it is worth noting that the expected NHB under perfect information is higher than the case with 2 subgroups. Likewise, EVPI for 3 subgroups is greater than the EVPI observed for the 2 subgroups (2.4 versus 2.2 net-QALYs). In this case, a decision-maker might have no preference for either specifications under current information, as they both give the same NHB. However, further research has more value if additional information is planned to be collected in order to resolve uncertainty around conditional parameters for 3 instead of 2 subgroups.

Table 3.4. Value of heterogeneity. A stylized example with three subgroups.

Iteration	Population			Subgroup 1			Subgroup 2			Subgroup 3		
	T1	T2	Max	T1	T2	Max	T1	T2	Max	T1	T2	Max
1	15	13	15	12	18	18	17	9	17	14	16	16
2	9	15	15	9	15	15	9	15	15	9	15	15
3	12	12	12	16	10	16	8	13	13	16	12	16
4	14	12	14	19	9	19	10	15	15	17	9	17
5	12	12	12	8	11	11	15	12	15	10	13	13
Average	12.4	12.8	13.6	12.8	12.6	15.8	11.8	12.8	15	13.2	13	15.4
<p>E(NHB) with current information = 12.8 E(NHB) with perfect information = 13.6 EVPI = (13.6 – 12.8) = 0.8</p> <p>$E(NHB)_s$ with current information = (12.8 x 0.25) + (12.8 x 0.5) + (13.2 x 0.25) = 12.9 $E(NHB)_s$ with perfect information = [(15.8 x 0.25) + (15 x 0.50) + (15.4 x 0.25)] = 15.3 EVPI_s = (15.3 - 12.9) = 2.4</p>									<p>Static Value (with current information): A. Three versus one subgroup $E(NHB)_s - E(NHB) = 12.9 - 12.8 = 0.1$ B. Three versus two subgroups $E(NHB)_s - E(NHB) = 12.9 - 12.9 = 0$</p> <p>Dynamic Value (with perfect information): A. Three versus one subgroup $E(NHB)_s - E(NHB) = 15.3 - 13.6 = 1.7$ B. Three versus two subgroups $E(NHB)_s - E(NHB) = 15.3 - 15.1 = 0.2$</p>			

3.7. Discussion

Cost-effectiveness evidence is increasingly being used to inform decision-making about adoption of new health technologies, in response health systems' need to allocate resources efficiently. In most cases the efficiency of this process can be improved by identifying subsets of the population where the new intervention is cost-effective from those where it is not. This chapter argues that the decision rule for adoption or rejection of new interventions in different subgroups should be based on expected NHBs. Furthermore, by using NHBs, the potential best specification for resource allocation can be easily identified for different levels of disaggregation. The relevance of defining and selecting subgroups is emphasized and it is proposed that the selection process under current information should follow an efficiency criterion for the best specification. The efficiency frontier for subgroups is presented as a tool to identify those that are most relevant. However, this does not mean that other specifications should be eliminated from the analysis. Indeed, they might produce more benefits if further research is undertaken to resolve the current uncertainty.

One pragmatic reason for developing this framework is the fact that most cost-effectiveness studies carry out subgroup analysis to explore heterogeneity. This work offers a systematic approach to undertaking such analysis. A second reason relates to the limitation of implementing decisions at individual level in a collectively funded NHS. Two elements are highlighted here. First, generating guidelines with high levels of disaggregation is difficult due to our lack of understanding of the total variability in cost-effectiveness between patients. Moreover, even if those guidelines were produced, there would be associated costs that would hinder their implementation (these costs have been presented in this chapter as a transaction cost function). Second, patients (or doctors on their behalf) might not choose the most cost-effective treatment (Eddy, 1991). Even in the situation where we have perfect and complete information at individual level (and an optimal efficient decision can be made at individual level), a proportion of the population could still face not having access to new, more expensive (and effective) treatments due to budget constraints. This is not an argument in favour of subgroups analysis, since subgroup-based decision-making face the same problem of denying treatment to some patients.

However, the restrictions imposed by social constraints are better accepted when they affect groups of the population than when they affect individuals who are not able to identify themselves as member of a particular subset (Lee and Ishii-Kuntz, 1987).

As mentioned in Chapter 2, recent contributions in the field of health econometrics have demonstrated that the joint distribution of potential outcomes can be estimated using information about patient choices for treatment (Basu et al., 2007, Basu, 2011). It has been suggested that when patients make decisions for treatment, they consider not only the information publicly available and provided by the health professionals but also other individual characteristics that cannot be observed by a social decision-maker. Access to this information could help characterize the heterogeneity in treatment effects, which could be used to generate more accurate cost-effectiveness estimates. The challenge for a social decision-maker is to identify the proportion of patients for whom a new strategy is (cost-)effective based on observable and measurable characteristics, not only for the purposes of comparative effectiveness research but also to take into account the opportunity cost of alternative resource allocations⁴¹. Thus, subgroup analysis has a key role even when individual treatment effects and, as a result, individual NB can be estimated.

The value of considering heterogeneity when making decisions in healthcare is not a new concept. The relevant literature has been presented in Chapter 2. This chapter extends previous work in this field, making explicit two dimensions of value, called here *static* and *dynamic*. While the importance of the static value is intuitive and easy to realize, this could be less obvious for the dynamic value. The dynamic value corresponds to the additional (population) health that is expected to be achieved when an infinite sample is collected to resolve current uncertainty for a particular level of disaggregation compared to the average case. This value can be disentangled in two sources: (i) the value of an infinite sample to estimate conditional parameters⁴², and (ii) the value of estimating conditional parameters with an infinite sample. The first one relates to the fact that, under uncertainty, expected maximum

⁴¹ This is especially relevant for entities such as NICE in the UK.

⁴² This refers to parameters estimated conditional to the subgroup category determined by a particular specification

NHB are higher than maximum expected NHB (i.e. EVPI greater than zero). This is true at any level of disaggregation and it does not constitute, in itself, a gain due to heterogeneity. The second is the value due to heterogeneity under perfect information. Because the value of estimating conditional parameters is also observed under current information, it is to be anticipated that the difference between the expected maximum NHB and the maximum expected NHB (i.e. EVPI) is smaller than the average (or the previous level of disaggregation). If a particular specification does not have enough informative capacity or because the amount of data (sample size) to examine its effect is too limited, EVPI can increase instead of decrease.

The concepts behind the static and dynamic VoH are closely related to the current metrics proposed in this field, such as EVPI and EVIC. First, dynamic value relates to the value of resolving uncertainty in order to update a guideline in the future that considers different decisions for different subgroups. Total EVIC represents only the static value of heterogeneity and can be understood as though each individual were representative of a subgroup of homogeneous patients. This idea is consistent with two previous reports that have calculated EVIC from cost-effectiveness models (Basu and Meltzer, 2007, van Gestel et al., 2012). However, some confusion remains in the literature in terms of the degree of correlation between EVIC and EVPI. While van Gestel et al. (2012) have stated that EVIC and EVPI are not correlated, Basu and Meltzer (2010) have indicated that EVIC can also be expressed as an extension of EVPI. Basu and Meltzer (2010) point out that because EVPI corresponds to the value of having an infinite sample, this quantity can also be expressed as “the difference between the expected per patient outcome realized based on decisions made with such perfect information minus the expected outcome realized based on current information” (page 4). Since EVPI is defined on the basis of the expected per patient outcome, then it is appropriate to define EVIC as an extension of EVPI. However, when EVPI is estimated, most studies express this as the value of an infinite sample to obtain the true **mean** parameter, which is calculated by averaging the maximum NHB across many possible (average) realizations simulated by a PSA. In other words, by estimating the EVPI this way, one does not derive any estimate of the value of individualized information (e.g. individual treatment effect). Therefore, it is reasonable to argue that, in practice, and using current methodological tools, EVIC

does not capture EVPI. While EVPI depends on the variability of the expected NHB within individuals (which depends on the sample size), EVIC depends on the variability between individuals (which relies on the true variability among patients). An extended and formal explanation of this issue is presented in Appendix 1.

In addition, EVIC for specific parameter(s) ($EVIC_{\theta_i}$) has been proposed as an informative metric to represent the value of eliciting information about one (or a few) parameter(s) (Basu and Meltzer, 2007). This quantity has also been interpreted as the value of implementing a subgroup-based policy (van Gestel et al., 2012). In fact, $EVIC_{\theta_i}$ is similar to the static value of heterogeneity. The advantage of the $EVIC_{\theta_i}$ approach is that it can provide an estimate of the static value for a set of several parameters, which can be less computationally demanding than the approach presented in this chapter. However, it does not provide specific information as to which decisions should be made in specific subgroups defined on the basis of the combinations of such parameters. Furthermore, it has been proposed that $EVIC_{\theta_i}$ can be used to rank parameters to help health professionals to prioritize information that should be elicited from individuals (Basu and Meltzer, 2007). This is similar to the efficiency frontier for subgroup analysis presented here, where alternative specifications are compared in terms of their static value. However, while $EVIC_{\theta_i}$ contrasts different levels of disaggregation without making them explicit, the efficiency frontier provides more detailed information for policy making.

Another source of possible confusion can arise from the interpretation of $EVIC_{\theta_i}$ and EVPPI. Both share characteristics that are useful for comparisons between different parameters. However, whilst $EVIC_{\theta_i}$ provides information on the value of using individual information about θ_i , EVPPI gives information about the value of resolving the uncertainty of the effect of θ_i on the net health outcome. In other words, while $EVIC_{\theta_i}$ or the static value, are useful for making decisions across subgroups under current information; EVPPI provides information about which parameters should be investigated in the future. Another possible source of confusion is the uncertainty of the specifications that define subgroups (e.g. moderators of treatment). For example, we might know for sure that our patient has a genetic polymorphism (complete information) but we might be uncertain about the effect of the polymorphism on the (net) health outcome. In a more realistic world, we are

uncertain about the effect of the polymorphism on the outcome but also whether the patient has the polymorphism (since genetic tests are not 100% sensitive and specific). It is therefore important to emphasize that $EVIC_{\theta_i}$ assumes that the information at individual level is 100% accurate. In other words, it represents the value of using information 100% specific and 100% sensitive at individual level (van Gestel et al., 2012). In principle, the uncertainty around the value that a covariate takes can be represented as another EVPPI for the diagnostic of the condition. EVPPI calculation is computationally demanding for the population analysis and may be even more difficult in the case of subgroups. This work assumes that EVPPI methods can be applied as an extension to subgroup analysis, which might not be the case due to correlation between subgroups. Applied work is needed to tease out the strengths and limitations of the EVPPI approach to subgroup analysis.

An important concept mentioned in this chapter is that subgroup specific parameters can be exchangeable or non-exchangeable. One parameter (θ_i) is exchangeable if the information used to estimate $\theta_{i|s}$ conditional for one subgroup can be used to estimate $\theta_{i|(1-s)}$ for another subgroup. The $EVPI_s$ estimated as the weighted average of the EVPI for each subgroup captures the uncertainty given by both exchangeable and non-exchangeable parameters. Therefore, $EVPI_s$ is the informative metric with which to address the question of whether further research should be conducted in order to update a guideline for the whole population, considering that different decisions can be made in different subgroups with future information. However, we might be interested in conducting information only in the subgroup. In this case, we should choose the one that offers the highest population EVPI. The aim in the future would be to synthesize the new information and update the guideline for the whole population. If this is the purpose, the EVPI estimate, as presented in this chapter, underestimates the real EVPI for that subgroup, because it does not take into account the value of (previous) information provided by another subgroup. The extent to which parameters relating to one subgroup add value of information to decisions relating to another subgroup is given by the degree of exchangeability of the parameters relating to this second subgroup. This issue offers an opportunity for future research.

It is important to highlight that the approach to decision-making that underlies and motivates this work assumes a health authority to be able to: (i) undertake a systematic review and analysis of the current scientific evidence; and (ii) prescribe the most cost-effective treatment for patients. While this is a reasonable characterization of the current scenario in the UK NHS, it is not the case in other countries. Even in the UK, there has been deliberation of a move to the decision-making approach from a centralized to a devolved process. This might include modifications of the current process in order to satisfy patient preferences and choices for treatment (Department of Health, 2010). It was mentioned that choices for treatment could be informative about the joint distribution of potential outcomes. Therefore, a decision-making process that formally considers these choices might be consistent with the societal aims. The role of choices for treatment in a devolved healthcare system is a matter that requires further research.

Overall, while analysing data to obtain average cost-effectiveness estimates does not recognize the underlying heterogeneity, a systematic analysis of subgroups in CEA offers relevant information to central or local authorities for making better decisions in allocating resources. The value of heterogeneity has been presented highlighting two dimensions of the decision-making problem. While the static value informs whether different decisions must be made in different subgroups, the dynamic value reveals whether heterogeneity should be considered in future research. Application of this framework to real data is needed to reveal the feasibility of its implementation in practice, which is the motivation for the next chapter of this thesis.

4. CHAPTER 4:

AN APPLICATION OF THE VALUE OF HETEROGENEITY FOR COST-EFFECTIVENESS SUBGROUP ANALYSIS

4.1. Introduction

The previous chapter developed a methodological framework for cost-effectiveness subgroup analysis. The main points of this approach can be summarized in three key messages. First, several specifications supported by biological rationale should be compared in terms of their potential gains. This comparison offers an opportunity to address the selection of subgroups considering an efficiency criterion. Second, the optimal number of subgroups is defined at the maximum point of stratification, such that the additional benefits are greater than the additional (transaction) costs. Third, the value of heterogeneity has two dimensions, the additional expected net benefits according to the available information (static value) and the potential expected net benefits that can be achieved if the remaining uncertainty is resolved for a particular level of disaggregation (dynamic value).

The aim of this chapter is to apply the methods described in Chapter 3 in a real cost-effectiveness study in order to test its feasibility, identify its practical strengths and weaknesses as well as presenting the potential impact in the decision-making process. The case study was carried out with data collected alongside a multicentre clinical trial performed in the UK to explore alternative treatments in patients with coronary heart disease (CHD).

The chapter starts by presenting the clinical and cost-effectiveness study used as a motivational example followed by the description of the methods. Next, results are presented emphasizing the key points of the new approach. The chapter ends with a discussion of the main findings, their potential applications, practical strengths and weaknesses of the methods applied here and final conclusions.

4.2. Motivational example: RITA-3

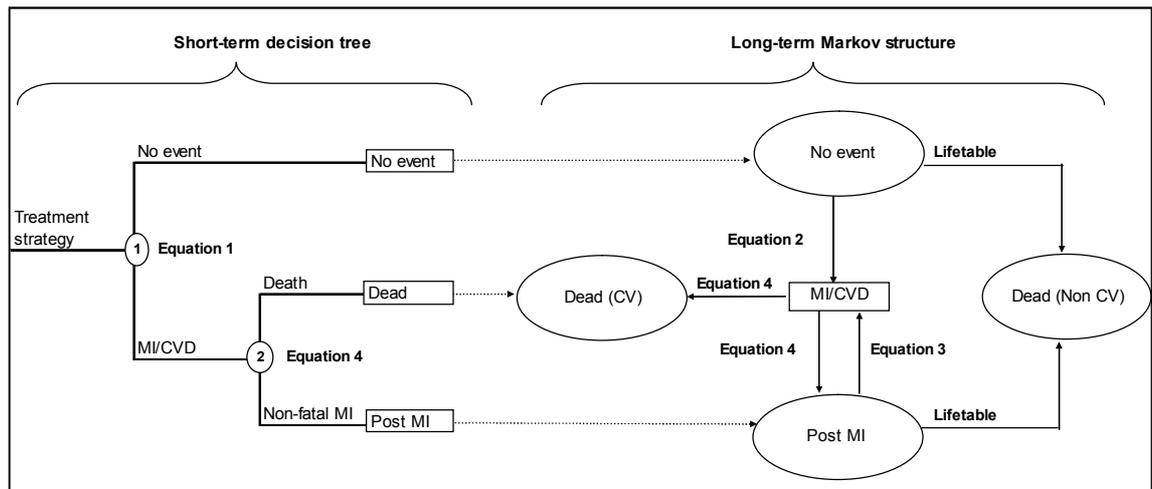
4.2.1. Clinical effectiveness and cost-effectiveness

Non-ST-elevation acute coronary syndrome (NSTE-ACS) is one type of CHD that is associated with high incidence of complications and mortality (Anderson et al., 2007). The third Randomised Intervention Trial of unstable Angina (RITA-3) is one of the trials performed to compare the efficacy of an intensive versus a conservative strategy in patients with NSTE-ACS (Fox et al., 2002). After a five-year follow-up, the study demonstrated that an intensive strategy was able to reduce the risk of the primary outcome, which was defined as a combined endpoint of death and myocardial infarction (MI) (primary outcome: OR 0.78; 95% CI 0.61 to 0.99) (Fox et al., 2005).

Henriksson et al. (2008) reported in a subsequent study, a cost-effectiveness analysis carried out alongside the RITA-3 trial. The study was primarily based on the individual patient data of the trial, which were used to populate a decision analytic model. The structure of the model is presented in Figure 4.1. In addition to standard cost-effectiveness analyses, this study aimed to explore heterogeneity across patients. Thus, the analysis was conducted to produce cost-effectiveness estimates for five subgroups defined on the basis of a predicted baseline risk score. The baseline risk score was estimated from a logistic regression used to predict the probability of presenting the primary outcome using a set of individual characteristics. This approach is consistent with the previous clinical report (Fox et al., 2005). In order to operationalize this in practice, each subgroup was characterized as an average patient profile such that the predicted probability is equal to the average predicted probability of that subgroup.

The study showed that the invasive treatment was not cost-effective at a threshold of £20,000/QALY gained (ICER=21,943). The subgroup analysis showed that the invasive treatment was cost-effective only in two subgroups, both with the highest risk (£11,898/QALY gained and £10,476/QALY gained respectively). Patient profiles and cost-effectiveness results are presented in Table 4.1. Details about the clinical and cost-effectiveness study are presented in Appendix 2.

Figure 4.1. Structure of the RITA-3 decision model.



Note: Reproduced with permission (Henriksson et al., 2008)

Table 4.1. Incremental cost-effectiveness ratios (ICERs) of the RITA-3 cost-effectiveness study for subgroups.

	1 st quartile	2 nd quartile	3 rd quartile	4 th quartile, lower eight (4a)	4 th quartile, upper eight (4b)
Age	45	52	52	61	66
Diabetes	No	No	No	No	Yes
Previous MI	No	No	Yes	Yes	Yes
Smoker	No	Yes	No	Yes	No
Heart rate	72	82	82	87	97
ST depression	No	No	Yes	Yes	Yes
Angina (grade 3 or 4)	Yes	No	Yes	No	No
Sex	Female	Male	Male	Male	Male
Left bundle branch block	No	No	No	No	No
ICER (£/QALY)	53,760	22,949	21,325	11,957	12,750

Note: Each subgroup is characterized as an average patient defined by a set of individual characteristics. The probability of each average profile is tantamount to the average predicted probability of the corresponding subgroup. (table reproduced with permission from (Henriksson et al., 2008)).

4.2.2. Limitations of RITA-3 subgroup analysis

Heterogeneity in baseline risk and treatment effect were analysed and reported in the original study (Henriksson et al., 2008). However, the approach adopted might not necessarily be the most adequate for the purposes of – for example - clinical practice. One of the limitations in the RITA-3 cost-effectiveness analysis is that subgroups were defined based on a baseline risk score estimated from the data. Although this

approach is internally valid, it fails in terms of external validity. For example, doctors are limited in defining a high risk patient on the basis of the risk score used in this study. Instead, they are more likely to estimate the baseline risk based on a well established risk score system, for example, the GRACE risk score as recommended by NICE guidelines (NICE Guideline 94, 2010).

In addition, the analytical approach used a representative profile for each subgroup. For example, for subgroup 1 (lowest average risk) the profile was a female patient, 45 years old with severe angina. On the other hand, the highest risk subgroup was defined by a 66 years old male patient with diabetes, previous MI, no severe angina but presenting ST depression at baseline. As explained above, these profiles were defined according to the baseline risk score, such that the predicted risk score of each specific profile equalled the mean risk score of that subgroup. In some cases, patients in the real world can be very similar to the profiles presented in the report and can be categorised in one particular subgroup; however, in many other cases patients cannot be categorised. Thus, according to the elements presented in Chapter 3, this analysis failed to be explicit in its consideration of how this subgroup definition can be operationalized in practice.

Furthermore, the baseline risk score was estimated for the primary outcome, i.e. the composite endpoint of cardiovascular death or myocardial infarction during the index hospitalization. It is clear that, for decisions based on cost-effectiveness, the relevant outcome is the ratio between incremental costs and benefits (or NHB) estimated for a longer time horizon. Thus, an analysis based on discrete (baseline) covariates (or their combinations) could offer a better understanding of the baseline risk heterogeneity and its effect on the relevant outcomes for decision-making, for example, net QALYs.

4.3. Methods

A subgroup cost-effectiveness analysis was undertaken based on the individual patient data collected in the RITA-3 trial, including baseline risk characteristics, treatment effect, HRQoL and costs. Details of these data have been reported previously (Henriksson et al., 2008, Epstein et al., 2008, Kim et al., 2005). In this section, five main points that must be considered in cost-effectiveness subgroup analysis are described, namely: definition of subgroups; sources of heterogeneity; selection of subgroup specifications; cost-effectiveness analysis; and value of further research.

All results are presented as population values. The size of the population of patients with NSTEMI-ACS was estimated considering an annual incidence of 59,756 patients, a time horizon of 10 years and a discount rate of 3.5% per year. The incidence corresponds to the UK estimate available from the British national statistics (Office for National Statistics). The relevant population for this case study was estimated in 556,723 patients according to the formula (3.17) in Chapter 3. For simplicity, the annual incidence was assumed to be the same across different subgroups.

Following the arguments developed in Chapter 3, definitions of subgroups considered three main elements: first, specifications must offer feasibility for future implementation in practice; second, specifications must have biological plausibility; and third, no ethical or equity concerns should hinder a potential implementation of decisions based on those specifications. A fourth logical constraint is the lack of information in the data either to include other specifications or to examine further combinations of specifications. All covariates that have been reported as significant predictors of the primary outcomes were considered for the analysis. This is because the heterogeneity that can be captured and represented through the decision model is only due to this set of covariates. However, there is no guarantee that all the covariates will be equally significant in their ability to explain variability of the cost-effectiveness outcome, as they did for the clinical primary outcome. A regression analysis was conducted to explore the effect of each covariate on the probability that the new strategy was cost-effective for a particular patient. The analysis was performed in STATA 12.

4.3.1. Sources of Heterogeneity

As mentioned, the original cost-effectiveness model developed by Henriksson et al. (2008) was specified on the basis of eight equations, four to estimate the transition probabilities of the decision model and four to estimate costs and utilities. For the purpose of this application study, it was decided to keep the original decision model to undertake the subgroup analysis as described in Chapter 3, rather than exploring the effect of alternative structural assumptions in the model.

This analysis incorporates baseline risk heterogeneity through the prediction of individual parameters, including probabilities, costs and utilities. These individual estimates are conditional to the set of covariates considered in the regression equations. Thus, for example, only age and severity of angina were significant covariates for predicting the composite event (Equation 1). On the other hand, all covariates were significant predictors of the failure to the treatment in the follow-up estimation (Equation 2). Equations are presented in Appendix 2.

The impact of treatment effect heterogeneity was also explored in this study. The effects of the interaction terms between treatment and the covariates of interest were analysed in all equations. Heterogeneity in preferences was not considered because of the lack of information about community preferences in the general literature⁴³. Neither was heterogeneity in costs exclusively considered. All covariates used to predict the expected individual costs were also risk predictors at baseline. Therefore, heterogeneity in costs in this case is a result of baseline risk heterogeneity.

4.3.2. Selection of Subgroups

Total expected population NHB were estimated for each specification. Details of the implementation are described below. Results were plotted to represent the efficiency frontier for subgroup analysis, according to the framework presented in Chapter 3. For illustrative purposes, a simplified guideline is presented for each level of disaggregation. These guidelines are constructed on the basis of the specification that provides the highest NHB and showed as a clinical algorithm.

⁴³ The difficulty of exploring heterogeneity for community preferences has been discussed in Chapter 2.

4.3.3. Implementation of the trial based Cost-Effectiveness Analysis

The model estimated the individual NHB of an invasive and conservative intervention conditional to the characteristics of each patient recruited in the trial. In other words, instead of choosing a representative profile for each subgroup as in the original report (Henriksson et al., 2008), the profile of each patient was used to estimate expected costs and benefits of both strategies, providing an estimate of the counterfactual. The mean costs, QALYs and ICERs are calculated as the average across individual estimates. For subgroup analysis, the mean values are obtained as the average across patients that are members of one particular category (e.g. diabetics). This technique also provides an opportunity to estimate the Expected Value of Individualized Care (EVIC). According to the definition of Basu and Meltzer (2007) and the subsequent application by van Gestel *et al.* (2012), total EVIC can be calculated as the difference between the average of the maximum individual NHB and the maximum of the average NHB (see Appendix 1 for details). In addition, parameter-specific EVIC can also be estimated from this data using the method explained by van Gestel et al. (2012) (see Appendix 3).

4.3.4. Value of further research

As mentioned in Chapter 3, one of the main contributions of this analytical framework is to consider uncertainty as one of the dimensions of the value of heterogeneity. The characterization of parameter uncertainty requires consideration of the standard errors of the point estimates obtained from the regression models into the decision model (equations 1-4, costs and utility). This was undertaken using the Cholesky decomposition of the covariance matrix estimated from such equations. Next, a PSA was conducted in order to propagate the parameter uncertainty through the decision model. One thousand iterations were run for each patient, so that an individual profile of uncertainty was obtained. As a consequence, the PSA produced 1,810,000 iterations (four matrices of 1,000 by 1,810): two for expected costs and two for expected QALYs of the invasive and conservative strategy respectively. These matrices provided the whole data needed for the subgroup analysis planned in this chapter. The model was implemented in Excel 2007.

The uncertainty relating to the overall population was estimated by averaging each of the 1,000 iterations across the 1,810 individuals, producing a unique vector of 1,000 iterations. Because the purpose was only to estimate uncertainty at this stage (no heterogeneity), the seed that generates the set of random parameters was set so that each iteration of the PSA was estimated in every individual with a unique set of random parameters. The vectors of 1,000 possible realizations each (two for costs and two for benefits) provide the results to estimate the maximum average NHB and the average of the maximum NHB, and as a consequence, the total expected value of perfect information (EVPI).

For subgroups, the expected NHB and expected maximum NHB, were estimated using the same approach but separately for each specific subgroup. For example, in order to analyse diabetes, each matrix⁴⁴ was divided into two, one including patients with diabetes (244 individuals) and another with patients without diabetes (1,566 subjects), i.e. one matrix of 1,000 by 244 and other of 1,000 by 1,566. Later, NHBs for both treatments were calculated for diabetic and non-diabetic patients combining the information of costs and QALYs. Finally, the maximum average NHB and the average across the maximum NHB were estimated for both sub-populations. Subgroup analysis for the rest of the specifications followed the same principles described for diabetes. These estimates provided the basis to calculate the static and dynamic value of heterogeneity. For simplicity, all the results shown in this application study were calculated for a threshold (λ) value of £20,000/QALY.

The transaction costs between specifications on the cost-effectiveness frontier will be discussed considering the type of guidelines resulting from the subgroup cost-effectiveness analysis.

⁴⁴ This refers to each of four matrices: (i) matrix for costs of the invasive strategy, (ii) matrix for costs of the conservative strategy, (iii) matrix of QALYs for invasive strategy and (iv) matrix of QALYs for conservative strategy.

4.4. Results

4.4.1. Definition of subgroups

All covariates that were significant in predicting the primary clinical outcome were considered. Average marginal effects were estimated from a *logit* model to explore the effect of each covariate on the probability that the new strategy is cost-effective for a particular individual at a cost-effectiveness threshold of £20,000 per QALY. The dependent variable is binary taking the value of 1 if the individual expected NHB of the invasive strategy is the maximum and zero if the individual expected NHB of the conservative strategy is the maximum. The analysis was conducted using the command *margeff* in STATA 12 (Bartus, 2005). Table 4.2 shows the average marginal effects from the multivariable analysis.

Table 4.2. Average marginal effects of nine covariates on the probability that the invasive strategy is the most cost-effective at a threshold of \$20,000 per QALY. Results based on a multivariable logit model.

	dF/dx	Std Error	z (P > z)
Diabetes	0.448	0.021	20.67 (p<0.001)
Previous MI	0.270	0.016	16.29 (p<0.001)
Smoker	0.359	0.015	23.89 (p<0.001)
ST depression	0.177	0.014	12.34 (p<0.001)
Left Bundle Branch Block	0.622	0.025	24.47 (p<0.001)
Severe Angina	-0.082	0.013	-6.31 (p<0.001)
Sex	0.192	0.012	14.90 (p<0.001)
Age	0.009	0.0006	15.15 (p<0.001)
Pulse	0.04	0.002	19.29 (p<0.001)

The *logit* model showed that all covariates are significantly associated with a greater probability of the invasive strategy being cost-effective, with diabetes the variable with the highest magnitude of effect. The only exception is severe angina, which is significant but negatively associated⁴⁵. Further, in the univariate analysis angina was the only covariate that was not significant. This might suggest a limited capacity of angina to explain the heterogeneity of decisions. Regardless, these results support the

⁴⁵ This finding seems inconsistent with the clinical report that showed that severe angina is associated with higher risk of presenting the primary outcome, which could be interpreted as more capacity to benefit within these patients. This can be explained by the fact that patients in this category incur higher costs under the invasive strategy (see Table A2.3 in Appendix 2) and, as a consequence, is less cost-effective.

use of all these covariates in the subsequent analysis because they might provide evidence for different decisions between subgroups.

Three of these covariates were excluded from the analysis for reasons other than their ability to explain heterogeneity. Sex and age have been excluded, because decisions that differentiate reimbursement based on any of those specifications are subject to ethical criticism. In the case of pulse, which corresponds to a numerical variable, there is no consensus about how to categorise it and it would be very difficult to implement alternative decisions based on an arbitrary definition. This is also applicable to age. However, their suitability as baseline risk predictors would indicate that they should be used as part of the estimation of a well documented baseline risk score system, for example, TIMI or GRACE (Antman et al., 2000, Muthappan et al., 2012). Thus, subgroups based on alternative cut-offs in the score can be explored. Unfortunately, the data available for this study did not provide enough information to calculate any well established baseline risk-score. For this reason, an additional subgroup analysis based on the baseline risk score used by the original cost-effectiveness study (Henriksson et al., 2008) was used to illustrate this point. Next, a brief justification for the selected specifications is provided.

4.4.1.1. Diabetes

Diabetes has been recognized as a major baseline risk factor affecting CHD. In particular for patients with NSTEMI-ACS, diabetes might explain differences in cost-effectiveness outcomes because of higher rates of mortality or myocardial infarction (Luscher et al., 2003). The biological plausibility is founded on the fact that patients with diabetes have, on average, higher levels of inflammation, which is etiologically related with more unstable atheromatous plaques (Creager et al., 2003). Thus, diabetics have increased rates of acute CHD, a higher number of complications and greater mortality rates as well as higher likelihood of restenosis and future coronary events (Lincoff, 2003, Stone et al., 1989). Therefore, the co-morbidity “diabetes” is acknowledged to be a relevant condition in cardiovascular pathophysiology and it seems entirely appropriate to explore whether the cost-effectiveness of an invasive strategy varies between patients with and without this condition. No ethical concerns seem relevant and the specification can be easily operationalized in practice.

4.4.1.2. ST depression

ST depression is an electrocardiographic marker of myocardial ischemia that indicates partial obstruction of the coronary network. In addition, ST depression is associated with increased future cardiac events and mortality (Nyman et al., 1993, Hyde et al., 1999, Krone et al., 1993, Savonitto et al., 1999). Moreover, it has been reported that patients with ST depression have more capacity to benefit from an early revascularization than those without this marker (Diderholm et al., 2002). There are no limitations of implementation since electrocardiography is considered a basic element of the coronary examination at the first evaluation in emergency. No ethical concerns should be raised, because it is a condition that might provide information about which patients can perceive more benefit from the invasive strategy.

4.4.1.3. Previous MI

The antecedent of a previous MI guarantees the fact that the patient has coronary heart disease and is associated with a higher rate of complications and mortality due, for example, to ventricular arrhythmias or left ventricular dysfunction (Bigger et al., 1984, Schulze et al., 1975). It is highly likely that by providing early revascularization patients will benefit more in clinical terms but also extra costs might be avoided because the procedure may also be needed after the early stage of the episode. This rationale leads to us to consider that, on average, an invasive strategy can be cost-effective only in patients with previous MI. This specification can be easily operationalized in practice and no ethical concern should limit its potential implementation.

4.4.1.4. Left Bundle Branch Block

Left bundle branch block (LBBB) is an electrocardiographic anomaly that can occult classical electrocardiographic manifestations of acute myocardial infarction (AMI). In addition, the diagnostic of the new LBBB in the course of an acute coronary syndrome is highly suggestive of myocardial infarction (Sgarbossa, 1996). An invasive strategy is considered effective and cost-effective in patients with AMI. Therefore, if an important proportion of patients with LBBB corresponds to patients with AMI, then is highly likely that NST-ACS patients with LBBB will have benefits greater than the average. Similar to ST depression, another specification

based on electrocardiography, this specification is easy to implement in practice and no ethical concerns should be raised.

4.4.1.5. Smoking

Tobacco is a well acknowledged independent risk factor of CHD (Lloyd-Jones et al., 2006). Its effect on the coronary bed includes vasomotor dysfunction, inflammation and modification of the lipid profile (Ambrose and Barua, 2004). In the context of NST-ACS, clinical guidelines recommend smoking be considered in the risk stratification (Anderson et al., 2007). More recently it has been reported that the treatment effect of an early invasive strategy in patients with NST-ACS is greater in smokers than in non-smokers (Aune et al., 2010). Therefore, clinical and biological evidence suggest that patients who smoke might have more capacity to benefit from an invasive strategy. From an ethical point of view, a valid concern could be to generate alternative recommendations in favour of smokers. Because this is an unresolved issue, from the analyst's point of view, it is considered appropriate to examine subgroups based on this specification.

4.4.1.6. Severe Angina

Angina is a consequence of the lack of blood flow to some areas of the heart and its severity is associated with the degree of coronary injury. In the clinical RITA-3 report angina proved to be a good predictor of the primary outcome (Fox et al., 2005). Therefore, if patients with severe angina have more capacity to benefit, an early intervention might also prevent future complications, avoiding extra costs. Thus, there is clinical and biological support to explore this covariate. Neither ethical concerns nor operationalization problems are identified for this specification. Angina was considered severe if patients presented with angina grade 3 or 4⁴⁶.

4.4.2. Cost-effectiveness results

4.4.2.1. Average

The invasive strategy was considered not cost-effective at $\lambda=20,000/\text{QALY}$ (ICER=21,960). The expected NHBs achieved by the most cost-effective strategy (conservative strategy) are 4,397,388 net-QALYs. If further research is undertaken

⁴⁶ Angina degree 3 or 4 refers to the presence of typical cardiogenic thoracic pain associated to clinical signs of pulmonary oedema or cardiogenic shock respectively (Anderson et al., 2007).

to resolve the current uncertainty, the expected NHBs are 4,408,143 net-QALYs. The EVPI, i.e. the potential gains of resolving the uncertainty are 10,755 net-QALYs.

4.4.2.2. *Expected Value of Individualized Care (EVIC)*

The total population EVIC was estimated at 14,349 net-QALYs. This corresponds to the difference between the expected maximum individual NHBs (4,411,737 net-QALYs) and the maximum expected NHBs (4,397,388 net-QALYs). This can be interpreted as the expected gains that can be achieved if decisions were based on individualized cost-effectiveness estimates. This between-patient heterogeneity, from which EVIC is estimated, is conditional on the information that has been considered in the decision model, which includes a set of observable covariates that were good predictors of relevant parameters of the model. The individualized analysis also provided evidence that the new strategy should be implemented in 591 patients of the sample (32.65%). The subgroup analysis in the next section aims to identify these patients through observable characteristics.

4.4.2.3. *Subgroup analysis: The two subgroup case*

The expected NHBs under current and perfect information were estimated for six binary specifications. Results are presented in Table 4.3. Severe angina was the only covariate that did not differentiate decisions between subgroups. As a consequence, the maximum expected NHBs are the same as the average (4,397,388 net-QALYs). Diabetes showed the highest expected NHBs under current and perfect information among all covariates, 4,403,199 and 4,412,177 net-QALYs respectively. This provides the greatest static value for all binary covariates, 5,811 net-QALYs (4,403,199 minus 4,397,388). A guideline based on this finding is shown in Figure 4.2. Furthermore, the potential health gains that can be obtained by resolving uncertainty (EVPI) conditional for diabetic and non-diabetic patients is lower than average (8,978 versus 10,755 net-QALYs). However, the absolute value of health under perfect information is higher than average. This provides evidence that the value of resolving uncertainty conditional to heterogeneity (dynamic value) has two sources: first, the value of an infinite sample in estimating conditional parameters, which is revealed by a positive EVPI; and second, the value of estimating

conditional parameters with an infinite sample, which is revealed by the fact that the EVPI is lower than the average.

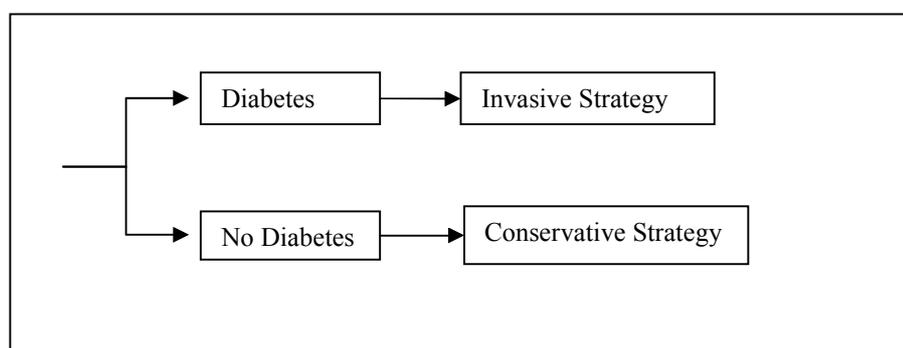
Table 4.3. Cost-effectiveness subgroup analysis of RITA-3 trial: The two subgroup case.

Specification (proportion of patients with the condition)	NHB (current information)	NHB (perfect information)	EVPI	ICER (with the condition)	ICER (without the condition)
Average	4,397,388	4,408,143	10,755	21,943	
Diabetes (0.13)	4,403,199	4,412,177	8,978	15,808	23,528
Previous Myocardial Infarction (0.27)	4,400,926	4,411,501	10,574	18,558	23,877
ST depression (0.36)	4,398,950	4,411,199	12,248	19,452	23,765
Left Bundle Branch Block (0.035)	4,399,530	4,410,850	11,320	14,628	22,402
Smoker (0.32)	4,399,896	4,411,278	11,382	19,016	23,721
Severe Angina (0.35)	4,397,388	4,410,526	13,138	22,704	21,495

Note: NHB and EVPI are presented in QALYs (net of costs) for a threshold value of £20,000/QALY gained. The ICER is presented in sterling pounds per QALY gained.

For patients without previous MI, expected NHBs were 4,400,926 net-QALYs, providing a positive static VoH of 3,538 net-QALYs. However, the increase in NHBs provided by this specification is lower than was observed in the case of diabetes. In terms of the potential health gains if perfect information were revealed, the expected maximum NHBs are 4,411,501 net-QALYs leading to an EVPI of 10,574 net-QALYs. Specifications such as depression of the segment ST, LBBB and being a smoker also provided a positive static value (1,562; 2,142 and 2,508 net-QALYs respectively). However, these values are lower than can be obtained by making decisions based on the two previous specifications.

Figure 4.2. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The two subgroup case



4.4.2.4. Subgroup analysis: the four subgroup case

Pairs of binary covariates were analysed as single specifications of four subgroups. The results are presented in Table 4.4. They show that under current information the combination of diabetes and LBBB achieves the maximum expected NHBs (4,404,566 net-QALYs). This corresponds to a positive static value of 7,178 net-QALYs compared to the average case and 1,367 net-QALYs compared to the specification diabetes for two subgroups. The results also showed that diabetes was part of the five specifications with the highest expected NHBs. This suggests an important informative role of diabetes for understanding heterogeneity in comparison with other covariates⁴⁷. Previous MI, LBBB and smoking seem to be the next most informative covariates. They provided the next highest expected NHB either in combination with diabetes or between them. As expected from the analysis for the two subgroups, severe angina was not informative in combination with any of the previous covariates. However, it provided evidence for different decisions in combination with depression of ST segment leading to a modest gain in expected NHBs. Thus, although at this level of disaggregation, severe angina does not offer major gains, it keeps its potential when further disaggregation is examined.

In terms of dynamic value, results are consistent with the analysis for the two subgroups where diabetes provided the highest value. In this case, the specifications with the highest NHBs under perfect information were in combination with diabetes. Likewise, EVPI in all these specifications was lower than average. A clinical guideline constructed at this level of disaggregation is presented in Figure 4.3. In addition to this analysis, decision-makers might be interested in selecting subgroups where further research is worth conducting. For example, let us consider the situation where there are four subgroups defined by the combination of diabetes and LBBB. In this case, the subgroup of patients that are non-diabetic and do not have LBBB produces the maximum health gains when uncertainty is resolved (EVPI 6,767 net-QALYs). A list of EVPI per subgroup is provided in Appendix 3.

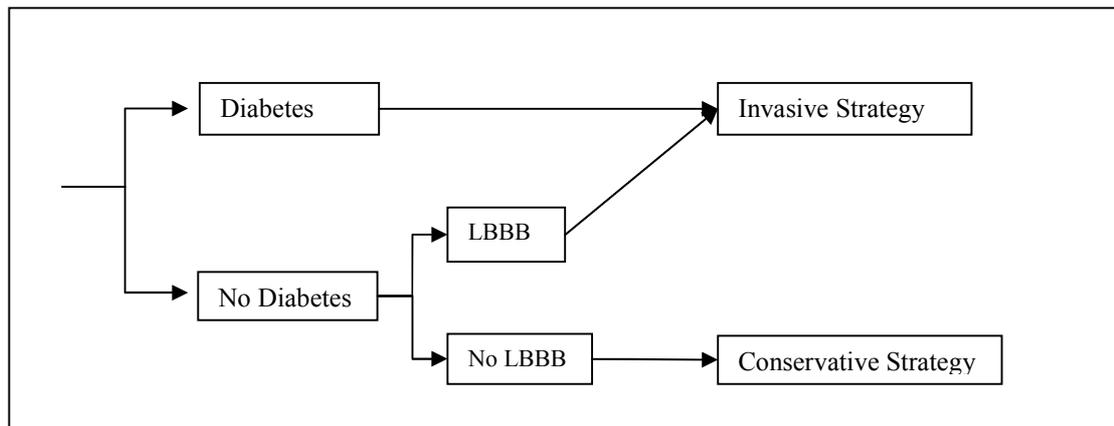
⁴⁷ The informative capacity of one covariate or specification refers to its ability for disaggregating the population such that more individuals are classified correctly, i.e. as they would have been according to their individual expected incremental cost-effectiveness ratio.

Table 4.4. Cost-effectiveness subgroup analysis of RITA-3 trial: The four subgroup case.

Specification	NHB (current information)	NHB (perfect information)	EVPI	ICER			
				A & B	A & No B	No A & B	No A & No B
Average	4,397,388	4,408,143	10,755	21,943			
(A) Diabetes and (B) LBBB	4,404,566	4,412,841	8,275	11,526	16,111	15,553	24,013
(A) Diabetes and (B) Previous MI	4,403,714	4,413,271	9,556	14,899	16,484	19,720	25,554
(A) Diabetes and (B) Smoking	4,403,455	4,413,444	9,988	14,223	16,429	19,881	25,990
(A) Diabetes and (B) Severe Angina	4,403,199	4,412,590	9,391	17,212	14,813	24,344	23,071
(A) Diabetes and (B) ST depression	4,403,199	4,412,897	9,698	15,013	16,434	20,661	25,586
(A) Previous MI and (B) LBBB	4,402,052	4,412,169	10,117	13,969	18,897	15,111	24,373
(A) LBBB and (B) Smoking	4,401,288	4,412,090	10,802	12,308	15,380	19,283	24,366
(A) ST depression and (B) Smoking	4,400,939	4,412,249	11,310	16,642	21,107	20,697	25,672
(A) Previous MI and (B) ST depression	4,400,929	4,412,673	11,746	19,439	16,877	20,533	26,812
(A) Previous MI and (B) Smoking	4,400,926	4,412,664	11,738	15,994	19,833	20,435	26,211
(A) Previous MI and (B) Severe Angina	4,400,926	4,411,979	11,052	19,824	17,480	25,075	23,314
(A) ST depression and (B) LBBB	4,400,214	4,411,976	11,762	13,249	19,748	15,297	24,392
(A) Smoking and (B) Severe Angina	4,399,896	4,411,627	11,731	19,376	18,804	24,732	23,127
(A) LBBB and (B) Severe Angina	4,399,530	4,411,306	11,776	16,421	13,137	23,195	21,946
(A) ST depression and (B) Severe Angina	4,399,396	4,411,566	12,170	20,408	18,860	24,481	23,360

Note: Abbreviations used in this table are: MI (myocardial infarction), ST depression (depression of the ST segment), LBBB (left bundle branch block), ICER (Incremental cost-effectiveness threshold), NHB (Net Health Benefits) and EVPI (Expected Value of Perfect Information)

Figure 4.3. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The four subgroup case



4.4.2.5. Additional subgroup analysis

The analysis conducted so far provides a partial understanding of how to characterize heterogeneity of the population using certain informative covariates. Further analysis can be guided by this knowledge, particularly exploring combinations of covariates that have shown informative capacity. Ultimately, the analysis can be extended to the study of the combinations of the whole set of observable covariates in one single specification, in order to produce a single guideline.

Table 4.5 presents the results for six additional specifications. Four of them were defined on the basis of the alternative combination between diabetes, previous MI, smoking, LBBB and ST depression to produce eight subgroups. From these, the specification that combined diabetes, previous MI and smoking reached the maximum expected NHBs (4,405,519 net QALYs). The analysis for the other three specifications showed that previous MI, smoking or ST depression do not add more information to the combination defined by diabetes and LBBB (zero static value). A guideline based on this specification is showed in Figure 4.4. Details of these results are presented in Appendix 3.

Table 4.5. Subgroup analysis for additional combinations

Specification	NHB (current information)	NHB (perfect information)	EVPI
(1) Diabetes & Previous MI & Smoking (8 subgroups)	4,405,519	4,414,587	9,068
(2) Diabetes & Previous MI & LBBB (8 subgroups)	4,404,566	4,413,892	9,326
(3) Diabetes & Smoking & LBBB (8 subgroups)	4,404,566	4,414,198	9,632
(4) Diabetes & LBBB & ST depression (8 subgroups)	4,404,566	4,413,820	9,254
(5) Diabetes & Previous MI & Smoking & LBBB (16 subgroups)	4,406,788	4,415,294	8,505
(6) All covariates (49 subgroups)	4,408,359	4,416,806	8,447

Note 1: NHB and EVPI expressed as QALYs net of costs for the population.

Note 2: For the eight subgroups case the specification that combines diabetes, previous MI and smoking provides the highest NHB. Specifications (2), (3) and (4) have the same expected NHB as the single combination between diabetes and LBBB (see Table 4.4) (zero static value).

Note 3: Specification (6) is defined as multiple combinations between six covariates, which produce 64 theoretical subgroups. However, the trial did not provide data for the whole subgroups. Therefore, these results are constrained to 49 subgroups.

Figure 4.4. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The eight subgroup case

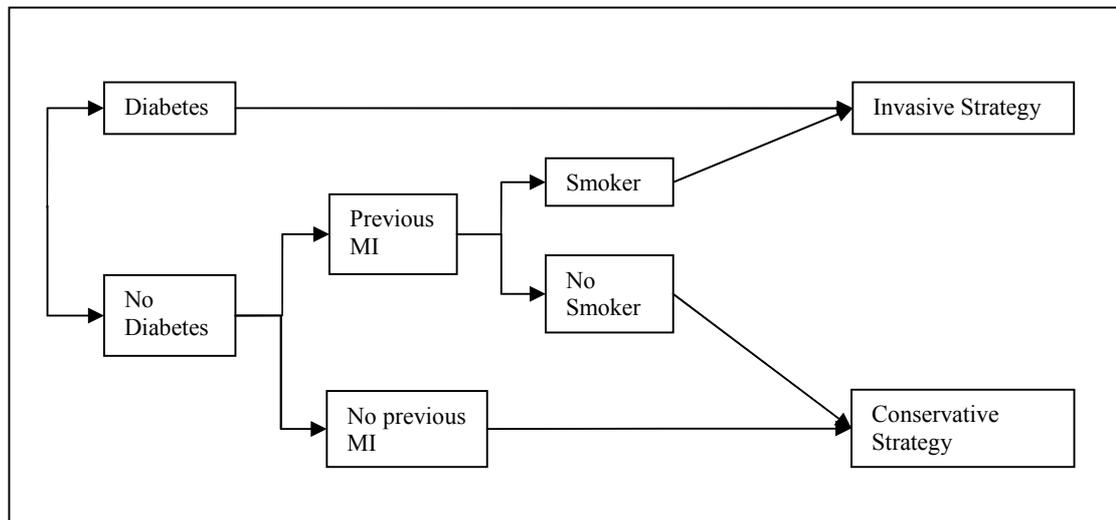


Figure 4.5. Guideline for the management of patients with Non ST elevation Acute Coronary Syndrome: The sixteen subgroup case

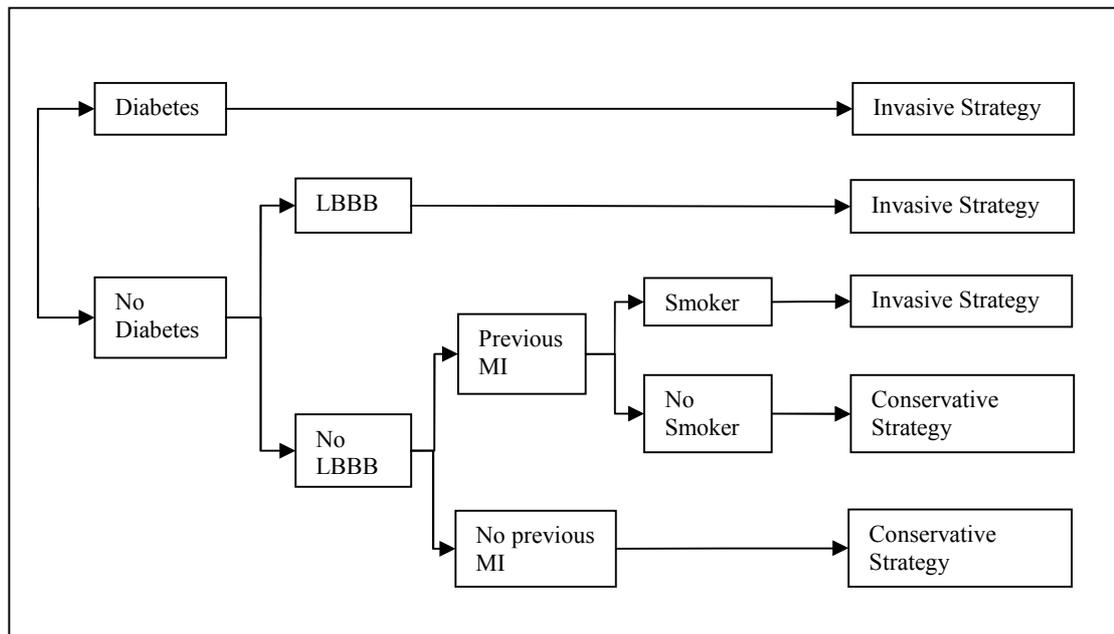
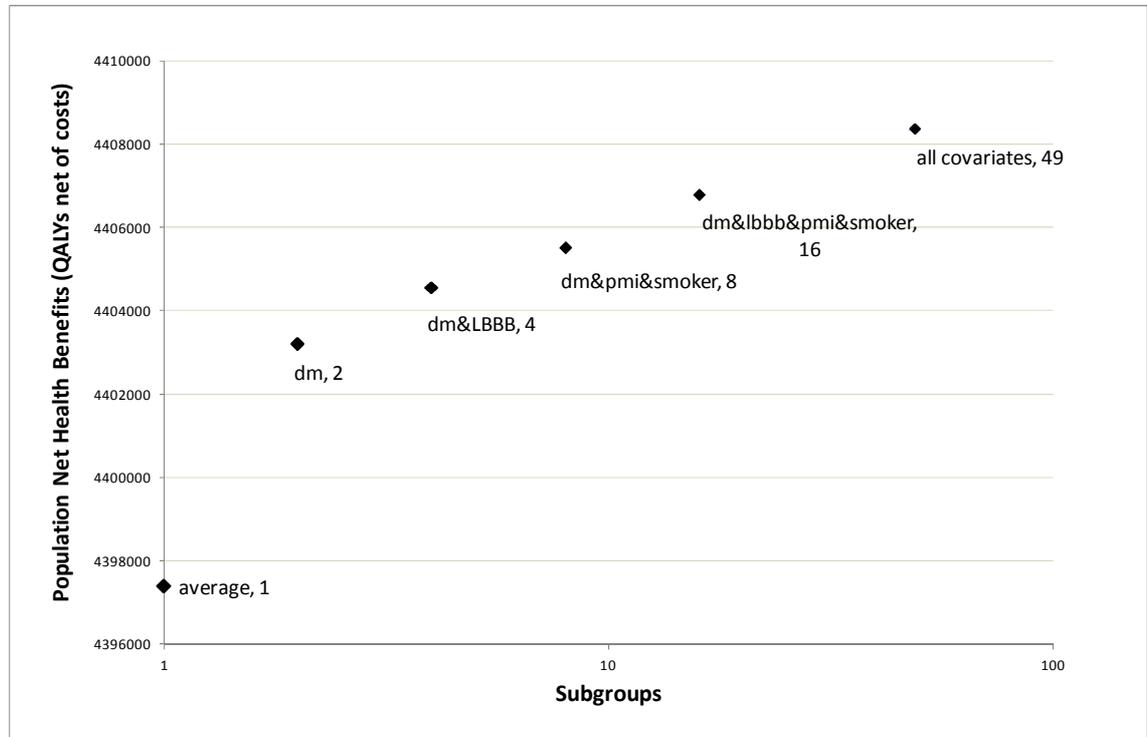


Table 4.5 also shows the estimations of two additional specifications. One of them is based on the combination of diabetes, previous MI, LBBB and smoking leading to sixteen subgroups. This specification provided higher expected NHBs (4,406,788 net-QALYs) than combinations examined in previous analyses. The corresponding guideline is presented in Figure 4.5. The final specification presented in Table 4.5 corresponds to the combination of all (six) observable covariates. Although this specification produces 64 potential subgroups, the trial only provided data for 49 subsets of patients (details in Appendix 3). This analysis showed the highest expected NHBs, which correspond to a static value of 10,971 net-QALYs. In other words, it takes account of 76.5% of the total EVIC (10,971/14,349). The efficiency frontier for subgroup analysis presented in Figure 4.6 shows the maximum expected NHB for each level of disaggregation. As suggested in Chapter 3, the curve shows diminishing marginal returns of net health. In other words, higher net health gains are achieved at lower levels of disaggregation. The guideline that can be constructed from the evidence under current information with the specification for 49 subgroups is presented in Figure 4.7.

It seems feasible that all guidelines presented for this subgroup analysis could be implemented in a clinical setting. They are constructed on the basis of well known

and well accepted clinical characteristics. Therefore, it is argued that the increase in NHBs is higher than the transaction costs in each case.

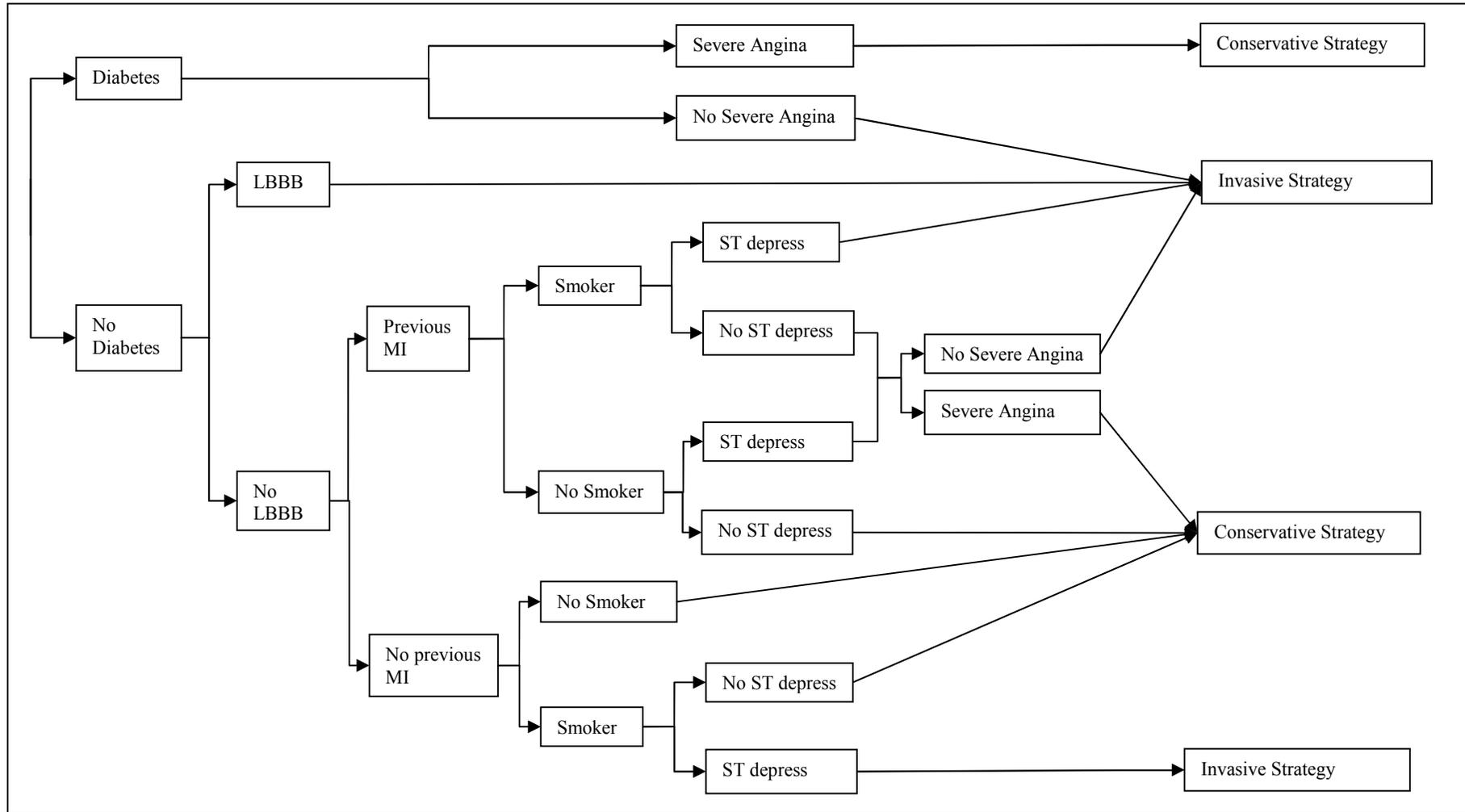
Figure 4.6. Efficiency Frontier for Subgroup Analysis. Specifications that maximise the expected Net Health Benefits for each level of disaggregation.



Note: Each point is labelled with the name of the specifications and the number of subgroups. The abbreviations are: dm (diabetes mellitus), dm&LBBB (diabetes mellitus and left bundle branch block), dm&pmi&smoker (diabetes mellitus, previous myocardial infarction and smoker), all covariates (diabetes mellitus, previous myocardial infarction, smoker, left bundle branch block, ST depression and angina). The x axis is presented in a logarithmic scale to facilitate visualization.

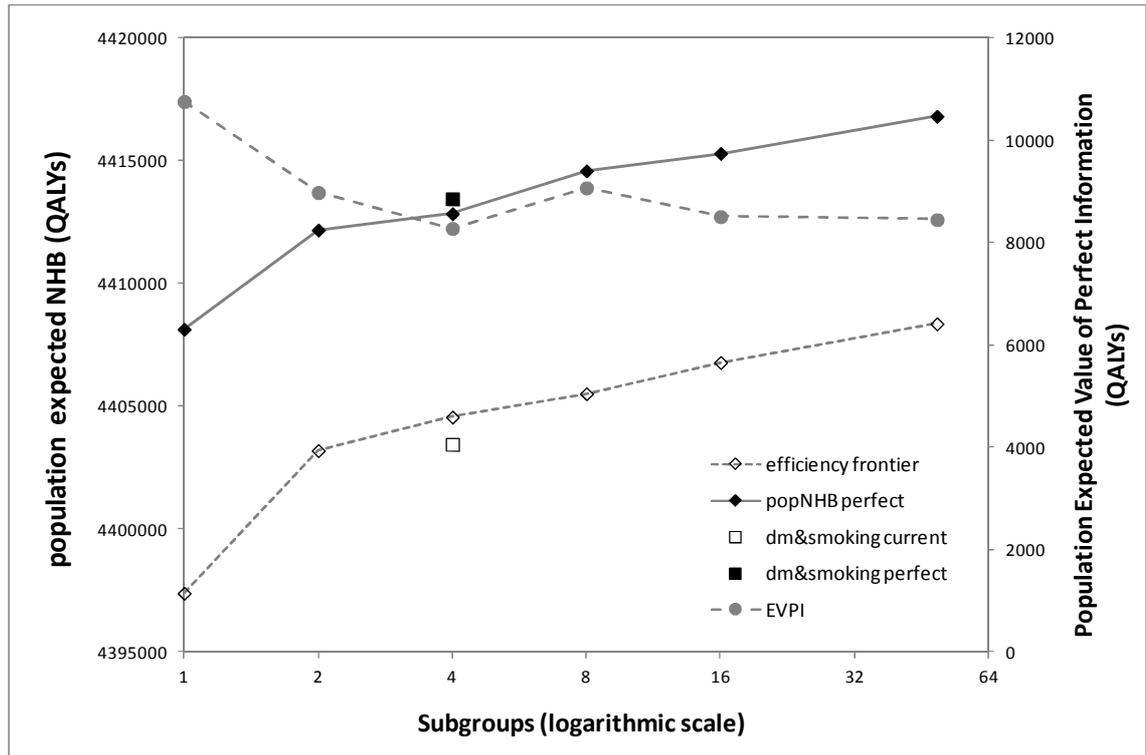
In terms of dynamic value, all specifications showed positive values with respect to the baseline and the previous level of disaggregation. The highest estimates under perfect information were always given by the same specification on the efficiency frontier, except for one case. The specification defined by diabetes and smoking (4 subgroups) is not on the frontier but it provides more expected NHBs under perfect information than the specification on the frontier defined by diabetes and LBBB. This was referred to in Chapter 3 as a non-proportional dynamic value and suggests that further research conditional to the former specification has more value than the latter. The estimates under current and perfect information as well as the EVPI are presented in Figure 4.8.

Figure 4.7. Guideline for the specification that combines six covariates



These results show that the specification that combined all covariates provided the highest expected NHBs under current and perfect information observed in this analysis. The distance between empty and filled points in the graph corresponds to the EVPI, which is depicted in filled circles. These results also show a trend in which the EVPI decreases as more subgroups are considered in the analysis.

Figure 4.8. Expected Net Health Benefits under current and perfect information for different level of disaggregation.



Note: Expected Net Health Benefits (NHB) expressed in QALYs. Empty diamonds represent the expected NHB achieved with the specification that maximises net health in one particular level of disaggregation. The dotted line across those points illustrates the efficiency frontier for subgroup analysis (as shown in Figure 4.6). Filled diamonds represent the expected NHB that might be achieved if an infinite sample is collected for the corresponding specification on the efficiency frontier. The empty and filled square correspond to the expected NHB under current and perfect information for the specification defined by diabetes and smoking (dm&smoking).

As explained in section 4.4.1 a baseline risk score might be a more informative specification than those examined hitherto in this chapter. Table 4.6 shows the results of the subgroup analysis based on the specification used in the original RITA-3 cost-effectiveness study. The invasive strategy is considered cost-effective only in those patients whose risk score was above the percentile 75. The expected NHBs under current information are 4,407,074 net-QALYs, which is equivalent to 9,686 net-QALYs. The differences between the ICERs presented in Table 4.6 and those reported

by Henriksson et al. (2008) (Table 4.1) can be explained by the different methodology. In this study the estimates correspond to the average across individuals, whereas Henriksson et al. (2008) reported the estimate of an average individual (see section 4.2.1). The results under perfect information illustrate the positive dynamic value, which is even greater than the specification of sixteen subgroups but lower than the specification for all covariates.

Table 4.6. Subgroup analysis for the baseline risk score

NHB (current information)	NHB (perfect information)	EVPI	ICER				
			Percentile (0-25)	Percentile (25-50)	Percentile (50-75)	Percentile (75-87.5)	Percentile (87.5-100)
4,407,074	4,416,926	9,852	34,055	24,925	20,129	17,974	15,813

Finally, as expected and explained in Chapter 3, EVIC for parameters were equivalent to the estimations of the static value of heterogeneity. For example, the estimation of the specific EVIC for diabetes was 5,803 net-QALYs, which is approximately the same as the static value reported in this chapter (5,811 net-QALYs). It is argued that small differences between estimates of static value and parameter specific EVIC are explained by approximations alongside the calculation process. Appendix 4 provides an explanation of the parameter specific EVIC and a demonstration of its equivalence with the static value concept.

4.4.3. Analysis of other sources of heterogeneity

In addition to the results presented above, the relevance of considering treatment effect heterogeneity was also studied in this chapter. Several alternative models were studied and compared with original equations. Table 4.7 shows the Akaike Information Criterion (AIC) for the equations where interactions between covariates and treatment can be included in the model. This excludes Equation 4 and the equation for utilities at randomisation because they did not include the treatment term. The values presented for the model with interaction corresponds to the smallest AIC observed among several models studied with different interaction terms. In all cases the equations without interaction terms were considered more appropriate than alternative specifications

including interactions between covariates and treatment. Given these characteristics of goodness of fit and parsimony the analysis was restricted to the original equations.

Table 4.7. Akaike Information Criterion (AIC) of equation models with and without interaction with treatment effect.

	Model without interaction	Model with interaction
Equation 1	460	463
Equation 2 (3)	1739	1742
Costs during the index hospitalization	36,242	36,252
Cost first year after the index hospitalization	35,837	35,838
Change in utility at follow-up	546	547

4.5. Discussion

Chapter 3 argued in favour of the role of subgroup analysis in exploring heterogeneity in the context of cost-effectiveness studies. This chapter shows an application of the methods presented in that previous chapter using the economic evaluation of the RITA-3 trial as a motivational example. In this study a subgroup analysis based primarily on six binary specifications (diabetes, previous MI, smoking, ST depression, LBBB and severe angina) was performed. They were examined as single binary specifications and also in combination. This led to a specification of 64 possible subgroups when all covariates were simultaneously combined. The results showed that as long as the analysis progressed in disaggregation, higher expected NHBs (under current information) could be achieved when informative specifications were considered. Further, the expected health gains between two points (static value) were higher at low numbers of subgroups and decreased as the analysis increased in disaggregation. This can be expressed as diminishing marginal returns of net health. In terms of dynamic value, these results are consistent with the hypothesis that the value of collecting further research increases when it is conditional to heterogeneity. In addition, results are compatible with the hypothesis that when the analysis has been performed using informative specifications the general trend is towards a decrease in the EVPI as more subgroups are considered.

It is important to emphasize that these results depend on the decisions across different subgroups, which will vary depending on the threshold used. For simplicity, in this study all interpretations have been made based on a threshold of £20,000 per QALY. In the event that decision-makers require results for different threshold values, this analysis can be extended to examine the static and dynamic value of each relevant specification for a set of threshold values.

Another important element of this framework is the role of transaction costs in the selection between levels of disaggregation. In this chapter, transaction costs have not been estimated, which can be considered a limitation of this study. However, it is argued that the guidelines presented here do not impose additional complexity on other schemes of risk stratification presented in the literature for these patients, and as a consequence, it is unlikely to determine difficulties in their implementation. Thus, even the guideline that combines 49 subgroups can be reasonably implemented in practice because the transaction costs are likely to be very low. In contrast, the restriction of transaction costs could be observed for the specification based on the baseline risk score. It can be argued that the implementation of such guidelines requires the validation of that specific baseline risk score by the clinical community. For this reason it is very likely that the transaction costs are higher than the additional benefits.

As pointed out previously in Chapter 3, the dynamic value has two sources; first, the value of an infinite sample to estimate conditional parameters and second, the value of estimating conditional parameters⁴⁸ with an infinite sample. While the first relates to the classical concept of Value of Information, the second refers to the value of an estimation of parameters for specific subgroups under perfect information. As explained before, if the specification is informative, the effect of heterogeneity is to be expected, not only on the estimation of perfect information but also under current information. As a consequence, the EVPI should decrease. This chapter provides empirical evidence of this argument. It shows that as the level of disaggregation increases the expected NHBs under perfect information increase and, in general, the EVPI decrease. This is true for the most informative specifications, i.e. those that lie on the efficiency frontier.

⁴⁸ Footnote 42 in Chapter 3.

The chapter also includes an estimation of EVIC. Total EVIC (1,981 net QALYs) represents the potential gains if individual information about the set of covariates considered in the model is used to make individualized decisions. This can also be interpreted as the maximum achievable static value according to the information considered in this study. This chapter shows that the specification that combines all covariates accounts for more than 75% of this value, which can be considered a significant improvement of the efficiency achieved by the clinical guideline.

As explained in Chapter 3, EVIC is independent of EVPI since it depends only on the level of heterogeneity presented in the random sample, whereas EVPI depends on the sample size. Results of this study are consistent with this statement since it provides different expected values for EVIC (1,981) and EVPI (1,484). On the other hand, parameter-specific EVIC has been suggested as a good instrument to inform policies based on subgroups (van Gestel et al., 2012). The framework developed in this thesis is entirely consistent with such a statement because parameter-specific EVIC is equivalent to the static value of heterogeneity. However, this approach has an advantage in that it offers a more detailed understanding of the decisions that should be made in specific subgroups. This is considered paramount for the construction of clinical guidelines.

The framework presented in this thesis proposes to select subgroups based on an efficiency criterion, the maximisation of the maximum expected NHBs. A similar answer can be achieved by analysing the rate of misclassification in each subgroup. For example, let us assume that a guideline based on two subgroups is implemented. It is known from the analysis at individual level that the invasive strategy is expected to be cost-effective in 591 patients (32.65%) and not-cost-effective in 1,219 patients (67.35%). From these, a guideline based on previous MI misclassifies 550 patients (30.39%). In contrast, a guideline based on diabetes misclassifies 457 patients (25.25%). The conclusion is that a guideline based on diabetes is more accurate than a guideline based on previous MI. This is exactly the same result obtained applying the decision rule proposed by this thesis. The additional benefit of performing this analysis is that it provides information about the distribution of misclassifications under each specification. For example, in patients who should receive the invasive strategy, diabetes misclassifies 402 and previous MI misclassifies only 320. On the other hand, for patients who should not receive the invasive strategy, diabetes misclassifies 55,

whereas previous MI misclassifies 230. Thus, if the decision-maker is concerned about misclassifications in patients who should receive the invasive strategy, previous MI is a better specification than diabetes. This is an advantage of this type of analysis that cannot be performed, for example, with the estimation of EVIC for specific parameters.

One of the relevant contributions of this chapter is to provide a demonstration of an analytical technique able to generate estimates that are consistent with the theory. In the stage of implementation of the method, several approaches were considered; however, only one provided adequate estimates. One alternative was to estimate specific parameters of the original equations for specific subgroups. For example, in order to estimate cost-effectiveness in diabetic patients, equations might have been estimated using only the diabetic patients in the trial. A second alternative was to modify original equations, only including treatment effect and the covariate of interest. Third, the original model could be used to obtain results based on representative profiles of the subgroups of interest, as used in the original cost-effectiveness study (Henriksson et al., 2008). These three approaches presented some drawbacks: first, if original equations were modified in their initial structure, an additional source of *structural* uncertainty would have been introduced⁴⁹. Second, parameters estimated based on smaller sample sizes are less consistent and efficient than estimates from bigger samples, leading to biased cost-effectiveness results and less adequate representation of uncertainty. Third, the definition of a representative profile is not an easy task and is still an approximation of the reality that can be misleading. Finally, the estimation of the overall NHBs assumes a linear relationship between NHBs for specific subgroups and the average NHB. This assumption cannot be held under any of these alternative approaches. Hence, an approach based on the estimation of cost-effectiveness at individual level was finally preferred. This technique has been used in the past (Nease and Owens, 1994, Sculpher, 1998) and also in the cost-effectiveness study of RITA-3, but only to describe the heterogeneity across individuals. As mentioned, the method uses individual patient profiles, which are propagated through the decision model producing individual cost-effectiveness results. It is assumed that the information about the covariates is perfect. Thus, the overall NHBs (i.e. the average of the individual NHBs) are linearly correlated with estimates in specific subgroups. In addition, the characterization of uncertainty is

⁴⁹ Structural uncertainty has been defined and explained in Chapter 1.

also achieved properly under this approach. The use of the same set of random parameters for each of the 1,000 iterations produces a correct representation of the parameter uncertainty without including additional sources of variability such as individual patient variability.

This study has also explored the possibility of considering other sources of heterogeneity, such as treatment effect heterogeneity. In this chapter, variations of treatment effect were explored adding interaction terms between baseline risk covariates and treatment into the regression models. The fact that different model specifications might produce completely different results might be due to the effect of the interaction(s) terms or because of the new structure of the model, that might not necessarily provide consistent or efficient estimates. It is important to clarify that the methods applied in this study addressed the characterization of *parameter* uncertainty only. In contrast, structural assumptions such as alternative specifications of the model, corresponds to an additional source of uncertainty called structural uncertainty (Jackson et al., 2011, Bojke et al., 2009a). In this case, structural assumptions have been supported on the basis of statistical tools, for example, Akaike information criterion (AIC), parsimony and models with minimum collinearity. This might still be subject to criticism, since other statistical methods could have been explored, for example generalized linear models instead of ordinary least squares in the case of costs and utilities. Nevertheless, because there is not a clear guideline in this respect, the approach used here was considered valid and appropriate. This area might be further explored in the future, and the use of methods to deal with structural uncertainty, such as model averaging, is worth taking into account (Jackson et al., 2009).

In this study, five covariates of interest were used to predict costs (sex, age, severe angina, previous MI and ST depression). It is important to highlight the fact that heterogeneity in costs can be presented in two ways: first, when the specification is defined exclusively based on costs and without regard for any biological support; and second, when the specification has been demonstrated to be a risk factor for relevant health outcomes and it also explains differences in costs. The case presented in this study is the second situation. While considerations of heterogeneity based solely on costs can raise ethical questions, it can be considered appropriate to take into account specifications with both costs and clinical support. In fact, if patients with a particular

risk factor have lower relative risk than patients without the risk and costs are the same, there is risk of concluding that a new technology is cost-effective only in the subgroup with the condition, whereas this might not be the case if differences in costs are properly taken into account. For example, let us compare two technologies with average incremental costs and benefits being £20,000 and 1 QALY. Next, let us assume differences based purely on effects in two subgroups: 1.3 QALY in the subgroup with a risk factor (lower relative risk) and 0.8 QALYs in the subgroup without risk factors. This leads to the following estimates: $\text{£}20,000/1.3\text{QALYs}=15,384$ and $\text{£}20,000/0.8=25,000$ indicating that the new treatment is cost-effective in the subgroup with the risk factor only. But, if costs differences are taken into account (assuming that patients at higher risk are £10,000 more costly) the estimates are $\text{£}25,000/1.3\text{QALYs}=19,230$ and $15,000/0.8\text{QALYs}=18,750$. If the threshold is assumed to be £20,000 per QALY, the omission of heterogeneity in costs could deny a cost-effective strategy to patients with less capacity to benefit because of the assumption that they incur more costs.

One of the limitations of this study relates to the exchangeability of information between subgroups. The estimation of the EVPI for one subgroup assumes that no additional information will be provided by another subgroup when further research is undertaken. This is a limitation when decision-makers are focused on one particular subgroup. However, if the target is the original population, the overall EVPI_s presented here is the metric of interest to update the guideline in the future. In contrast, if the objective is to collect information only in one subgroup, and then to synthesize evidence in the future, the EVPI, as estimated in this chapter, underestimates the real value of further research. An alternative estimate for each subgroup should be estimated in order to account for the exchangeability of information between subgroups. This requires the development of methods to adjust the EVPI_s presented here, which provides an opportunity for further research. On the other hand, if the objective is to conduct further research to update a guideline in one specific subgroup only, the EVPI estimated for that specific subgroup (without adjustment), as presented here, is the correct estimate.

This study has been performed on the basis of individual patient level data. The advantage of this type of data is that it provides complete information about each individual in the sample. Consequently, not only static and dynamic VoH can be

estimated, but also EVIC and specific-parameter EVIC. Recently, van Gestel et al. (2012) have estimated EVIC through individual patient level simulation in a case study for cost-effectiveness in glaucoma. The data they produced by simulation can also be used to apply the methods presented here. This represents a good example of how to conduct the analysis when trial-based data is not available and researchers only have aggregated data. However, some additional elements need to be considered. One of the characteristics of such data is that individual profiles are drawn from prior information about the distribution of observable covariates in the population. A common limitation of that type of data is that the structure of the variance/covariance matrix of such covariates is not always available, which leads to less representative profiles. In contrast, the data used in this chapter consider these correlations because the units of observation (patients) can be considered a truly representative sample of the target population (Fox et al., 2002).

As mentioned in Chapter 3 this approach aimed to study and understand heterogeneity in terms of its additional value in current decisions, but also in terms of resolving the remaining parameter uncertainty. However, it does not provide the methods to prioritize the type of study that ought to be undertaken, or the size of the studies that represent a good use of the limited resources. In order to deal with this issue, a future research agenda should include the adaptation of the expected value of perfect information for parameters and the expected net benefit of sampling information to subgroup analysis. This has not been undertaken here, not only because of the complexity that an adaptation of such methods would demand but also because it is considered to be outside of the scope of the main purpose outlined in the first chapter, which is to study elements of value of heterogeneity in cost-effectiveness analysis.

4.6. Conclusions

The value of heterogeneity framework is a novel analytical approach whose practical implementation has been demonstrated in this chapter. An approach based on the cost-effectiveness estimated at individual level in the context of individual patient data produced an adequate method for obtaining robust and interpretable estimates. Either the static or the dynamic value of heterogeneity can be properly presented to decision-

makers for different subgroup specifications in order to provide additional elements to central and local health authorities for the decision-making process.

In addition, the argument that resolving uncertainty conditional to heterogeneity provides more net health benefits has been empirically demonstrated in this chapter, and the idea that understanding heterogeneity can decrease the additional expected net health benefits that can be achieved when uncertainty is resolved (i.e. EVPI) has also been shown here.

Finally, it must be emphasized that a systematic application of this analytical approach requires not only additional effort in terms of modelling but also making explicit those justifications that give support to the selection of subgroup specifications that are being considered.

5. CHAPTER 5

INDIVIDUAL DECISIONS AND SOCIAL VALUE:

Alternative decision-making approaches and the value of heterogeneity in the era of individualized care

5.1. Introduction

Previous chapters have examined some important elements of the value of considering heterogeneity for decision-making in a collectively funded health system. As introduced in Chapter 2, the overall value corresponds to the difference between the total expected NB with complete information⁵⁰ and the expected NB for the average population, which has been called EVIC⁵¹. The EVIC can be estimated from the available data (under incomplete information) and interpreted as the maximum static value that can be achieved when decisions at individual level are implemented based on individual cost-effectiveness. However, as explained in Chapter 3, decisions at individual level, i.e. decisions that account for the overall value revealed by EVIC, cannot be implemented due to high transaction costs. As a consequence, recommendations based on subgroups can be considered an appropriate policy not only because they account for some proportion of this static value but also because they offer the opportunity to increase value through further research to resolve decision uncertainty (dynamic value).

The arguments developed in previous chapters are consistent with a framework of a centralized process of decision-making. However, both the concepts of decentralizing decision-making and individualization of care have grown in popularity in the last few years (Mishra, 2011, Malandrino and Smith, 2011, Hewitt, 2011, Tursz et al., 2011, Trusheim et al., 2007). For example, in the US individualization has been broadly

⁵⁰ Complete information (formally presented in Chapter 2) corresponds to a theoretical (unachievable) state of knowledge where each subject can be distinguished from other based on a set of individual observable characteristics. It is unachievable because, as explained in Chapter 2, there are characteristics that are unobservable, unmeasurable or unknowable.

⁵¹ Specifically, this concept of EVIC corresponds to the EVIC with cost internalization.

focused on how to achieve personalized medicine⁵² (Hamburg and Collins, 2010) with significant emphasis being placed upon patient choice (US Public Law Sec 6301, 2010). In the UK, patient choice has been part of the policy agenda over the last decade with regard to access and utilization of healthcare facilities (McCabe et al., 2008). More recently, a proposal for reform of the NHS has considered the empowerment of local authorities and General Practitioners (GPs) so that they can make decisions that were previously made centrally. Under this new setting, central authority will devolve decisions at individual-patient level (i.e. doctors, patients or both) in order to give explicit consideration to preferences and treatment choices expressed by patients (Department of Health, 2010).

These changes in the policy agenda represent a strong motivation to study issues around the individualization of care. In this thesis, the term “decisions at individual level” has been used to indicate individual recommendations based on cost-effectiveness predicted specifically for one patient⁵³. This prediction is based on a set of all observable and measurable characteristics, which might include individual preferences as defined in Chapter 2. A second type of decisions at individual level is what in this chapter is termed an “individualized decision”. These decisions are made under Individualized Care (IC), which corresponds to an interdisciplinary model of healthcare that considers two elements: (i) all patient heterogeneity that can be observed; and (ii) individual choices (Happ, 2010). In this context individual choices correspond to ex-post choices revealed for example through Shared Decision-making⁵⁴ (Charles et al., 1997). This definition of IC is consistent with the use that this term has had in fields such as nursing (Gerrish, 2000, Suhonen et al., 2002) and medicine (Baker et al., Peppercorn et al., 2011). Therefore, an individualized decision can be considered one type of decision at individual level where patient choice is taken into account.

⁵² Personalized Medicine has been defined as the discipline that allows health professionals to prescribe “the right treatment to the right patient at the right dose at the right time”. It has been argued that this goal can be achieved by revealing observable biological or clinical characteristics (Thompson, 2011).

⁵³ In principle, this prediction should be based on complete information. However, in practice, it is estimated conditional to the current understanding of the joint distribution of potential outcomes. In other words, because the current knowledge is limited to a set of observable characteristics, and we still face the challenge of understanding the remaining non-observed heterogeneity, decisions at individual level are in practice decisions at subgroup level.

⁵⁴ Shared decision-making model has become the more accepted decision-making model at doctor-patient level, Its main characteristic is that the patient and the doctor share information about knowledge, preferences and values in order to conclude with a mutual agreement in terms of the treatment.

Following the line of reasoning developed in Chapter 3, this chapter argues that decisions at individual level can only be implemented as individualized decisions, i.e. allowing the patients to make choices without restrictions. In other words, the impossibility of implementing individual level decisions in a centralized system can only be solved through a decentralized decision process. Although the idea of making unrestricted decisions for individual patients seems to receive support in the current health policy scenario, the discussion should also consider the fact that a health system built on the principles of public universal coverage⁵⁵ faces budgeting constraints that make it unlikely that unrestrictive choices can be given to the whole community. This is the source of potential conflict within the NHS, that aims to maximise population health and, at the same time, is pursuing freedom of choice for its patients.

This chapter examines individualization of care in two domains. The first is the extent to which a decision-making approach that incorporates unrestricted choices conflicts with the objective of a collectively funded NHS. It is argued that permitting unrestricted choices depends, for the most part, on a positive judgement, the value of which can be expressed as an expected population health loss due to unrestricted choices. Second, it acknowledges that a normative judgment is also relevant for individualization of care. However, it does not define whether restrictions should be implemented or not. This chapter develops the argument that normative judgments must be made to define the way in which values of health (or HRQoL) are considered in the analysis. Although this is an issue that invites a broader discussion in the economic evaluation field, its role in a policy for individualization of care is essential.

The current chapter is structured in the following manner. First, the conceptual fundamentals of IC are introduced, emphasizing its economic rationale. Second, the conflict between society and the individual in public decision-making is presented, offering an empirical solution. Third, a normative judgment is examined on the basis of the definition of the maximand of the social planner's function. Fourth, a conceptual framework is formulated, structured in six alternative approaches to decision-making. Fifth, the positive and normative elements of the framework are applied to a real case

⁵⁵ The World Health Assembly has launched an agenda to promote universal coverage in all Member States of the World Health Organization. The resolution 58.33 from 2005 states "everyone should be able to access health services and not be subject to financial hardship in doing so". In countries such as UK the universal coverage is pursued to be achieved as a goal of the public sector.

study. Finally, the discussion section highlights the implications of this framework for policy making, with emphasis on the gaps that need further research.

5.2. *Individual decisions and social value*

5.2.1. Economic rationale for pursuing Individualized Care

As discussed in Chapter 2 the main motivation for situating IC at the centre of healthcare arises from an ethical imperative. The autonomy of the patient is the main ethical principle that refers to the right to make voluntary, well informed and rational decisions from all available alternatives (Gerrish, 2000)⁵⁶. Empirical evidence has shown that many actors in the health field have validated this idea, for example, health professionals, healthcare managers, patients and families (Radwin and Alster, 2002). However, although not always acknowledged, the idea of IC also has an economic rationale.

Let us consider a hypothetical scenario where there is perfect and complete information, the health authority does not impose any restriction on access to treatments (it does not offer any coverage either) and there is no income inequality across individuals (although individual pockets impose constraints). In this case individuals choose a new treatment as long as the benefits are greater than the consumption they have to sacrifice⁵⁷. In addition, some people might be willing to forgo some of their own consumption to pay for health interventions that provide additional health in other individuals of the society. This altruistic behaviour can be managed by a social (public) insurer who, given the full information available, will give full coverage only to those patients for whom the benefits are greater than the price and will make the transfers from altruistic individuals to sick people in order to facilitate access to more expensive treatments. However, neither complete nor perfect information exist in healthcare. Most of the studies of comparative effectiveness provide (uncertain) average estimates about treatment. Thus, if a particular treatment is expected to be cost-effective, i.e. the average expected benefits are higher than the expected benefits forgone, the system will recommend the treatment for all patients, including those for whom the treatment is not effective or is

⁵⁶ This is also the same principle to support the implementation of shared decision-making (Gulland, 2011).

⁵⁷ The value of the benefits revealed by individuals under this scenario does not vary due to ability to pay (because it is assumed no income inequality).

even harmful. The result of this policy can be a large efficiency loss as a result of moral hazard (Basu, 2011)⁵⁸. It also imposes a limitation to effect efficient transfers from healthy people to sick people because the budgets coming from altruistic pockets can be wrongly allocated to patients not in need of the treatment. Considering specific information about subgroups can decrease this loss. At the maximum, when all heterogeneity is revealed, the true marginal benefits curve can be represented and the efficiency loss becomes zero⁵⁹.

Individualized decisions are consistent with the aim of decreasing the loss due to moral hazard, because they seek to reveal the joint distribution of potential outcomes. As mentioned in Chapter 2, this understanding could be achieved by exploring observable heterogeneity but also modelling revealed patient choices, because they bring information about unobserved expected individual gains. Therefore, if treatment selection under individualized decisions is consistent with the maximization of individual NHBs, unrestricted choices should be implemented in the NHS⁶⁰. Chapter 1 introduced the argument that the monopoly of the demand side of healthcare is a requirement to achieve efficiency. Here, it is important to emphasize that individualized decisions (unrestricted choices) do not contravene this argument; rather it is a consistent approach to efficiency insofar as costs and benefits are considered.

However, there is no guarantee that treatment selection is consistent with a unique decision rule, for example, the maximization of individual expected **net** health. In fact, patients might want to maximize their individual expected health or something else. If the role of a social planner is to maximize population health subject to budget constraint (maximising individual NHBs) whereas patients maximize another factor, for example **health**, we face a conflict that is represented and analysed in the following section.

⁵⁸ Basu refers to moral hazard in the sense that people will demand according to the information for the average population, which can be considered an inappropriate behaviour because it fails in making consideration of individual level information. This use of the term might conflict with the usual understanding of moral hazard in economics, where the person who make a choice does not take any risk because the costs are borne by someone else.

⁵⁹ The marginal benefit curve represents the true demand curve of the patients for the treatment. This curve reveals the individual treatment effect heterogeneity (Basu, 2011).

⁶⁰ The reason why the maximisation of net health benefits is an objective that the health system should pursue has been introduced and argued in the first chapter of this thesis.

5.2.2. Society versus the individual: centralized or devolved decisions?

5.2.2.1. *Definition of the conflict between the society and the individual*

Eddy (1991) presents this problem as follows (page 1,450):

“The conflict can arise whenever an individual receives a disproportionate amount of resources (e.g. services) without replenishing them (e.g. paying for them). Failure to replenish a resource can occur directly (e.g. the individual is billed for a service but does not pay the bill) or indirectly (e.g. an insurance premium does not anticipate the cost of the service). When either occurs, other people can be harmed in two main ways, depending on whether and how the resources eventually are replenished. If the funds are not replenished, the harm is to health; other people do not get the benefits of the lost services. If the funds eventually are replenished, the harm is financial; other people must pay to replenish the funds. In both cases, the ‘other people’ are what we call society”.

From a collectively funded NHS point of view, where all costs related to health are covered by the health system’s budget, the definition of a *disproportionate amount of resources* depends on the value of the benefits they provide which is revealed when they are compared with the opportunity cost of the health budget (i.e. the health forgone due to displaced activities in the health system). An important concern associated with the implementation of IC in healthcare is that patients and doctors might choose some treatments that could be considered *disproportionate*. This introduces conflict between societal interests versus individual interests. This conflict occurs when the societal decision-maker rejects a new treatment, usually one that is more effective and more expensive. For the purposes of this analysis, it is assumed that there is no conflict when the new treatment is considered cost-effective since free choices can be implemented⁶¹. Next, an analytical approach is proposed to address the question of whether a decentralized process for healthcare decision-making can be considered a good use of limited resources of a collectively funded NHS. Although this description of the health

⁶¹ Indeed, another conflict could arise if a health authority wants to compel patients to receive only one treatment, the most cost-effective, which would exclude other less effective treatments from the whole set of alternatives. The fact that patient might choose less effective (and cheaper) treatments can be inconsistent with the objective of maximizing expected health (since it can only be achieved by choosing the alternative that provides the maximum expected health). However, it is still an empirical question the extent to which individuals are considering other important (unobserved) aspects in their decisions. These elements (if they were observed) might ultimately reveal that the selection is such that maximizes health.

system is a very good representation of a public national health system (e.g. the NHS in the UK) it also applies to any system (public or private) whose objective is health maximisation of a heterogeneous population in terms of health risk profiles.

5.2.2.2. *The value of unrestricted choices: base case analysis*

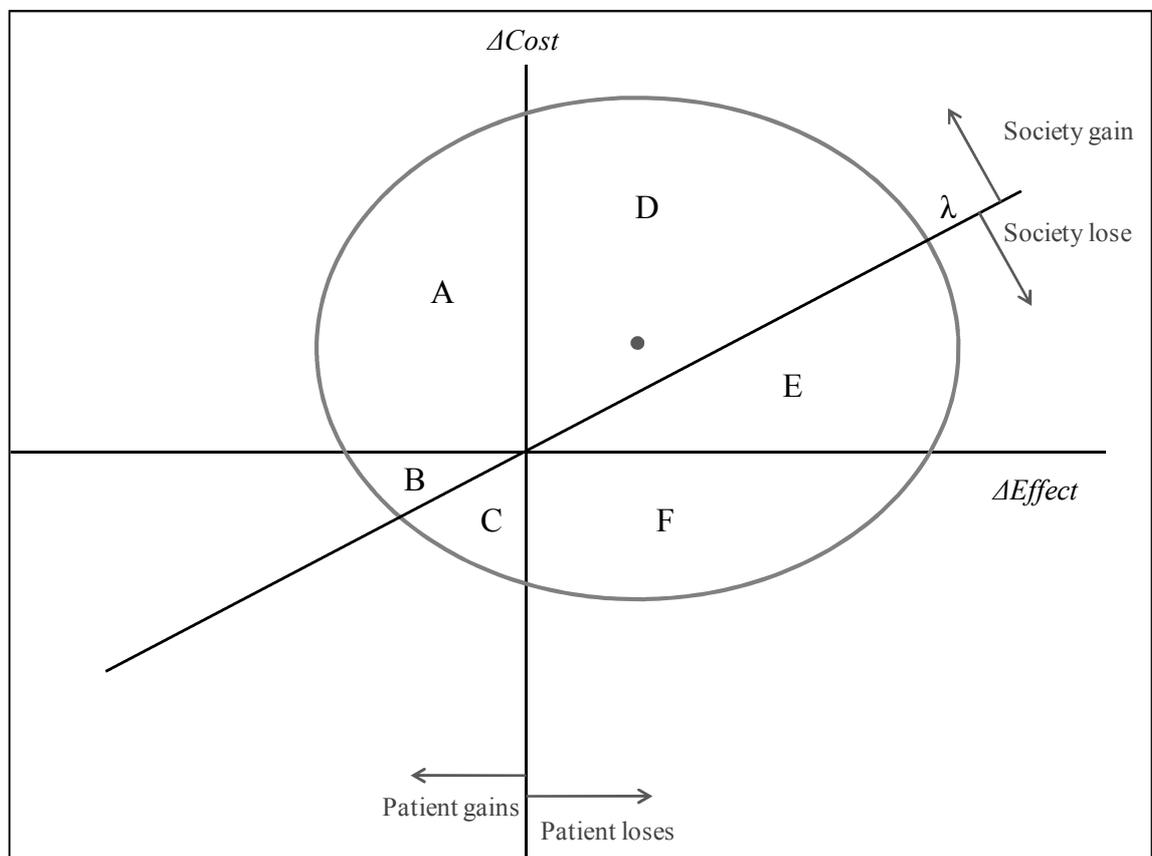
Let us consider the problem faced by the social planner when the decision based on average cost-effectiveness is to reject the new intervention. Figure 5.1 shows a hypothetical cloud of points (oval on the cost-effectiveness plane) that represents the joint distribution of potential outcomes (assuming each point represents one individual). Six areas are defined based on incremental costs, incremental benefits and the cost-effectiveness threshold. The area A contains individuals whose incremental costs are positive (the new treatment is more expensive) and the incremental benefits are negative (new treatment is less effective). The opposite area, F, is that occupied by individuals whose incremental costs are negative (new treatment is cheaper) and incremental benefits are positive (new treatment is more effective). Further, area D includes those individuals whose incremental costs and benefits are positive, but their individual incremental cost-effectiveness ratios (*i*ICERs) are greater than the cost-effectiveness threshold (λ) (i.e. their individual NHBs are less than zero). Area C contains patients with negative incremental costs and benefits whose *i*ICERs are lower than λ . Area B is similar to C but the *i*ICER is greater than λ . Likewise, E is equal to D but the *i*ICER is lower than λ .

If the “average ICER” leads to a decision to reject the new strategy, the net gains of restricted choices for the society can be characterized as the difference between the gains and the losses. The gains correspond to the density of the distribution that lies above λ (A+B+D) whereas the area below λ represents the loss (E+F+C). Thus, the net societal gain of restricting choices is (A+B+D) - (E+F+C).

As an alternative to central decision-making, the social planner can implement a decentralized decision-making process where patients and doctors (assuming doctors as perfect agents of the patients) jointly decide on treatment without restrictions imposed by the central authority. This decentralized approach assumes that: (i) patients (and doctors) will use as much individual information as possible so that they can make a

good prediction of the position of the patient in the cloud⁶²; and (ii) they will choose in a way that maximizes their expected health benefits, where health is measured with the same metric used by the social planner (e.g. QALY). Assumption (ii) is how the source of conflict for the NHS has been initially operationalized in this analysis, since by choosing the most effective treatment (irrespective of costs), using Eddy’s words, an individual might receive “*a disproportionate amount of resources (e.g. services) without replenishing them (e.g. paying for them)*”. This assumption will be relaxed later in this chapter. If this is the case, the net gain for the society under a decentralized process is $(A+B+E+F) - (C+D)$. Table 1 presents the sub-areas of the distribution in terms of gains or losses.

Figure 5.1. Social versus individual conflict: the case of rejection of a technology



Note: The line λ represents the cost-effectiveness threshold. The central dot represents the average incremental costs and benefits of a new strategy compared with the standard. The oval represents the joint distribution of potential outcomes of individuals in terms of incremental costs and effects.

Patients can be situated in one of the following areas according to their individual incremental costs and effects
 A new treatment is more expensive and less effective
 B new treatment is cheaper and less effective

⁶² This includes observable characteristics and revealed choices. If revealed choices are informative of the joint distribution of potential outcomes, the challenge is for analysts to estimate this joint distribution using patient’s ex-post choices.

C new treatment is cheaper and less effective and the incremental net health benefits are greater than zero
D new treatment is more expensive and more effective and the incremental net health benefits are smaller than zero
E new treatment is more expensive and more effective
F new treatment is cheaper and more effective

Table 5.1. Sub-areas of the joint distribution of potential outcomes as societal gains or losses under alternative decisions.

Rejection from a centralized decision-making process		Decentralized (individualized) decision-making-process	
<i>Area</i>	<i>Societal gain or loss</i>	<i>Area</i>	<i>Societal gain or loss</i>
A+B	Gain	A+B	Gain
E+F	Loss	E+F	Gain
D	Gain	D	Loss
C	Loss	C	Loss

A rational solution to this conflict is shown below. If the net gains of allowing individuals to choose are greater than the net gains derived from rejecting the new treatment then no restrictions to the new treatment should be imposed.

$$\text{Net gain due to decentralized process} \quad (A+B+E+F) - (C+D) \quad (5.1)$$

$$\text{Net gain due to centralised process} \quad (A+B+D) - (E+F+C) \quad (5.2)$$

A decentralized decision-making process is better than a centralized rejection if

$$[(A+B+E+F) - (C+D)] > (A+B+D) - (E+F+C)$$

$$E + F - C - D > D - E - F - C$$

$$2E + 2F > 2D$$

$$E + F > D$$

More widely, a decentralized process is a rational option as long as:

$$E + F \geq D \quad (5.3)$$

There are two additional elements of this problem. First, it might be the case where the implementation of guidance that rejects a new (more effective) treatment leads to some additional costs (C_r). C_r includes: (i) the time spent in giving special considerations to conflicting views; (ii) costs of reviewing additional information submitted to the regulatory agency; (iii) additional effort to disseminate and monitor the fulfilment of a guideline that does not have the approval of the whole clinical community; and (iv) the

resources needed to deal with some societal discomfort created by the fact that some patients will not have access to an effective treatment. Second, as mentioned above, a decentralized decision-making process aims to take into account as much individual information as possible, which is the result of the exchange of information between patients and health professionals and other relevant actors. It seems totally appropriate to assume this process as the implementation of the shared decision-making model (SDM) (Charles et al., 1997). It is very likely that this process is not without costs. These costs include the extra time of health professionals and extra costs to reveal heterogeneity (C_{SDM}). If both, $C_r \geq 0$ and $C_{SDM} \geq 0$, then they should be subtracted from the net societal gain from centralized rejection and decentralized decisions, respectively. Thus, *decentralized decisions are the best option if:*

$$E + F - C_{SDM} \geq D - C_r \quad (5.4)$$

$$E + F - D \geq C_{SDM} - C_r \quad (5.5)$$

The expression to the left in equation (5.5) represents the additional value of adopting decentralized decisions instead of rejecting the new (non cost-effective) treatment. The expression to the right of the inequality corresponds to the additional cost of implementing shared decision-making and individualized decisions with respect to the costs of rejecting the treatment for the whole patient population. If the additional value of decentralized over centralized decisions is greater than the additional costs of implementing decentralized over centralized decisions, then decentralized decisions can be considered a good use of resources.

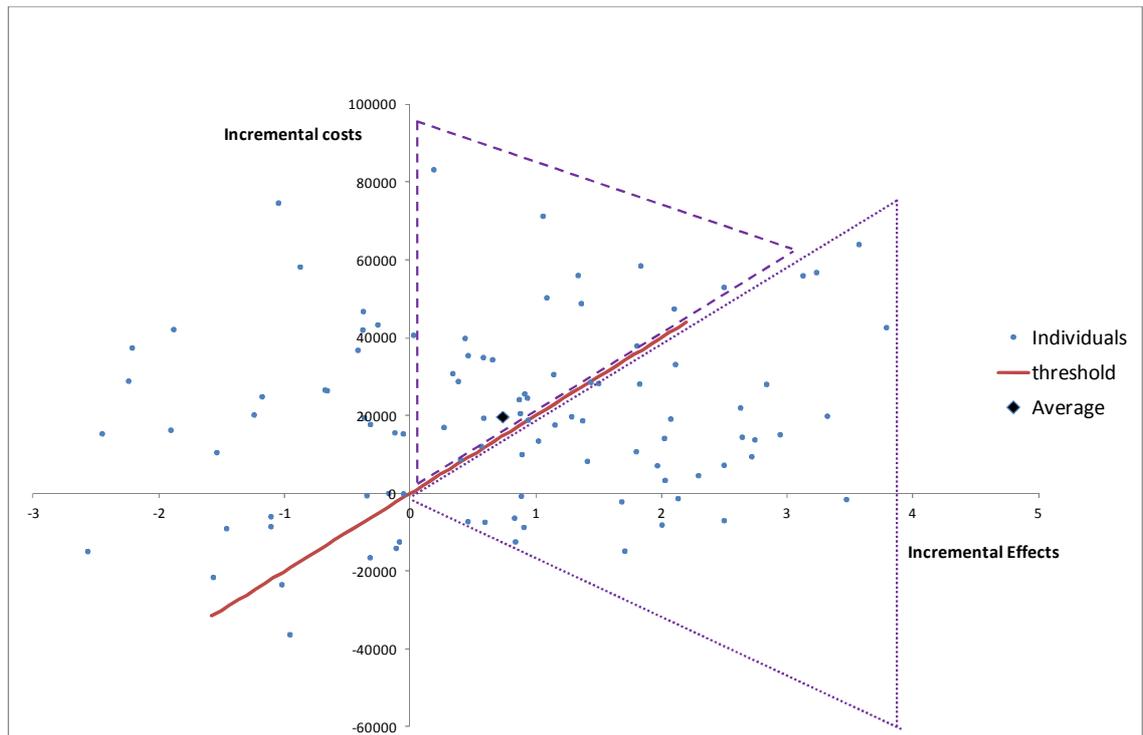
The problem of asymmetry of information has been implicitly considered in the solution presented here. If decisions are made under average current information, it is very likely that some relevant information of the individual patients is not considered. However, if decisions are devolved to the individual level, in the context of SDM, asymmetry of information should no longer constitute a problem since all relevant and available information will be taken into account. Once again, this relies on the assumption that doctors are perfect agents for their patients.

From this point onwards, area D will be called NNE quadrant (north north east) and (E+F) will be called NSE area (north south east). A stylized example is shown in Figure 5.2. Each area corresponds to the sum of the incremental individual Net Health Benefits ($iNHB$). The value of leaving individuals to choose can be calculated as the difference between the weighted averages of the $iNHB$ in the NSE and NNE. The weights for NSE and NNE (w_s and w_n , respectively) are the proportion of the densities that such areas represent, with respect to the whole joint distribution. A population estimate can be obtained multiplying this amount by the relevant population as explained in Chapter 3 (equation 3.18).

As shown in Figure 5.2, the value of implementing unrestricted choices instead of rejecting the new treatment is 28,000 QALYs. Thus, decentralized decisions can be made if the additional costs of SDM compared to centralized decisions ($C_{SDM} - C_r$) are less than £560,000,000. It could be assumed that the implementation of SDM is very unlikely to cost more than that amount. Therefore, decentralized decisions should be implemented even though the costs due to rejection are zero ($C_r=0$). Because the NNE and NSE areas depend on the cost-effectiveness threshold, the difference in costs ($C_{SDM} - C_r$) can be plotted as a function of λ (this will be explained in detail in Section 5.4 and shown in Figure 5.4).

This example shows the value of decentralized decisions as health gains. However, it might be the case that the value is negative, indicating expected losses. If decision-makers are willing to implement decentralized decisions assuming the loss, the expected health forgone represents the social value of unrestricted choices for healthcare (this concept will be examined in detail in Section 5.4).

Figure 5.2. The societal versus individual conflict: A stylized example.



Note 1: The graph shows 100 simulated points distributed on the four quadrants of the cost-effectiveness plane. The diagonal line represents a cost-effectiveness threshold of 20,000 per unit of benefit. The black diamond illustrates the mean incremental costs and benefits. The dashed and dotted triangles frame the areas of conflict, previously defined as NNE and NSE respectively.

Note 2: Estimation:

1. Area NSE (dotted triangle) has 39 individuals (39%). The average incremental individual net health benefit is 1.36 QALYs.
2. Area NNE (dashed triangle) has 27 individuals (27%). The average incremental individual net health benefit is 0.89 QALYs.
3. The difference between NSE and NNE is $(1.36 \times 0.39) - (0.89 \times 0.27) = 0.28$ QALYs (net of costs)
4. Assuming a relevant population of 100,000 patients the value of decentralized decisions is 28,000 QALYs.
5. The additional cost of shared decision-making compared to transaction costs due to rejection should be less than the value of decentralized decisions (28,000 QALYs) converted into monetary terms, i.e. 560,000,000 (28,000x20,000).

5.2.2.3. *The value of unrestricted choices: An extended analysis*

As explained in the previous section, this method assumes that patient choices are consistent with the maximization of health. Further, it assumes that the health outcome is the same as that which matters for the society, for example, QALYs. However, it might be the case that patients' (and doctors') choices are driven by a different outcome. Even if they are interested in maximising expected health, QALYs might not represent the only element in their individual preferences. In other cases, choices are directly explained by considerations other than maximization of expected health gains. For this reason, the method presented in the previous section needs to be extended in order to capture these alternative heuristics.

The assumption that patients and society focus on the same health outcome can be relaxed by estimating which individuals change their decisions when they seek to maximise another outcome. Thus, in those subjects who rejected the new strategy (based on the rule of maximising QALYs), which implied societal gains (areas A and B), a change in their decision will impose societal losses. As a consequence, they should be subtracted from the A+B density. Likewise, in those individuals whose adoption of the new treatment represented a societal gain (E+F), a change in their treatment choice (based on a different decision rule) will represent a societal loss. Analogously, for areas C and D, where rejection and adoption, respectively, were understood as societal losses, a change in the decision implies a societal gain. These changes affect only the net gains due to a decentralized process, which can be expressed as:

$$[(A_i+B_i+E_i+F_i) - (A_j+B_j+E_j+F_j)] - [(C_i+D_i) - (C_j+D_j)] \quad (5.6)$$

where i represents the density of individuals whose decision is the same under any heuristic, i.e. the new decision rule or the maximand used by the social planner (e.g. QALY), and j represents the density of individuals whose decision changed, when they apply an alternative decision rule. In the case of a centralized process the areas are the same as in equation (5.2), however, they must be expressed as a sum of individuals i and j as follows:

$$(A_i + A_j + B_i + B_j + D_i + D_j) - (E_i + E_j + F_i + F_j + C_i + C_j) \quad (5.7)$$

This analysis needs the identification of the outcome based on which individuals will make decisions, i.e. the new maximand. Expected costs and benefits must be estimated for both outcomes simultaneously, in order to identify which individuals keep or change their decisions. Finally, as in the previous case, a devolved decision-making process is appropriate if net gains due to decentralized decisions are greater than net gains of those that are centralized, i.e. the magnitude of equation (5.6) is greater than the magnitude of equation (5.7).

$$[(A_i + B_i + E_i + F_i) - (A_j + B_j + E_j + F_j)] - [(C_i + D_i) - (C_j + D_j)] > \quad (5.8)$$

$$(A_i + A_j + B_i + B_j + D_i + D_j) - (E_i + E_j + F_i + F_j + C_i + C_j)$$

$$E_i + F_i + C_j \geq A_j + B_j + D_i \quad (5.9)$$

If the transaction costs derived from the implementation of centralized and decentralized decisions are also considered in the expression (5.9), the solution is equivalent to equations (5.4) and (5.5).

$$E_i + F_i + C_j - C_{SDM} \geq A_j + B_j + D_i - C_r \quad (5.10)$$

$$E_i + F_i + C_j - A_j - B_j - D_i \geq C_{SDM} - C_r \quad (5.11)$$

The analytical approach presented in this section supports the argument that the implementation of centralized versus decentralized decisions must consider a positive judgment, which is the net health that can be gained or forgone when patients make their own choices. A stylized example of this extended approach is presented in Appendix 5 and its implementation in a real case study is illustrated in section 5.4.

5.2.3. Social value orientation: A normative judgment for individualization of care

The need for a detailed examination of the implications of individualized care and unrestricted choices is motivated by the role given to patient autonomy in the current policy debate. Autonomy of patients is a well established principle of social value (NICE, 2005). However, although it has been considered one of the central tenets in the conceptualization of a healthcare system, the extent to which the system responds to this principle is not unbounded. Society values individual autonomy, insofar as its consequences do not affect other important objectives of the society, such as other individuals' wellbeing or health (Oshana, 2003). Previous sections showed how the decision of implementing restricted versus unrestricted choices can be supported by a consideration of efficiency. In other words, the social value judgment required to implement unrestricted choices (value of individual choices for treatment) is not normative but positive, i.e. it must take into account the amount of health that is forgone due to decentralized decisions. However, individualization of care is not only about the implementation of unrestricted choices. It also requires an analysis of the role of the individual in the definition of the outcome that the social planner aims to maximise, which ultimately relies on a normative judgment.

This section develops a categorization that takes into account two perspectives. First, a microeconomic perspective that considers characteristics of welfarism and extra-welfarism, and second, a behavioural perspective that takes alternative social value orientations into account⁶³ (Offerman et al., 1996, Ubel et al., 2000, Messick and McClintock, 1968). Both perspectives are combined, producing three categories that will be formalized in the framework presented in the next section.

As defined in Chapter 2, *preferences* refer to the rational element of judgment that guides the ranking of health states. They are represented as a set of values elicited from a particular population whose welfare, utility or health wants to be maximised. The use of preferences in economic evaluation is founded on the grounds of the neoclassical welfarist approach, which places the maximization of utility (as the only argument of

⁶³ Social value orientation is a social psychology theory of choice. In healthcare decision-making can be defined as the judgments and assumptions held by the social planner in terms of the objectives that members of the society should pursue when they make choices that affect others.

the function) in the centre of welfare theory (Hurley, 2000)⁶⁴. In particular for health economic evaluation, the stream that assumes QALYs as utilities (also called welfarist) relies on the assumption that values (utilities) have been revealed by individuals under conditions of uncertainty, for example, using the “standard gamble” tool (Drummond et al., 2005). An alternative approach uses health outcomes in monetary terms, for example with methods of contingent valuation. An element of frequent confusion relates to who should provide such values, patients or society? From a welfarist perspective, the benefits should be constructed as utilities of the affected individuals, i.e. the patients (Gandjour, 2010). If the sum of individual benefits is greater than the sum of costs (including the disutility of some other patients) the new treatment should be adopted and no transference of benefits between gainers and losers need to be made (Greenberg and Weimer, 1996). This view is consistent with an individualistic social value orientation, where each individual is interested in his/her own welfare/utility only.

Several guidelines of cost-effectiveness analysis have recommended the use of societal values in the construction of QALYs (Gold et al., 1996, NICE, 2008). One of the main arguments in favour of this position is that a representative sample of individuals of the society can express their preferences under the “veil of ignorance” (Harsanyi, 1953, Harsanyi, 1955, Rawls, 1971), providing the adequate set of values to construct QALYs⁶⁵ (Gold et al., 1996, Drummond et al., 2005). A second argument is raised from the fact that values obtained from patients for one particular health state are greater than those elicited from the general population (Sackett, 1978; Boyd,1990)⁶⁶. Subsequently, using patient (overrated) values would lead to lower estimates of incremental effectiveness with a subsequent higher incremental cost-effectiveness ratio, which

⁶⁴ Besides utility as the only argument and preferences on the basis of the utility maximization, the welfarist approach assumes that individual are the best judges of their welfare or utility and that utility derives from the consequences (outcomes) and is not related to the process of how the outcome is obtained.

⁶⁵ The ‘veil of ignorance’ argument is constructed on the basis of a hypothetical scenario where the roles of the individuals in the society are redistributed and reassigned (Harsanyi, 1953). Then, the true morality of an issue can be judged by rational individuals under this state of blindness previous to the new assignment. Under a later conceptualization given by Rawls, people covered by a ‘thicker’ veil of ignorance do not have any clue about which talents or personal capacities will be endorsed (or they do not know the probability of being assigned some specific role in the society) and therefore they would reveal values taking the position of the worst-off in the society which leads to a more egalitarian distribution of well-being as a consequence of the decisions they make.

⁶⁶ Higher values from patients are explained because patients suffer adaptation to their health conditions (they adjust their expectations to the new circumstances) (Gold, Ubel 2000, Loewenstein 1999) and/or because of cognitive dissonance, which refer to the fact that patients do not admit how low is their quality of life (Ubel, 2000 referring Festinger 1959).

ultimately affects patients (Gold et al., 1996). A third argument states that decisions for the use of public resources should use public preferences (Gold et al., 1996). It is fundamental to acknowledge that these arguments do not support QALYs as utilities; they only claim that the construction of QALYs based on community values is appropriate.

QALYs based on community preferences have been better defined as “capabilities that good health brings us” (Brouwer and Koopmanschap, 2000), (page 447). This notion finds its grounds in the extra-welfaristic approach (Brouwer et al., 2008, Culyer, 1990). Extra-welfarism has some important roots in Sen’s capabilities approach (Sen, 1993), especially on the use of outcomes other than utilities. For Sen, the welfarist approach based on utilities seems too narrow, since it is too focused on the emotional responses of people to particular valuable states (or dimensions of wellbeing) rather than people’s actual capabilities of achieving such states (Brouwer et al., 2008). However, in terms of the interpretation of QALYs, the “extra-welfarist QALY” (also called the “health QALY”) is not a measure of the whole capabilities set but only their health determinants⁶⁷. In addition, the extra-welfarist approach acknowledges that an indication of preferences is still needed to undertake comparative assessments of health gains (or capabilities gains) because the maximisation process relies ultimately on a cardinal outcome index⁶⁸. Finally, the use of community preferences avoids the problem of adaptive preferences (i.e. patients value their health state better because they suffer adaptation). Thus, health outcomes based on community values guaranties that two individuals with different degrees of adaptation, but with the same capabilities, have the same priority of access to limited resources⁶⁹.

⁶⁷ This contrasts with an alternative view that suggests the QALY as a measurement of the whole capability set (Cookson, 2005). However, as Cookson acknowledges, the estimation of the “capability QALY” needs further methodological development: “*Until such methods have been developed, however, standard QALY instruments may be most useful in contexts where the health determinants of capability are of primary concern to decision-makers*” (page 826).

⁶⁸ This assertion highlights another important difference between Sen’s capability approach and extra-welfarism; while the latter retains the notion of maximisation of the expected benefit (“health QALY”) the former imposes supremacy of equity and distributional concerns (Coast et al., 2008).

⁶⁹ This idea has received some empirical support. Lowenstein and Ubel (2008) reported that while adapted patients reveal normal levels of happiness they also refer willingness to make big efforts to recover normal health and therefore there is no reason to categorize them in different levels of access to healthcare.

Community based QALYs finds its significance as a socially legitimate set of values that can be used to make cardinal representation of the HRQoL. QALYs constructed from this value set are a good example of an outcome adopted by a Paternalist decision-maker. This is defined on the basis of a cooperative social value orientation, where individuals pursue increased health for both themselves and the rest. In its translation to social choice, the paternalist decision-maker imposes an equity judgment in the sense of not favouring any person over another, which becomes a fundamental normative judgment for individualization of care. Further, under this approach the social planner seeks to maximise some outcome that is not individual utility but health, which is consistent with the grounds of extra-welfarism.

The third value orientation considered is altruism. Altruists seek to maximise the welfare of others irrespective of the consequences on themselves. In the healthcare evaluation problem this concept can be applied to a decision-maker who focuses his/her analysis on the expected health of a particular group of patients. Under this approach, the decision-maker can also incorporate for example, differences between adapted and non-adapted patients⁷⁰ (Dolan and Kahneman, 2008). This approach can be operationalized using QALYs constructed on the basis of values elicited from a representative group of patients suffering the condition under assessment, which is a legitimate set of values for a sub-population (non-individualistic). Therefore, it is more consistent with the view of maximising health rather than utilities⁷¹, which adheres to the principles of extra-welfarism.

In this thesis, is argued that the definition of using patient versus societal values should be the result of a normative value judgment. This contrasts with most of the arguments in favour or against the use of patient values in the literature, which are mainly pragmatic. For example, Gold et al. (1996) argued that the use of values elicited from patients might produce results that seem less cost-effective and could result in patients being denied treatment. Appendix 6 shows the scenario where this statement holds but

⁷⁰ This idea has been raised in the literature that advocates for the use of experienced utility. According to Dolan and Kahneman if adaptation is not considered in resource allocation, then patients who have suffered adaptation will be in advantage because their gains in experience utility will be (by adaptation) greater. As a consequence they claim: “*decisions will then be made as if adapted patients’s gains in experienced utility count for more than the gains of patients who adapt less*” (page 220).

⁷¹ These values might be obtained by time trade-off for example. The motivation to use these values adheres to the idea of having a preference based system (Torrance, 2006), though the assumptions of utility do not hold and the outcome becomes a measure of patient health.

also other three cases where patient values do not generate higher ICERs than societal values. On the other hand, as explained in Chapter 2, if we are interested in considering heterogeneity in preferences, it is more feasible to implement individualized decisions based on patient values than with community values.

5.3. Alternative approaches to decision-making: a descriptive framework

This section presents the description of six alternative approaches to decision-making, defined on the basis of two dimensions: the level where decisions are made; and the value orientation of the decision-making approach. The framework assumes a healthcare system where all costs fall inside the health budget and an economic assessment follows the rules of cost-effectiveness analysis.

The level of decisions can be centralized or devolved. This definition must respond to a positive judgment, as presented in section 5.2.2. A centralized process is characterized by the production of mandatory guidelines where heterogeneity can be taken into account through the implementation of differential decisions for different subgroups. Under this approach, the health authority is allowed to restrict the adoption of interventions that were considered not cost-effective. On the other hand, a devolved health system is characterized by decentralized decisions made at individual level. It is expected that patients and doctors consider as much individual information as possible so that they can make a good prediction of the individual expected benefit. Under this scheme patients make choices among all available alternatives, something that should result from the transference of information between patient and health professional in the context of a shared decision-making model. As explained before, the consideration of choices in the decentralized process imposes a challenge on the analyst's task of characterizing the joint distribution of potential outcomes, which should incorporate those choices in the estimation.

In the normative domain, this framework focuses on the definition of the maximand of the function that the decision-maker aims to maximise. Three categories have been defined on the basis of social value orientations. First, a paternalistic decision-maker pursues the maximisation of population health, adopting a health metric constructed on the basis of a socially legitimate set of values, for example, QALYs based on

community values. Second, an altruistic decision-maker seeks to maximise a health outcome constructed on the basis of a set of values elicited from a representative group of patients affected by the disease under evaluation. Finally, the welfarist decision-maker is consistent with the neoclassical welfare theory, in that the objective is to maximise total utility as the sum of individual utilities. A summary of the six approaches is presented in Table 5.2.

Table 5.2. Summary of six approaches for decision-making in healthcare

		Solution responds to an efficiency criterion →	
		Centralized Mandatory guidelines produced at central level Restriction to interventions that are not cost-effective Heterogeneity is taken into consideration providing different recommendations for different subgroups	Devolved Devolved decisions at individual level using as much observable information as possible Patients allowed to choose between all alternatives Choices are assumed to be revealed from a shared decision-making process
Solution responds to a value orientation to health outcomes criterion ↓	Paternalist (cooperative, extra-welfaristic) e.g. Community health state values	Paternalistic decision-maker and centralized decisions	Paternalistic decision-maker and devolved decisions
	Altruist (altruist, extra-welfarist) e.g. Patient health state values (not utilities)	Altruistic decision-maker and centralized decisions	Altruistic decision-maker and devolved decisions Patient preference heterogeneity can be operationalized in this approach.
	Welfarist (individualist) e.g. Patient health state utilities, individual WTPs	Welfarist decision-maker and centralized decisions	Welfarist decision-maker and devolved decisions Patient preference heterogeneity can be operationalized in this approach.

5.4. *Alternative approaches to decision-making: a case study*

5.4.1. Base case analysis

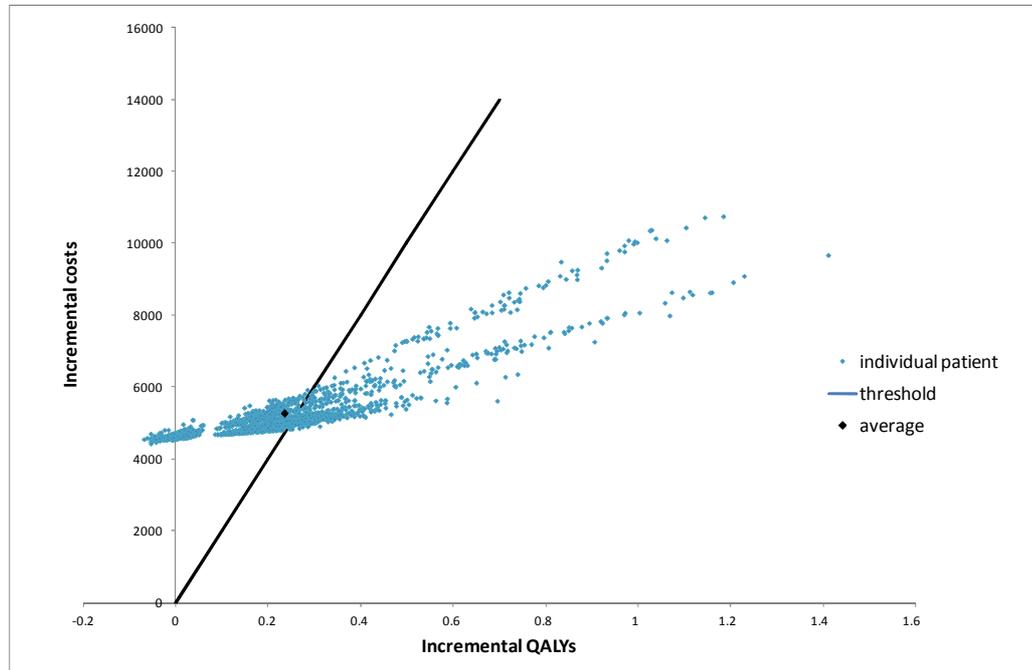
In this section, it is shown how the framework proposed in this thesis can be used to address a decision problem in a real case study. The study RITA-3 presented in detail in Chapter 4 offers an opportunity to illustrate the concepts developed in this chapter. In brief, the RITA-3 study compared the efficacy of an intensive versus a conservative treatment strategy for the early management of patients with non-ST-elevation acute coronary syndrome. The results on average showed that an invasive strategy is efficacious but is not cost-effective at a threshold of £20,000 per QALY. Further, the cost-effectiveness study showed significant heterogeneity between patients, which was analyzed in Chapter 4 in the context of subgroup analysis. In this chapter, the decision problem is examined in terms of the alternative approaches that a decision planner could adopt.

A normative and a positive dimension are considered in this analytical framework, namely, social value orientation (paternalistic, altruistic or utilitarian) and the level of the decision (centralized versus devolved). On one hand, the definition of the social value orientation responds to a purely normative judgment, for example, the comparison of cost-effectiveness estimates with community values, patient values or patient utilities. RITA-3 was conducted in the UK and its cost-effectiveness analysis was consistent with national guidelines that recommend values from the general community, therefore a paternalist value orientation is taken. On the other hand, in terms of the level of the decision, individual expected incremental costs and QALYs were estimated from RITA-3 (see Chapter 4). Thus, the question as to whether to implement a decentralized decision process is consistent with an appropriate use of limited resources can be addressed in this case where the average estimate suggests the rejection of the invasive strategy.

The individual incremental costs and QALYs were estimated from the data described in Chapter 4. Figure 5.3 shows the cloud that represents the joint distribution of potential outcomes where each point corresponds to one of the 1,810 patients of the trial.

According to the methods applied in chapter 4, each point is the average point across 1,000 iterations of a probabilistic sensitivity analysis.

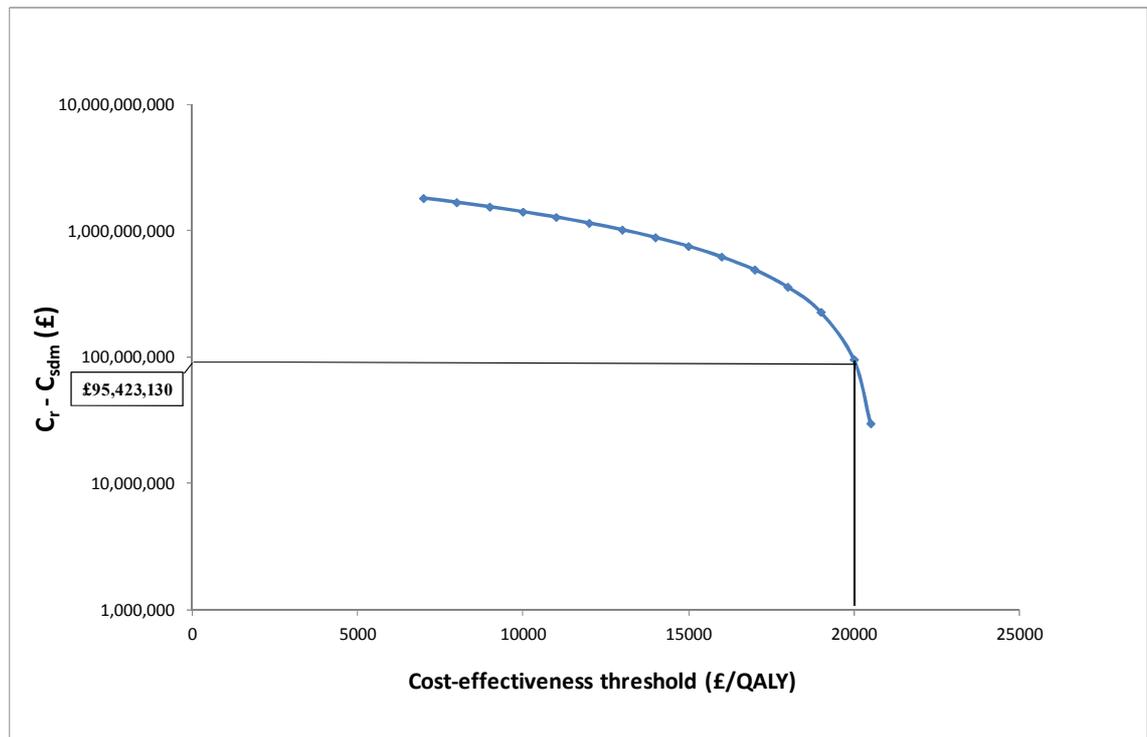
Figure 5.3. Estimated joint distribution of potential outcomes from RITA-3 trial.



In accordance with the method explained in Section 2 of this chapter, the average *i*NHB of the NNE quadrant is 0.099 net QALYs and its density represents 58% of the whole joint distribution (weighted average 0.057 net QALYs). Likewise, the average *i*NHB of the NSE area is 0.146 net QALYs and its density is 33.7% of the joint distribution (weighted average 0.049 net QALYs). As a consequence, the difference between NSE and NNE, estimated for a relevant population of 556,723 patients for the next ten years (see Chapter 4), is -4,771 net QALYs (27,454 – 32,225). This amount in monetary terms corresponds to -£95,423,130 and the result suggests that an approach based on decentralized decisions cannot be supported, as it produces an expected loss of -£95,423,130. In other words, decentralized decisions can be considered only if the additional cost of implementing a guideline based on rejection compared to shared-decision-making ($C_r - C_{sdm}$) in the next ten years is greater than £95,423,130, which is very unlikely. This amount can be plotted for different values of the cost-effectiveness threshold. The line drawn in Figure 5.4 represents a boundary for centralized and decentralized decisions. The values above the line can be considered pro decentralized decisions whereas the values below correspond to losses from decentralized decisions.

As mentioned in section 5.2.2, the expected loss represents the social value (imposed by a decision-maker) of implementing decentralized decisions. In reality, the true value of implementing unrestricted choices is not just the difference between NSE and NNE areas but this difference minus the costs of implementing decentralized decisions (C_{sdm}) and plus the costs due to rejection (C_r).

Figure 5.4. Minimum cost of restriction to consider decentralized decisions as the best option.



Note: $C_r - C_{sdm}$ represents the additional costs of implementing a guideline based on restrictions compared to the costs of a shared decision-making process. Results are shown in a logarithmic scale for different values of cost-effectiveness threshold (λ). The base reference case when λ is £20,000/QALY is illustrated.

5.4.2. Extended method

This section analyses the case when a different outcome is maximised at individual level. RITA-3, as is the case with many emergency health problems, describes a clinical setting where conducting shared decision-making is unfeasible. Instead, it is very likely that the intervention will be the one that is prescribed by the doctor, who only takes into account the clinical evidence. Consistent with that idea and for illustrative purposes, it is assumed that doctors will make their prescriptions maximising life years. It also

assumes that doctors have knowledge about the effect of individual characteristics on the health outcomes, which is tantamount to maximising individualized life years.

The RITA-3 model allows us to estimate the expected individual life years' increment. The analysis is focused on finding those individuals who would receive a different strategy if the rule of maximising life years was applied. According to equation (5.9) the densities of interest are E_i , F_i , C_j , A_j , B_j and D_i . The study showed the incremental life years to be positive in all patients. Thus, all patients in the NW quadrant (area A) would receive invasive treatment. More formally, A_j is equal to A. Individuals in the remaining areas do not change their decisions. In terms of the areas, this can be expressed as E_i plus F_i is equal to E plus F (NSE) and D_i is equal to D (NNE). B_j and C_j (equivalent to B and C in the previous analysis) are equal to zero.

The average *i*NHB of the NW quadrant is 0.25 QALYs and its density is 8.23% (weighted average of 0.02 QALYs). Applying equation (5.11) the results show that a decentralized process is associated with an expected loss of 16,397 QALYs ($NSE - NNE - NW = 27,454 - 32,225 - 11,626$).

5.5. Discussion

Decisions based on cost-effectiveness analysis usually face the conflict derived from rejecting treatments that are deemed effective but not cost-effective. This conflict is usually observed and debated in countries where decisions are made at central level, after an exhaustive assessment of the evidence available. This chapter examined the individualization of care in two domains: first, the extent to which the implementation of a decentralized decision-making process (allowing for unrestricted patient choices) determines health losses in a collectively funded health system and second, the normative judgments that social planners must make when individualization is being implemented in the health system.

This chapter developed an analytical approach to addressing the question whether the implementation of unrestricted choices in a collectively funded health system represents a good use of limited resources. In the case that unrestricted choices lead to health losses, the method allows analysts to present the expected health that the society is sacrificing in order to satisfy such social value. This method contributes to the

understanding of the potential conflict derived from rejection of new technologies and provides an important empirical element of judgment for decision-makers.

In addition, a case study has been used to illustrate the practical implication of this method. This example assumed the structure presented in Chapter 4 to estimate the individual net outcomes for each treatment (invasive and conservative). However, as explained in detail in that chapter, the estimation of expected costs and benefits for both treatments (factual and counterfactual) did not consider individual treatment effects. Only recently, Basu (2011) has provided some empirical illustration of how to estimate individual treatment effects applied to health. However, those methods require data from adaptive trials where individuals are able to reveal choices for treatments, something that cannot be implemented with the information available for this study. Further research is needed in the cost-effectiveness field to introduce individual treatment effects in the analysis, which will offer a potentially better estimation of the joint distribution and an opportunity to improve the estimations of the method presented in this chapter. Another limitation of this study is the context of the health problem. As an emergency problem, patients (and doctors) do not have enough time to reveal choices according to SDM process. Therefore, emergency is one of the areas where this type of analysis is limited *a priori* because of the clinical scenario. However, despite the context, the example offers an opportunity to illustrate the concepts here developed. As a matter of fact, the results support restricted choices even though decentralized decisions could have been implemented.

One important element related to the joint distribution and beyond the challenges related to its estimation is the assumption that patients and doctors will be able to predict the position of the patient on the cloud presented in Figure 5.1. It has been argued that a good prediction depends on the process whereby patients and doctors share and reveal as much individual level information as possible, or at least, the information that was used to estimate the joint distribution. Two elements emerge from this point. First, there is still the need for a centralized agency responsible for the evidence assessment. Doctors and other health professionals are not expected to have access to evidence and the skills to undertake an adequate assessment. Therefore, a decentralized process still relies on the systematic assessment of the available evidence in order to provide doctors (and patients) with the correct information limiting bias and

misleading interpretations. Furthermore, the agency is able to gather information that doctors and patients are not able to access (e.g. early information given by the industry). In addition, the agency should be equipped with the tools to undertake an assessment of such information with the highest quality standards of the country or jurisdiction. Second, it is assumed that health professionals act as perfect agents of the patients. Given the asymmetry of information between doctors and patients, it is a necessary condition that doctors can provide the adequate information but also can consider specific information about the patient (clinical characteristics and preferences) in order to give proper assistance to the decision of the patient. This has been conceptualized here as the implementation of shared decision-making (Charles et al., 1997). Any incentive to health professionals determining behaviour other than that in the interest of the patient is a failure that might alter this prediction.

Another important assumption of this method is that patients will decide according to the rule of maximising health. The base case presented in this chapter assumes the health metric is the same for patients and the healthcare system, for example, QALYs constructed from community preferences. This assumption is restrictive in the sense that the decision of the patient might not be based on the outcome defined under (social) normative judgments⁷². In response, an extension of the method has been proposed in order to undertake the same analysis, but considering that patients could choose based on a different outcome. On the other hand, there is some empirical evidence suggesting that there are certain patients that could choose those treatments that are cost-effective. It has been reported that while most patients value “having choices” it is not clear that all patients are interested in “making choices”(Ogden et al., 2008), relying ultimately on some expert opinion. Others have referred to certain patients, in particular in some societies (e.g. Japan), have a stronger sense of social duty when express their choices which would affect their decisions (Iyengar and Lepper, 1999).

Despite the possibility of considering alternative outcomes for individual decisions, this approach still has a limitation. The method assumes that individuals make decisions according to the axioms of expected utility theory (von Neumann and Morgenstern,

⁷² This sentence is written in conditional because patient might (or might not) choose according to QALYs. Under a shared decision-making process doctors explain the effect of the treatment based on the information he/she has received from the national agency. In this context QALY is a main part of the outcome's assessment and can be very influential on patient's final decisions.

1947), i.e. the maximization of the expected value of the outcome⁷³. It has been well studied that when people make decisions under uncertainty they are, in general, risk averse to gains and risk-seeking to losses, something that has been formalized in Prospect Theory⁷⁴ (Kahneman and Tversky, 1979, Kahneman and Tversky, 1992). In other words, they do not assume a linear utility function between gains (or losses) and value. An area of future research should explore how these descriptive elements of choice can be adapted into the model presented in this chapter.

On the other hand, the definition of alternative approaches to decision-making, such as the consideration of unrestricted choices, has determined the need for paying attention to social values. Here, it has been argued that social planners need to make a normative judgment about the outcome that will be maximised. Individualization of care can be seen from a classical welfarist perspective to one that is extra-welfarist, depending on the social value orientation adopted. Thus, three approaches have been defined (paternalist, altruistic and welfarist). The lack of consideration of other social values in this categorization is not an exclusion of such values but only a practical approach to offer a useful conceptualization that we can consider a starting point. Examination of other social values (e.g. equity) requires special attention and should be studied in future research.

An important social value explored in this chapter is choice. It has been proposed that the value of choice can be expressed as the expected health loss due to implementation of decentralized decisions. This sets out an efficiency-choice trade-off that can be interpreted as the health forgone in order to implement unrestricted choices. If autonomy is valuable insofar as individual decisions do not affect other subjects of the population, the questions faced by the social planner are: what is the magnitude of

⁷³ The expected value is the result of a linear combination of possible outcomes and their probabilities.

⁷⁴ Prospect Theory has two main characteristics. First, individuals set out a reference point from where they evaluate prospects in terms of gains and losses. Second, in contrast to EUT where the expected utility is the weighted average between outcomes and probabilities, PT relies on some value (utility) that is a function of the outcome and some weighting function that weights the probabilities. Kahneman and Tversky in their formalization of PT proposed a value function that is concave in gains and convex in losses. Further, the concave part of the curve is flatter than the convex. This shape is explained by the fact that patients are risk averse when they face uncertain decisions about gains and they are risk seeking when face decisions about losses. The weighting function represents the fact that, for example, an increment of a probability from 0.001 to 0.051 is different that from 0.5 to 0.55. It has been suggested that subjects overweight small probabilities and underweight higher probabilities. This means that individuals who face very rare events tend to overreact as well as they tend to underreact to medium or large probabilities.

health that the society should be willing to forgo in order to satisfy the autonomy of patients? And which patients would deserve this special attention? It seems that these questions are the ultimate value judgments related to the decision about implementing unrestricted choices of treatment.

The framework and method developed in this chapter is a consequence of the author's interest in the understanding of the value of heterogeneity. Hence, there are important links with previous contributions in this field. One of these is the relationship with the metrics of the static value of heterogeneity, such as EVIC. The magnitude of total EVIC indicates the degree of observable heterogeneity that has an impact on decisions. Thus, if the total estimated EVIC is high, the (predicted) joint distribution of potential outcomes is likely to lie widespread on the cost-effectiveness plane. If this is the case and the new treatment is considered not cost-effective on average, the decision-maker should ask whether a decentralized decision-making process is as efficient as (or more efficient than) implementing the rejection. The method presented here is a major contribution to addressing this question.

Finally, this framework is consistent with a healthcare system that bears all health related costs. In other systems, for example, when part of the health expenditure relies on private (individual) pockets or when there is a private market for health insurance, the solutions provided in this work might not be adequate. The implications of different market failures on the reimbursement decision of a public health authority will be addressed in the final chapter of this thesis.

6. CHAPTER 6:

DISCUSSION AND CONCLUSIONS

6.1. Summary of the thesis contribution

Health has been recognized as the main objective of health systems. In the context of limited budget, an optimal allocation of resources becomes a central goal of the decision-making process. Cost-effectiveness analysis has been used to support decisions about reimbursement of new health interventions to be consistent with the efficiency goal. These decisions are usually made under uncertainty and with limited knowledge about heterogeneity. The value of considering future research to resolve current decision uncertainty has been well established (Claxton and Posnett, 1996). Likewise, the value of making different decisions across individuals has also been recognized (Coyle et al., 2003, Basu and Meltzer, 2007). However, the bridge between these two dimensions within a common framework for decision-making remains to be examined. This was one of the main motivations of this work. Furthermore, this thesis has identified that individual choices can be a potential additional source of value for the health system, raising the need for deeper understanding of choice as part of a decision process in a collectively funded health system.

The main purpose of this thesis was to contribute to the understanding of the value of heterogeneity in cost-effectiveness analysis for healthcare decision-making. The objective was addressed through four main steps. First, a literature review was conducted to describe the standard approaches in the assessment of heterogeneity in cost-effectiveness and the methods for realizing its value. Preferences and choices were also examined, in order to connect the study of heterogeneity with the policy agenda of the individualization of care. As a result, gaps and requirements for research were identified. Second, a novel analytical framework was produced to conduct cost-effectiveness subgroup analysis. This chapter contributed to the toolbox in CEA, introducing the consideration of several elements in a systematic manner, such as

criteria for identification, selection and analysis of subgroup specification. In addition, the value of resolving uncertainty was introduced as one dimension of the value of heterogeneity. Third, the feasibility of such a framework was assessed through its application to a real case study. This exercise offered an opportunity to implement a stochastic approach to estimations at individual level, as well as to connect these methods with previous contributions in the field. Finally, the thesis addressed issues on how individualized care can be incorporated in a collectively funded national health system. In particular, it examined the cases where choices made at individual level can be implemented and analysed the normative judgments that any decision-maker should take into account with attempts to implement individualized care. In the remaining part of this section, further details of each chapter are presented.

Chapter 2 defines total heterogeneity or total variability as the differences in outcomes between individuals. Observed heterogeneity, also commonly referred simply as heterogeneity, is defined as the proportion of variability that can be explained by observable and known characteristics. In particular, for the study of between-patient heterogeneity, which is the main focus of this thesis, it recognises four sources of heterogeneity that should be addressed: baseline risk; treatment effect; preferences; and costs. It points out the concerns raised in the assessment of effectiveness, where the significance of treatment effects between subgroups is constrained by recommendations of evidence based medicine, which is anchored in the principles of statistical inference (i.e. a p-value cut-off point usually set at 0.05). In the case of cost-effectiveness, adherence to the principles of expected utility theory makes these concerns less relevant. However, the operationalization of disaggregated decisions implies several transaction costs that impose constraints in their implementation.

Chapter 2 also describes and analyses the methods proposed in the literature to examine the value of considering heterogeneity in healthcare decision-making. Most of them rely on the net benefits framework to express value. In other words, the value of heterogeneity is expressed as the expected net health (e.g. life years or QALYs) that can be gained if decisions take into account subgroup or individual level information. It was observed that no previous literature considers the relevance of parameter uncertainty when addressing heterogeneity, which motivated important part of the work developed in Chapter 3.

One of the elements further explored is the study of heterogeneity in preferences and choices. Both concepts have been used very broadly, even in the field of CEA, and this offers an opportunity for a clearer definition. The definition of choice restricts the concept to decisions for a particular treatment. They can be ex-ante (treatment preference) or ex-post (revealed choice) depending whether a critical scrutiny process has been performed (see Chapter 2). The concept of preferences, instead, was circumscribed to the analysis whereby CEA adheres to the principles of utility and microeconomic theory. Preferences determine the value of health as a result of an exercise that forces people to make decisions between alternative health states, causing them to reveal their preferences. One of the interesting findings provided by the previous literature is that preferences and revealed choices are not well correlated. Therefore, the significant progress achieved and still ongoing in the field of patient preferences in the last few years needs to be complemented with further understanding of revealed choices. Two areas were identified: one relates to the role of choices in understanding the joint distribution of potential outcomes; and the second refers to the understanding of how revealed choices can be implemented without restrictions in a collectively funded health system. The second area motivated the work conducted in Chapter 5.

Chapter 2 is an organized synthesis of current knowledge. It is the result of the identification and understanding of the relevant elements that should be considered in the study of heterogeneity in CEA, which include the definition and clarification of concepts, terminology, sources and methods to explore heterogeneity. It is highlighted the need of considering and revealing heterogeneity, which is claimed in this thesis as another source of uncertainty of the decision problem. To the author's knowledge, there is no published article that has combined all these elements and presented them in the manner seen in Chapter 2. Therefore, Chapter 2 can be considered a contribution, in itself, to this field.

Chapter 3 introduces a systematic approach to conducting cost-effectiveness subgroup analysis. It starts by recognizing that the maximum value of heterogeneity is achieved when decisions are performed at individual level. This value can be expressed as the expected value of individualized care (EVIC) when it refers to the decisions based on the expected NHB. However, it argues that the effort to operationalize individual level

decisions implies transaction costs such that it is not feasible to implement in a centralized decision-making process. Therefore, it proposes that subgroup analysis becomes the best approach to take into account heterogeneity. This motivates the development of an analytical framework to guide the characterization of partial degrees of heterogeneity, addressing issues of identification and selection of subgroups according to the efficiency goals of a health system. Chapter 3 can be considered a substantial contribution, in the sense that it creates a systematic approach to subgroup analysis and adds important elements to the understanding of heterogeneity in CEA.

Three elements can be highlighted as important contributions of Chapter 3. They address issues about the selection of subgroups, the optimal number of subgroups and uncertainty as an additional source of value of heterogeneity. In terms of selection of subgroups, Chapter 3 suggests that different specifications should be compared in terms of the additional expected NHB they provide. As a consequence, an efficiency frontier for subgroup analysis can be represented from the set of points corresponding to the specifications with higher expected net benefits for each level of disaggregation. The additional value given by different decisions in different subgroups, in terms of net benefits, is termed static value of heterogeneity. Although the concept of static value has been presented before, for example, as EVIC (Basu and Meltzer, 2007) or stratification (Coyle et al., 2003), this chapter adds two important elements. First, an adequate characterization of heterogeneity must be undertaken examining all feasible specifications at different levels of disaggregation. Second, a frontier can be defined from the set of points that produce the greatest expected NHB at each level of disaggregation, which is expected to show diminishing marginal returns of health.

In terms of the optimal number of subgroups, Chapter 3 suggests that the level of disaggregation that can be implemented (i.e. the number of subgroups that can be considered) depends on the expected benefits and the transaction costs derived from the additional level of disaggregation. They include the effort to implement subgroup level decisions, additional resources to reveal heterogeneity and the effort to enforce the fulfilment of guidelines. The comparison between these two elements provides a judgment of efficiency that supports the definition of the optimal number of subgroups that should be considered.

A third element presented in Chapter 3 is the incorporation of uncertainty as a dimension of value of heterogeneity. It argues that the additional value of making different decisions for different patients is partially achieved by the information available, but there is also a second source of value, which is the potential gain of resolving uncertainty. This has been termed the dynamic value of heterogeneity and is expressed as the additional expected net benefits that might be achieved if conditional parameters (conditional for the subgroup) are estimated under perfect information. The dynamic value is a comparison between two adjacent points of disaggregation in terms of the expected NHB under perfect information. It is expected that this value increases as more subgroups are considered, which is consistent with the value of considering conditional parameters estimated with an infinite sample. In contrast, the value of an infinite sample to estimate the conditional parameters, expressed as EVPI for a particular specification, can increase or decrease depending on the informative capacity of each specification. It is expected that the EVPI related to the specifications on the efficiency frontier (which are, by definition, the most informative) decreases with higher levels of disaggregation. This is because of the significant increase of static value at the same time as dynamic value, leading to a lower difference between expected NHB under perfect and current information. However, the limited information and knowledge might affect the estimation of the efficiency frontier, i.e. a frontier formed by specifications with little informative capacity, which leads to higher EVPI.

Chapter 4 is a practical piece of work where methods proposed in Chapter 3 have been applied. It has used the data of a trial-based cost-effectiveness analysis, showing that a new invasive treatment for the management of patients with non-ST-elevation acute coronary syndrome was not cost-effective when compared with conservative management at a threshold of £20,000 per QALY. It shows how subgroup analysis can be performed to obtain estimations of static and dynamic value. The method implemented was based on the estimation of individual cost-effectiveness, i.e. expected costs and benefits for each treatment for each individual, which represents an estimation of the joint distribution of potential outcomes. From this point, where total static value is estimated (EVIC), the analysis aimed to characterize heterogeneity through subgroup analysis in order to account for most of this value in a feasible manner.

The analysis was carried out examining six relevant binary variables, which in combination led to 64 potential subgroups. One important finding that emerged from this exercise was that analyses based on a large number of subgroups do not necessarily imply complex guidelines and they are worth conducting. Another element is the progressive understanding that can be achieved through the analysis in terms of the informative capacity of the specifications used. For example, severe angina was informative at higher levels of disaggregation, but it did not provide any extra value at lower levels. This allows analysts to guide the selection of specifications for a further level of disaggregation, which can be limited to the combination of the most informative specifications. A third fundamental issue is that the decision model must be able to capture as much heterogeneity as possible and only one model structure should be used to estimate all outputs (i.e. average, subgroup and individual outputs). Finally, Chapter 4 has illustrated the relevance of estimating the dynamic value in decisions about future research. This includes the value of research to make different decisions in different subgroups in the future and the possibility of prioritizing research in one subgroup only.

Chapter 5 can be considered a step forward in the examination of the value of heterogeneity. Previous chapters have been constructed on the basis of a centralized process where decisions are made in order to maximize the health of the population. This implies that some effective but very expensive treatments can be restricted to some groups of patients, a restriction supported by the idea that the reallocation of those additional resources to other activities of the health system provides more health. This can, of course, be improved using subgroup analysis where restrictions would apply to potentially smaller groups of patients. It was also mentioned in previous chapters that decisions at individual level would produce the highest population health. The arguments provided in Chapter 5 recognize this idea and suggests that the only manner for implementing decision-making at individual level, consistent with the aim of a collectively funded health system, is allowing patients and doctors to make choices without restriction (individualized decisions). The challenge is then to identify when unrestricted choices can be considered a policy consistent with the objective of the health system, or at least acceptable according to certain social values judgments.

Chapter 5 has developed an analytical approach to address this issue. Basically, gains and losses for society are analysed as a comparison between decisions undertaken at central level (with potential restrictions) versus decentralized (without restrictions). The estimation of the joint distribution of potential outcomes is a basic input for conducting this analysis. Acting on the assumption that the difference between social planners and patients is that the latter will maximise health instead of net health, the method provides an estimate of the net health that the social planner would forgo were he to decide to implement unrestricted choices. This amount can be conceptualized as the value of unrestricted choices for treatment. Although some seed of this development can be found in the earlier work of Basu and Meltzer (2007), this approach can be considered new in that it uses the whole of the joint distribution for a comparative analysis across the cost-effectiveness plane. This compares gains and losses under alternative decision rules, i.e. decisions rules adopted by society and individuals⁷⁵. This chapter also extends the analysis to explore different decision rules and incorporates transaction costs that have not been discussed before in the literature (costs related to rejection and costs related to implementation of unrestricted choices). Thus, this approach offers a novel method to address the question of whether unrestricted choices can be implemented in a collectively funded NHS and therefore it can be considered a substantive contribution to the decision-making process in healthcare.

The method developed in Chapter 5 provides elements for a positive judgment about the implementation of individualization of care. Indeed, it is argued that the question of whether restricted or unrestricted choices should be implemented ought to consider an estimate of the health forgone associated to decentralization of these decisions. However, individualization is also about normative judgement, though not related to the question of restricted or unrestricted choices. Chapter 5 examines the fundamental normative judgements that should be considered for an agenda of individualization. It argues that this should consider the type of “individual” values that will be used to operationalize the maximand of the objective function. Paternalistic, altruistic and welfarist decision-makers are defined on the basis of three social value orientations: cooperativism; altruism; and individualism, respectively. They can be operationalized as

⁷⁵ The closer estimate to this approach is the EVIC without cost internalization (EVIC-noCI). However, although they are similar in estimating the value on the individual side, EVIC-noCI compares with the value obtained on average at societal level.

values of representative individuals of the community (paternalistic), a representative sample of patients (altruistic) or values obtained directly from individuals (welfarist). Because the selection between these alternatives does not respond to any empirical result, either of them can be adopted based on a purely normative value judgment.

6.2. Discussion and opportunities for further research

The contributions generated in this thesis rely on certain assumptions that have been pointed out throughout the previous chapters. This section summarizes the main assumptions and discusses the generalisability of these methods in scenarios where they do not hold. In addition, it highlights areas where further research would be beneficial.

The first condition assumed is that the healthcare system is collectively financed and all costs fall within the health budget. The notion of a collectively funded system implies transferences from wealthy people to poor people that makes it possible to offer a minimum “essential” or “statutory” package of healthcare services according to the needs of patients, something that has been an objective of many countries (Jost, 2005). It has been discussed that cost-effectiveness is also a useful tool to define this essential package in situations when the system allows people to have access to complementary⁷⁶ insurance (Smith, 2012). It is argued that the methods proposed in this thesis, as an extension of CEA, are also expected to apply under those scenarios.

Some caveats can be highlighted in terms of the applicability of classical CEA and the methods proposed here, in health systems where out-of-pocket co-payments are considered⁷⁷. It has been argued that direct out-of-pocket expenditure for healthcare introduces changes in the social welfare function that a social planner aims to maximise. Hoel (2007) presents the case where the social welfare function is determined by health and income related utilities, and people only have access to healthcare services covered in the essential package. In this case a classical cost-effectiveness decision rule applies. However, when people can buy services not covered by healthcare out-of-pocket, the welfare function includes the disutility due to the disease (whose treatment is not

⁷⁶ Complementary insurance refers to partial or total coverage of services that can or cannot be part of the essential package (Mossialos and Thompson, 2004).

⁷⁷ This refers to the case where each person has the alternative of buying -out-of-pocket- treatments that are not covered in the essential package.

covered) and the additional utility given by the treatments financed out-of-pocket in those individuals who are able to afford them. Hoel (2007) argued that these modifications of the social welfare function in presence of out-of-pocket expenditure lead to a different decision rule, which might limit the application of classical cost-effectiveness methods. As a consequence, the methods proposed here might also be limited.

Some countries have implemented substitute private insurance schemes (Sapelli and Vial, 2003). This type of insurance is an alternative to the public statutory package and it exempts, at least, part of the taxes or premiums associated to the essential package. In this case, a risk selection or cream skinning process arises since young healthy citizens opt out of the public system and move to the private (Roberts et al., 2008). Cost-effectiveness methods can still be useful in addressing resource allocation within the public sector. However, their use for decisions at the level of the whole population might be limited. The extent to which the methods developed in this thesis can be applied to healthcare systems with substitute private insurance schemes is a limitation of this work and an opportunity for future research.

Another element that has been highlighted throughout all the chapters is the relevance of a better understanding of the joint distribution of potential outcomes. Previous literature has emphasized that these estimates are relevant, since they provide the evidence needed for individualized care and personalized medicine (Basu, 2009). It has also been argued that individualized care is more efficient than subgroup analysis (Basu, 2011). This thesis has been developed in accordance with those principles but at the same time it highlights the role of the understanding of the joint distribution to support decisions conducted by social planners in socialized health systems.

In Chapter 5 the joint distribution is used to assess whether decentralized decisions can be implemented in the health system. It assumes that unrestricted choices will follow predictable rational judgments, based on individual characteristics as suggested previously (Basu, 2009, Basu, 2011). If the expected loss is considered too high, decisions based on subgroups provide the best approach. A limitation of the analysis carried out in this thesis was the estimation of the joint distribution of potential outcomes which was entirely based on observable characteristics. Future research

should improve this estimation, accounting for some part of the unobserved heterogeneity. This can be addressed using new statistical techniques to estimate individual treatment effects such as marginal treatment effects (Basu, 2011) or person-centred treatment effects (Basu, 2012). A better estimation of joint distribution is not only important to conduct the analysis presented in Chapter 5 but also to obtain a better estimation of EVIC and the proportion of the static value that can be achieved through a guideline based on subgroups.

Chapter 3 and its application in Chapter 4 have largely discussed the role of parameter uncertainty when making decisions about subgroup analysis. It has been explained that the dynamic value is in part the value of an infinite sample to estimate conditional parameters (value of information) and partially due to the value of estimating conditional parameters with an infinite sample (value of heterogeneity). Once a particular level of disaggregation has been defined, EVPI can be used to address the question of whether further research is worthwhile. This further research aims to obtain more precise estimates of parameters conditional to specific subgroups. It could be undertaken in the whole population, but it could also be developed in one subgroup, the one that offers the greatest value. In addition, in a further stage EVPPI could reveal that the uncertainty in that specific subgroup is mainly explained by one parameter. It was discussed in Chapter 3 that the estimation of EVPI presented in this thesis does not account for the exchangeability between subgroups and might underestimate the real value. This is a limitation of the methods presented here. However, if the aim is to compare EVPI across subgroups to prioritize further research, given that the exchangeability is symmetric, the additional value –per person- should be equal. Therefore, the difference is ultimately shown by the proportion of the subgroup in the population. If the bigger subgroup shows the higher EVPI, this should be selected for further research.

The analytical approach presented in chapter 5 addressed interesting issues around choice. The assumption of the base case was that social planners and individuals make choices based on the same maximand, except that for social interests the maximand is net of costs. An extended approach showed that the maximand of the individual can be different, though it has to be identifiable and quantifiable. Furthermore, the analysis assumed that either social planners or patients maximize the expected value according

to the principles of expected utility theory under risk neutrality. Although this assumption seems reasonable for the purposes of a social planner (Arrow and Lind, 1970), it might not be the case for individuals who value probabilities and consequences differentially⁷⁸ (Kahneman and Tversky, 1979, Tversky and Kahneman, 1981). Thus, while the decision rule applied for the social planner is entirely appropriate, the rule of maximization of an expected outcome might be a limitation of this analysis. Alternative specifications of the expected utility function to predict individual decisions should consider attitudes to risk and the individual reference point. This is a challenging field for future research.

Finally, the idea of predicting individual decisions from an empirical point of view has been largely addressed in the field of behavioural economics. Ariely (2009) suggests that people are predictably irrational when they make decisions. In other words, there are some identifiable elements in people's behaviour that allow us to predict their decisions. Those elements include some attributes that are known by a public decision planner but there are others that have not been considered in public decision-making. Discrete choice experiments, for example, have revealed that individuals value several dimensions and not only those related to health (Ryan et al., 2001). The magnitude of effect of those attributes on patients' choices for treatments might be used to generate predictions about individual choices. As a result, a patient-level decision rule based on a particular set of attributes might be generated for a particular condition and applied to the current framework. Further research should shed lights on the feasibility of using health and non-health related attributes in the prediction of individual decisions as well as their consistency with other decision rules.

6.3. Recommendations for analysts and decision-makers

The research performed in this thesis has produced new knowledge and has improved the understanding of several elements that are useful for the decision-making process in healthcare systems. This section presents a number of recommendations that can be considered by analysts and decision-makers in the future.

⁷⁸ This argument refers to Kahneman & Tversky's research on Prospect Theory. For details see footnote 73 in Chapter 5.

The healthcare system has an opportunity to increase the health gained as a consequence of their decisions about new interventions if they take heterogeneity into account. In the context of a centralized decision-making process, the analysis should characterize the total value of making different decisions for different patients. For this, analysts should construct decision models through which the heterogeneity between individuals can be captured, characterized and represented. This model provides the basis to estimate the joint distribution of potential outcomes, i.e. the distribution of individual expected costs and benefits for alternative treatments. Whenever possible this should be undertaken using individual patient data, as shown in this thesis. Further, the structure of the decision model should not vary for different subgroups, and the conditional parameters (conditional for each subgroup) should be obtained from the same source of information. Results of this analysis should show the total static value (EVIC), the estimated efficiency frontier for subgroups and the proportion of the total static value they account for.

The consideration of heterogeneity in CEA should not be limited to the number of subgroups that will be examined or the different specifications that will be examined at each level of disaggregation. The sources of heterogeneity such as baseline risk, treatment effect, costs and preferences should also be explicitly considered. In particular for preferences, there are a number of published studies that have incorporated heterogeneity in patient preferences. However, from an extra-welfarist perspective, CEA usually uses societal preferences to construct health outcomes. To the author's knowledge, there is only one theoretical paper raising the relevance of considering heterogeneity in societal preferences (Sculpher and Gafni, 2001). Thus, analysts should make an additional effort when studying heterogeneity in CEA, which is, to further understand the factors that can explain heterogeneity in societal preferences.

The estimates of uncertainty for different levels of disaggregation are other important results that should be presented. They include the dynamic value for alternative specifications and the EVPI for specific subgroups. Decision-makers should examine these results to make decisions about further research. They should focus on specifications that provide high dynamic value and, simultaneously, on the EVPI associated to that specification and in each subgroup. If the desire is to prioritize further

research in one subgroup only, decisions-makers should choose the subgroup where further research offers the highest expected value.

Recommendations based on subgroups might not be easily adopted by decision-makers. One of the reasons is the concern about the strength of evidence under subgroup analysis, which leads to a higher probability of spurious findings. A second element is the transaction costs of implementing a complex guideline, and a third factor relates to ethical or equity concerns. A judgement based on those elements might lead to the rejection of a consideration of heterogeneity. This thesis recommends that all these elements should be considered in the analysis: the former through an adequate characterization of uncertainty; the second through a formal consideration of transaction costs; and the third making an explicit consideration of the ethical constraints. The possibility of rejecting the consideration of heterogeneity in the decision-making process should be examined in the knowledge of the health that could have been gained through different decisions for different patients (e.g. EVIC).

In the case where a new technology is expected not to be cost-effective according to current information, decision-makers have an opportunity to assess the value of implementing unrestricted choices in the healthcare system. The magnitude of the health that is expected to be forgone because of a decentralized decision process can be estimated as in Chapter 5 and presented for public scrutiny. This is especially important in jurisdictions where individual choice is promoted as one of the central tenets of the health system. Because individualization of care also raises the discussion about whose values are being considered to make these decisions, analysts must be explicit in providing information as to whether a paternalistic, altruistic or welfarist point of view has been adopted.

6.4. *Final conclusion*

This thesis has contributed to the understanding of the value of heterogeneity in cost-effectiveness analysis in the following aspects. First, it incorporates the dimension of parameter uncertainty into the study of heterogeneity in cost-effectiveness as part of a systematic approach to conducting subgroup analysis. Second, it demonstrates how this theoretical analytical framework can be implemented in practice, using individual

patient data through the estimation of cost-effectiveness estimates at individual level. Third, it presents a broad conceptual framework aimed at helping social decision-makers address the policy agenda about individualization of healthcare. This discusses the normative elements that should be considered when decisions at individual-level require implementation. In addition, it develops an analytical approach to studying the conflict between decisions at societal and individual level. This has been conceptualized in this thesis as the value of individual choices for healthcare treatments. Furthermore, a set of opportunities for future research have been identified alongside this thesis. They can also be considered a contribution for the development of this area of knowledge in the future.

7. APPENDIX 1

The expected Value of Individualized Care (EVIC) and the Expected Value of Perfect Information (EVPI): A source of confusion

A1. Understanding the source of confusion

Basu and Meltzer introduced the expected value of individualized care (EVIC) for first time in 2007. It was presented as the difference between the value achieved under an individualized care model, where values of specific parameters (θ) are elicited directly from patients and used to make decisions at individual level, and the value attained under a paternalistic model, where decisions are based on the expected value of the distribution $p(\theta)$. This was expressed as:

A.1

$$EVIC = \int_{\theta \in \Theta} \left\{ \max_j NB(\theta) p(\theta) d\theta \right\} - \max_j \int_{\theta \in \Theta} NB(\theta) p(\theta) d\theta$$

The authors presented EVIC as a metric to represent the value of individual preferences. However, they pointed out that although θ initially represented a vector of patient preferences, EVIC might be estimated for any other parameter of interest in Θ . Furthermore, the expression in A.1 assumes that decisions made under either the individualized care or paternalistic models are both consistent with the criterion of maximising expected net benefits. Thus, because costs are considered in the decision they called this EVIC formulation “with cost-internalization”. Alternatively, EVIC without cost-internalization ($EVIC_{(NO-CI)}$) described the same concept but when a criterion based on maximising expected benefits was applied (disregarding costs). The expression used by the authors was:

$$EVIC_{(NO-CI)} = \int_{\theta \in \Theta} \left[\left\{ \max_j B_j(\theta) \right\} - (1/\lambda) C_{k'(\theta)} \right] p(\theta) d\theta - \left\{ \max_j \int_{\theta \in \Theta} B_j(\theta) p(\theta) d\theta \right\} - (1/\lambda) \int_{\theta \in \Theta} C_k(\theta) p(\theta) d\theta$$

A.2

Where B_j and C_j are benefits and costs, respectively, for the strategy j , and k is the strategy that maximizes expected health benefits, k' refers to k for each level of θ , in other words, the strategy that maximises expected health at individual level.

More recently, van Gestel et al. (2012) have conducted a study to explore the feasibility of the EVIC framework. Because of the apparent similarity between EVIC and EVPI they show a comparative table explaining that no- correlation exists between these two metrics: “The value of EVPI does not predict the value of EVIC, nor vice versa” (page 15). This seems to contradict Basu and Meltzer in a methods review paper published in 2010 (Basu and Meltzer, 2010). In this report, the authors provide an alternative expression either for EVPI and EVIC, expressing the former as:

$$EVPI = \iint \max\{\bar{y}_1, \bar{y}_0\} dG(\bar{Y}_1, \bar{Y}_0) - \max\{E(\bar{Y}_1), E(\bar{Y}_0)\} \quad \text{A.3}$$

This formula expresses the value of having an infinite sample that can resolve the uncertainty. Indeed, an infinite sample contains all patients; therefore, EVPI can be expressed in terms of individual patient outcomes estimated from such unsurpassable data. \bar{Y}_k (for k strategies; $k=0,1$) are the average health outcomes of a representative sample and $dG(\bar{Y}_1, \bar{Y}_0)$ is the joint distribution that contains the whole uncertainty around \bar{Y}_k . The pair (\bar{y}_1, \bar{y}_0) is one particular realization from the joint distribution $dG(\bar{Y}_1, \bar{Y}_0)$. Subsequently, the first part of the formula (affected by the integral) corresponds to the expected per patient outcome consequence of the decision under such perfect information (infinite sample) and the second part is the expected outcomes under current information.

Next, Basu and Meltzer (2010) offer an alternative expression for EVIC as an extension of EVPI:

$$EVIC = \sum_s \left[\iint \max\{\bar{y}_1^s, \bar{y}_0^s\} dG^s(\bar{Y}_1^s, \bar{Y}_0^s) - \max\{E(\bar{Y}_1), E(\bar{Y}_0)\} \right]$$

which can also be written in its extended form,

$$EVIC = \sum_s \left[\left(\int \max\{\bar{y}_1^s, \bar{y}_0^s\} dG^s(\bar{Y}_1^s, \bar{Y}_0^s) - \max\{E(\bar{Y}_1^s), E(\bar{Y}_0^s)\} \right) + (\max\{E(\bar{Y}_1^s), E(\bar{Y}_0^s)\} - \max\{E(\bar{Y}_1), E(\bar{Y}_0)\}) \right]$$

A.4

They explain that the first rounded parenthesis corresponds to the EVPI for each subgroup (s) and the second part represents the value of making specific decision for subgroups, what in this chapter has been called static value of heterogeneity. Thus, Basu and Meltzer (2010) formally expand their previous definition of the concept of EVIC. In addition, the idea that EVPI is part of EVIC was not denied in the authors' former paper (Basu and Meltzer, 2007) because in the example they provided in this report, EVPI was implicitly zero. In their example, EVIC was estimated to reflect the value of elicit patient preferences, which can be argued as perfect information.

According to the latter definition, EVIC is a function of EVPI and both metrics are perfectly correlated. This is the source of confusion with the statement provided by van Gestel et al. (2012). In this section it is argued that this confusion is more apparent than real and depends on the conceptualization of what perfect information represents and how it can be measured in practice.

A2. Resolving the confusion

One of the important conditions of the latter definition of EVIC provided by Basu and Meltzer (2010) is the assumption of an infinite sample for the subset "s" so that the true value of $(\bar{y}_1^s, \bar{y}_0^s)$ can be estimated. According to this condition, EVIC actually contains EVPI and reflects either the static or dynamic value of heterogeneity. However, because the idea of having an infinite sample is practically unachievable, it is not feasible to estimate this "super" EVIC proposed by Basu and Meltzer. Instead, we can only attempt to estimate a more modest EVIC from $(\hat{y}_1^s, \hat{y}_0^s)$ where (\hat{y}_k^s) is an estimate of (\bar{y}_k^s) .

Perfect information is usually estimated as the mean of the maximum NHB across many possible realizations simulated through a probabilistic sensitivity analysis. Each realization represents one potential scenario of the reality, on average. Thus, EVPI is estimated in practice from propagating the uncertainty around the average and not from

individual patients. As proposed in this chapter, EVPI can also be estimated for subgroups in order to represent the uncertainty of the parameters only, i.e. the uncertainty around $(\hat{y}_1^s, \hat{y}_0^s)$.

EVIC, as estimated by van Gestel et al. (2012) and also by Basu and Meltzer (2007) in their case-study example, results in a useful metric to represent only the value given by the heterogeneity of the population. If we consider that EVIC is not estimated with an infinite sample, and the (uncertain) average effect of the treatment and other covariates considered in the decision model are taken as unbiased estimates, then EVIC only reflects differences between patients. In addition, because the magnitude of EVPI depends on the size of the sample from which average effects were estimated and EVIC depends on the different characteristics of patients, it is very likely that both metrics are not correlated in practice.

8. APPENDIX 2:

Clinical effectiveness and cost-effectiveness of RITA-3 trial

A2.1. Clinical background

Coronary heart disease (CHD) is one of the main health problems in healthcare systems and a field in which most clinical research has been developed. As a consequence, it is one of the areas that has concentrated a vast amount of research and where most innovation has been produced. These technologies are potentially beneficial to the whole target population, though it is highly unlikely that every individual would receive the same magnitude of these benefits (mostly assumed to be the average benefit) because the population is heterogeneous in nature. This has led manufacturers, researchers and policy makers to be increasingly interested in exploring how the treatment effect and cost-effectiveness vary across subgroups of patients with CHD (Griffin et al., 2007).

Non-ST-elevation acute coronary syndrome (NSTEMI-ACS) is one type of CHD that is associated with a high incidence of complications and mortality (Anderson et al., 2007). The standard management of NSTEMI-ACS includes anti-platelet (aspirin), anti-thrombotic (heparin) and anti-ischemic therapy (beta-blockers). Angiography with intent to perform revascularization is usually recommended for high risk patients, to be delivered as promptly as possible. This therapeutic strategy has been termed “invasive strategy” which contrasts with the “conservative strategy” characterized by medical management only⁷⁹ (Anderson et al., 2007). In the UK, NICE has recommended an invasive strategy for patients with high and moderate risk (predicted 6-month mortality above 3% according to the GRACE risk score system) and conservative treatment for patients with low risk (NICE Guideline 94, 2010).

The third Randomised Intervention Trial of unstable Angina (RITA-3) is one of the trials undertaken to compare the efficacy of an intensive versus a conservative strategy in patients with NSTEMI-ACS. It was a multicentre study that recruited 1810 patients from

⁷⁹ Conservative strategy has also been called “selective invasive management” to express the fact that invasive interventions are also considered when patients develop rapid impairments in the following few hours after being admitted in the emergency.

45 centres in the UK (Fox et al., 2002). Of those recruited, 895 patients were randomized to the intervention and 915 were allocated to the conservative strategy. After a five-year follow-up, the study demonstrated that an intensive strategy was able to reduce the risk of the composite outcome of death and myocardial infarction (MI) (primary outcome: OR 0.78; 95% CI 0.61 to 0.99) (Fox et al., 2005).

The study also conducted a subgroup analysis defined on the basis of their predicted probability of experiencing the primary outcome (baseline risk score). A set of nine covariates were selected to specify a logistic regression predictive model of the primary outcome⁸⁰. Patients were categorized into four quartiles according to their baseline risk score. In addition, the highest risk subgroup was divided into two in order to obtain more detailed information about high risk patients. The treatment effect was reported for these final 5 subgroups, showing a significant effect concentration on high-risk patients (Table A2.1). These positive results for treatment effect were supported by a second study that showed improvement in Health Related Quality of Life (HRQoL) (Kim et al., 2005).

Table A2.1. Treatment effect by baseline risk subgroup in the trial RITA-3. Treatment effect estimated upon the combined primary outcome and expressed as Odds Ratio (OR) and 95% confidence interval (95% CI).

Risk group	OR (95% CI)
1 st quartile	0.96 (0.44 – 2.10)
2 nd quartile	1.10 (0.62 – 1.95)
3 rd quartile	0.8 (0.49 – 1.30)
4 th quartile, lower eight (4a)	0.76 (0.44 – 1.35)
4 th quartile, upper eight (4b)	0.44 (0.25 – 0.76)
Total	0.78 (0.61 – 0.99)
*Interaction test (p=0.0039)	

Note: Primary outcome is the combined rate of death and non-fatal myocardial infarction. Risk groups were defined on the basis of a (baseline risk) score estimated from a predictive logistic model for the primary outcome (1st quartile is the lowest risk and 4b is the highest risk). Source: Fox K.A., Poole-Wilson P., Clayton T.C. et al. Lancet 2005;366:914-20.

⁸⁰ The covariates were selected from a larger predefined set of variables collected at baseline using stepwise technique. They included demographic characteristics, co-morbidities, symptoms and signs recorded at baseline and electrocardiographic signs of coronary heart disease.

A2.2. Cost-Effectiveness

Henriksson et al. (2008) report in a subsequent study, a cost-effectiveness analysis performed alongside the RITA-3 trial. This is the only study available for this decision problem using the UK NHS perspective. The study was primarily based on the individual patient data of the trial which were used to populate a decision analytic model. Costs were expressed in pounds Sterling and outcomes in quality adjusted-life years (QALYs). QALYs were constructed using the set of values of the health states defined by the EQ-5D instrument for the UK population (Dolan et al., 1996). Both costs and QALYs were discounted using a rate of 3.5% per annum and uncertainty was characterized using probabilistic sensitivity analysis.

The decision model consisted of two parts: first, a short-term decision tree representing the index hospitalization, i.e. the period of time between the admission and the discharge. Second, a Markov model characterising the long-term period after hospitalization. The model is shown in Figure 4.1.

A set of four statistical equations were used to estimate the transition probabilities during the index hospitalization as well as between Markov states. Equation 1 was specified to estimate the probability of occurrence of the combined outcome (death or MI) during the index hospitalization. This outcome was categorised as death or non-fatal MI. The latter led the patient to the “Post-MI” state. This categorization was defined based on the conditional probability of dying given the combined outcome, estimated by Equation 4. Both models, 1 and 4, were estimated using logistic regression. Thus, the decision tree originated three states that defined the starting point of the Markov structure. These states were: no event, post MI and death, the latter being represented twice in the Markov model in order to distinguish between cardiovascular and non-cardiovascular etiologies.

For the Markov model, Equation 2 was specified to estimate the transition probability between “no event” and the combined outcome (death or MI) in a subsequent period. Each period had the same length of one year. A *Weibull* proportional hazards model was used to estimate the transition probability, which was calculated considering time-dependency for the following periods. An immediate categorization between

cardiovascular death and post-MI was done for those patients who achieved the combined outcome. Thus, the box (MI/CVD) in Figure 1 does not represent a health state but rather an intermediate virtual classification that has to be categorized in death or MI using Equation 4, similar to that in the decision tree. The hazard estimated from Equation 2 was increased every ten years by updating the age covariate. This variation was implemented using tunnel states for the first 5 years, after that, the hazard was assumed constant. Next, the transition between the Post-MI state and death or a new MI was estimated through Equation 3, also using a *Weibull* model. Similar to Equation 2, Equation 3 also considered decreasing risk with respect to the time from the previous MI. Thus, time dependency was incorporated for the first five years after MI. Once again, Equation 4 was used to categorize patients in either death or post-MI health state. Non-cardiovascular death was estimated using UK lifetime tables. Finally, two equations were specified to estimate costs, one at index admission and one at follow-up, and two other equations to estimate the Health Related Quality of Life (HRQoL) values, also called utilities or QALY weights, at randomization and follow-up. Equations 1 to 4 are presented in Table A2.2. Equations to estimate costs and health utilities are presented in Table A2.3. and Table A2.4. respectively.

Table A2.2. Estimated short-term and long-term risks of the composite endpoint of cardiovascular death or myocardial infarction and the predicted proportion of composite events being non-fatal.

Explanatory Variables	OR of composite endpoint of myocardial infarction or cardiovascular death during the index hospitalization (Eq 1)			HR of composite endpoint of myocardial infarction or cardiovascular death from hospital discharge until end of trial (Eq 2 and 3)			OR of composite endpoint being non-fatal (Eq 4)		
	OR	95% lower limit	95% upper limit	HR	95% lower limit	95% upper limit	OR	95% lower limit	95% upper limit
Age (for every 10 years over 60)	1.731	1.262	2.374	1.777	1.499	2.108	0.699	0.520	0.941
Diabetes				1.905	1.359	2.672			
Previous Myocardial Infarction				1.471	1.087	1.990	0.492	0.286	0.847
Smoker				1.651	1.207	2.258			
Pulse (for every 5 beats per minute)				1.062	1.012	1.114			
ST depression				1.423	1.067	1.913			
Angina (grade 3 or 4)	1.893	1.086	3.299	1.323	0.988	1.771			
Male				1.372	1.007	1.869			
Left bundle branch block				1.977	1.169	3.344			
Randomized to early interventional strategy	1.520	0.864	2.675	0.621	0.464	0.830			
Ancillary parameter				0.579	0.505	0.664			
Composite endpoint during index hospitalization							3.040	1.614	5.726

Table A2.3. Estimated costs during the index hospitalization and the first year after index hospitalization

Explanatory variables	Costs during the index hospitalization		Costs first year after the index hospitalization	
	Cost	SE	Cost	SE
Constant	1778	295	2735	248
Age (for every 10 years over 60)	878	153		
Previous MI			724	262
ST depression	1224	268		
Angina (grade 3 or 4)			1034	247
Male	1035	264	586	242
Randomized to early interventional strategy	5654	256	-1106	233
Non-fatal MI during the index hospitalization	6221	972		
Dying during the index hospitalization	7947	1229		
MI during year			5467	804

Table A2.4. Estimated HRQoL values (utilities) at randomization and changes in HRQoL values

Explanatory variables	Utility at randomization		Change in utility at follow-up	
	Utility	SE	Utility	SE
Constant	0.692	0.015		
Diabetes	-0.051	0.021		
Previous MI	-0.044	0.016	-0.010	0.016
ST depression	-0.066	0.015		
Angina (grade 3 or 4)	-0.074	0.015		
Male	0.073	0.015		
Randomized to conservative strategy (4 months follow-up)			0.044	0.013
Randomized to interventional strategy (4 months follow-up)			0.038	0.017
Randomized to conservative strategy (12 months follow-up)			0.038	0.008
Randomized to interventional strategy (12 months follow-up)			0.018	0.015
Myocardial infarction during year			-0.035	0.022

9. APPENDIX 3:

Cost-effectiveness subgroup analysis of the RITA-3 trial: details per subgroup

Table A3.1. Cost-effectiveness subgroup analysis of RITA-3: The case of two subgroups.

Specification	Subgroup	NHB current information (population)	NHB perfect information (population)	EVPI (population)
Diabetes	Weighted average	7.90 (4,403,199)	7.92 (4,412,177)	0.0161 (8,978)
	Diabetes (0.13)	5.78 (433,793)	5.79 (434,958)	0.0155 (1,165)
	No diabetes (0.86)	8.24 (3,969,406)	8.25 (3,977,219)	0.0162 (7,813)
Previous Myocardial Infarction (PMI)	Weighted average	7.90 (4,400,926)	7.92 (4,411,501)	0.018 (10,574)
	PMI (0.27)	6.07 (935,981)	6.10 (940,334)	0.028 (4,353)
	No PMI (0.72)	8.6 (3,464,945)	8.62 (3,471,167)	0.015 (6,221)
ST depression	Weighted average	7.90 (4,398,950)	7.92 (4,411,199)	0.022 (12,248)
	ST depression (0.36)	6.61 (1,343,732)	6.65 (1,350,399)	0.032 (6,666)
	No ST depression (0.63)	8.63 (3,055,218)	8.65 (3,060,800)	0.015 (5,582)
Left Bundle Branch Block (LBBB)	Weighted average	7.90 (4,399,530)	7.92 (4,410,850)	0.020 (11,320)
	LBBB (0.035)	5.83 (114,895)	7.97 (115,152)	0.013 (256)
	No LBBB (0.965)	5.84 (4,284,635)	7.99 (4,295,698)	0.020 (11,064)
Smoker	Weighted average	7.90 (4,399,896)	7.92 (4,411,278)	0.020 (11,382)
	Smoker (0.32)	8.63 (1,555,901)	8.66 (1,561,304)	0.029 (5,404)
	No Smoker (0.68)	7.55 (2,843,995)	7.57 (2,849,974)	0.015 (5,978)
Angina (degree 3 or 4)	Weighted average	7.89 (4,397,388)	7.92 (4,410,526)	0.023 (13,138)
	Severe Angina (0.35)	6.21 (1,243,495)	6.24 (1,247,858)	0.021 (4,363)
	No Severe Angina (0.65)	6.24 (3,153,893)	8.86 (3,162,668)	0.024 (8,775)

Note: NHB for current and perfect information per person are presented with two decimals places. EVPI per person are presented with three decimals places. Differences between population values for the same number per person are explained by differences in decimals places that are not shown.

Table A3.2. Cost-effectiveness subgroup analysis of RITA-3: The case of four subgroups.

Specification	Subgroup	NHB (current information)	NHB (perfect information)	EVPI
Diabetes and PMI	Weighted average	7.91 (4,403,714)	7.92 (4,413,271)	0.017 (9,556)
	Diabetes and PMI (0.05)	4.69 (126,915)	4.70 (127,214)	0.011 (300)
	Diabetes and No PMI (0.08)	6.39 (306,873)	6.40 (307,802)	0.019 (928)
	No diabetes and PMI (0.23)	6.36 (809,099)	6.41 (813,570)	0.035 (4,470)
	No diabetes and No PMI (0.64)	8.91 (3,160,827)	8.92 (3,164,685)	0.010 (3,858)
Diabetes and ST depression	Weighted average	7.90 (4,403,199)	7.92 (4,412,897)	0.017 (9,698)
	Diabetes and ST depression (0.05)	4.61 (140,737)	4.62 (141,113)	0.012 (376)
	Diabetes and No ST depression (0.08)	6.57 (293,285)	6.59 (294,122)	0.035 (837)
	No diabetes and ST depression (0.31)	6.98 (1,204,659)	7.01 (1,209,941)	0.019 (5,282)
	No diabetes and No ST depression (0.56)	8.94 (2,764,518)	8.95 (2,767,721)	0.010 (3,203)
Previous MI and ST depression	Weighted average	7.90 (4,400,929)	7.92 (4,412,673)	0.021 (11,746)
	previous MI & ST dep (0.08)	4.66 (212,209)	4.68 (213,086)	0.019 (876)
	previous MI & No ST dep (0.2)	6.66 (723,771)	6.69 (727,424)	0.033 (3,652)
	No previous MI & ST dep (0.28)	7.19 (1,132,595)	7.19 (1,137,661)	0.032 (5,066)
	No previous MI & no ST dep (0.44)	9.51 (2,332,350)	9.51 (2,334,502)	0.008 (2,151)
ST depression and LBBB	Weighted average	7.90 (4,400,214)	7.90 (4,411,976)	0.021 (11,762)
	ST dep and LBBB (0.01)	4.01 (22,203)	4.01 (22,246)	0.007 (44)
	ST dep and No LBBB (0.35)	6.69 (1,321,529)	6.72 (1,328,339)	0.034 (6,810)
	No ST dep and LBBB (0.03)	6.55 (92,692)	6.56 (92,925)	0.016 (233)
	No ST dep and No LBBB (0.61)	8.72 (2,963,789)	8.74 (2,968,465)	0.013(4,675)
Diabetes and Left Bundle Brach Block (LBBB)	Weighted average	7.91 (4,404,566)	7.92 (4,412,841)	0.014 (8,275)
	Diabetes and LBBB (0.006)	3.94 (13,356)	3.95 (13,370)	0.004 (15)
	Diabetes and No LBBB (0.129)	5.86 (420,437)	5.88 (421,641)	0.016 (1,204)
	No diabetes and LBBB (0.029)	6.22 (101,539)	6.24 (101,828)	0.017 (289)
	No diabetes and No LBBB (0.836)	8.31 (3,869,233)	8.32 (3,876,002)	0.014 (6,767)

Table A3.2. (cont) Cost-effectiveness subgroup analysis of RITA-3: The case of four subgroups.

Specification	Subgroup	NHB (current information)	NHB (perfect information)	EVPI
Previous MI and LBBB	Weighted average	7.90 (4,402,052)	7.92 (4,412,169)	0.018 (10,117)
	Previous MI and LBBB (0.01)	3.78 (26,788)	3.79 (26,855)	0.009 (67)
	Previous MI and No LBBB (0.27)	6.18 (909,193)	6.21 (913,623)	0.030 (4,430)
	No Previous MI and LBBB (0.02)	6.98 (88,108)	7.00 (88,314)	0.016 (207)
	No Previous MI and No LBBB (0.7)	8.66 (3,377,963)	8.67 (3,383,377)	0.013 (5,413)
Previous MI and Smoking	Weighted average	7.90 (4,400,926)	7.92 (4,412,664)	0.021 (11,738)
	Previous MI and smoking (0.01)	6.45 (271,865)	6.46 (272,505)	0.015 (640)
	Previous MI and No smoking (0.27)	5.93 (664,116)	5.96 (668,150)	0.036 (4,035)
	No Previous MI and smoking (0.02)	9.30 (1,284,792)	9.33 (1,289,241)	0.032 (4,449)
	No Previous MI and No smoking (0.7)	8.24 (2,180,153)	8.25 (2,182,768)	0.009 (2,614)
LBBB and Smoking	Weighted average	7.90 (4,401,288)	7.92 (4,412,090)	0.019 (10,802)
	LBBB and smoking (0.007)	6.76 (27,054)	6.77 (27,075)	0.005 (22)
	LBBB and No smoking (0.028)	5.59 (87,842)	5.61 (88,104)	0.016 (262)
	No LBBB and smoking (0.317)	8.67 (1,528,847)	8.70 (1,534,381)	0.031 (5,534)
	No LBBB and No smoking (0.648)	7.64 (2,757,545)	7.65 (2,762,530)	0.013 (4,984)
Diabetes and Smoking	Weighted average	7.91 (4,403,455)	7.92 (4,413,444)	0.017 (9,988)
	Diabetes and smoking (0.03)	5.78 (106,816)	5.79 (106,995)	0.009 (178)
	Diabetes and No smoking (0.1)	5.77 (326,977)	5.79 (328,022)	0.018 (1,045)
	No diabetes and smoking (0.29)	8.95 (1,449,084)	8.99 (1,454,691)	0.034 (5,606)
	No diabetes and No smoking (0.58)	7.87 (2,520,578)	7.87 (2,523,736)	0.009 (3,159)
ST depression and Smoking	Weighted average	7.90 (4,400,939)	7.92 (4,412,249)	0.020 (11,310)
	ST dep and smoking (0.11)	7.17 (452,414)	7.19 (453,589)	0.018 (1,175)
	ST dep and No smoking (0.25)	6.38 (893,307)	6.41 (897,311)	0.028 (4,004)
	No ST dep and smoking (0.21)	9.42 (1,104,529)	9.45 (1,108,093)	0.030 (3,564)
	No ST dep and No smoking (0.43)	8.24 (1,950,689)	8.25 (1,953,256)	0.010 (2,567)

Table A3.3. Cost-effectiveness subgroup analysis of RITA-3: The case of eight subgroups.

Specification	NHB (current information)	NHB (perfect information)	EVPI	Subgroups				ICER
				Diabetes	Previous MI	Smoking	p	
Diabetes & Previous MI & Smoker	4,405,519	4,414,587	0.016 (9,068)					
			0.005 (29)	1	1	1	0.009	13,193
			0.013 (292)	1	1	0	0.039	15,423
			0.012(167)	1	0	1	0.024	14,792
			0.023 (810)	1	0	0	0.062	17,266
			0.018 (667)	0	1	1	0.067	16,569
			0.026 (2,396)	0	1	0	0.162	21,515
			0.026 (3,285)	0	0	1	0.224	21,389
			0.006 (1,418)	0	0	0	0.413	28,623
Diabetes & LBBB &ST depression	4,404,566	4,413,820	0.016 (9,254)	Diabetes	LBBB	ST depress		
			0.002 (4)	1	1	1	0.003	10,780
			0.006 (12)	1	1	0	0.003	12,349
			0.013 (397)	1	0	1	0.052	15,374
			0.019 (850)	1	0	0	0.077	16,677
			0.013 (55)	0	1	1	0.007	14,692
			0.019 (238)	0	1	0	0.022	15,867
			0.029 (4,984)	0	0	1	0.303	20,889
			0.009 (2,710)	0	0	0	0.533	26,341
PMI & LBBB & ST depression	4,402,052	4,413,387	0.020 (11,335)	Previous MI	LBBB	ST depress		
			0.004 (7)	1	1	1	0.003	12,866
			0.011 (64)	1	1	0	0.009	14,351
			0.020 (896)	1	0	1	0.079	17,095
			0.036 (3,719)	1	0	0	0.186	19,867
			0.009 (39)	0	1	1	0.007	13,441
			0.021 (188)	0	1	0	0.015	16,122
			0.030 (4,609)	0	0	1	0.276	20,854
			0.007 (1,808)	0	0	0	0.425	27,527

Note: NHB and EVPI are presented in QALYs (net of costs) for a threshold value of £20,000/QALY. The ICER is presented in sterling pounds per QALY gained.

Table A3.1.(cont) Cost-effectiveness subgroup analysis of RITA-3: The case of eight subgroups.

Specification	NHB (current information)	NHB (perfect information)	EVPI	Subgroups			p	ICER
				Diabetes	Previous MI	Smoking		
Diabetes & Previous MI LBBB	4,404,566	4,413,892	0.016 (9,326)					
			0.002 (3)	1	1	1	0.002	10,735
			0.012 (316)	1	1	0	0.046	15,204
			0.006 (13)	1	0	1	0.004	12,149
			0.020 (949)	1	0	0	0.082	16,783
			0.013 (81)	0	1	1	0.011	14,984
			0.021 (220)	0	0	1	0.019	15,960
			0.036 (4,391)	0	1	0	0.218	20,070
			0.009 (3,350)	0	0	0	0.618	26,071
			Diabetes & Smoking & LBBB	4,404,566	4,414,198	0.016 (9,632)	Diabetes	Smoking
0.002 (2)	1	1				1	0.001	9,970
0.011 (195)	1	1				0	0.032	14,609
0.006 (15)	1	0				1	0.004	12,382
0.019 (1,063)	1	0				0	0.097	16,684
0.009 (29)	0	1				1	0.006	13,475
0.020 (272)	0	0				1	0.024	16,111
0.034 (5,494)	0	1				0	0.285	20,080
0.008 (2,558)	0	0				0	0.551	26,798

Note: NHB and EVPI are presented in QALYs (net of costs) for a threshold value of £20,000/QALY. The ICER is presented in sterling pounds per QALY gained.

Table A3.2. Cost-effectiveness subgroup analysis of RITA-3: The case of sixteen subgroups.

Specification	NHB (current information)	NHB (perfect information)	EVPI	Subgroups				ICER
				Diabetes	LBBB	Previous MI	Smoking	
Diabetes & Previous MI & Smoking & LBBB	4,406,788	4,415,294	8,505	1	1	1	1	10,393
				1	1	1	0	11,127
				1	1	0	1	9,135
				1	1	0	0	12,965
				1	0	1	1	13,708
				1	0	1	0	15,608
				1	0	0	1	15,050
				1	0	0	0	17,611
				0	1	1	1	13,480
				0	1	1	0	15,182
				0	1	0	1	13,473
				0	1	0	0	16,929
				0	0	1	1	16,639
				0	0	1	0	22,184
				0	0	0	1	21,666
				0	0	0	0	29,421

Table A3.3. Cost-effectiveness subgroup analysis of RITA-3: The case of forty-nine subgroups.

Specification	Subgroup						ICER
	Diabetes	Previous MI	LBBB	smoker	ST depress	angina	
All covariates	1	1	0	1	0	0	11,483
	1	0	0	0	0	0	18,296
	1	1	0	0	0	0	14,690
	1	0	1	0	0	0	12,166
NHB current information 4,408,359	1	1	1	0	0	0	11,127
	1	0	0	1	0	0	15,258
	1	0	0	0	1	0	14,834
	1	1	0	0	1	0	13,652
	1	0	1	0	1	0	11,812
NHB perfect information 4,441,842	1	0	0	1	1	0	12,512
	1	1	0	1	1	0	9,135
	1	1	1	1	1	0	10,393
	1	1	0	0	0	1	16,842
	1	0	1	0	0	1	18,063

Table A3.3. (cont) Cost-effectiveness subgroup analysis of RITA-3: The case of forty-nine subgroups.

Specification	Subgroup						ICER
	Diabetes	Previous MI	LBBB	smoker	ST depress	angina	
All covariates	1	0	0	1	0	1	17,516
	1	1	0	1	0	1	14,583
	1	0	0	0	1	1	17,570
	1	0	1	0	1	1	13,109
	1	1	0	0	1	1	17,518
	1	0	0	1	1	1	15,400
	1	1	0	1	1	1	14,779
	1	0	0	0	0	1	21,743
	0	1	1	1	1	0	12,305
	0	0	1	0	1	0	12,789
	0	1	1	0	0	0	12,946
	0	0	1	1	0	0	13,152
	0	0	1	1	0	1	13,966
	0	1	0	1	1	0	14,136
	0	1	1	1	1	1	14,650
	0	1	1	0	1	1	15,645
	0	1	0	1	1	1	16,127
	0	1	1	0	0	1	16,376
	0	0	1	0	1	1	16,426
	0	1	0	1	0	0	16,892
	0	1	0	0	1	0	17,596
	0	0	0	1	1	0	17,753
	0	1	0	1	0	1	18,574
	0	0	1	0	0	1	19,022
	0	0	1	0	0	0	19,181
	0	0	0	1	1	1	19,845
	0	1	0	0	1	1	21,575
	0	1	0	0	0	0	22,180
	0	1	0	0	0	1	23,355
	0	0	0	1	0	0	23,817
	0	0	0	0	0	1	24,037
	0	0	0	1	0	1	24,828
	0	1	0	1	0	1	26,066
	0	0	0	0	0	0	32,644
	0	0	0	0	0	1	37,831

10. APPENDIX 4:

Estimation of the Expected Value of Individualized Care (EVIC) and parameter-specific EVIC

The total EVIC can be estimated as the difference between the mean of the maximum expected individual NHB and the maximum between the average iNHB. Table A4.1 shows the estimation of total population EVIC for the case study presented in Chapter 4. It shows the iNHB for each strategy and the maximum between them. Population estimates were calculated multiplying the average per person value times the expected present value of the population (556,723 patients) for the next 10 years.

Table A4.1. Estimation of the total Expected Value of Individualized Care (EVIC)

Patient	NHB interventional strategy	NHB conservative strategy	Max iNHB
1	9.23	9.37	9.37
2	12.62	12.51	12.62
:	:	:	:
:	:	:	:
1809	11.40	11.48	11.48
1810	7.87	7.89	7.89
Average (per person)	7.87	7.89	7.92
Population	4,384,267	4,397,267	4,411,737
Max _j (E ₀ iNHB)	4,397,267		
E ₀ (max iNHB)	4,411,737		
EVIC	4,411,737– 4,397,267= 14,349 net-QALYs		

According to Basu and Meltzer (2007), the parameter-specific EVIC corresponds to the difference between the total EVIC and the EVIC for the remaining parameters other than θ_i .

$$EVIC_{\theta_i} = EVIC - \int_{x \in \theta_i} p_i EVIC(\theta_i = x) d\theta_i$$

In this expression, p_i corresponds to the marginal probability distribution of the parameter of interest (θ_i). The integral represents EVIC estimated when θ_i is

fixed at any constant value x and then averaged across all possible values that θ_i can take. When the difference between total EVIC and the second expression of the equation is calculated, the result expresses the expected value of having individualized information about the parameter θ_i . The above expression has been simplified by van Gestel et al. as:

$$EVIC_{\theta_i} = EVIC - \left(mean_{\theta_i} \left[mean_{\theta_c} \left\{ \max_j iNHB_j(\theta_c/\theta_i) \right\} \right] - mean_{\theta_i} \left[\max_j \{ mean_{\theta_c} iNHB_j(\theta_c/\theta_i) \} \right] \right)$$

With θ_i being the parameter of interest and θ_c a vector of the remaining parameters. Following this idea, Table A4.2 shows the parameter-specific EVIC for diabetes.

Table A4.2. Estimation of the diabetes specific Expected Value of Individualized Care ($EVIC_{diabetes}$)

max expected in patients with no diabetics	4,587,882
max expected in patients with diabetics	3,217,893
average max expected	4,403,199
expected max in patients with no diabetics	4,597,096
expected max in patients with diabetics	3,222,158
average expected max	4,411,745
EVIC	14,349
EVIC for diabetes	14,349 – (4,411,745-4,403,199) 14,349 – 8,546 5,803

The diabetes-specific EVIC (5,803 net-QALYs) is almost the same as the static value of heterogeneity in this example (5.811 net-QALYs). It is argued that these values are the same and the small differences are explained by approximations along the estimation process. A demonstration is presented in Box A4.3.

Box A4.3. Demonstration of the equivalence between EVIC for specific parameters and static value of heterogeneity

Static value is estimated from individual NHB as in Chapter 3 (equation (3.8)):

$$\text{Static value} = \sum_{s=1}^S w_s (\max_j E_{\theta} iNHB_s) - \max_j E_{\theta} iNHB$$

According to van Gestel et al. (2012) EVIC for parameter i (θ_i) that defines s subgroups (where $s=1,2,\dots,S$) can be expressed as

$$EVIC_{\theta_i} = EVIC - \left[\sum_{s=1}^S w_s (E_{\theta} \max_j iNHB_s) - \sum_{s=1}^S w_s (\max_j E_{\theta} iNHB_s) \right]$$

$$EVIC_{\theta_i} = [E_{\theta} \max_j iNHB - \max_j E_{\theta} iNHB] - \left[\sum_{s=1}^S w_s (E_{\theta} \max_j iNHB_s) - \sum_{s=1}^S w_s (\max_j E_{\theta} iNHB_s) \right]$$

However, the expected value across the maximum individual net benefits is the same as the weighted average of the averages of the maximum individual net health benefits for each subgroup

$$E_{\theta} \max_j iNHB = \sum_{s=1}^S w_s (E_{\theta} \max_j iNHB_s)$$

then, EVIC for parameters can be expressed as:

$$EVIC_{\theta_i} = \left[\sum_{s=1}^S w_s (E_{\theta} \max_j iNHB_s) - \max_j E_{\theta} iNHB \right] - \left[\sum_{s=1}^S w_s (E_{\theta} \max_j iNHB_s) - \sum_{s=1}^S w_s (\max_j E_{\theta} iNHB_s) \right]$$

$$EVIC_{\theta_i} = \sum_{s=1}^S w_s (\max_j E_{\theta} iNHB_s) - \max_j E_{\theta} iNHB$$

which is equal to the static value of heterogeneity.

11. APPENDIX 5:

Stylized example of the extended method

This appendix presents a stylized example to illustrate an extension of the method presented in section 5.2.2. The aim is to examine when at individual level doctors and patients make choices according to an argument that differ from the social planner's maximand. It is assumed that the social planner applies a decision rule in order to maximize QALYs net of costs. The base case assumes that patients and doctors choose the treatment that provides the maximum expected QALYs. This does not restrict the analysis to QALYs, the only constraint is that the metric has to be the same for patients and social planners. The extension of the method relaxes this restriction.

Table A5.1 shows the individual level data of six subjects. Incremental costs, incremental benefits and incremental net health benefits are presented for each individual. It also presents the area of the cost-effectiveness plane where each subject is located. In addition, it shows a hypothetical set of numbers that represents incremental gains in a second outcome. This alternative outcome could have been estimated through the decision model used for the cost-effectiveness study, though this is not a requirement for the analysis. Indeed, the only relevant piece of information is whether the individual makes the same choice as he/she would have chosen according to QALYs. Figure A5.1 shows the distribution of the points in the cost-effectiveness plane.

The ICER for this population is £71,666 per QALY (£11,466 / 0.16 QALYs), which means that the new intervention should not be considered cost-effective for a reference threshold of £20,000 per QALY. Table A5. 2 shows the results of the base case analysis and the extended analysis. The base case analysis shows that there is a positive value in decentralizing decisions (for simplicity let us assume that transaction costs are zero). It is estimated, for a hypothetical relevant population of 10,000 people, in 200 QALYs or £4,000,000 (Table A5.2). The extended analysis shows that were a second decision rule considered, one individual would change his decision. Subject 1 in Table A5.1, who lies on the NW quadrant, would decide in favour of the new intervention. The individual net benefits of this subject were not a matter of conflict under the previous decision rule, because it was assumed that he/she would reject the treatment. However, given the

change of the decision, his/her net benefits now represent a loss for the society. Under the new scenario, the value of decentralizing a decision is negative. For the same relevant population the expected loss of decentralized decisions (assuming zero transaction costs) is 1,800 QALYs or £36,000,000 (Table A5.2).

Table A5.1. Illustrative example of a hypothetical population of six individuals and their joint distribution of potential outcomes. Individual net health benefits are estimated for a cost-effectiveness threshold of £20,000 per QALY.

	Incremental Costs (£)	Incremental QALYs	Individual Net Health Benefits	Area	Incremental alternative outcome	Is the decision the same as using QALYs?
1	7,400	-0.98	-1.35	NW	-0.61	Yes
2	13,800	-0.56	-1.25	NW	0.54	No
3	16,600	0.57	-0.26	NNE	0.66	Yes
4	13,600	0.34	-0.34	NNE	0.47	Yes
5	10,600	0.96	0.43	NSE	0.86	Yes
6	6,800	0.63	0.29	NSE	0.79	Yes
average	11,466	0.16	-0.41	NNE		

Figure A5.1. Graphical representation of the joint distribution (hypothetical example). The dots represent individuals. The line in the NE quadrant represents the cost-effectiveness threshold of 20,000/QALY gained.

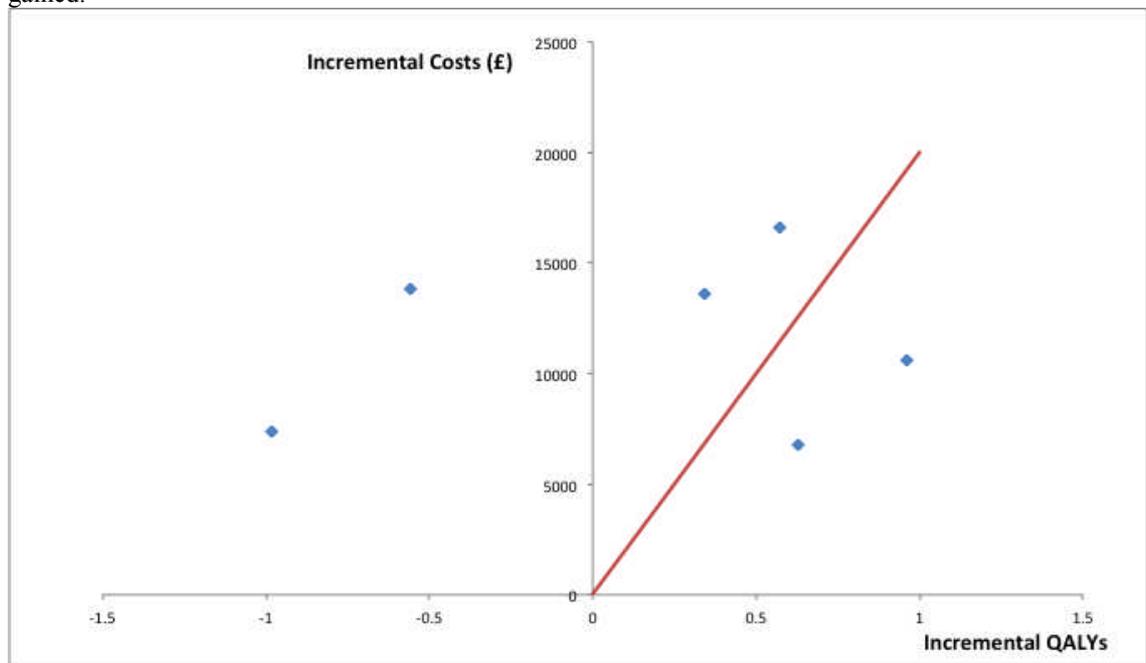


Table A5.2. Results of the base case and extended analysis.

	Area	Average iNHB	Density	Expected iNHB
Base case analysis	NNE (Subjects 3 & 4)	$\frac{1}{2} \times (0.26 + 0.34) = 0.3$	1/3	0.1
	NSE (Subjects 5 & 6)	$\frac{1}{2} \times (0.43 + 0.29) = 0.36$	1/3	0.12
<p>Result:</p> <p>$E + F - D \geq 0$ (assuming $C_{sdm} - C_r = 0$)</p> <p><i>In this case: $F = 0$; $E = NSE$; $D = NNE$</i></p> <p>$NSE - NNE = 0.12 - 0.1 = 0.02$ net QALYs For a relevant population of 10,000 people is 200 QALYs or £4,000,000</p>				
Extended Analysis	Changes from NW (NW*)(Subject 2)	1.25	1/6	0.208
	<p>Result:</p> <p>$E_i + F_i + C_j - A_j - B_j - D_i \geq 0$ (assuming $C_{sdm} - C_r = 0$)</p> <p><i>In this case: $C_j = B_j = F_i = 0$; $A_j = NW^*$; $E_i = NSE$; $D_i = NNE$</i></p> <p>$NSE - NNE - NW^* = 0.12 - 0.1 - 0.208 = -0.18$ net QALYs For a relevant population of 10,000 people is 1,800 QALYs or £36,000,000</p>			

12. APPENDIX 6:

Societal versus patient values

Table A6.1 shows four scenarios comparing ICERs for societal and patient values. Treatment 1 corresponds to standard care and is assumed to maintain the patient in the same health state, which is valued as 0.4 and 0.6 by the society and patients respectively. Treatment 2 provides improvement of quality and/or quantity of life. For simplicity, costs remain the same (£1,000 and £50,000).

Scenario 1 is similar to the example in Gold et al. (1996). Here treatment 2 is assumed to see patients recover to full health. Consequently, the ICER based on societal values is lower. Scenario 2 presents the case where treatment 2 gives a partial improvement in quality (0.4 versus 0.7 for social values and 0.6 versus 0.9 for patient values). It is assumed that there is proportionality between social and patient values for each treatment, i.e. in both cases the difference between treatments is 0.3. In this case, the ICER is the same for both cases. Scenario 3 shows a similar situation but assumes that the value a patient attaches to a better health state is higher (0.4 versus 0.7 for social values and 0.6 versus 0.95 for patient values). In this scenario the ICER based on patient values is lower. Scenario 4 presents the case where there are not only quality improvements but also quantity increases (10 years with treatment 1 and 15 years with treatment 2). Assuming the same weights as in Scenario 2, the ICER in this case is also lower with patient values. Finally, Scenario 5 illustrates the case where there are improvement in length and quality of life, but the quality increment is higher for patients than social values. In this case, the ICER estimated with patient values is even lower than Scenario 4.

Table A6.1. Cost-effectiveness estimates using societal and patient values for 5 alternative scenarios

	Cost 1	Cost 2	Years 1	Years 2	Weight 1	Weight 2	QALY 1	QALY 2	ICER
SOCIETAL									
Scenario 1	1000	50000	10	10	0.4	1	4	10	8,166
Scenario 2	1000	50000	10	10	0.4	0.7	4	7	16,333
Scenario 3	1000	50000	10	10	0.4	0.7	4	7	16,333
Scenario 4	1000	50000	10	15	0.4	0.7	4	10.5	7,538
Scenario 5	1000	50000	10	15	0.4	0.7	4	10.5	7,538
PATIENTS									
Scenario 1	1000	50000	10	10	0.6	1	6	10	12,250
Scenario 2	1000	50000	10	10	0.6	0.9	6	9	16,333
Scenario 3	1000	50000	10	10	0.6	0.95	6	9.5	14,000
Scenario 4	1000	50000	10	15	0.6	0.9	6	13.5	6,533
Scenario 5	1000	50000	10	15	0.6	0.95	6	14.25	5,939

Note 1: Scenario 1 assumes no improvement in quantity of life but full recovery of quality of life.

Note 2: Scenario 2 assumes no improvement in quantity of life but partial and proportional recovery in quality of life

Note 3: Scenario 3 assumes no improvement in quantity of life and partial recovery in quality of life. In this case the gains in quality are overrated by patients.

Note 4: Scenario 4 assumes improvement in quantity of life as well as proportional gains in quality of life.

Note 5: Scenario 5 assumes improvement in quantity of life and greater incremental gains in quality of life in patients than society.

13. List of Abbreviations

CBA	Cost Benefit Analysis
CEA	Cost-Effectiveness Analysis
CER	Comparative Effectiveness Research
EVPI	Expected Value of Perfect Information
EVPII	Expected Value of Perfect Information for parameters
EVSI	Expected Value of Sampling Information
ENBS	Expected Net Benefit of Sampling
EVIC	Expected Value of Individualized Care
HRQoL	Health Related Quality of Life
HTA	Health Technology Assessment
IC	Individualized Care
ICER	Incremental Cost-Effectiveness Ratio
iICER	Individual Incremental Cost-Effectiveness Ratio
INB	Incremental Net Benefits
NB	Net Benefits
iNB	individual Net Benefits
NHB	Net Health Benefits

NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NST-ACS	Non-ST Acute Coronary Syndrome
PM	Personalized Medicine
PSA	Probabilistic Sensitivity Analysis
QALY	Quality Adjusted Life Year
SDM	Shared Decision-making
TINB	Total Incremental Net Benefits
UK	United Kingdom
VoI	Value of Information
VoH	Value of Heterogeneity

14. REFERENCES

- Ades, A. E., Lu, G. & Claxton, K. 2004. Expected value of sample information calculations in medical decision modeling. *Medical Decision-making*, 24, 207-27.
- Aickin, M. & Gensler, H. 1996. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *American Journal of Public Health*, 86, 726-8.
- Ambrose, J. A. & Barua, R. S. 2004. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *Journal of the American College of Cardiology*, 43, 1731-7.
- Anderson, J., Adams, C., Antman, E., Bridges, C., Califf, R., Casey, D., Chavey, W., F., F., Hochman, J., Levin, T., Lincoff, M., Peterson, E., Theroux, P., Wenger, N. K. & Wright, S. 2007. ACC/AHA 2007 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction. *Circulation*, 166.
- Antman, E. M., Cohen, M., Bernink, P. J. L. M., McCabe, C. H., Horacek, T., Papuchis, G., Mautner, B., Corbalan, R., Radley, D. & Braunwald, E. 2000. The TIMI Risk Score for Unstable Angina/Non-ST Elevation MI: A Method for Prognostication and Therapeutic Decision-making. *JAMA*, 284, 835-842.
- Ariely, D. 2009. *Predictably Irrational*, London, Harper Collins.
- Arrow, K. 1963. Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53, 941-973.
- Arrow, K. & Lind, R. 1970. Uncertainty and the Evaluation of Public Investments Decisions. *American Economic Reviews*, 60, 364-378.
- Assmann, S., Pocock, S., Enos, L. & Kasten, L. 2000. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*, 355, 1064 - 1069.
- Aune, E., Endresen, K., Roislien, J., Hjelmestaeth, J. & Otterstad, J. E. 2010. The effect of tobacco smoking and treatment strategy on the one-year mortality of patients with acute non-ST-segment elevation myocardial infarction. *BMC Cardiovascular Disorders*, 10, 59.
- Baker, J. N., Hinds, P. S., Spunt, S. L., Barfield, R. C., Allen, C., Powell, B. C., Anderson, L. H. & Kane, J. R. Integration of palliative care practices into the ongoing care of children with cancer: individualized care planning and coordination. *Pediatric Clinics of North America*, 55, 223-50.
- Baron, R. M. & Kenny, D. A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bartus, T. 2005. Estimation of marginal effects using margeff. *The Stata Journal*, 5, 309-329.
- Basu, A. 2009. Individualization at the heart of comparative effectiveness research: the time for i-CER has come. *Medical Decision-making*, 29, NP9-NP11.
- Basu, A. 2011. Economics of individualization in comparative effectiveness research and a basis for a patient-centered healthcare. *Journal of Health Economics*, 30, 549-59.
- Basu, A. 2012. Estimating person-centered treatment (PET) effects using instrumental variables. *NBER working papers*, Working paper 18056, 1-36.

- Basu, A., Heckman, J. J., Navarro-Lozano, S. & Urzua, S. 2007. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics*, 16, 1133-57.
- Basu, A. & Meltzer, D. 2007. Value of Information on Preference Heterogeneity and Individualized Care. *Med Decis Making*, 27, 112-127.
- Basu, A. & Meltzer, D. 2010. Quantitative methods for valuing comparative effectiveness information. *Biopharmaceutical report*, 17, 2-10.
- Bender, R. & Lange, S. 2001. Adjusting for multiple testing--when and how? *Journal of Clinical Epidemiology*, 54, 343-9.
- Bigger, J. T., Jr., Fleiss, J. L., Kleiger, R., Miller, J. P. & Rolnitzky, L. M. 1984. The relationships among ventricular arrhythmias, left ventricular dysfunction, and mortality in the 2 years after myocardial infarction. *Circulation*, 69, 250-8.
- Bland, J. M. & Altman, D. G. 1995. Multiple significance tests: the Bonferroni method. *BMJ*, 310, 170.
- Boardman, A., Greenberg, D., Vining, A. & Weimer, D. 2006. *Cost-Benefit Analysis: Concepts and Practice*, New Jersey, Pearson Prentice Hall.
- Bojke, L., Claxton, K., Sculpher, M. & Palmer, S. 2009a. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*.
- Bojke, L., Claxton, K., Sculpher, M. & Palmer, S. 2009b. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health*, 12, 739-49.
- Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M. & Tsuchiya, A. 2005. Should patients have a greater role in valuing health states? *Applied Health Economics & Health Policy*, 4, 201-8.
- Brazier, J., Ratcliffe, J., Salomon, J. & Tsuchiya, A. 2007. *Measuring and valuing health benefits for economic evaluation*, Oxford, Oxford University Press.
- Briggs, A. 2001. Handling Uncertainty in Economic Evaluation and presenting the results. In: DRUMMOND, M. & MCGUIRE, A. (eds.) *Economic Evaluation in Healthcare: merging theory with practice*. New York: Oxford University Press.
- Briggs, A., Mihaylova, B., Sculpher, M., Hall, A., Wolstenholme, J., Simoons, M., Deckers, J., Ferrari, R., Remme, W. J., Bertrand, M., Fox, K. & On Behalf of the Europa Trial Investigators 2007. Cost-effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study. *Heart*, 93, 1081-1086.
- Briggs, A. & Sculpher, M. 1998. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*, 13, 397-409.
- Briggs, A., Sculpher, M. J. & Claxton, K. (eds.) 2006. *Decision Modelling for Health Economic Evaluation*, Gosport, Hampshire: Oxford University Press.
- Briggs, A. H. 2000. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*, 17, 479-500.
- British Social Attitudes Survey. 2009. Available: <http://www.esds.ac.uk/government/bsa/>.
- Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A. & Peters, T. J. 2004. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology*, 57, 229-36.
- Brookes, S. T., Whitley, E., Peters, T. J., Mulheran, P. A., Egger, M. & Davey Smith, G. 2001. Subgroup analyses in randomised controlled trials: quantifying the

- risks of false-positives and false-negatives. *Health Technology Assessment (Winchester, England)*, 5, 1-56.
- Brouwer, W. B., Culyer, A. J., Van Exel, N. J. & Rutten, F. F. 2008. Welfarism vs. extra-welfarism. *Journal of Health Economics*, 27, 325-38.
- Brouwer, W. B. & Koopmanschap, M. A. 2000. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of Health Economics*, 19, 439-59.
- Charalambous, A., Katajisto, J., Valimaki, M., Leino-Kilpi, H. & Suhonen, R. 2010. Individualised care and the professional practice environment: nurses' perceptions. *International Nursing Review*, 57, 500-7.
- Charles, C., Gafni, A. & Whelan, T. 1997. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Social Science & Medicine*, 44, 681-92.
- Claxton, K. 1999a. Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals. *Health Economics*, 8, 269-74.
- Claxton, K. 1999b. The irrelevance of inference: a decision-making approach to the stochastic evaluation of healthcare technologies. *Journal of Health Economics*, 18, 341-64.
- Claxton, K., Palmer, S., Longworth, L., Bojke, L., Griffin, S., Mckenna, C., Soares, M., Spackman, E. & Youn, J. 2011a. Uncertainty, evidence and irrecoverable costs: Informing approval, pricing and research decisions for health technologies. *CHE research paper 69* [Online]. Available: <http://www.york.ac.uk/che/news/che-research-paper-69/>.
- Claxton, K., Paulden, M., Gravelle, H., Brouwer, W. & Culyer, A. J. 2011b. Discounting and decision-making in the economic evaluation of health-care technologies. *Health Econ*, 20, 2-15.
- Claxton, K. & Posnett, J. 1996. An economic approach to clinical trial design and research priority-setting. *Health Economics*, 5, 513-24.
- Claxton, K., Sculpher, M. J. & Culyer, A. J. 2007. Mark versus Luke? Appropriate Methods for the Evaluation of Public Interventions. *Centre for Health Economics Discussion Paper*, 31. University of York, York.
- Claxton, K. P. & Sculpher, M. J. 2006. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *PharmacoEconomics*, 24, 1055-68.
- Coast, J., Smith, R. D. & Lorgelly, P. 2008. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. *Social Science & Medicine*, 67, 1190-8.
- Conti, R., Veenstra, D. L., Armstrong, K., Lesko, L. J. & Grosse, S. D. 2010. Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost-effectiveness, and future research priorities. *Medical Decision-making*, 30, 328-40.
- Cook, D., GebSKI, V. & Keech, A. 2004. Subgroup analysis in clinical trials. *Med J Aust*, 180, 289 - 291.
- Cookson, R. 2005. QALYs and the capability approach. *Health Economics*, 14, 817-29.
- Coyle, D., Buxton, M. J. & O'brien, B. J. 2003. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Economics*, 12, 421-7.
- Creager, M. A., Luscher, T. F., Cosentino, F. & Beckman, J. A. 2003. Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: Part I. *Circulation*, 108, 1527-32.

- Cui, L., Hung, H., Wang, S. & Tsong, Y. 2002a. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat*, 12, 347 - 358.
- Cui, L., Hung, H. M. J., Wang, S. J. & Tsong, Y. 2002b. ISSUES RELATED TO SUBGROUP ANALYSIS IN CLINICAL TRIALS. *Journal of Biopharmaceutical Statistics*, 12, 347 - 358.
- Culyer, A. J. 1989. The normative economics of healthcare finance and provision. *Oxford Review of Economic Policy*, 5, 34-56
- Culyer, A. J. 1990. The internal market: An acceptable means to a desirable end. *Centre for Health Economics Discussion Paper*, 67, University of York, York.
- Department of Health. 2009. Implementation of the right to choice and information set out in the NHS Constitution. Available: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_093458.pdf.
- Department of Health 2010. Equity and excellence: Liberating the NHS. *In: DEPARTMENT OF HEALTH (ed.) White paper*. London.
- Diderholm, E., Andren, B., Frostfeldt, G., Genberg, M., Jernberg, T., Lagerqvist, B., Lindahl, B. & Wallentin, L., 2nd 2002. ST depression in ECG at entry indicates severe coronary lesions and large benefits of an early invasive treatment strategy in unstable coronary artery disease; the FRISC II ECG substudy. The Fast Revascularisation during InStability in Coronary artery disease. *European Heart Journal*, 23, 41-9.
- Dolan, P., Gudex, C., Kind, P. & Williams, A. 1996. The time trade-off method: results from a general population study. *Health Economics*, 5, 141-54.
- Dolan, P. & Kahneman, D. 2008. Interpretations of Utility and their implications for the valuation of health. *The economic journal*, 118, 215-234.
- Dolan, P. & Olsen, J. A. 2002. *Distributing healthcare*, Oxford, Oxford University Press.
- Donaldson, C. 1999. Valuing the benefits of publicly-provided healthcare: does 'ability to pay' preclude the use of 'willingness to pay'? *Social Science & Medicine*, 49, 551-63.
- Dowie, J. 1998. Towards the equitably efficient and transparently decidable use of public funds in the deep blue millennium. *Health Economics*, 7, 93-103.
- Drummond, M. 2012. Twenty years of using economic evaluations for reimbursement decisions: What have we achieved? *Centre for Health Economics Discussion Paper*, 75. University of York. York.
- Drummond, M., Barbieri, M., Cook, J., Glick, H. A., Lis, J., Malik, F., Reed, S. D., Rutten, F., Sculpher, M. & Severens, J. 2009. Transferability of Economic Evaluations Across Jurisdictions: ISPOR Good Research Practices Task Force Report. *Value in Health*, 12, 409-418.
- Drummond, M., Sculpher, M., Torrence, G., O'Brien, B. & Stoddart, G. (eds.) 2005. *Methods for the Economic Evaluation of Healthcare Programmes* Oxford University Press.
- Eddy, D. 2009. Health technology assessment and evidence-based medicine: what are we talking about? *Value in Health*, 12, S6-S7.
- Eddy, D. M. 1991. Clinical decision-making: from theory to practice. The individual vs society. Is there a conflict? *JAMA : the journal of the American Medical Association*, 265, 1449-1450.
- Epstein, D. M., Sculpher, M. J., Clayton, T. C., Henderson, R. A., Pocock, S. J., Buxton, M. J. & Fox, K. A. A. 2008. Costs of an early intervention versus a

- conservative strategy in acute coronary syndrome. *International Journal of Cardiology*, 127, 240-246.
- Eunetha Work Package 8. *Handbook on Health Technology Assessment capacity building*, Barcelona (Spain), Catalan Agency for Health Technology Assessment and Research. Catalan Health Service. Department of Health Autonomous Government of Catalonia; 2008.
- Evans, W. E. & Relling, M. V. 2004. Moving towards individualized medicine with pharmacogenomics. *Nature*, 429, 464-8.
- Feinstein, A. 1998. The problem of cogent subgroups: a clinicostatistical tragedy. *J Clin Epidemiol*, 51, 297 - 299.
- Fenwick, E., Claxton, K. & Sculpher, M. 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics*, 10, 779-87.
- Fenwick, E., Claxton, K. & Sculpher, M. 2008. The value of implementation and the value of information: combined and uneven development. *Medical Decision-making*, 28, 21-32.
- Fox, K. A. A., Poole-Wilson, P., Clayton, T. C., Henderson, R. A., Shaw, T. R. D., Wheatley, D. J., Knight, R. & Pocock, S. J. 2005. 5-year outcome of an interventional strategy in non-ST-elevation acute coronary syndrome: the British Heart Foundation RITA 3 randomised trial. *The Lancet*, 366, 914-920.
- Fox, K. A. A., Poole-Wilson, P. A., Henderson, R. A., Clayton, T. C., Chamberlain, D. A., Shaw, T. R. D., Wheatley, D. J. & Pocock, S. J. 2002. Interventional versus conservative treatment for patients with unstable angina or non-ST-elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. *The Lancet*, 360, 743-751.
- Gabler, N., Duan, N., Liao, D., Elmore, J., Ganiats, T. & Kravitz, R. 2009. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials*, 10, 43.
- Gail, M. & Simon, R. 1985. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*, 41, 361-72.
- Gandjour, A. 2010. Theoretical foundation of patient v. population preferences in calculating QALYs. *Medical Decision-making*, 30, E57-63.
- Gerrish, K. 2000. Individualized care: its conceptualization and practice within a multiethnic society. *Journal of Advanced Nursing*, 32, 91-9.
- Ginnelly, L., Claxton, K., Sculpher, M. J. & Golder, S. 2005. Using value of information analysis to inform publicly funded research priorities. *Applied Health Economics & Health Policy*, 4, 37-46.
- Gold, M. R., Siegel, J. E., Russell, L. B. & Weinstein, M. C. 1996. *Cost-effectiveness in Health and Medicine*, Oxford University Press.
- Goodman, N. 1947. The problem of counterfactual conditionals. *Journal of Philosophy*, 44, 113-128.
- Goodman, S. N. 1999a. Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine*, 130, 995-1004.
- Goodman, S. N. 1999b. Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, 130, 1005-13.
- Gray, A., Rivero-Arias, O., Leal, J., Dakin, H. & Ramos-Goñi, J. M. 2012. How important is parameter uncertainty around the UK EQ-5D-3L value set when estimating treatment effects? Presented at the 3rd joint CES/HESG Meeting, Marseille, France [Online]. Available: <http://www.ces-asso.org/sites/default/files/Gray%20et%20al%20HESG%20submitted%201.pdf>.

- Greenberg, B. & Weimer, V. 1996. *Cost Benefit Analysis: Concepts and Practice* New Jersey, Prentice-Hall, Inc.
- Griffin, B., Rimmerman, C. & Topol, J. 2007. *The Cleveland Clinic Cardiology Board Review*, Philadelphia, Lippincott Williams & Wilkins.
- Griffin, S., Claxton, K. & Sculpher, M. 2008. Decision analysis for resource allocation in healthcare. *Journal of Health Services & Research Policy*, 13 Suppl 3, 23-30.
- Griffin, S., Welton, N. J. & Claxton, K. 2010. Exploring the Research Decision Space: The Expected Value of Information for Sequential Research Designs. *Med Decis Making*, 30, 155-162.
- Griffin, S. C., Claxton, K. P., Palmer, S. J. & Sculpher, M. J. 2011. Dangerous omissions: the consequences of ignoring decision uncertainty. *Health Economics*, 20, 212-24.
- Groot Koerkamp, B., Weinstein, M. C., Stijnen, T., Heijnenbrok-Kal, M. H. & Hunink, M. G. 2010. Uncertainty and patient heterogeneity in medical decision models. *Medical Decision-making*, 30, 194-205.
- Grossman, M. 1972. On the concept of health capital and the demand for health. *Journal of Political Economy*, 80, 223-255.
- Grouin, J.-M., Coste, M. & Lewis, J. 2005. Subgroup Analyses in Randomized Clinical Trials: Statistical and Regulatory Issues. *Journal of Biopharmaceutical Statistics*, 15, 869 - 882.
- Gulland, A. 2011. Welcome to the century of the patient: The Salzburg statment on shared decision-making. *British Medical Journal*, 342, 792-796.
- Hamburg, M. & Collins, F. 2010. The path to personalized medicine. *New England Journal of Medicine*, 363, 301-304.
- Happ, M. B. 2010. Individualized care for frail older adults: challenges for healthcare reform in acute and critical care. *Geriatric Nursing*, 31, 63-5.
- Harsanyi, J. C. 1953. Cardinal utility in welfare economics and in the theory of risk taking. *J Political Economy*, 61, 434-35.
- Harsanyi, J. C. 1955. Cardinal welfare, individualistics ethics, and interpersonal comparisons of utility. *J Political Economy*, 63, 309-21.
- Hauck, W. W., Anderson, S. & Marcus, S. M. 1998. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials*, 19, 249-56.
- Heckman, J. J., Urzua, S. & Vytlacil, E. J. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics*, 88, 389-432.
- Heckman, J. J. & Vytlacil, E. J. 2007. Econometric evaluation of social programs, Part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In: HECKMAN, J. J. & LEAMER, E. (eds.) *Handbook of Econometrics*. Oxford: Elsevier.
- Henderson, R. A., Pocock, S. J., Sharp, S. J., Nanchahal, K., Sculpher, M. J., Buxton, M. J. & Hampton, J. R. 1998. Long-term results of RITA-1 trial: clinical and cost comparisons of coronary angioplasty and coronary-artery bypass grafting. Randomised Intervention Treatment of Angina. *Lancet*, 352, 1419-25.
- Henriksson, M., Epstein, D. M., Palmer, S. J., Sculpher, M. J., Clayton, T. C., Pocock, S. J., Henderson, R. A., Buxton, M. J. & Fox, K. A. A. 2008. The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. *Heart*, 94, 717-723.

- Hernandez, A., Boersma, E., Murray, G., Habbema, J. & Steyerberg, E. 2006. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J*, 151, 257 - 264.
- Hernandez, A. V., Steyerberg, E. W. & Habbema, J. D. 2004. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology*, 57, 454-60.
- Hewitt, R. E. 2011. Biobanking: the foundation of personalized medicine. *Current Opinion in Oncology*, 23, 112-9.
- Hoel, M. 2007. What should (public) health insurance cover? *Journal of Health Economics*, 26, 251-62.
- Holland, P. 1986. Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-960.
- Hoomans, T., Fenwick, E. A. L., Palmer, S. & Claxton, K. 2009. Value of Information and Value of Implementation: Application of an Analytic Framework to Inform Resource Allocation Decisions in Metastatic Hormone-Refractory Prostate Cancer. *Value in Health*, 12, 315-324.
- Hurley, J. 2000. An overview of the normative economics of the healthcare sector. In: CULYER, A. J. & NEWHOUSE, J. (eds.) *Handbook of Health Economics*. Amsterdam: Elsevier Science.
- Hyde, T. A., French, J. K., Wong, C. K., Straznick, I. T., Whitlock, R. M. & White, H. D. 1999. Four-year survival of patients with acute coronary syndromes without ST-segment elevation and prognostic significance of 0.5-mm ST-segment depression. *American Journal of Cardiology*, 84, 379-85.
- Iezzoni, L. I., Shwartz, M., Moskowitz, M. A., Ash, A. S., Sawitz, E. & Burnside, S. 1990. Illness severity and costs of admissions at teaching and nonteaching hospitals. *JAMA : the journal of the American Medical Association*, 264, 1426-31.
- Inahta. 2012. *International Network of Agencies for Health Technology Assessment* [Online]. Available: <http://www.inahta.net/> [Accessed May 2012].
- Ioannidis, J. P. 2005. Why most published research findings are false. *PLoS Medicine / Public Library of Science*, 2, e124.
- Ispor. 2012. *Pharmacoeconomic guidelines around the world* [Online]. Lawrenceville, NJ: International Society for Pharmacoeconomics and Outcomes Research. [Accessed May 2012].
- Iyengar, S. & Lepper, M. 1999. Rethinking the value of choice: Cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, 76, 349-366.
- Jackson, C. H., Bojke, L., Thompson, S. G., Claxton, K. & Sharples, L. D. 2011. A framework for addressing structural uncertainty in decision models. *Medical Decision-making*, 31, 662-74.
- Jackson, C. H., Thompson, S. G. & Sharples, L. D. 2009. Accounting for uncertainty in health economic decision models by using model averaging. *J R Statist Soc*, 172, 383-404.
- Jost, T. 2005. *Healthcare Coverage Determinations*, Maidenhead, Open University Press.
- Kahneman, D. & Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.
- Kahneman, D. & Tversky, A. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of risk and uncertainty*, 5.

- Kent, D. M. & Lindenauer, P. K. 2010. Aggregating and disaggregating patients in clinical trials and their subgroup analyses. *Annals of Internal Medicine*, 153, 51-2.
- Kim, J., Henderson, R. A., Pocock, S. J., Clayton, T., Sculpher, M. J., Fox, K. A. & Investigators, R.-T. 2005. Health-related quality of life after interventional or conservative strategy in patients with unstable angina or non-ST-segment elevation myocardial infarction: one-year results of the third Randomized Intervention Trial of unstable Angina (RITA-3). *Journal of the American College of Cardiology*, 45, 221-8.
- Kind, P. 2005. Valuing health outcomes: ten questions for the insomniac health economist. In: SMITH, P., GINNELY, L. & SCULPHER, M. (eds.) *Health Policy and Economics*. Berkshire: Open University Press.
- Klose, T. 1999. The contingent valuation method in healthcare. *Health policy*, 47, 97-123.
- Kraemer, H., Stice, E., Kazdin, A., Offord, D. & Kupfer, D. 2001. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *Am J Psychiatry*, 158, 848 - 856.
- Kraemer, H., Wilson, G., Fairburn, C. & Agras, W. 2002. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry*, 59, 877 - 883.
- Kravitz, R., Duan, N. & Braslow, J. 2004. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q*, 82, 661 - 687.
- Krone, R. J., Greenberg, H., Dwyer, E. M., Jr., Kleiger, R. E. & Boden, W. E. 1993. Long-term prognostic significance of ST segment depression during acute myocardial infarction. The Multicenter Diltiazem Postinfarction Trial Research Group. *Journal of the American College of Cardiology*, 22, 361-7.
- Kuntz, K. M. & Goldie, S. J. 2002. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. *Medical Decision-making*, 22, 218-27.
- Lagakos, S. 2006. The challenge of subgroup analyses - reporting without distorting. *New England Journal of Medicine*, 354, 1667 - 1669.
- Lee, G. & Ishii-Kuntz, M. 1987. Social interaction, loneliness, and emotional well-being among elderly. *Research on Aging*, 9, 459-482.
- Lincoff, A. M. 2003. Important triad in cardiovascular medicine: diabetes, coronary intervention, and platelet glycoprotein IIb/IIIa receptor blockade. *Circulation*, 107, 1556-9.
- Lingsma, H., Rozenbeek, B., Steyerberg, E. & Investigators, I. 2010. Covariate adjustment increases statistical power in randomized controlled trials. *Journal of Clinical Epidemiology*, 63, 1391; author reply 1392-3.
- Lloyd-Jones, D. M., Leip, E. P., Larson, M. G., D'agostino, R. B., Beiser, A., Wilson, P. W., Wolf, P. A. & Levy, D. 2006. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation*, 113, 791-8.
- Loewenstein, G. & Ubel, P. 2008. Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, 92, 1975-1810.
- Lord, S. J., GebSKI, V. J. & Keech, A. C. 2004. Multiple analyses in clinical trials: sound science or data dredging? *Med J Aust*, 181, 452-4.
- Luscher, T. F., Creager, M. A., Beckman, J. A. & Cosentino, F. 2003. Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: Part II. *Circulation*, 108, 1655-61.

- Malandrino, N. & Smith, R. J. 2011. Personalized medicine in diabetes. *Clinical Chemistry*, 57, 231-40.
- Manca, A., Rice, N., Sculpher, M. J. & Briggs, A. H. 2005. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics*, 14, 471-485.
- Mccabe, C., Claxton, K. & Culyer, A. J. 2008. The NICE cost-effectiveness threshold: what it is and what that means. *PharmacoEconomics*, 26, 733-44.
- Mcclosey, D. & Ziliak, S. 1996. The standard error of regressions. *Journal of Economic Literature*, XXXIV, 97-114.
- Mechanic, R., Coleman, K. & Dobson, A. 1998. Teaching hospital costs: implications for academic missions in a competitive market. *JAMA : the journal of the American Medical Association*, 280, 1015-9.
- Messick, D. M. & McClintock, C. G. 1968. Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, 4, 1-25.
- Meyer, U. A. 1994. Pharmacogenetics: the slow, the rapid, and the ultrarapid. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 1983-4.
- Mishra, L. 2011. Healthcare reform: how personalized medicine could help bundling of care for liver diseases. *Hepatology*, 53, 379-81.
- Moffett, J. K., Torgerson, D., Bell-Syer, S., Jackson, D., Llewlyn-Phillips, H., Farrin, A. & Barber, J. 1999. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. *BMJ*, 319, 279-83.
- Mossialos, E. & Thompson, S. (eds.) 2004. *Voluntary Health Insurance in the European Union*, Brussels: World Health Organization.
- Muthappan, P., Rogers, A. & Eagle, K. 2012. Using the GRACE risk scores in everyday practice. *American Journal of Medicine*, 125, 144-7.
- National Health and Hospitals Reform Commision 2009. A Healthier Future for all Australians - Final Report June 2009. Commonwealth of Australia.
- Nease, R. F., Jr. & Owens, D. K. 1994. A method for estimating the cost-effectiveness of incorporating patient preferences into practice guidelines. *Medical Decision-making*, 14, 382-92.
- Nice 2005. Social value judgments: Principles for the development of NICE guidelines. 2nd ed ed.
- Nice. 2008. *Guide to the methods of technology appraisal* [Online]. National Institute for Health and Clinical Excellence. Available: www.nice.org.uk [Accessed May 2012].
- Nice Guideline 94. 2010. *Unstable Angina and NSTEMI: the early management of unstable angina and non-ST-segment-elevation myocardial infarction* [Online]. National Institute for Health and Clinical Excellence. [Accessed December 2011].
- Nixon, R. M. & Thompson, S. G. 2005. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Economics*, 14, 1217-1229.
- Nyman, I., Areskog, M., Areskog, N. H., Swahn, E. & Wallentin, L. 1993. Very early risk stratification by electrocardiogram at rest in men with suspected unstable coronary heart disease. The RISC Study Group. *Journal of Internal Medicine*, 234, 293-301.
- O'brien, B. J. & Sculpher, M. J. 2000. Building uncertainty into cost-effectiveness rankings: portfolio risk-return tradeoffs and implications for decision rules. *Medical Care*, 38, 460-8.

- O'hagan, A., McCabe, C., Akehurst, R., Brennan, A., Briggs, A., Claxton, K., Fenwick, E., Fryback, D., Sculpher, M., Spiegelhalter, D. & Willan, A. 2005. Incorporation of uncertainty in health economic modelling studies. *Pharmacoeconomics*, 23, 529-36.
- Offerman, T., Sonnemans, J. & Schram, A. 1996. Value orientations, expectations and voluntary contributions in public goods. *The economic journal*, 106, 817-845.
- Office for National Statistics. Available: <http://www.ons.gov.uk/ons/index.html> [Accessed January 2012].
- Ogden, J., Daniells, E. & Barnett, J. 2008. The value of choice: development of a new measurement tool. *British Journal of General Practice*, 58, 614-618.
- Oshana, M. 2003. How much should we value autonomy? In: PAUL, E. F., MILLER, F. D. & PAUL, J. (eds.) *Autonomy*. Cambridge: Cambridge University Press.
- Oxman, A. & Guyatt, G. 1992. A consumer's guide to subgroup analyses. *Annals of Internal Medicine*, 116, 78 - 84.
- Parker, A. & Naylor, C. 2000. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. *Am Heart J*, 139, 952 - 961.
- Pauly, M. V. 1995. Valuing health benefits in monetary terms. In: SLOAN, F. A. (ed.) *Valuing Healthcare. Costs, Benefits and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge: Cambridge University Press.
- Peppercorn, J. M., Smith, T. J., Helft, P. R., Debono, D. J., Berry, S. R., Wollins, D. S., Hayes, D. M., Von Roenn, J. H., Schnipper, L. E. & American Society of Clinical, O. 2011. American society of clinical oncology statement: toward individualized care for patients with advanced cancer. *Journal of Clinical Oncology*, 29, 755-60.
- Phelps, C. E. & Mushlin, A. I. 1988. Focusing technology assessment using medical decision theory. *Medical Decision-making*, 8, 279-89.
- Philips, Z., Claxton, K. & Palmer, S. 2008. The half-life of truth: What are appropriate time horizons for research decisions? *Med Decis Making*, 28, 287-299.
- Phillips, C. V. 2001. The economics of 'more research is needed'. *International Journal of Epidemiology*, 30, 771-6.
- Pliskin, J., Shepard, D. & Weinstein, M. C. 1980. Utility functions for life years and health status. *Operations Research*, 28, 206-224.
- Pocock, S., Assmann, S., Enos, L. & Kasten, L. 2002. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine*, 21, 2917 - 2930.
- Preference Collaborative Review Group 2008. Patients' preferences within randomized trials: systematic review and patient level meta-analysis. *BMJ*, 337, a1864.
- Radwin, L. E. & Alster, K. 2002. Individualized nursing care: an empirically generated definition. *International Nursing Review*, 49, 54-63.
- Rawls, J. 1971. *A theory of justice*, Cambridge, Harvard University Press.
- Roberts, M. J., Hsiao, W., Berman, P. & Reich, M. 2008. *Getting Health Reform Right*, New York, Oxford University Press.
- Rothman, K. J. 1990. No adjustments are needed for multiple comparisons. *Epidemiology*, 1, 43-6.
- Rothwell, P. 2005. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*, 365, 176 - 186.
- Rumsfeld, D. 2002. DoD News Briefing - Secretary Rumsfeld and Gen. Myers. In: DEFENSE, U. D. O. (ed.).
- Ryan, M., Bate, A., Eastmond, C. J. & Ludbrook, A. 2001. Use of discrete choice experiments to elicit preferences. *Quality in Healthcare*, 10 Suppl 1, i55-60.

- Sapelli, C. & Vial, B. 2003. Self-selection and moral hazard in Chilean health insurance. *Journal of Health Economics*, 22, 459-76.
- Savonitto, S., Ardissino, D., Granger, C. B., Morando, G., Prando, M. D., Mafri, A., Cavallini, C., Melandri, G., Thompson, T. D., Vahanian, A., Ohman, E. M., Califf, R. M., Van De Werf, F. & Topol, E. J. 1999. Prognostic value of the admission electrocardiogram in acute coronary syndromes. *JAMA : the journal of the American Medical Association*, 281, 707-13.
- Schulze, R. A., Jr., Rouleau, J., Rigo, P., Bowers, S., Strauss, H. W. & Pitt, B. 1975. Ventricular arrhythmias in the late hospital phase of acute myocardial infarction. Relation to left ventricular function detected by gated cardiac blood pool scanning. *Circulation*, 52, 1006-11.
- Sculpher, M. 1998. The cost-effectiveness of preference-based treatment allocation: the case of hysterectomy versus endometrial resection in the treatment of menorrhagia. *Health Economics*, 7, 129-42.
- Sculpher, M. 2008a. Subgroups and Heterogeneity in Cost-Effectiveness Analysis. *Pharmacoeconomics*, 26, 799.
- Sculpher, M. 2010. Reflecting heterogeneity in patient benefits: the role of subgroup analysis with comparative effectiveness. *Value in Health*, 13 Suppl 1, S18-21.
- Sculpher, M., Fenwick, E. & Claxton, K. 2000. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics*, 17, 461-77.
- Sculpher, M. & Gafni, A. 2001. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Economics*, 10, 317-24.
- Sculpher, M. J. 2008b. *Briefing paper for methods review workshop on identifying subgroups and exploring heterogeneity* [Online]. London: on behalf of the Institute's Decision Support Unit. Available: <http://www.nice.org.uk/media/4A6/37/SubgroupsBriefingPaper.pdf> [Accessed February 2012].
- Sculpher, M. J., Pang, F. S., Manca, A., Drummond, M. F., Golder, S., Urdahl, H., Davies, L. M. & Eastwood, A. 2004. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technology Assessment (Winchester, England)*, 8, iii-iv.
- Sculpher, M. J., Petticrew, M., Kelland, J. L., Elliott, R. A., Holdright, D. R. & Buxton, M. J. Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions. *Health Technology Assessment (Winchester, England)*, 2, i-iv.
- Sen, A. 1993. Capability and well-being. In: NAUSSBAUM, M. & SEN, A. (eds.) *The Quality of Life*. Oxford: Clarendon Press.
- Sen, A. K. 1980. Equality of what? In: MCMURRIN, S. (ed.) *The Tanner Lectures on Human Values*. Cambridge: Cambridge University Press.
- Sevdalis, N. & Jacklin, R. 2008. Interaction effects and subgroup analysis in clinical trials: more than meets the eye? *Journal of Evaluation in Clinical Practice*, 14, 919-922.
- Sgarbossa, E. B. 1996. Recent advances in the electrocardiographic diagnosis of myocardial infarction: left bundle branch block and pacing. *Pacing & Clinical Electrophysiology*, 19, 1370-9.
- Sleight, P. 2000. Debate: Subgroup analyses in clinical trials - fun to look at, but don't believe them! *Current Controlled Trials in Cardiovascular Medicine*, 1, 25 - 27.

- Smith, P. 2012. Incorporating financial protection into decision rules for publicly financed healthcare treatments. *Health economics*, doi(10.1002/hec.2774).
- Stallones, R. 1987. The use and abuse of subgroup analysis in epidemiological research. *Prev Med*, 16, 183 - 194.
- Stevens, W. & Normand, C. 2004. Optimisation versus certainty: understanding the issue of heterogeneity in economic evaluation. *Social Science & Medicine*, 58, 315-320.
- Steyerberg, E. W. 2009. *Clinical prediction models: a practical approach to development validation and updating*, New York, Springer.
- Stinnett, A. A. & Mullahy, J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision-making*, 18, S68-80.
- Stone, P. H., Muller, J. E., Hartwell, T., York, B. J., Rutherford, J. D., Parker, C. B., Turi, Z. G., Strauss, H. W., Willerson, J. T. & Robertson, T. 1989. The effect of diabetes mellitus on prognosis and serial left ventricular function after acute myocardial infarction: contribution of both coronary disease and diastolic left ventricular dysfunction to the adverse prognosis. The MILIS Study Group. *Journal of the American College of Cardiology*, 14, 49-57.
- Straus, S., Richardson, W. S., Glasziou, P. & Haynes, B. 2011. *Evidence based medicine: How to practice and teach it*, Churchill Livingstone Elsevier.
- Sugden, R. & Williams, A. 1978. *The principles of practical cost-benefit analysis*. Oxford: Oxford University Press.
- Suhonen, R., Valimaki, M. & Leino-Kilpi, H. 2002. "Individualised care" from patients', nurses' and relatives' perspective--a review of the literature. *International Journal of Nursing Studies*, 39, 645-54.
- Sun, X., Briel, M., Busse, J. W., Akl, E. A., You, J. J., Mejza, F., Bala, M., Diaz-Granados, N., Bassler, D., Mertz, D., Srinathan, S. K., Vandvik, P. O., Malaga, G., Alshurafa, M., Dahm, P., Alonso-Coello, P., Heels-Ansdell, D. M., Bhatnagar, N., Johnston, B. C., Wang, L., Walter, S. D., Altman, D. G. & Guyatt, G. H. 2009. Subgroup Analysis of Trials Is Rarely Easy (SATIRE): a study protocol for a systematic review to characterize the analysis, reporting, and claim of subgroup effects in randomized trials. *Trials [Electronic Resource]*, 10, 101.
- Sun, X., Briel, M., Walter, S. D. & Guyatt, G. H. 2010. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*, 340, c117.
- Thompson, C. A. 2011. Regulations, economics hindering adoption of personalized medicine. *American Journal of Health-System Pharmacy*, 68, 372-4.
- Torgerson, D. & Torgerson, C. (eds.) 2008. *Torgerson Designing Randomized Trials in Health, Education and Social Sciences: An introduction*, Hampshire: Palgrave Macmillan.
- Torrance, G. W. 2006. Utility measurement in healthcare: the things I never got to. *Pharmacoeconomics*, 24, 1069-78.
- Trusheim, M. R., Berndt, E. R. & Douglas, F. L. 2007. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nature Reviews. Drug Discovery*, 6, 287-93.
- Tsuchiya, A. & Williams, A. 2001. Welfare economics and economic evaluation. In: DRUMMOND, M. & MCGUIRE, A. (eds.) *Economic evaluation in healthcare*. Oxford: Oxford University Press.

- Tursz, T., Andre, F., Lazar, V., Lacroix, L. & Soria, J. C. 2011. Implications of personalized medicine--perspective from a cancer center. *Nature Reviews Clinical Oncology*, 8, 177-83.
- Tversky, A. & Kahneman, D. 1981. The framing of decisions and the psychology of choice. *Science*, 211, 453-8.
- Ubel, P., Richardson, J. & Menzel, P. 2000. Societal value, the person trade-off, and the dilemma of whose values to measure for cost-effectiveness analysis. *Health Economics*, 9, 127-136.
- Us Public Law Sec 6301. 2010. *Establishment: Patient-Centered Outcomes Research Institute* [Online]. Available: www.pcori.org [Accessed May 2012].
- Van Gestel, A., Grutters, J., Schouten, J., Webers, C., Beckers, H., Joore, M. & Severens, J. 2012. The role of expected value of individualized care in cost-effectiveness analyses and decision-making. *Value in Health*, 15, 13-21.
- Von Neumann, J. & Morgenstern, O. 1947. *Theory of games and economic behaviour*, NJ, Princeton University Press.
- Walker, S., Sculpher, M., Claxton, K. & Palmer, S. 2012. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value in Health*, 15, 570-579.
- Wang, R., Lagakos, S., Ware, J., Hunter, D. & Drazen, J. 2007. Statistics in medicine - reporting of subgroup analyses in clinical trials. *N Engl J Med*, 357, 2189 - 2194.
- World Health Organization 2000. *The World Health Report: Health systems improving performance*, Geneva, World Health Organization.
- Yusuf, S., Wittes, J., Probstfield, J. & Tyroler, H. 1991. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *Journal of the American Medical Association*, 266, 93 - 98.
- Zarate, V., Kind, P. & Chuang, L. H. 2008. Hispanic valuation of the EQ-5D health states: a social value set for Latin Americans. *Value in Health*, 11, 1170-7.
- Zarate, V., Kind, P., Valenzuela, P., Vignau, A., Olivares-Tirado, P. & Munoz, A. 2011. Social valuation of EQ-5D health states: the Chilean case. *Value in Health*, 14, 1135-41.
- Zizek, S. 2006. Philosophy, the "unkown knowns", and the public use of reason. *Topoi*, 25, 137-142.