HEALTH ECONOMICS RESEARCH GROUP (HERG)
BRUNEL UNIVERSITY

Current Use and Potential Value of Cost-effectiveness Analysis in U.S. Health Care: The Case of Medicare National Coverage Determinations

A thesis submitted to Brunel University for the degree of Doctor of Philosophy

James D. Chambers

January, 2012

Preface

Acknowledgments

I am extremely grateful to my supervisors, Professor Martin Buxton and Dr. Joanne Lord for their attentive supervision, guidance and encouragement throughout the development of this thesis. I am also grateful to Professor Stephen Morris who provided supervision throughout the early part of my study. I feel very fortunate to have had the opportunity to work with each one of them and their guidance and advice has been truly invaluable.

To my Mother and Father, I owe you a debt of gratitude that I can never hope to repay. Without your sacrifice and unwavering support I could not have contemplated pursuing a PhD. I would like to dedicate this thesis to you.

I am fortunate to work at the Center for the Evaluation of Value and Risk in Health (CEVR) at Tufts Medical Center in Boston. I am particularly grateful to Dr. Peter Neumann who has been hugely supportive and has allowed me the flexibility to continue my studies. Also, I owe Sarah Bliss, a colleague at CEVR, thanks for reviewing a draft of my thesis.

Published and presented work

Parts of the empirical aspect of this thesis have been published and presented. In 2010 a paper was published in Medical Decision Making pertaining to the research presented in Chapter 5. The citation is as follows:

> Chambers JD, Neumann PJ, Buxton MJ. Does Medicare have an implicit cost-effectiveness threshold? Med Decis Making. 2010 Jul-Aug;30(4):E14-27. Epub 2010 Jun 15.

In October, 2011, I received confirmation that a paper related to the empirical work presented in Chapter 6 submitted was accepted for publication in Medical Care. The provisional citation is as follows

> Chambers JD, Morris S, Neumann PJ, Buxton MJ. Factors Predicting Medicare National Coverage: An Empirical Analysis. Medical Care. Accepted October, 2011

I presented parts of the empirical work presented in Chapter 7 at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) meeting in Baltimore in 2011. The citation for this presentation is as follows:

> Chambers JD, Cohen JT, Neumann PJ, Lord J, Buxton MJ. Using Cost-Effectiveness Information to Allocate Medicare Resources – How Much Health for the Money? ISPOR 16th Annual International Meeting, Baltimore May 24th, 2011. Podium presentation

Role of the Center for the Evaluation of Value and Risk in Health at Tufts Medical Center

Included in the research presented in Chapter 6, was a variable taken from the Tufts Medical Center National Coverage Determinations (NCD) Database. This variable pertained to an assessment of the clinical evidence presented in each NCD. I was not involved in the data collection process for this variable. All other research presented in Chapter 6 is my own.

This research presented in Chapter 7 was conducted as part of a larger project performed at the Center for the Evaluation of Value and Risk in Health (CEVR) at Tufts Medical Center in Boston. The broader study considered all interventions available to Medicare beneficiaries, not limited to those pertaining to NCDs, and was funded by the Commonwealth Fund. My role in the project was to perform the requisite research pertaining to CMS NCDs.

Additional data collection was required for research presented in this chapter not collected for the broader project conducted at Tufts Medical Center. This included data required for consideration of expenditures in the year following first use of an intervention and the impact of a reallocation of expenditures on their distribution across patient populations. All research presented in this chapter is my own work, including the data collection, analysis, and reporting.

Abstract

There is a growing recognition that we cannot afford the provision of all new health care technologies, even those that are proven to be beneficial. This is increasingly true in the US, where health care spending is on an unsustainable upward trajectory. US health care spending is greatly in excess of that of other countries; however, with respect to key health metrics, the US health care system performs relatively poorly. Despite this, unlike many other developed countries economic evaluation, and more specifically cost-effectiveness evidence, is used sparingly in the US health care system. Notably, the Centers for Medicare and Medicaid Services (CMS), administrators of the Medicare programme, state that cost-effectiveness evidence is not relevant to coverage decisions for medical technology and interventions evaluated as part of National Coverage Determinations (NCDs). The empirical aspect of this thesis evaluates the current use and potential value of using cost-effectiveness evidence in CMS NCDs. A database was built using data obtained from NCD decision memoranda, the medical literature, a Medicare claims database, and Medicare reimbursement information. The findings of the empirical work show that, CMS's stated position notwithstanding, cost-effectiveness evidence has been cited or discussed in a number of coverage decisions, and there is a statistically significant difference between positive and non-coverage decisions with respect to cost-effectiveness. When controlling for factors likely to have an effect on coverage decisions, the availability of cost-effectiveness evidence is a statistically significant predictor of coverage. In addition, the quality of the supporting clinical evidence, the availability of alternative interventions, and the recency of the decision are statistically significant variables. Further, when hypothetically reallocating resources in accordance with cost-effectiveness substantial gains in aggregate health are estimated. It is shown that using cost-effectiveness to guide resource allocation has an effect on resource allocation across patient populations and types of technology.

Table of Contents

List of tables

List of Figures

Abbreviations

AETMIS	Agence d'évaluation des technologies et des modes d'Intervention en Santé
AHRQ	Agency for Healthcare Research and Quality
AMA	American Medical Association
AMCP	Academy of Managed Care Pharmacy
AN-DRG	Australian National Diagnosis Related Groups
AU	Australia
AuSCT	Autologous Stem Cell Transplantation
BBA	Balanced Budget Act
BMG	Federal Ministry of Health (Germany)
BMI	Body Mass Index
CADTH	Canadian Agency for Drugs and Technologies in Health
CAN	Canada
CBO	Congressional Budget Office
CCOHTA	Canadian Coordinating Office for Health Technology Assessment
CDR	Common Drug Review
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CEDAC	Canadian Expert Drug Advisory Committee
CER	Comparative-Effectiveness Analysis
CEVR	Center for the Evaluation of Value and Risk in Health
CI	Confidence Interval
CMS	Centers for Medicare & Medicaid Services
CNV	Classic subfoveal lesions
CPAP	Continuous Positive Airway Pressure
CPI	Consumer Price Index
CPT	Current Procedural Terminology
CU	Consumers Union

CVZ	College voor Zorgverzekeringen
DBS	Deep Brain Stimulation
DERP	Drug Effectiveness Review Project
DoD	Department of Defense
DoH	Department of Health
DPN	Diabetic Peripheral Neuropathy
ECP	External Counterpulsation
EQ-5D	EuroQol Group 5-Dimension Self-Report Questionnaire
ESA	Erythropoiesis Stimulating Agents
ESRD	End-Stage Renal Disease
FDG	2-Fluorodeoxy-D-Glucose
FOBT	Fecal Occult Blood Test
G-BA	Federal Joint Committee (Germany)
GDP	Gross Domestic Product
HCFA	Health Care Financing Administration
HDA	Health Development Agency
HEED	Health Economic Evaluations Database
HIV	Human immunodeficiency virus
HTA	Health Technology Assessment
HUI	Health Utilities Index
ICD	Implantable Cardioverter Defibrillator
ICD-9	International Classification of Diseases
ICER	Incremental Cost-Effectiveness Ratio
IDU	Intravenous Drug Use
iFOBT	Immunological Fecal Occult Blood Testing
INR	International normalised ratio
IQWiG	Institute for Quality and Efficiency in Health Care
ISPOR	International Society of Pharmacoeconomics and Outcomes Research
LAGB	Laparoscopic Adjustable Gastric Banding
LOPS	Loss of Protective Sensation

MBS	Medicare Benefits Schedule
MEDCAC	Medicare Evidence Development & Coverage Advisory Committee
MILP	Mixed Integer Linear Programme
MIPPA	Medicare Improvements for Patients and Providers Act
MRI	Magnetic Resonance Imaging
MSAC	Medical Services Advisory Committee
NA	Not applicable
NCD	National Coverage Determination
NEMA	National Electric Manufacturers Association
NETT	National Emphysema Treatment Trial
NHS	National Health Service
NHS EED	National Health Service Economic Evaluations Database
NICE	National Institute for Health and Clinical Excellence
NNWT	Noncontact Normothermic Wound Therapy
OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratio
OSA	Obstructive Sleep Apnea
OTA	Office of Technology Assessment
P&T	Pharmacy and Therapeutics
PBAC	Pharmaceutical Benefits Advisory Committee
PBS	Pharmaceutical Benefits Scheme
PCORI	Patient-Centered Outcomes Research Institute
PCT	Primary Care Trust
PET	Positron Emission Tomography
PhRMA	Pharmaceutical Research and Manufacturers Association of America
PPACA	Patient Protection and Affordable Care Act
PPP	Purchasing Price Parity
PTA	Percutaneous Transluminal Angioplasty
QALY	Quality Adjusted Life Year
R&D	Research and Design

ROC	Receiver Operating Characteristic
RYGBP	Roux-en-Y gastric bypass
SBU	Swedish Council of Technology Assessment in Health Care
SG	Standard Gamble
SGB	Sendi, Gafni, and Birch
SNM	The Society of Nuclear Medicine
SSA	Social Security Amendments
TA	Technology Assessment
TEC	Blue Cross Blue Shield Association Technology Evaluation Center
TLV	Tandvårds- och läkemedelsförmånsverket (Sweden)
TTO	Time-Trade Off
UK	United Kingdom
US	United States of America
USPSTF	US Preventative Services Task Force
VA	Department of Veterans Affairs
VBP	Value Based Pricing
VIF	Variance Inflation Facor
WHO	World Health Organization
WTP	Willingness To Pay

1. Introduction

There is growing recognition that we cannot afford the provision of all new health care technologies, even those that are proven to be beneficial. This is increasingly true in the US, where health care spending is on an unsustainable upward trajectory, and where current health care spending is twice that of many developed countries in terms of GDP per capita.

The US health care system performs poorly in comparison to others. Ranked 37th by the World Health Organization (WHO) in their global rankings of health care systems, and placed last in the Commonwealth Fund's 2010 rankings of health care systems in Australia, Canada, New Zealand, the United Kingdom, Germany, the Netherlands and the US, the US health care system has much room for improvement. Despite health care spending greatly in excess of that in other developed countries, average life expectancy in the US is shorter and infant mortality higher. Further, with respect to health care resources, the US has fewer physicians and hospital beds per capita compared to other developed countries. Most notable, however, is the lack of universal health insurance coverage in the US, with many US citizens having either no or insufficient health insurance.

There is increasing awareness that resource allocation must be addressed in a systematic rather than intuitive manner. One approach to the prioritisation of resources between competing interventions is to use economic evaluation to assess health care technology. In Chapter 2, I present the theory underpinning the use of economic evaluation, and more specifically cost-effectiveness analysis, to inform resource allocation. I provide a worked example illustrating how a cost-effectiveness decision rule can lead to the efficient allocation of scarce resources across multiple health care programmes. I describe the league table and mathematical programming approaches as two frameworks for implementing a cost-effectiveness decision rule. The practicality of both approaches is, however, restricted by the requirement for complete knowledge of the costs and benefits of available health care programmes. Therefore, it is necessary to have a benchmark

value, or decision rule, with which to interpret the findings of cost-effectiveness studies. The remainder of Chapter 2 focuses on the cost-effectiveness threshold. I present the various valuations of the cost-effectiveness threshold, including the threshold operated by the National Institute for Health and Clinical Excellence (NICE) in the UK and thresholds derived through retrospective evaluation of decisions made by various international agencies. I also present the advantages and disadvantages of hard vs. soft and explicit vs. implicit cost-effectiveness threshold valuations, and of cost-effectiveness acceptability curves, used to interpret cost-effectiveness evidence while conveniently evading the question of the value of the cost-effectiveness threshold.

In Chapter 3, I place the US health care system into an international context with respect to spending, abundance of health care resources, and key health statistics. Comparator countries were chosen on the basis that they help illustrate variation in how economic evidence and other factors are considered in the evaluation of health care technologies across jurisdictions. I chose the UK, Sweden, Australia, and Canada as examples of countries in which cost-effectiveness evidence plays a fundamental role in decision-making, and Germany and France as examples of countries in which cost-effectiveness evidence plays a lesser role. In spite of health care spending greatly in excess of spending in other countries, the US health care system performs poorly across a number of key health metrics. Despite an evident need to increase the return on health care spending, cost-effectiveness evidence is used only sporadically in the US health care system. Notably, and of particular relevance to the empirical aspect of this thesis, Medicare, the largest payer in the US, states that cost-effectiveness is not a factor considered in its coverage decisions. To provide insight into the resistance to cost-effectiveness evidence in the US health care system, I review the failed attempts by Medicare and the state of Oregon's Medicaid programme to incorporate cost-effectiveness evidence into decision-making. Finally, I discuss the implications of the recent Patient Protection and Affordable Care Act (PPACA) legislation for the future use of cost-effectiveness evidence in the US, and highlight recent instances in which the

Centers for Medicare and Medicaid Services (CMS) used cost-effectiveness evidence in coverage decisions for preventative care.

The foundations of my empirical work are presented in Chapter 4. I chose Medicare, the health insurance programme for Americans aged 65 years and over and those with certain disabilities, as the aspect of the US health care system on which to focus my research. Medicare is the largest payer in the US, providing coverage to 46 million Americans at a cost of $6 billion, approximately 5 % of GDP. Medicare coverage decisions for medical technology have far-reaching influence and are thought to affect private payers' coverage decisions. This research concerns CMS's national coverage policies, or National Coverage Determinations (NCD). National Coverage Determinations are binding to all regional Medicare Administrative Contractors (MACs), and are reserved for interventions deemed particularly controversial or projected to have a major impact on the Medicare programme.

The research objectives for the empirical component of this thesis are as follows:

Empirical Research: Part 1

1. To examine NCD decision memos to determine if the presented evidence review is consistent with CMS's stated position that cost-effectiveness evidence is not relevant to coverage decisions.
2. To determine if there is a difference between the cost-effectiveness of positive coverage decisions and non-coverage decisions.

Empirical Research: Part 2

1. To determine if cost-effectiveness is an independent predictor of coverage decisions included in NCDs when controlling for other factors likely to have an effect on coverage decisions.

Empirical Research: Part 3

1. To estimate potential gains in aggregate health achieved from reallocating expenditures between interventions covered as part of NCDs in a manner consistent with a cost-effectiveness decision rule.
2. To estimate the impact of reallocation on the distribution of expenditures across disease areas (oncology, cardiology, and other) and types of intervention (treatment, diagnostic, and other).

Also in Chapter 4, I describe a literature search I performed to identify studies to help inform the methodological approach for the empirical work. First, I performed a search to identify studies that evaluated the role of cost-effectiveness evidence in coverage and reimbursement decisions, or in recommendations for the efficient use of medical technology. I identified and reviewed studies that evaluated decisions made by NICE in the US, the PBAC in Australia, CEDAC in Canada, PHARMAC in New Zealand, and an HMO in the US. Second, I performed a search to identify studies that estimated efficiency gains from alternative approaches for resource allocation.

In the remainder of Chapter 4, I describe the development of the database used for the empirical work. Variables in the database include cost-effectiveness, quality of supporting clinical evidence, availability of alternative interventions, date of decision, coverage requestor, and type of intervention. I primarily generated the cost-effectiveness variable through literature searches, although on occasion a relevant cost-effectiveness ratio originated from the decision memo accompanying the NCD. The variable classifying the quality of the supporting clinical evidence was generated through review of the decision memo by two researchers at Tufts Medical Center using the US Preventative Service Task Force (USPSTF) guidelines for grading evidence. I generated the remaining variables from the information presented in decision memos. Additional variables were required for the third piece of empirical work, including incremental cost and incremental QALY gain data, the cost of the intervention in the year following its first use, the existing utilisation rate, and the size of the eligible patient population. The

incremental cost and incremental effectiveness data typically originated from the cost-effectiveness studies; the existing utilisation rate and size of the eligible patient population from a Medicare claims database; and the additional cost data required for estimation of the intervention in the year following first use from Medicare reimbursement codes.

Chapter 5 describes the first piece of my empirical work. As noted, CMS state that cost-effectiveness is not a factor it considers when making NCDs. The first objective of the research presented in chapter 5 was to examine NCD decision memos to identify instances when cost-effectiveness evidence is cited or discussed, thus assessing the consistency of CMS's behaviour with its stated position on the use of cost-effectiveness evidence. I reviewed each decision memo (n=140) for discussion or citation of cost-effectiveness evidence relevant to the included coverage decisions (n=255). On 14 occasions, a coverage decision was associated with either discussion of cost-effectiveness evidence or a citation of a relevant cost-effectiveness study. Twelve of the 14 coverage decisions were positive, and notably, in each instance the estimate of cost-effectiveness was favourable (maximum ICER of $27,161 per life year gained). The second objective of this research was to determine if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness. I supplemented the estimates of cost-effectiveness identified in decision memos with a series of literature searches to identify published estimates of cost-effectiveness relevant to included coverage decisions. For 64 coverage decisions, an associated cost-effectiveness estimate was identified. Findings show that CMS are covering interventions not cost-effective by traditional standards; nine covered interventions are associated with an ICER greater than $100,000 per QALY and three with an ICER greater than $500,000 per QALY. I used a Mann Whitney U test to determine a statistically significant difference between positive coverage decisions and non-coverage decision with respect to their cost-effectiveness, suggesting that interventions subject to positive coverage decisions tend to be associated with more favourable cost-effectiveness evidence.

Chapter 6 describes the second piece of empirical work. This research builds on Chapter 5 and evaluates whether, when controlling for factors that are likely to have an effect on Medicare coverage decisions, cost-effectiveness, or the availability of cost-effectiveness evidence, is statistically significantly associated with the coverage decision. In addition to cost-effectiveness, I accounted for a number of aspects thought to be relevant to decision-making, with independent variables such as the quality of the supporting clinical evidence, the availability of alternative interventions, intervention type, origin of the request for coverage, and date of the decision included in the model. I estimated the model using binomial logistic regression, regressing the coverage decision (positive/non-coverage) against the independent variables. I performed univariate and multivariate regressions. Compared with interventions estimated to be dominant, those with no associated estimate of cost-effectiveness were approximately five times less likely to receive a positive coverage decision. Interventions associated with good quality supporting clinical evidence were six times more likely to receive a positive coverage decision compared with those associated with insufficient evidence. Compared to interventions with no available alternative, those with an available alternative were approximately eight times less likely to be associated with a positive coverage decision. Finally, coverage decisions made in 2006-2007 were approximately 10 times less likely to be associated with a positive coverage decision than those made in 1999-2001, with interventions considered in more recent time periods increasingly less likely to be associated with a positive coverage decision.

While the findings are insufficient to conclude that CMS coverage decisions are consistent with cost-effectiveness, it is notable that the availability of cost-effectiveness evidence estimating the intervention to be dominant is associated with the coverage decision. In addition, this research provides insight into the 'reasonable and necessary' coverage criterion, suggesting that CMS operate evidence-based coverage policy and highlighting that the availability of alternatives is relevant to decision-making. The

findings show that, when controlling for other factors, CMS became more restrictive with respect to coverage over the time period considered.

Chapter 7 describes the third piece of empirical work. In the research presented in Chapter 5, I determined that CMS cover a number of interventions not cost-effective by traditional standards. Coverage of interventions with high ICERs is inefficient as it consumes considerable resources and produces marginal health gains. The research presented in Chapter 7 considers the question of the inefficient use of resources and estimates what gains in aggregate health could be achieved from allocating resources using a cost-effectiveness rule. Specifically, the objectives of the research are to estimate potential gains in aggregate health from reallocating existing expenditures, and to estimate the impact of reallocation on the distribution of resources across disease areas and types of intervention. To reallocate expenditures between interventions, I simulated disinvestment in relatively cost-ineffective interventions and increased investment in cost-effective interventions through the manipulation of utilisation rates. The findings estimate that substantial gains in aggregate health are achievable from reallocating expenditures to maximise health while maintaining a net change in total expenditure of zero. Further, simply increasing the utilisation of dominant interventions was estimated to yield substantial aggregate health gains and cost-savings. The distribution of expenditures across disease areas and types of technology following reallocation of resources was different than the existing distribution.

Chapter 8 constitutes the final chapter of this thesis, in which I summarise the empirical work and discuss the key findings. Also, I discuss the limitations of this thesis and the steps that can be taken to further develop its empirical aspects. Finally, I describe how this thesis contributes to knowledge and its policy relevance.

This thesis will contribute to knowledge in a variety of ways. First, it provides the first systematic assessment of the cost-effectiveness of interventions evaluated by CMS

through NCDs, illustrating that CMS cover interventions that do not represent good value, although there is a statistically significant difference between positive and non-coverage decisions with respect to cost-effectiveness. Second, it provides the first empirical analysis of CMS NCDs that considers a variety of factors likely to have an effect on coverage decisions. This analysis shows that the quality of the supporting clinical evidence, the availability of alternative interventions, the date of the decision, and the availability cost-effectiveness evidence are associated with coverage decisions. Third, it provides the first attempt to estimate efficiency gains in the Medicare programme through the hypothetical reallocation of resources using cost-effectiveness evidence. The analysis shows that substantial gains in aggregate health are achievable from using a cost-effectiveness rule to inform resource allocation while maintaining existing levels of expenditures.

Given the recent passing of the PPACA legislation and the ongoing debate surrounding the future of the Medicare programme, this thesis is timely in terms of policy relevance. This research sheds light on the value of interventions offered in the Medicare programme and shows that CMS have on occasion included cost-effectiveness evidence in their review of the evidence base. The research illustrates the evidence-based nature of CMS coverage decisions and provides an insight into the interpretation of the 'reasonable and necessary' criteria operated by CMS. This type of research has the potential to lead to better and more consistent decision-making and increase the accountability of CMS. The research suggests that substantial gains in aggregate health are achievable from reallocating existing expenditures, and that cost-effectiveness evidence has the potential to inform more efficient resource allocation decisions.

2. Background and Theory

2.1. Introduction

In this chapter, I describe the role of economic evaluation in the allocation of scarce health care resources. Unlike other sectors of the economy, market forces cannot be relied upon for resource allocation in health care. Economic evaluation offers an approach to help choose between competing health care interventions to prioritise health care spending.

In the following sections, I describe two methods of economic evaluation, cost-effectiveness analysis and cost-benefit analysis, and the key differences between them. I provide a worked example using a scenario including multiple health care programmes to illustrate how adherence to a cost-effectiveness decision rule will result in efficient resource allocation. Two frameworks for implementing a cost-effectiveness decision rule are described; the league table approach and mathematical programming. These approaches are limited by the magnitude of the information requirements. Alternative frameworks such as the 'searching for the threshold' and the 'Sendi, Gafni, and Birch' are discussed as approaches that are not inhibited by these information requirements.

The requirement of a cost-effectiveness threshold to interpret cost-effectiveness evidence is highlighted and a worked example illustrates how a calibrated threshold can lead to efficient resource allocation. The last section of this chapter focuses on approaches for the derivation of the cost-effectiveness threshold, various valuations of thresholds that are used or have been proposed, and various criteria with which thresholds can be characterised. Cost-effectiveness acceptability curves (CEAC) are also shown as a method of presenting cost-effectiveness findings that avoids the need for a cost-effectiveness threshold.

2.2. The need for economic evaluation in health care

2.2.1. The scarcity of health care resources

Health care resources are scarce. Consequently, health care systems cannot provide all care that would potentially benefit patients. Decisions must be made as how best to allocate available resources to meet health care system goals. Often referred to as the prioritisation or rationing of resources, resource allocation can be emotive and contentious.

While currently at the forefront of health care policy debate, the prioritisation of health care resources is not only a recent concern. It has always been the case that doctors have had to judge the reasonableness of the care they provide. (Ubel, 2001; Ubel & Goold, 1997) Williams, 2002 states, "*Time and effort and other health care resources devoted to one patient could not be devoted to another, so they had to decide how to allocate these scarce resources so as to do the most good, as they saw it*."(Williams, 2002)

How to allocate scarce health care resources remains a challenge. In many countries specialised institutions have been established to provide guidance on resource allocation, e.g., the National Institute for Health and Clinical Excellence (NICE) in the UK. A principal role of these institutions is to evaluate the evidence base for new and established interventions as a precursor to coverage and reimbursement decisions or to provide guidance to practitioners. It seems reasonable to ask why such institutions are necessary. Similar institutions do not exist in other industries in which market forces are relied upon to drive down costs and to efficiently allocate resources. In the following section, I will discuss why markets are insufficient and inappropriate for health care resource allocation.

2.2.2. Markets in health care

In many industries, markets are the mechanism used to ration goods. The forces that determine the price and quantity of goods in a market are supply and demand. Alone,

however, markets are insufficient to ration health care.

In the absence of government intervention, uncertainty, asymmetry of information, and risk preferences for health and health care precipitate the need and desire for health insurance. However, for three principal reasons markets do not work well in health care. (Donaldson et al. 2005;Donaldson 2008;Donaldson et al. 2008) First, insurance premiums are actuarially unfair. If health insurance was actuarially fair then premiums paid would be equivalent to expenditure incurred. However, in reality, health insurance is actuarially unfair, as the premiums are 'loaded' to cover administration costs and achieve profit. Consequently, some who would otherwise have obtained insurance will be priced out of the market. (Donaldson et al. 2005;Donaldson 2008) The second reason is 'moral hazard'. Moral hazard is the term used to describe how having insurance coverage changes individuals' behaviour. If a third party, i.e., the insurer, pays for health care, individuals have less incentive to avoid illness or injury and thus are more likely to use health care than if they were uninsured. This will result in cost inflation with health care becoming more expensive without a corresponding increase in health outcomes. (Donaldson et al. 2005;Donaldson 2008) Finally, a well-functioning health insurance market will set premiums in line with individual risk, i.e., low premiums for those at low risk and higher premiums for those at higher risk. This is problematic as individuals at higher risk tend to be those who can least afford insurance coverage. This results in a social problem, as those in most need of health care are those without health insurance coverage. (Donaldson et al. 2005;Donaldson 2008)

Despite circumventing, at least in part, some of the reasons for market failure, the establishment of publicly funded health care systems do not avoid the problem of scarcity and the need to prioritise available resources. Economic evaluation, as part of a broader health technology assessment (HTA) programme, is one mechanism by which a health care system can work toward achieving value from health care spending. Through the use of economic evaluation, including cost-effectiveness analysis, the relative value of

competing interventions can be used to guide resource allocation decisions. There is debate, however, as to the underlying economic framework that should be used when evaluating health care programmes. Debate has focused upon the welfarist and extra-welfarist frameworks; these are discussed below.

2.3. Economic evaluation in health care

Economic evaluation forms the basis of informing efficient resource allocation. Within the field of health economics, economic evaluation has been defined as a method of ''*ensuring that the value of what is being gained from an activity outweighs the value of what is being sacrificed''*. (Williams 1983)

2.3.1. Efficiency in health care

To understand the use of economic evaluation in health care, it is necessary to appreciate different concepts of efficiency. Palmer and Torgerson (1999) describe three concepts of efficiency; allocative, productive, and technical efficiency. (Palmer & Torgerson 1999) Each concept is described below.

Technical efficiency relates to the relationship between resources (capital and labour) and health outcome. An allocation of resources is technically efficient when the maximum health outcome is achieved from a set of input resources. Accordingly, a particular allocation of resources is technically inefficient when the same health outcome can be achieved with less of any one type of input. (Palmer & Torgerson 1999)

Productive efficiency relates to the relationship between health outcome and the cost of input resources. This differs from technical efficiency, which does not account for circumstances where the same health outcome can be achieved with a different

combination of inputs. By considering input costs it is possible to choose between different combinations of resources to maximise health outcome within a budget constraint. With respect to health care interventions, the consideration of productive efficiency facilitates evaluation of the relative value for money of interventions that generate directly comparable health outcomes. (Palmer & Torgerson 1999)

In contrast to technical and productive efficiency, allocative efficiency accounts for the efficiency with which outcomes are distributed across society. Resource use is allocatively efficient when any alternative allocation of resources results in at least one person being worse off. Absolute adherence to this principle is difficult, as doing so would preclude an allocation of resources resulting in many people benefiting in terms of health gain at the expense of few being made worse off. Consequently, the decision rule has been modified; resource allocation is allocatively efficient when the welfare of the community is maximised. Accordingly, allocative efficiency has its roots in welfare economics. (Palmer & Torgerson 1999)

2.3.2. The allocation of scarce resources

Health care decision makers can take two broad approaches when allocating scarce resources. The first approach is to eliminate system waste, i.e., to reduce spending without affecting the ability to produce specific health care outputs. Waste is present in varying degrees in all health care systems and its reduction is likely a significant source of savings in some health care systems. (Delaune & Everett 2008;Donaldson et al. 2008) One approach to waste reduction is to ensure that care is 'appropriate', i.e., that the potential benefits of care outweigh the potential harms. Efforts to evaluate the appropriateness of care have led to the development of frameworks based upon available clinical evidence and expert opinion to guide when care should be provided. (Brook et al. 1986;Brook 2009;Fitch et al. 2001) Examples include frameworks for gastrointestinal endoscopy, tympanostomy tubes, and coronary angiography. (Froehlich et al.

1997;Hemingway et al. 2008;Kleinman et al. 1994)

The second approach is to attempt to prioritise resources in order to maximise return, in terms of health gain, from investment. This requires maximising investment in health care programmes that generate most health benefit from investment and minimising investment in programmes that generate little, or no, return. (Donaldson et al. 2008) To achieve efficient health care spending the allocation of resources must be done in a rational manner. There is debate, however, as to the most appropriate method for health care resource allocation. (Drummond et al. 2005;Gafni & Birch 2006;Holm 1998;Ubel & Goold 1997)

Various criteria are likely to be considered when allocating scarce resources between health care programmes. The nature and strength of the available evidence, the potential impact of a decision on access to care, relative value for money, and the economic consequences of implementation are likely to be criteria in the decision-making process. (Drummond et al. 2005;Folland, Goodman, & Stano 2003;Keenan, Neumann, & Phillips 2006;Neumann 2005) Society's preferences for the allocation of resources are also important, and may not be in accordance with the maximisation of health. It has been shown that society typically values care for the elderly, the treatment of more severe diseases, and the avoidance of discrimination against people with chronic illnesses or disabilities highly. (Neumann 2005;NICE 2010c;Ubel 2001) It is a challenge for health care decision makers to concurrently account for each of these criteria and it may be necessary for a decision maker to trade off each factor against the magnitude of health gain. (Devlin & Sussex 2011) Across jurisdictions decision makers are, therefore, likely to have unique internal criteria and to allocate health care resources differently.

It is necessary to use a consistent decision-making framework. Without a framework with which to make decisions, a decision maker has to rely on judgement or to assume that what was done before was the best course of action. Relying on previous decisions

to guide resource allocation is rarely likely to yield more efficient decisions than the systematic consideration of relevant criteria. (Drummond et al. 2005) It has been suggested that decisions should be fully transparent, must rest upon criteria that stakeholders agree are relevant, and should be revisable following the availability of additional evidence. (Daniels 2000)

2.3.3. Welfarism vs. extra-welfarism

Welfare economics is a branch of economics that uses microeconomic techniques to examine individuals' preferences, the optimal allocation of resources, and the consequences of resource allocation on social welfare. (Birch & Donaldson 2003;Brouwer et al. 2008;Coast 2004;Johannesson 1995) Cost-benefit analysis has its theoretical foundations in welfare economics. (Jonsson 2009a) Cost-benefit analysis requires the consequences of a health care programme to be measured in monetary units, facilitating a comparison of costs and benefits in commensurate units. The decision rule is simply that if benefits outweigh costs, the health care programme should be implemented, and if not, it should not be implemented. With its theoretical underpinnings in welfare economics, cost-benefit analysis is conceptually appealing. The key criterion when considering a redistribution of resources is whether the redistribution represents a potential Pareto improvement (Kaldor-Hicks criterion) in social welfare, i.e., those who gain from a policy change compensate the losers and remain in a preferred position. (Gafni 2006;Sugden & Williams 1979) For example, if willingness to pay for health care is greatest for the wealthiest in society, then a policy change that results in greater health care for the wealthiest would be optimal, as the gainers could compensate the losers and remain better off. Such a policy would therefore benefit society as a whole, increasing overall social welfare. (Drummond et al. 2005) Cost-benefit analysis is established as the methodology of choice in sectors of the economy other than health care, including transportation and education. (Claxton et al. 2010)

2.3.4. Applying cost-benefit analysis to health care

There are a number challenges to applying cost-benefit analysis to health care. Placing a monetary value on health is associated with a number of difficulties. (Ryan et al. 2001) Various techniques exist, including; the human capital, revealed preference, and contingent valuation approaches (See Section 2.6.1.1 for further details). (Buxton 2005;Drummond et al. 2005a) Each of these approaches has strengths and weaknesses and none is universally accepted. Assigning a monetary value to health is contentious and many decision makers find it difficult, or even unethical, to depend on such valuations. (Weinstein & Feinberg 1980) Maybe the most notable challenge is that methods that place a monetary value on health intrinsically favour the wealthiest in society. (Gold et al. 1996) To date, cost-benefit analysis has been used infrequently to inform resource allocation in health care, with cost-effectiveness analysis the preferred approach. (Drummond et al. 2005)

2.3.5. Extra-welfarism and cost-effectiveness analysis

Given the challenges outlined above, an alternative to the welfarist framework has been favoured within health economics. (Weinstein & Stason 1977) There is a strong ethical appeal that life, particularly life enjoyed in good health, is different from other commodities. It is argued that healthy life is necessary in order to carry out all other activities and, therefore, should be awarded special moral importance. (Daniels 2008) Extra-welfarism, or the non-welfarist approach, has been embraced as the theory underpinning much of economic evaluation in health care. (Brouwer & Koopmanschap 2000;Coast 2004;Culyer 1989;Sugden & Williams 1979;Tsuchiya & Williams 2010) Extra-welfarism differs from welfarism, for rather than aiming to maximise social welfare, the objective is to maximise aggregate health, irrespective of initial health status, age, disability, or indeed, willingness or ability to pay. (Birch & Donaldson 2003;Coast 2004) The extra-welfarist approach may be considered ‘utilitarian’ as the distribution of health gains is not considered. (Lord, Laking, & Fischer 2004) However, as the extra-welfarist framework is uniquely focused on health maximisation, resultant resource

allocation may not be consistent with society's preferences. (Dolan & Cookson 2000;Nord et al. 1995;Ubel 2001) The extra-welfarist framework has been suggested to be morally superior to the welfarist approach. As the assigned value of health is independent of the distribution of wealth in society, it is considered by many to be 'income free'. (Weinstein & Manning, Jr. 1997) Others suggest, however, that adopting cost-effectiveness analysis in preference to cost-benefit analysis does not avoid considerations of income distribution. (Donaldson, Birch, & Gafni 2002;Gafni 2006)

The analytical method of choice when considering an extra-welfarist framework is cost-effectiveness analysis. Cost-effectiveness analysis estimates the relative cost per unit of health gained across competing interventions.(Folland, Goodman, & Stano 2003;Garber & Phelps 1997) This may be in terms of the cost per disease specific unit (e.g., reduction in tumour size, reduction in ulcer healing time, etc), the cost per life year, or the cost per QALY gained, often referred to as cost-utility analysis. Cost-utility analysis is the preferred methodology for many decision makers, including national HTA bodies in the UK, Australia, Canada, and Sweden among others (Section 3.3.1). The QALY incorporates both quality of life and life expectancy into a single unit of health, allowing for comparison across disease areas. (Brazier 2008;Drummond et al. 2005;Weinstein et al. 1996)

2.3.6. The equivalence of cost-effectiveness analysis and cost-benefit analysis

A key difference between cost-effectiveness analysis and cost-benefit analysis can be illustrated by comparing their objectives. While the objective of cost-effectiveness analysis is to determine the least costly way to achieve a goal, the objective of cost-benefit analysis is to determine whether the goal is worth achieving. (Bala, Zarkin, & Mauskopf 2002;Donaldson 1998) There have been, however, various attempts to align the two methods by grounding cost-effectiveness analysis in a welfarist framework. (Garber et al. 1996;Johannesson 1995;Johannesson & O'Conor 1997;Meltzer 1997;Phelps & Mushlin 1991)

Researchers have attempted to illustrate circumstances when cost-effectiveness analysis and cost-benefit analysis can be considered equivalent. (Bala, Zarkin, & Mauskopf 2002;Johannesson 1995) Johannesson M (2005) suggests that cost-effectiveness analysis can be interpreted as cost-benefit analysis when the willingness to pay (WTP) per unit of effectiveness is assumed to be constant and the same for everyone. (Johannesson 1995) Bala et al. (2002) suggest less restrictive conditions; WTP for health gain needs to be the same in each patient subgroup, and that the WTP per unit of health gain and the magnitude of health gain achieved from treatment for a random individual in a given subgroup are independent random variables. (Bala, Zarkin, & Mauskopf 2002;Johannesson 1995)

It has been suggested that an explicit valuation of a unit of health outcome, e.g., the QALY, converts an ICER calculation into a quasi-net benefit criterion. (Drummond et al. 2005) However, the net benefit approach is 'quasi' cost-benefit analysis as its theoretical underpinnings remain grounded in an extra-welfarist rather than a welfarist framework; society's valuation of a QALY is preceded by the assumption that the sole objective of resource allocation is QALY maximisation. Further, it also does not avoid the question of how to value a QALY. (Brouwer et al. 2008)

2.4. Using economic evidence to inform resource allocation

The primary purpose of economic evaluation is to help decision makers address problems due of scarcity in health care resources. (Bryan, Williams, & McIver 2007) Sweden, Australia, and Ontario, Canada were among the first jurisdictions to use economic evaluation to inform health care resource allocation. (O'Donnell et al. 2009) Now, many countries have institutions that consider economic evidence for new and established interventions as a precursor to coverage and reimbursement decisions, or to issue recommendations for the efficient use of health care technology. (Clement et al. 2009;ISPOR 2011;Raftery 2008) Cost-effectiveness analysis, with QALYs as the preferred outcome measure, is the predominant methodology with the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia and NICE in the UK among HTA agencies that require such information. (NICE 2008a;PBAC 2010)

Clinical trials often provide incomplete or insufficient information for decision makers. It is often infeasible for a clinical trial to include all competing interventions, be of sufficient duration, and to include all relevant endpoints. (Drummond et al. 2005) Consequently, the use of decision analytic models has become commonplace as decision makers synthesise evidence into a single analytic framework to inform resource allocation.

Economic evaluation in health care has been defined as, "*the comparison of alternative options in terms of their costs and consequences*". (Drummond et al. 2005) Costs and consequences should be evaluated over an appropriate time horizon, i.e., one over which costs and consequences are likely to differ and thus are discounted accordingly. The fundamental aspects of an economic evaluation are presented in Figure 1.

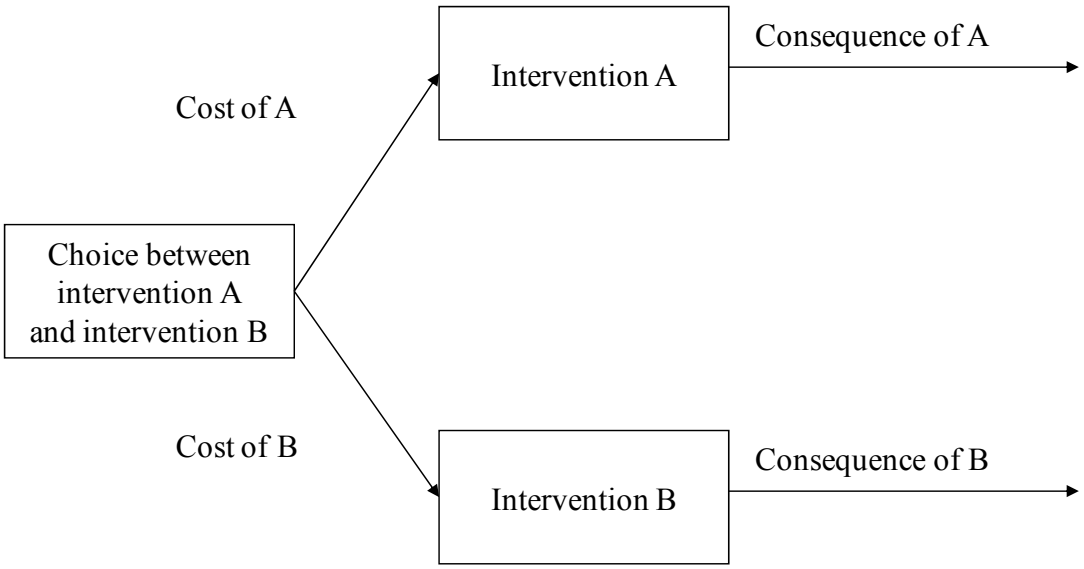


Figure 1. The fundamentals of economic evaluation (Drummond et al. 2005)

An economic model's structure is dependent on the clinical analytical problem. For example, decision trees are appropriate for acute clinical conditions, i.e., those for which a clinical resolution is reached in a short timeframe. Alternatively, Markov models can be used for chronic conditions where patients progress through a number of clinical, or Markov, states over a longer period of time. (Barton, Bryan, & Robinson 2004;Briggs, Claxton, & Sculpher 2006;Drummond et al. 2005) Other more complex structures are appropriate in certain circumstances. (Caro, Moller, & Getsios 2010;Duintjer Tebbens et al. 2008)

Cost-effectiveness is typically expressed using an incremental cost-effectiveness ratio (ICER). An ICER is the ratio of the difference in costs to the difference in effects between two competing interventions. The ICER can be formulated as follows:

$$ICER = \frac{Cost_{Intervention\ A} - Cost_{Intervention\ B}}{Effect_{Intervention\ A} - Effect_{Intervention\ B}}$$

The findings of cost-effectiveness analysis can be presented using a cost-effectiveness plane with each quadrant representing a potential outcome (Figure 2). Interpretation is clear when study findings fall in quadrant II or IV. In quadrant II the intervention is dominant, i.e., more effective and less costly than its comparator. In quadrant IV the intervention is dominated, i.e., less effective and more costly than its comparator. However, in order to interpret a study outcome when it falls in quadrants I (intervention is more effective and more costly than the comparator) or III (intervention is less effective and less costly than the comparator) is less straightforward as the decision maker must have some preference for the value of their chosen unit of health gain (e.g., a QALY). This valuation is referred to as the cost-effectiveness threshold (Section 2.6). (Buxton 2005;Eichler et al. 2004;McCabe, Claxton, & Culyer 2008)

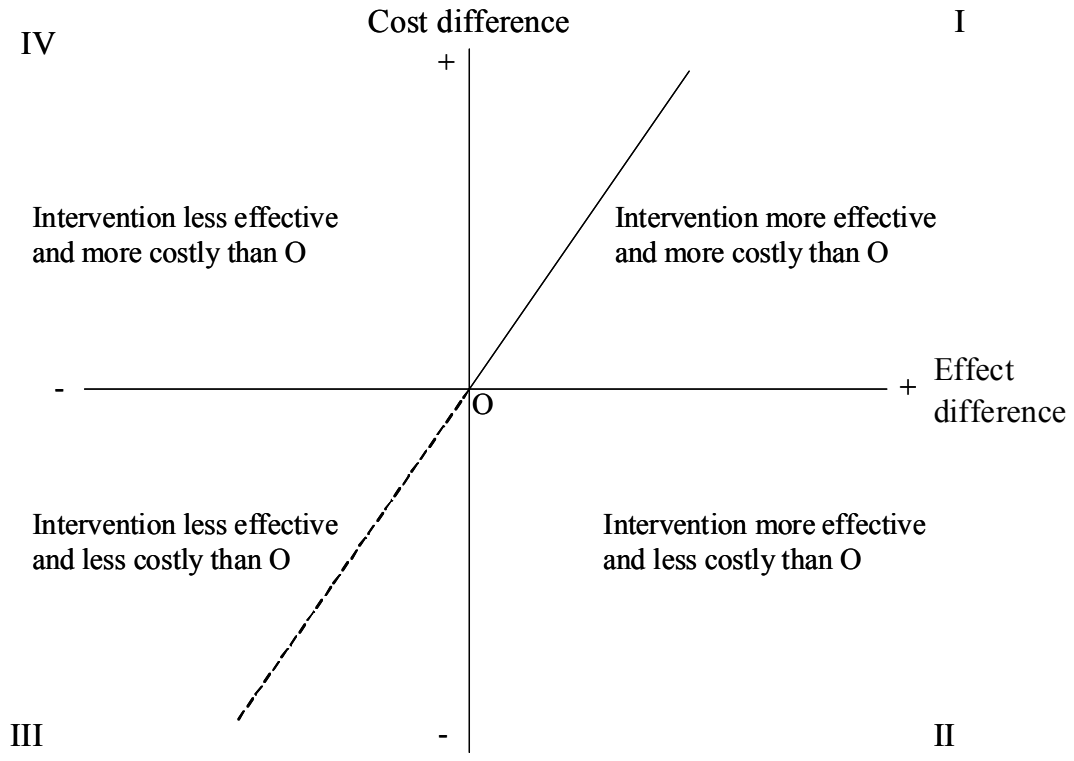


Figure 2. Cost-effectiveness plane (Black 1990)

2.5. Cost-effectiveness analysis

A worked example of using a cost-effectiveness decision rule based upon Lord et al. (2004) is presented below. (Lord, Laking, & Fischer 2004)

Included in this scenario are six health care programmes, $P_j (j = 1, 2, 3, 4, 5, 6)$, indicated for patients with a particular diagnosis (type P patients). E_j is the health gain associated with programme P_j and is measured in QALYs. C_j is the cost of programme P_j and is measured in monetary units ($). Table 1 and Figure 3 present a numerical example for 10,000 type P patients.

Three assumptions are necessary for this hypothetical example to be valid. (Johannesson & Weinstein 1993) First, available programmes are mutually exclusive, i.e., a patient may receive only one of them. As each patient must receive treatment, one of the available health care programmes must be 'best supportive care'. Second, health care programmes are perfectly divisible, i.e., a health care programme can be partially implemented while maintaining the characteristics of the entire programme. Third, each health care programme demonstrates constant returns to scale, i.e., costs and benefits are proportional to the scale of implementation. Therefore, treating half the eligible patient population incurs half the costs and yields half the benefits as would be the case if the entire patient population was treated. By maintaining these assumptions the cost-effectiveness frontier is piecewise-linear and convex in shape (Figure 3).

Table 1. Costs and effects of health care programmes available for four groups of patients

Patient Group	Health care programmes	Mean per patient		Total for group		Incremental analysis		
		Cost ($)	Effect (QALYs)	Cost ($ million)	Effect (QALYs)	Cost ($ million)	Effect (QALYs)	ICER ($ per QALY)
P-type	P_1	2,000	0.050	20	500	Baseline		
(n=10,000)	P_2	3,200	0.065	32	650	Ruled out through simple dominance		
	P_3	2,300	0.070	23	700	3	200	15,000
	P_4	3,350	0.106	33.5	1,060	10.5	360	29,167
	P_5	4,100	0.107	41	1,070	Ruled out through extended dominance		
	P_6	4,350	0.116	43.5	1,160	10	100	100,000
Q-type	Q_1	10,000	4.000	10	4,000	Baseline		
(n=2000)	Q_2	18,000	4.200	18	4,200	8	200	40,000
	Q_3	29,000	4.362	29	4,362	11	162	67,901
R-type	R_1	1,500	0.200	75	10,000	Baseline		
(n=100,000)	R_2	1,610	0.210	80.5	10,500	5.5	500	11,000
	R_3	1,700	0.212	85	10,600	4.5	100	45,000
S-type	S_1	100	0.005	10	500	Baseline		
(n=200,000)	S_2	105	0.007	10.5	700	0.5	200	2,500
	S_3	130	0.012	13	1,200	2.5	500	5,000

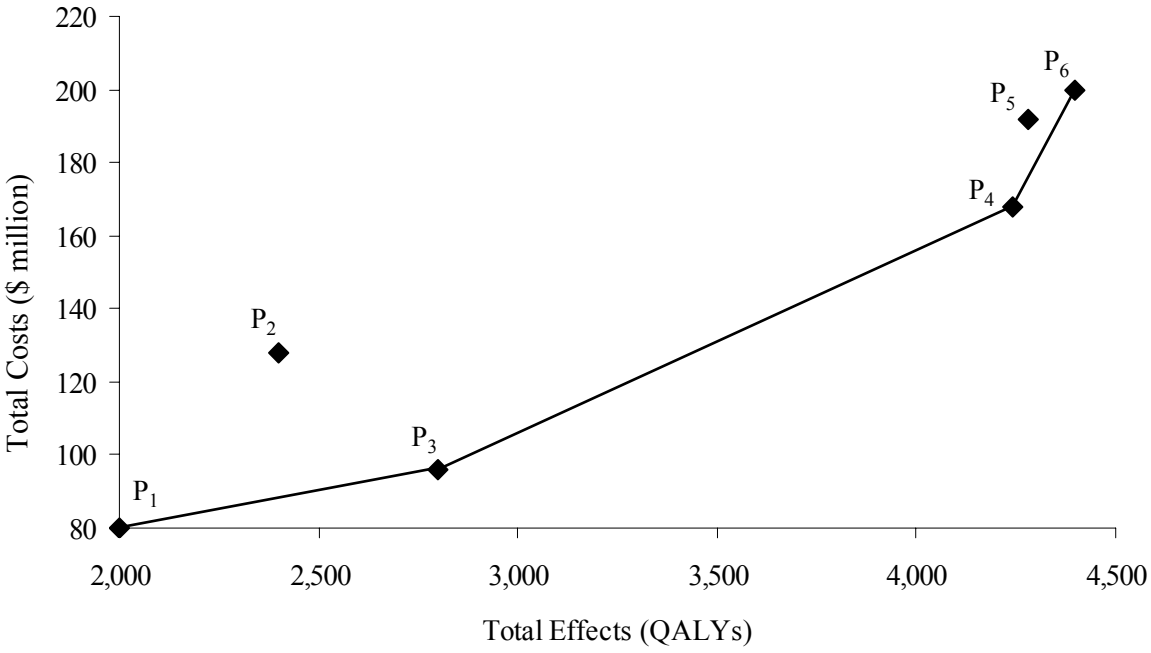


Figure 3. Costs and effects of programmes available to 'type P' patients

The decision maker's problem is how to allocate health care resources in a manner that maximises health gain. Two programmes can be immediately eliminated from consideration. First, P_2 is both more costly and less effective than P_3 ($3,200 vs. $2,300; 0.065 QALYs vs. 0.070 QALYs) and thus can be eliminated through 'simple dominance'. Second, P_5 can be eliminated through 'extended dominance', as a combination of two other health care programmes, P_4 and P_6, will yield greater benefits at less cost. If all 10,000 patients received P_5, 1,070 QALYs would be gained at a cost of $41 million. However, if 5,000 patients are treated with P_4 (530 QALYs; $16.75 million) and 5,000 with P_6 (580 QALYs; $21.75 million), aggregate QALYs gained will be greater and achieved at a lower cost (1,110 QALYs; $38.5 million). The remaining non-dominated health care programmes form a 'cost-effectiveness frontier', represented by the solid lines in Figure 3. The decision maker will maximise health gain by operating at a point on the frontier. The point on the cost-effectiveness frontier where the decision maker acts, and hence the composition of the health care programmes offered (where x% of patients will receive one programme and (100-x)% will receive the corresponding programme), is determined by the available budget. As the available budget increases or decreases, the

decision maker will move from left to right or from right to left on the frontier, respectively.

Table 1 includes the results of an incremental analysis of the available interventions. As described above, an ICER represents the additional cost required to generate an additional unit of health gain, e.g., a QALY. With respect to Figure 3, this represents the cost of purchasing an additional QALY by moving patients between adjacent non-dominated treatment options on the cost-effectiveness frontier. The ICER of programme P_k (k>1) in relation to the preceding programme P_j (j<k) can thus be defined as follows:

$$ICER_k = \frac{C_k - C_j}{E_k - E_j}$$

The calculated ICER is equivalent to the slope that joins points P_k and P_j in the cost-effectiveness space.

2.5.1. Resource allocation across multiple patient populations

In practice, a decision maker typically must make resource allocation decisions for multiple patient populations, each with multiple health care programmes available to them. To illustrate this scenario, Table 1 includes interventions for four patient populations; P-type, Q-type, R- type, and S-type patients. The cost-effectiveness frontiers for each of these patient populations can be drawn together in absolute cost-effectiveness space using incremental effectiveness and incremental cost for the x and y axes, respectively, with the least costly options for each population together at the origin (Figure 4). It has been noted, however, that having the least costly options together at the origin implies the availability of the 'costless bullet' (zero-cost, zero-effective option). (Briggs & Fenn 1997) Consistent with the assumptions presented above, each frontier in Figure 4 is piecewise linear and convex in shape.

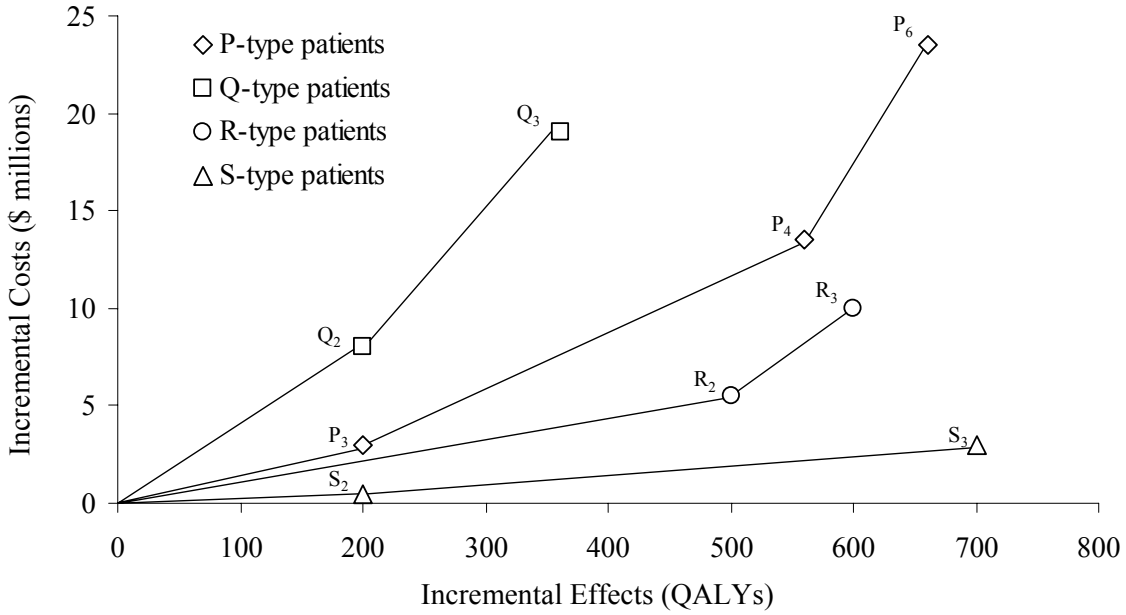


Figure 4. Cost-effectiveness space with cost-effectiveness frontiers for each patient population

2.5.1.1 League table approach

Weinstein and Zeckhauser first demonstrated that total health can be maximised (assuming perfect divisibility and constant returns to scale) through a league table approach. (Weinstein & Zeckhauser 1973) In a league table approach, all non-dominated interventions are ranked in order of their cost-effectiveness. Starting with the most cost-effective programme, health care programmes are implemented in rank order until the available budget is exhausted. This can be demonstrated numerically by extending the worked example (Table 2). The initial set of health care programmes are the least expensive (P_1, Q_1, R_1, S_1), and yield 15,000 QALYs at a cost of $115 million. The remaining health care programmes are ranked in order of cost-effectiveness with S_2 the marginal programme. When the health care budget is increased, patients are switched from S_1 to S_2, yielding in 15,200 QALYs at a cost of $115.5 million. If the budget is not

exhausted, the health care programme with the next lowest ICER is implemented. This process continues until the budget is exhausted.

The value of the ICER of the marginal programme, or the shadow price of the budget constraint, is referred to as the cost-effectiveness threshold. The threshold is a function of the costs and benefits of the available health care programmes and the available budget. Increasing the health care budget will result in the implementation of additional health care programmes with increasingly higher ICERs. Accordingly, the value of the cost-effectiveness threshold will also increase.

In practice, the league table approach is limited by the magnitude of data requirements. To construct a league table it is necessary to have complete knowledge of the costs and benefits of each available health care programme. If available data are incomplete then the approach will likely not lead to a QALY maximising resource allocation. Consequently, league tables are seldom used in practice.

Table 2. League table of available health care programmes

Set of health care programmes delivered	Marginal Programme	Incremental analysis			Total set of programmes	
		Cost ($ million)	Effect (QALYs)	ICER ($ per QALY)	Cost ($ million)	Effect (QALYs)
P_1, Q_1, R_1, S_1	-	-	-	-	115	15,000
P_1, Q_1, R_1, S_2	S_2	0.5	200	2,500	115.5	15,200
P_1, Q_1, R_1, S_3	S_3	2.5	500	5,000	118	15,700
P_1, Q_1, R_2, S_3	R_2	5.5	500	11,000	123.5	16,200
P_3, Q_1, R_2, S_3	P_3	3	200	15,000	126.5	16,400
P_4, Q_1, R_2, S_3	P_4	10.5	360	27,167	137	16,760
P_4, Q_2, R_2, S_3	Q_2	8	200	40,000	145	16,960
P_4, Q_2, R_3, S_3	R_3	4.5	100	45,000	149.5	17,060
P_4, Q_3, R_3, S_3	Q_3	11	162	67,901	160.5	17,222
P_6, Q_3, R_3, S_3	P_6	10	100	100,000	170.5	17,322

2.5.2. Mathematical programming

2.5.2.1 Overview

Mathematical programming is a theoretical approach for the comprehensive allocation of health care resources. An algorithm is used to allocate resources using rules consistent with those required for the league table approach described above. The formulation of the mathematical programming problem, as illustrated by Epstein DM et al, is presented below.

2.5.2.2 Mathematical programming formulation

The objective is to identify the pattern of resource use that maximises gross health benefit subject to the following constraints:

1. Budget constraint is not exceeded;
2. Each independent health care programme must be selected;
3. Each population group within the health care programme receives one and only one treatment.

The mathematical programming problem can be formulated as follows:

$$\max_{\Psi}\left(\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\sum_{i=1}^{I_k} e_{ijk}(t)x_{ijk}\right)$$

$\psi = (x_{ijk},\ i=1\ldots I_k, j = 1\ldots J_k, k=1\ldots K)$

$$\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\sum_{i=1}^{I_k} c_{ijk}(t)x_{ijk} \le \delta$$

$0 \le x_{ijk} \le 1 \qquad i = 1\ldots I_k, j = 1\ldots J_k, k=1\ldots K$

$$\sum_{j=1}^{J_k} x_{ijk} = 1 \qquad i = 1\ldots I_k, k=1\ldots K$$

It is assumed there are K health care programmes, and that each health care programme k (k: = 1…. K) has I_K population groups (i: = 1… I_K) and J_K treatments (j: = 1… J_K).

Variable x_{ijk} varies between zero and one ($0 \le x_{ijk} \le 1$). When $x = 0$ no proportion of population group i is allocated treatment j in health care programme k. When $x = 1$, the entire population group i is allocated treatment j in health care programme k.

Costs and benefits from treatment are evaluated over the model's time horizon (T). The model is static, i.e., the proportion of patients receiving a treatment does not change as a function of time. Variable *t* is the time index variable, where *t*: = 1… *T*.

Variable c_{ijk}(t) is the incremental cost of treatment *j* in health care programme *k* if the treatment is given to all members of population group *i* (both the pre-existing and newly diagnosed patient population) at time *t* (N.B. c_{ijk}(t) must be discounted to a fixed time point). Incremental cost, c_{ijk}(t), is defined as the difference in cost between each treatment *j* (*j*>1), and a comparator treatment (usually current care (j=1) for which costs are defined as zero).

Therefore, the total incremental cost in year *t* of all health care programmes and treatments can be formulated as follows:

$$C_{(t)} = \sum_{k=1}^{K} \sum_{j=1}^{J_k} \sum_{i=1}^{I_k} x_{ijk} c_{ijk}(t) \qquad \text{t} = 1 \ldots T$$

and, over the time horizon, the total incremental cost is bounded by the budget constraint δ.

Treatment benefit $e_{ijk}(t)$ is measured in quality adjusted life years (QALYs). The treatment benefit of each treatment *j* (*j*>1) is estimated in year *t* relative to a comparator treatment (*j*=1) for which treatment benefit is defined to be zero (N.B. e_{ijk}(t) must already be discounted to a fixed time point).

It is assumed that the cumulative incremental gain in QALYs is known only over the time horizon of the model. It is denoted by the time-invariant parameter b_{ijk}, where b_{ijk} is the gross benefit of treatment j in health care programme k if the treatment is applied to all of population group i.

Therefore, the total incremental QALY gain relative to current care may be formulated as follows:

$$B = \sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J_k}\sum_{i=1}^{I_k} e_{ijk}(t)\, x_{ijk} = \sum_{k=1}^{K}\sum_{j=1}^{J_k}\sum_{i=1}^{I_k} b_{ijk}\, x_{ijk}$$

Thus, above is the formulation of the problem of maximising health benefit while assuming that both costs and benefits show constant returns to scale.

2.5.2.3 Relaxation of assumptions in the mathematical programming formulation

Relaxation of perfect divisibility assumption

In the initial problem, it is assumed that a treatment can be partially implemented in the eligible patient population. This raises equity concerns as only a proportion of the eligible patient population receive treatment. Through mathematical programming, it is possible to relax the assumption of perfect divisibility to disallow the divisibility of health care programmes. This essentially imposes an additional "horizontal equity" constraint on the model that requires that people with equal need should have equal access to treatment. The effect of this constraint is that decision variables are binary for population i and health care programme j. This additional constraint can be formulated as follows:

$$x_{ijk} \in \{0,1\} \qquad i = 1\ldots I_k,\ J = 1\ldots J_k,\ k = 1\ldots K$$

The problem is now defined as a 0-1 mixed integer linear programme (MILP). Computationally complex, this method is only feasible for a limited number of decision variables taking binary variables. (Birch & Donaldson 1987;Birch & Gafni 1992;Epstein et al. 2007)

Accommodating alternative budget rules

Conventional methods of cost-effectiveness analysis assume no constraints with respect to the timing for which the budget can be spent. Typically, however, budgets exist over 12-month periods. Mathematical programming can account for multiple alternative budget rules. First, the available budget can be divided over the time period of the analysis. As per Epstein DM et al, the problem is illustrated below using a time horizon of 15 years. The constraint can be formulated as follows:

$C(t) = \delta/15 \qquad t = 1\ldots.15$

Second, in the case of this worked example, the total available budget must be exhausted within a 5-year period. In this scenario, health care programmes incurring a cost after the 5-year period are permitted only if other programmes generate the necessary cost-savings to offset this cost. This can be formulated as follows:

$$\sum_{t=1}^{5} C(t) = \delta$$

$C(t) = 0 \qquad t = 6\ldots.15$

Relaxation of constant returns to scale

The above approach is insufficient to allow the constant returns to scale assumption to be relaxed, i.e., that the costs and benefits of a health care programme are not proportional to the scale of its implementation. As relaxation of the constant returns to scale assumption requires nonlinearity, one approach to relax this assumption is to employ non-linear

programming. Non-linear programming increases the flexibility of the model and allows the benefits of a health care programme to vary with the scale of programme implementation. (Al, Feenstra, & Hout 2005;Elbasha & Messonnier 2004;Stinnett & Paltiel 1996) Another potential approach is mixed integer programming. This approach is appropriate when a number of the unknown variables are required to be integers and can be used when both the perfect divisibility and the constant returns to scale assumptions require relaxation. Such an approach is, however, computationally difficult. (Lord, Laking, & Fischer 2004;Stinnett & Paltiel 1996) This methodology can be expanded upon to include additional constraints, including the supply of health care professionals and the lead time necessary for training. (Earnshaw & Dennett 2003;Lord, Laking, & Fischer 2004;Sendi et al. 2003)

2.5.2.4 Summary

As illustrated above, mathematical programming facilitates relaxation of Johannesson and Weinstein's assumptions and provides an approach for comprehensive resource allocation. However, data requirements are equivalent to those for the league table approach, i.e., complete knowledge of the costs and benefits of available health care programmes, inhibiting the use of mathematical programming when used for the analysis of more than a limited number of programmes. Computational complexity also inhibits its use in practice when considering a multitude of health care programmes.

2.5.3. A cost-effectiveness threshold decision rule

The practical use of the league table and mathematical programming approaches is hindered by the magnitude of data requirements. Full information regarding the costs and benefits of each health care programme is required to comprehensively allocate health care resources. In reality, these data requirements prevent decision makers from implementing these methods and alternative approaches have been relied upon. (Drummond et al. 2005) The use of a cost-effectiveness threshold is one of these approaches. (Buxton 2005;Eichler, et al. 2004;McCabe, Claxton, & Culyer 2008)

If knowledge of the costs and benefits of all available interventions is unknown, a comprehensive ranking of health care programmes is unachievable. Consequently, a benchmark value, or cost-effectiveness threshold (λ), is required in order to interpret cost-effectiveness evidence. The cost-effectiveness threshold should represent a decision maker's valuation of a unit of health. If a health care programme is estimated to have an ICER greater than the threshold, it is not deemed representative of an efficient use of health care resources; alternatively, if the ICER falls below the threshold, it is deemed to be sufficiently cost-effective and thus representative of an efficient use of resources. Therefore, a health care intervention will be adopted only if the ICER of the programme is less than the specified cost-effectiveness threshold. That is:

$$ICER_j \leq \lambda^e$$

and

$$ICER_k > \lambda^e \text{ for all } k > j$$

λ^e = Estimated threshold

Using the example presented in Table 1 to evaluate P-type patients, if the estimated value of the threshold was $25,000 per QALY, all P-type patients would receive P_3, i.e., the intervention with the highest ICER less than λ^e. Accordingly, if the value of λ^e was increased to $30,000 then P-type patients would receive P_4 as this would now be the intervention with the highest ICER less than λ^e. As demonstrated by Johannesson and Weinstein (1993) and Lord et al. (2004), implementation of the threshold decision rule will lead to a QALY maximising allocation of health care resources providing the following conditions are met (Johannesson & Weinstein 1993;Lord, Laking, & Fischer 2004):

1. The threshold is correctly calibrated;
2. Health care programmes demonstrate perfect divisibility;
3. Health care programmes demonstrate constant returns to scale;

4. ICERs are correctly calculated.

2.5.4. Importance of the correct calibration of the threshold

As stated above, correct calibration of the threshold is a necessary condition for the threshold decision rule to result in a QALY maximising allocation of resources. The consequence of incorrect threshold calibration can be demonstrated by considering P-type patients in Table 1. (Johannesson & Weinstein 1993;Lord, Laking, & Fischer 2004) If the estimated threshold is not equal to the true threshold ($\lambda^e \neq \lambda$), i.e., the value of the threshold in a perfectly calibrated system, the application of a threshold decision rule will lead to inefficient resource allocation. To illustrate, assume that the correctly calibrated threshold is \$30,000 per QALY. If the decision maker uses an estimated threshold of \$25,000 per QALY, the decision rule will lead to the selection of P_3 (the health care programme with the highest ICER less than the estimated threshold ($ICER_j \leq \lambda^e$)). Treating all 10,000 p-type patients with P_3 would result in an aggregate gain of 700 QALYs at a cost of \$23 million. However, treating patients with P_3 will not exhaust the budget. If all p-type patients had instead been treated with P_4, the health care programme with the highest ICER less than the correctly calibrated threshold, aggregate health gain would have been 1,060 QALYs, 360 more than if patients had been treated with P_3. Similarly, overestimating λ will lead to an inefficient use of resources. If λ^e was estimated as \$100,000 per QALY the decision maker would implement P_6 for p-type patients. However, as the available budget (\$33.5 million) is insufficient to provide P_6 to all p-type patients (\$43.5 million is required) not all patients would receive care. In this case aggregate health gain would be 893 QALYs, 167 less than if p-type patients were treated with P_4.

2.5.5. Net benefit approach

If a decision maker places an explicit value on health, e.g., the QALY, it is possible to reorganise the ICER equation to present the same decision in terms of net benefit. The ICER equation is as follows:

$$ICER = \frac{Cost_{Intervention\ A} - Cost_{Intervention\ B}}{Effect_{Intervention\ A} - Effect_{Intervention\ B}}$$

When using a threshold decision rule the equation can be rearranged as follows:

$$\lambda^e = \frac{Cost_{Intervention\ A} - Cost_{Intervention\ B}}{Effect_{Intervention\ A} - Effect_{Intervention\ B}}$$

The threshold can essentially be considered an 'exchange rate' to convert units of effectiveness into a monetary value. Rearranging the equation results in the following:

$$0 \leq \lambda^e * (Cost_{Intervention\ A} - Cost_{Intervention\ B}) - (Effect_{Intervention\ A} - Effect_{Intervention\ B})$$

If net benefit is positive, implementing the intervention is an efficient use of resources; if negative, implementing the intervention is deemed a cost-ineffective use of resources. When considering multiple health care programmes, the threshold rule is equivalent to selecting the programme with the highest positive net benefit. This is presented geometrically in Figure 5. Line I_0 has a slope equal to the threshold (λ^e), in this case \$40,000 per QALY. The health care programme that will lead to QALY maximising resource allocation is that the greatest distance below I_0 on each cost-effectiveness frontier. In this scenario, the combination of health care programmes that yields the greatest amount of health gain are those shaded points on each frontier (P_4, Q_2, R_2, and S_3). (Ament & Baltussen 1997;Lord, Laking, & Fischer 2004)

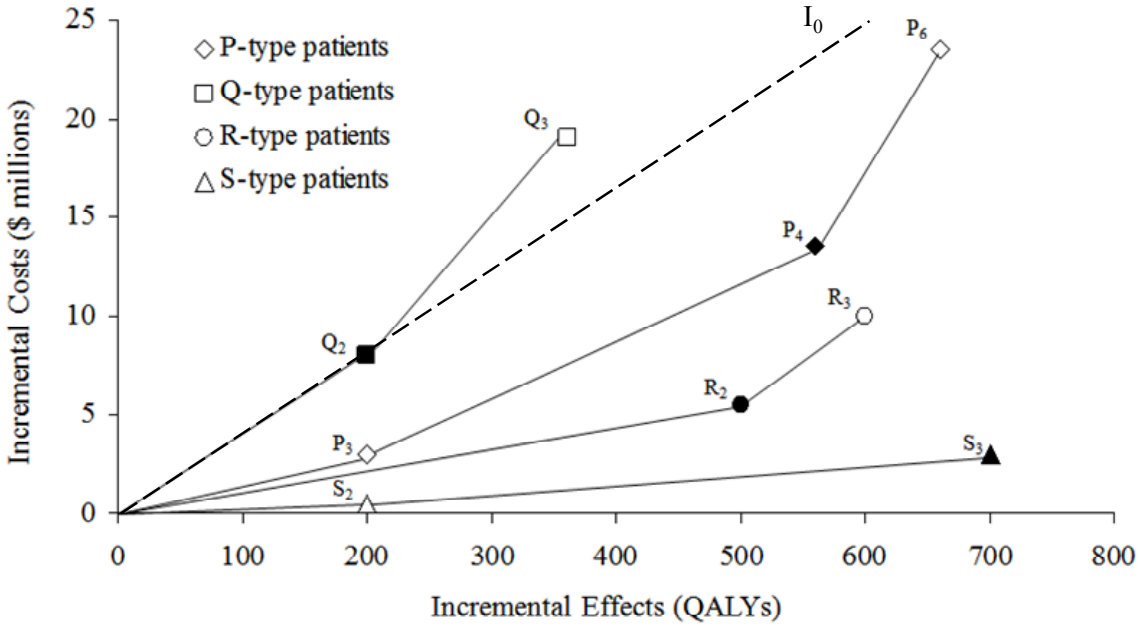


Figure 5. Resource allocation using a cost-effectiveness threshold as a decision rule

2.5.6. Relaxation of assumptions

Lord et al. (2004) illustrate that by relaxing the assumption of perfect divisibility, the threshold rule becomes unreliable. (Lord, Laking, & Fischer 2004) Although for the most part the perfect divisibility assumption affects only the marginal health care programme, knock-on effects to other cost-effectiveness frontiers may lead to inefficiencies when adopting the threshold rule. At the national level, however, where the budget is very large in relation to individual programmes, perfect divisibility is not an unrealistic assumption. (Lord, Laking, & Fischer 2004)

Relaxing the constant returns to scale assumption is potentially more problematic. In situations where non-constant returns to scale exist, the costs and benefits associated with the implementation of a health care programme will vary with the scale of implementation, i.e., the cost-effectiveness frontier would no longer be piecewise linear. When non-constant returns to scale exist, to estimate aggregate health gains and incurred

costs it is necessary to estimate the entire shape of the cost-effectiveness frontier, i.e., the costs and benefits of implementing each programme across all patients in a population. If non-linearities exist the opportunity cost of implementing a programme, and thus the value of the cost-effectiveness threshold, will vary at the margin. Consequently, when the constant returns to scale assumption does not hold, using a fixed threshold to guide resource allocation may lead to a suboptimal resource allocation. (Lord, Laking, & Fischer 2004)

2.5.7. Using cost-effectiveness information with incomplete information

2.5.7.1 Searching for a threshold

Culyer et al. (2007) use the example of NICE and the National Health Service (NHS) in the UK to illustrate how in the absence of complete information of the costs and benefits of available health care programmes, a decision maker should act as a 'threshold searcher' rather than setting a cost-effectiveness threshold. (Culyer et al. 2007)

As discussed above, if a decision maker with complete knowledge of the costs and benefits of all available health care programmes was to rank them in order of their cost-effectiveness and prioritise resources to the most cost-effective, the cost-effectiveness threshold would equal the ICER of the least cost-effective health care programme that the health care system could afford. In the UK the government is responsible for setting the NHS budget and, therefore, in this scenario the threshold is, by extension, set indirectly by the government. If the cost-effectiveness threshold was set independently by NICE, the health care budget would be a function of the threshold, as all health care programmes with an ICER less than or equal to the threshold would be implemented. Because setting the cost-effectiveness threshold is outside of NICE's mandate, NICE should instead act as a 'threshold searcher'. (Culyer et al. 2007;Karlsson & Johannesson 1996)

Figure 6 illustrates the hypothetical scenario of a health care system in which the costs and benefits of all health care programmes are known. The curve Ha represents the marginal health gain of currently implemented health care programmes and OE represents current expenditure. Ea represents the marginal health gain, or ICER, associated with the least cost-effective health care programme currently implemented. The ICER of this programme represents the shadow price of the budget constraint, i.e., the cost-effectiveness threshold.

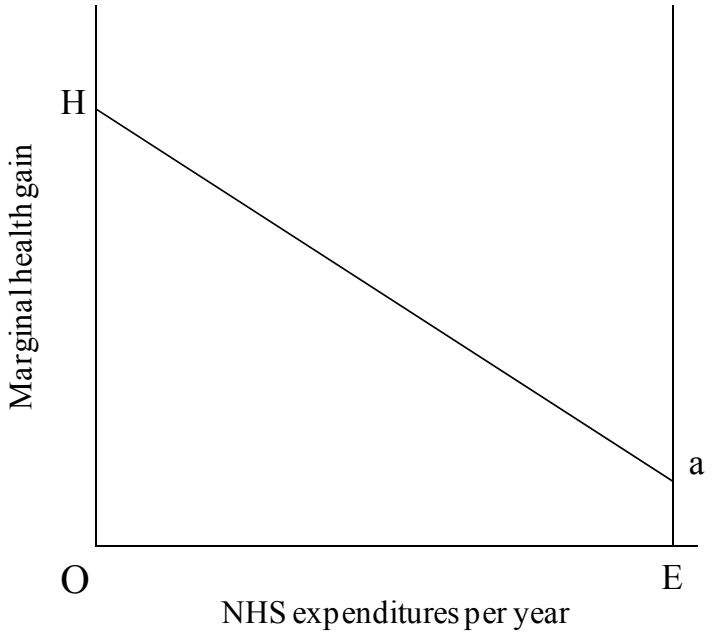


Figure 6. Marginal health gain associated with available interventions in the NHS

Figure 7 illustrates the more realistic analytical problem faced by NICE. As in Figure 6, curve Ha represents the marginal health gain of currently implemented health care programmes and OE the current expenditure. Curve cf represents available health care programmes not currently implemented in the NHS. A composite marginal health gain curve, Hde, is the horizontal sum of the curves Ha and ce and combines all available health care programmes, i.e., those currently provided in the NHS and those that are available but not currently provided. It is apparent from Figure 7 that there are three 'threshold' values relevant to the decision maker. The first is Ea, which represents the

marginal health gain of the least cost-effective health care programme currently implemented. The second is Ec, the marginal health gain associated with the most cost-effective health care programme not available in the current health care system. The third is Eb, which lies on curve Hde and is the threshold that would exist in a perfectly calibrated health care system. Implementation of a health care programme with a marginal health gain greater than Eb (in the range EE''), along with the displacement of a health care programme with a marginal health gain less than Eb (in the range E'E), will increase efficiency. To maximise health gained from current expenditure, the optimal solution is to disinvest in health care programmes that fall in the range E'E on curve Ha and substitute them for health care programmes that fall between EE'' on curve cf. It is suggested that NICE's role is to act within this range, referred to as the "zone of substitution". (Culyer et al. 2007;McCabe, Claxton, & Culyer 2008) In order to increase efficiency NICE should invest/disinvest in health care interventions that fall in the zone of substitution until such a point that additional investment/disinvestment will not result in efficiency gain. Consequently, rather than a constant fixed valuation, the value of the cost-effectiveness threshold should fluctuate as a function of the cost-effectiveness of currently available health care programmes, both implemented and not, and the health care budget.

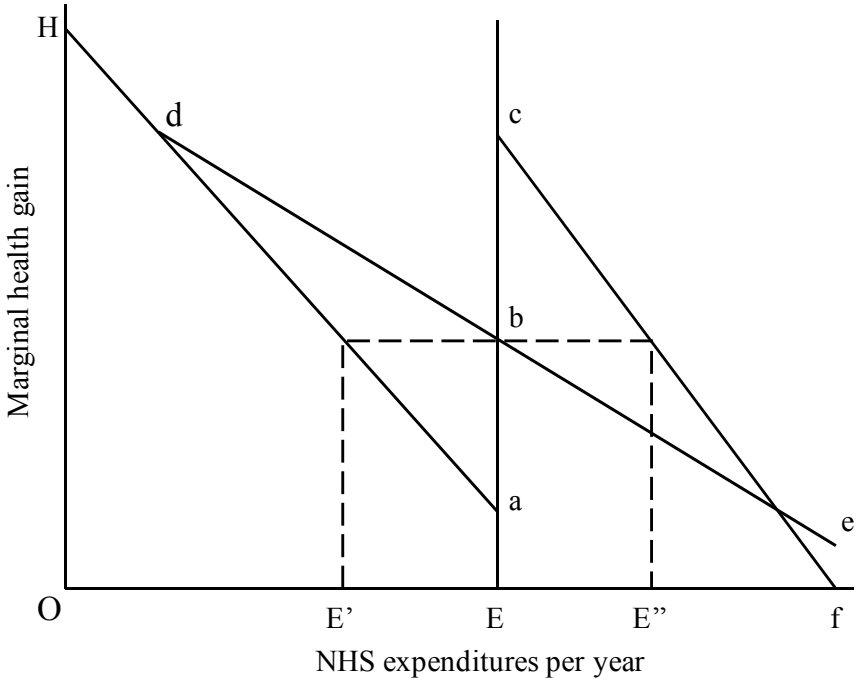


Figure 7. Representation of all health care programmes available to the NHS

2.5.7.2 The Sendi, Gafni and Birch (SGB) method

As illustrated above, indivisibilities, non-constant returns to scale, or a poorly calibrated threshold may result in the application of a threshold rule leading to suboptimal resource allocation. Proposed methods, such as integer programming, have the potential to account for these problems, but they lack feasibility, as they require complete knowledge of the costs and benefits of all available health care programmes. Sendi, Gafni and Birch (SGB) (2002) proposed a simple decision rule that eliminates these information requirements. (Sendi, Gafni, & Birch 2002) According to the SGB decision rule, a new health care programme should be implemented only after a less effective programme, or set of programmes, of equal cost is cancelled. By acting in accordance with the SGB decision rule, incremental improvements in efficiency would be achieved each time a new health care programme was implemented. However, the SGB method would only be optimising resource allocation if the marginal programme(s), i.e., the least cost-effective

programme(s) available, were displaced. It is unclear, however, how the marginal health care programmes would be identified.

Lord et al. (2004) demonstrate that the SGB decision rule may lead to an inefficient allocation of health care resources when the constant returns to scale assumption is invalid. Lord et al. show, however, that by estimating the costs and benefits of implementing the programme across all subgroups of the eligible patient population, the SGB decision rule can be improved upon and lead to a more efficient allocation of health care resources. (Lord, Laking, & Fischer 2004)

2.5.8. Disinvestment of health care programmes

The league table (Section 2.5.1.1), searching for the threshold (Section 2.5.7.1), and SGB (Section 2.5.7.2) approaches all require a disinvestment of existing health care programmes along with investment in more cost-effective programmes. The term disinvestment has been defined as, *"The processes of (partially or completely) withdrawing health resources from existing healthcare practices, procedures, technologies, or pharmaceuticals that are deemed to deliver little or no health gain for their cost, and thus do not represent efficient health resource allocation"*. (Elshaug et al. 2007) It has been suggested that up to 40% of patients do not receive treatments of proven effectiveness and up to 25% of treatments are unnecessary or even harmful. (Smith 1991;White 1995)

There are many challenges associated with the disinvestment of health care programmes. Notably, it has been suggested that there is a kink in consumers' threshold value for cost-effectiveness in health care, i.e., that society's willingness-to-accept (WTA) monetary compensation to forgo a health programme is greater than their willingness-to-pay (WTP) for a new health care programme of equivalent benefit. (O'Brien et al. 2002) This

phenomenon has implications for decision makers attempting to allocate resources in a manner consistent with societal preferences.

There is an increasing international awareness of the need to disinvest in health care programmes that are a relatively poor use of health care resources. A prominent example is NICE's database of "do not do" recommendations in which health care programmes determined to have little or no benefit are listed. (NICE 2010b) International experience, notably in Australia and Canada, has illustrated the challenges of identifying and removing candidates for disinvestment. (Elshaug, Hiller, Tunis, & Moss 2007;Elshaug et al. 2009;Garner & Littlejohns 2011)

2.5.9. Criticisms of cost-effectiveness analysis

With its theoretical underpinnings in an extra-welfarist framework, the objective of cost-effectiveness analysis is to allocate resources in order to maximise aggregate health. Given this sole objective, cost-effectiveness analysis does not account for other issues that may be important to the decision maker, e.g., equity concerns or societal preferences for health care. (Coast 2004;Dolan & Cookson 2000;Drummond et al. 2005;Nord et al. 1995;Ubel 2001) A disadvantage of the extra-welfarist framework is that it does not help inform resource allocation across industries. Although possible to compare the cost per QALY/life year ratios across the health care industry and some others (e.g., transport), cost-effectiveness analysis does not generate the necessary information to fully allocate resources at the national level. (Claxton et al. 2010) As described in Section 2.3.5, cost-effectiveness analysis considers a single unit of health outcome that typically has a single value across the population. It is possible, therefore, that cost-effectiveness analyses result in less favourable findings for treatments for the elderly or the severely ill as these populations are unable to accrue as much health as young healthy individuals. (Neumann & Greenberg 2009;Torrance 1986;Ubel 2001) It should be noted, however, that decision-makers that use cost-effectiveness evidence to inform resource allocation commonly account for criteria other than cost-effectiveness in their decisions, including equity and

distributional factors. (Anell & Persson 2005;Lopert 2009;NICE 2008a;Ramsberg et al. 2004)

Much of the criticism concerning cost-effectiveness analysis is that the information generated is insufficient to inform resource allocation when the decision maker is operating within a budget constraint, i.e., the question of affordability is not addressed. (Trueman, Drummond, & Hutton 2001) Birch and Gafni illustrate how solely considering cost-effectiveness evidence is insufficient to fully inform resource allocation. Without consideration of the underlying budget constraint, implementing a new health care programme with a positive incremental cost while continuing to provide access to all other available care will inevitably increase the overall cost of the health care system. (Birch & Gafni 2006a;Birch & Gafni 2006b) The same authors illustrate how accepting a programme based upon its ICER provides an indication of how much health gain is achieved from a unit of investment without any consideration as to the programmes forgone to pay for it. Thus, accepting a programme based solely on cost-effectiveness may not lead to an overall increase in efficiency. (Birch & Gafni 2006b) It has been illustrated, however, that concurrent consideration of cost-effectiveness and affordability can lead to health maximising decision resource allocation. (Culyer et al. 2007;Garber & Phelps 1997;Nuijten & Rutten 2002;Sendi & Briggs 2001;Trueman, Drummond, & Hutton 2001)

It has been suggested that the use of the ICER approach has led to an increase in overall health care expenditure in a number of countries. (Birch & Gafni 2004;Gafni & Birch 2003;Laupacis 2002) However, increased expenditure on health care is a global phenomenon and, to the best of my knowledge, the influence of using cost-effectiveness has not been evaluated. (OECD 2011) Indeed, Gold and Bryan suggest that in the US the alternative may be the case, i.e., that the failure to implement a health technology appraisal process has allowed health care costs to grow much more than would have been the case if a formal process of economic evaluation had been implemented. (Gold &

Bryan 2007)

A major drawback to using cost-effectiveness analysis is that, unlike for cost-benefit analysis, a clear decision rule does not always exist. For the most part, new health care programmes are associated with a positive incremental cost-effectiveness ratio (ICER), i.e., the programme is both more effective and costly than the comparator. As noted above, in the absence of a clear decision rule, it is necessary to compare cost-effectiveness findings against a benchmark value, or cost-effectiveness threshold. However, the value of the threshold and how it should be determined is a matter of controversy and debate. (Buxton 2005;Eichler et al. 2004)

In the next section, I review the approaches for the derivation of cost-effectiveness thresholds, suggested valuations, and various characteristics of them.

2.6. The cost-effectiveness threshold

In the absence of complete knowledge of the costs and benefits of available health care programmes, it is necessary to have a benchmark value, or decision rule, with which to interpret the findings of cost-effectiveness studies. This value, often referred to as the cost-effectiveness threshold, is of great importance. An incorrectly calibrated threshold value may have implications for decisions regarding the implementation of new technologies, disinvestment of existing technologies, and for the allocative efficiency of health care spending (Section 2.5.4).

Broadly, there are two schools of thought with respect to how the cost-effectiveness threshold should be valued. First, the cost-effectiveness threshold should reflect society's valuation of health. (Hirth et al. 2000) Second, the cost-effectiveness threshold should reflect the value of health care programmes displaced by the implementation of new programmes, i.e., the threshold should represent the opportunity cost with respect to health forgone. (McCabe, Claxton, & Culyer 2008)

The following sections describe approaches for valuing the cost-effectiveness threshold, examples of cost-effectiveness thresholds, characteristics of cost-effectiveness thresholds, and cost-effectiveness acceptability curves.

2.6.1. Deriving the cost-effectiveness threshold

2.6.1.1 Assigning a monetary value to life

Three broad approaches have been proposed to assign a monetary value to life; the human capital approach, the compensating wage method (revealed preference), and contingent valuation.

Human capital approach

The human capital approach values life by placing a monetary valuation on healthy time based upon an individual's future earnings. Therefore, use of a health care programme can be considered an investment in an individual's human capital. This concept of human capital can be used as the sole basis of valuing health improvements or as a method of valuing benefits accrued from using a health care programme (i.e., changes in productivity). (Drummond et al. 2005) The human capital approach is associated with a number of difficulties. Market wage rates typically vary, and it is difficult to account for variations due to race or gender. Also, it does not value healthy time not sold for a wage and thus may undervalue a health care programme's benefits. (Drummond et al. 2005)

The compensating wage method (revealed preference)

The compensating wage, or revealed preference, method is often used to estimate the value individuals' place on risk tradeoffs. This method evaluates individuals' employment behaviour to estimate the value they place on life. Essentially, the revealed preference method works as follows; two individuals are employed in identical jobs with the exception that one carries a higher risk of death or injury. The riskier job provides a higher salary than the lower risk job. The value the individual places on life can be estimated by multiplying the wage differential by the inverse of the difference in probability of death or injury. (See Appendix 1 for a worked example) (Brannon 2005; Drummond et al. 2005)

This form of estimation is consistent with welfare economics as it is based upon individuals' choices regarding trade-offs between the risk of death or injury and income. A strength of this method is that it is based on actual decisions involving health and money, rather than hypothetical scenarios. (Drummond et al. 2005) This method, though, does have a number of drawbacks. Individuals may not make rational employment choices and may not accurately perceive risk. (Brannon 2005) It is argued that revealed preference studies are biased as the wage of a particular job is just enough to entice the marginal worker; others require a higher wage to accept the same risk. (Shogren & Stano 2002)

Contingent valuation studies

Contingent valuation studies are often used to estimate the demand for non-market goods. With respect to health, although individuals are likely to place an infinite amount of money on their life, they often value small changes in the risk of death. (Brannon 2005) Using contingent valuation, the value of life is estimated by determining how much individuals would be willing to pay to avoid a certain level of added risk.

An advantage of contingent valuation is that it allows consideration of hypothetical scenarios, which is useful when evaluating health care interventions. (Diener, O'Brien, & Gafni 1998) In addition, this technique incorporates individual preferences into the analysis, consistent with welfare economics. A criticism of the contingent valuation technique is that it, as does not require actual cash transactions, it may not reflect individuals' true preferences. (Drummond et al. 2005) Further, as scenarios are hypothetical, responses may not be thoughtful or informed. It is necessary to control for factors that may affect the contingent valuation estimation, e.g., individuals with higher incomes are likely to have a higher willingness to pay for health. Also, the individual must understand risk, i.e., if an individual were willing to pay the same amount for a reduction in risk of 1/10,000 as for 2/10,000 then it would suggest that they are willing to pay for a general reduction in risk rather than valuing changes. (Brannon 2005)

2.6.1.2 Linking the threshold to GDP per capita

A conceptually appealing approach is to benchmark the cost-effectiveness threshold to GDP per capita. This approach allows a country's wealth to be accounted for in the threshold valuation and moves away from the tendency of setting thresholds as round numbers. (Weinstein 1996) (Section 2.6.2) Although arguably, benchmarking the threshold to GDP reflects a preference for using 'convenient' as opposed to 'round' numbers.

The cost-effectiveness threshold has been benchmarked to GDP per capita on two notable occasions: first, in the report by the WHO Commission on Macroeconomics and Health entitled, "Macroeconomics and health: investing in health for economic development"; second, in Alan Williams' 2004 lecture entitled, "What could be nicer than NICE?". (Sachs 2001;Williams 2004)

WHO Commission on Macroeconomics and Health - Macroeconomics and health: investing in health for economic development

In 2000, the Commission on Macroeconomics and Health, a group consisting of leading economists and health experts, met to discuss placing health at the centre of the development agenda. As part of the resulting proposed strategy, it was suggested that health should be valued explicitly in order to facilitate the economic analysis of health care programmes. The Disability Adjusted Life Year (DALY) was the recommended health outcome measure. The DALY is a unit of health that accounts for the present value of future years of lifetime lost through premature mortality adjusted for the severity of the illness or injury. (Fox-Rushby & Hanson 2001)

The commission suggested that based upon individuals' lost economic well-being as a result of disease, DALYs should be valued at three times GDP per capita. The valuation

was justified as follows: "*this multiple of earnings reflects the value of leisure time in addition to market consumption, the pure longevity effect, and the pain and suffering associated with disease*". (Sachs 2001) The valuation was supported by a variety of economic analyses. (Cutler et al. 1997;Philipson & Soares 2001;Topel & Murphy 1997) Absent from the report, however, is a description of the original analysis used to arrive at this figure. It would appear that rather than being determined from formal analysis, the value was derived from an informal meta-analysis of referenced articles. This was confirmed through communication with the authors via email (Appendix 2). The three-times-GDP threshold has since been cited in a variety of sources. (Access Economics Pty Limited 2004;Baltussen, Knai, & Sharan 2004;Dhanasiri & Puliyel 2007;Hoffman & Jackson 2003)

This valuation of a DALY was elaborated on as part of the WHO-CHOICE (CHOosing Interventions that are Cost Effective) programme. (Tan-Torres et al. 2003) Rather than taking a single value, the threshold was suggested to exist over a range: interventions were deemed highly cost-effective if costing less than one GDP per capita to avert a DALY, cost-effective if costing between one and three times GDP per capita to avert a DALY, and not cost-effective if costing more than three times GDP per capita to avert a DALY. (Tan-Torres et al. 2003;WHO 2005)

Alan Williams: What could be nicer than NICE?

In 2004, Professor Alan Williams presented a lecture entitled "*What could be nicer than NICE?*" Featured was discussion of the appropriateness of NICE's valuation of the cost-effectiveness threshold. It was suggested that £30,000 per QALY was "*far too high*" and that the threshold should be no more than 1xGDP per capita, approximately £18,000 in 2004 GBP. (Williams 2004) Williams argued that this was a "*common sense*" approach and that as GDP per capita should provide for all the needs of the average citizen (food, shelter, transport, education etc), the cost-effectiveness threshold should be no greater than this amount. Although society could afford to pay more than 1xGDP per capita

annually for a few individuals' health needs, it could not do so for many. Williams conceded that setting the threshold in this manner lacks a theoretical rationale, although he asserted that this was also the case for the then existing valuation. (Williams 2004)

2.6.1.3 Insight from other industries

A country's government must allocate available resources between industries. These investment decisions can provide an insight into how life is valued across sectors of the economy. In various industries it is necessary to place a monetary value on a year of life, or a life saved, when making investment decisions. Referred to as the value of a statistical life (VSL), such valuation is often used when evaluating potential investments in safety measures. (Viscusi & Aldy 2003)

Consideration of the VSL across industries may inform the appropriate value of the cost-effectiveness threshold in health care. It is logical that valuations across industries should have some degree of consistency. Loomes (2010) states, *"the cost effectiveness threshold should be set at a level consistent with the value attached to life in other parts of the public sector*". (Loomes 2010) For resources to be allocated efficiently across industries, the marginal benefit per dollar spent should be the same across programmes. (Tengs et al. 1995)

In the transport industry, investments are often made in safety measures to prevent injuries and fatalities. Cost-benefit analysis is the analysis of choice in the transport industry and, as described in Section 2.3.3, requires both costs and benefits to be measured in monetary terms. (Claxton et al. 2010) As the outcome of interest when evaluating safety measures is the prevention of a fatality, a VSL is required. Included in Table 3 are selected examples of VSLs used in the transport industry in the UK and Europe.

Table 3. Valuation of a statistical life in road and the railway transport industry.

Industry	Country	Valuation of a statistical life	Year	Reference
Roads	Europe	€1.1 - €1.3 million	2000	European conference of ministers of transport. Economic Evaluation of Road Traffic Safety Measures: Conclusion of round table. (Quinet 2000)
Roads	UK	£1,876,830	2007	Department of transportation – Guidance documents. (Department for Transport 2009)
Railways	UK	£10.8 million	2003	Fatal train accidents on Britain's main line railways: End of 2004 analysis. (Evans 2007)

2.6.2. Valuations of the cost-effectiveness threshold

Since Kaplan and Bush proposed the first cost-effectiveness threshold in 1982, various valuations have been proposed. (Kaplan & Bush 1982) Valuations may be explicitly stated by decision makers, inferred from previous decisions, or proposed by researchers. While explicit and implicit thresholds exist only in jurisdictions where economic evidence plays a role in decision-making, threshold values have been proposed to facilitate the interpretation of cost-effectiveness evidence even in jurisdictions where cost-effectiveness evidence plays a minor role (Table 4).

A common characteristic of the cost-effectiveness thresholds presented in Table 4 is that they are round numbers. As noted by Weinstein in reference to the threshold ranges proposed by Kaplan and Bush (1982) and Laupacis et al. (1992), although using different currencies, the valuations are essentially the same, *"in real terms the thresholds have changed, but the appeal of round numbers is long lasting"*. (Weinstein 1996) Indeed, it is suggested that the endurance of the $50,000 per QALY threshold value in a US health care system setting is due in large part to it being a 'round number'. This assertion is supported by the fact that the second most commonly used threshold in cost-effectiveness studies performed in a US health care system setting is $100,000 per QALY. (Greenberg, Winkelmayer, & Neumann 2006) The notion of round numbers is often commented upon and used to illustrate the arbitrary nature of many threshold values. (Bridges, Onukwugha, & Mullins 2010;Eichler et al. 2004;Evans, Tavakoli, & Crawford 2004)

Table 4. Explicit, implicit and assumed ICER threshold values in other countries (adapted from Cleemput et al. 2008)

Country	Institution	Author	Year	ICER threshold
Explicit ICER threshold range				
UK	NICE	NICE	2008	£20,000 - £30,000 per QALY
Implicit ICER threshold values or ranges based on past allocation decisions				
Australia	PBAC	Henry et al.	2005	AU$52,000 (approx) per QALY
	PBAC	George et al.	2001	AU$42,000 - AU$76,000 per QALY
New Zealand	PHARMAC	Pritchard et al.	2002	NZ$20,000 per QALY
Canada	CEDAC	Rocchi et al.	2008	Range of acceptance: dominant to CAN$80,000 per QALY Range of rejection: CAN$31,000 to CAN$137,000 per QALY (Rocchi et al., 2008)
UK	NICE	Towse & Pritchard	2002	£30,000 per QALY
ICER threshold values or ranges proposed by individuals or institutions				
UK	NICE	Williams	2004	1x GDP per capita (approx £18,000)
USA	NA	Various (Section 2.6.4.1)	1970s-	$50,000 per QALY
	NA	Kaplan & Bush	1982	$20,000 - $100,000 per Well-Year
	NA	Goldman et al.	1992	$20,000 - $100,000 per QALY
	NA	Kanis et al.	2002	$60,000 per QALY ($30,000 per QALY when not accounting for future costs)
	NA	Braithwaite et al.	2008	$109,000 - $297,000 per QALY
The Netherlands	College voor Zorgverzekeringen (CVZ)	The Council for Public Health and Health Care	2006	€80,000 per QALY
Developing world	NA	WHO	2003	3x GDP per capita
New Zealand	PHARMAC	Pritchard et al.	2002	NZ$20,000 per QALY (Pritchard, 2002)
Canada	NA	Laupacis et al.	1992	CAN$20,000 to CAN$100,000 per QALY

2.6.3. Characteristics of a cost-effectiveness threshold

2.6.3.1 Implicit vs. explicit thresholds

The decision maker makes explicit thresholds public and must abide by them in their decisions. If a coverage decision were made not in accordance with the explicit cost-effectiveness decision rule, justification would be required. In contrast, implicit cost-effectiveness thresholds are not made public, so identification is possible only through retrospective evaluation of established coverage decisions. (Eichler et al. 2004)

There are various theoretical advantages to operating an explicit cost-effectiveness threshold. (Coast 2001) An explicit threshold helps ensure consistency and transparency while decreasing burden on the decision maker. This, in turn, may increase efficiency, equity, and public trust. However, there are also disadvantages associated with explicit thresholds. Setting an explicit threshold would inevitably be politically sensitive and would require considerable political will to implement. Also, as many health care decision makers are not familiar with health economics, they may not be comfortable with the use of a threshold as a sole criterion, or a principal criterion, for decision-making. (Buxton 2005;Neumann 2005) Further, setting an explicit threshold raises issues regarding pharmaceutical pricing. An explicit threshold gives manufacturers information regarding the payer's maximum willingness to pay for a unit of health gain, thus providing an incentive to price products in such a way that computed cost-effectiveness is equal to the threshold. (Claxton et al. 2008)

To the best of my knowledge, the only institution that currently operates an explicit cost-effectiveness threshold is NICE in the UK. (NICE 2008a) This was, however, not always the case; rather, NICE originally claimed that a cost-effectiveness threshold was not in operation. (National Institute for Health and Clinical Excellence (NICE) 2001;Towse, Pritchard, & Devlin 2002) As discussed in Section 2.6.3.2, NICE's threshold currently ranges from £20,000 to £30,000. This valuation has been subject to scrutiny, with much debate focused on the threshold when used in the context of cancer treatments. Indeed,

pressure on NICE was such that with respect to treatments for end of life care, supplementary advice was issued to the Appraisal Committees providing guidance for under what circumstances a cost-effectiveness ratio in excess of the £30,000 upper bound of the threshold was permissible. (NICE 2009c;Towse 2009)

It is inevitable that use of an implicit threshold will lead to an explicit threshold. Given a sufficient sample size, and a degree of consistency with respect to a fixed decision rule, the value of an operated threshold will be identifiable through retrospective analysis. Attempts have been made to estimate the value of the threshold operated by decision makers in a variety of jurisdictions. Researchers have evaluated coverage decisions and recommendations made by NICE in the UK, the PBAC in the Australia, and CEDAC in Canada. (Dakin, Devlin, & Odeyemi 2006;Devlin & Parkin 2004;George, Harris, & Mitchell 2001;Henry, Hill, & Harris 2005;Rocchi et al. 2008) These studies are described in Section 4.5.1.

2.6.3.2 Hard and soft cost-effectiveness thresholds

Thresholds can also be characterised with respect to their rigidity. A 'hard' threshold has a fixed valuation and is unaffected by other factors. A 'soft' threshold is flexible and may fluctuate within a fixed range depending on the nature of the decision makers' problem. 'Hard' thresholds have the advantage of being transparent, consistent and predictable. In contrast, 'soft' thresholds allow for factors unrelated to cost-effectiveness to be accounted for, e.g., societal preferences. (Eichler et al. 2004)

As illustrated in Table 4, the majority of suggested cost-effectiveness thresholds are 'soft' in nature. The thresholds described by Kaplan and Bush (1982), Laupacis et al. (1992), Goldman et al. (1992), the WHO-CHOICE programme guidelines, and Braithwaite et al. (2008) all have lower and upper values. (Braithwaite et al. 2008;Goldman et al. 1992;Kaplan & Bush 1982;Laupacis et al. 1992;Tan-Torres et al. 2003) In the UK, NICE

uses a 'soft' threshold, ranging from £20,000 to £30,000, when evaluating health care programmes. In NICE's Guide to the Methods of Technology Appraisal it states, "*consideration of the cost effectiveness of a technology is a necessary, but is not the sole, basis for decision-making*". (NICE 2008a) With respect to the threshold range, NICE's methods guidance state the following:

- *"Below a most plausible ICER of £20,000 per QALY gained, the decision to recommend the use of a technology is normally based on the cost-effectiveness estimate and the acceptability of a technology as an effective use of NHS resources."*
- *"As the ICER of an intervention increases in the £20,000 to £30,000 range, the Committee's judgement about the acceptability of the technology as an effective use of NHS resources will make explicit reference to the relevant factors listed above."*
- *"Above a most plausible ICER of £30,000 per QALY gained, the Committee will need to identify an increasingly stronger case for supporting the technology as an effective use of NHS resources, with regard to the factors listed above."*

'*Relevant factors*' include: the degree of certainty around the ICER; whether there are strong reasons to indicate that the assessment of the change in HRQL has been inadequately captured, and may therefore misrepresent the health utility gained; and the innovative nature of the technology. (NICE 2008a)

Rawlins and Culyer (2004) first presented the rationale for NICE's use of a soft threshold, which has been referred to as a 'smudge' (Figure 8). (Rawlins & Culyer 2004;Towse & Pritchard 2002) An intervention with an ICER below inflection point A will almost certainly be deemed acceptable on the grounds of cost-effectiveness; above inflection point B, it will be rejected on the grounds of cost-effectiveness and will need to be a strong case supporting the technology as an effective use of NHS resources; between A and B, it will unlikely be rejected on the grounds of cost-ineffectiveness alone and

other factors will be taken into account. (National Institute for Health and Clinical Excellence (NICE) 2008;Rawlins & Culyer 2004)

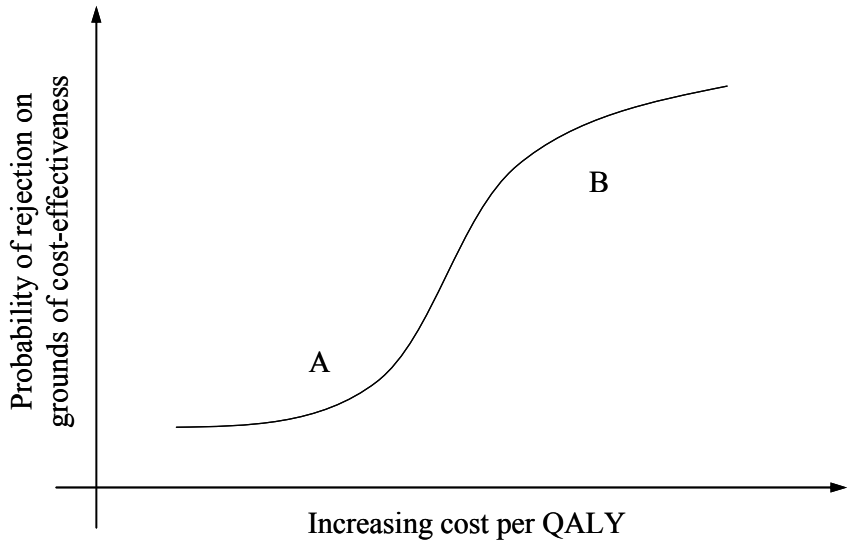


Figure 8. Probability of rejection based on the grounds of cost-effectiveness

As noted, NICE will consider other factors in the decision for technologies with an ICER above £20,000. Other than those factors stated in NICE's 2008 Guide to the Methods of Technology Appraisal, a number of special circumstances have been accounted for in technology appraisals. These special circumstances, as presented by Rawlins et al. (2010), are as follows: (Rawlins, Barnett, & Stevens 2010)

1. Severity of underlying illness.
2. End of life treatments:the public places special value on treatments that prolong life at the end of life.
3. Stakeholder persuasion, i.e., insights provided by stakeholders (e.g., patients, patient advocates, clinicians, NHS bodies, industry etc).
4. Significant innovation, i.e., the technology demonstrates distinct and substantive benefits not adequately captured in quality of life measures.

5. Disadvantaged populations: special priority to improving the health of the most disadvantaged members of the population.
6. Children: the assessment of improvement in quality of life is methodologically challenging; society would prefer to give 'the benefit of the doubt' to technologies affecting sick children.

Rawlins et al. (2010) presented some of the occasions when the above criteria were considered as part of the technology appraisal; these are presented below in Table 5.

Table 5. NICE's application of special circumstances in technology appraisals

Topic	ICER £ ('000s)	Severity	End of Life	Stakeholder persuasion	Significant innovation	Disadvantaged population	Children
Riluzole (motor *neurone disease*)	38-40	✓	✓	✓			
Trastuzumab (advanced breast cancer)	37.5	✓			✓		
Imatinib (chronic myeloid leukaemia)	36-65	✓			✓		
Imatinib (gastrointestinal stromal tumour)	NA	✓	✓		✓		
Pemetrexed (malignant mesothelioma)	34	✓	✓			✓	
Ranizumab (age-related macular degeneration)	>30			✓	✓		
Omalizumab (severe asthma)	>30	✓		✓	✓		
Sunitinib (advanced renal cancer)	50	✓	✓	✓	✓		
Lenalidomide (multiple myeloma)	43	✓	✓		✓		
Somatotropin (growth hormone deficiency)	NA			✓	✓		✓
Chronic subcutaneous insulin infusion (childhood Type 1 diabetes)	NA			✓			✓

2.6.4. ICER threshold values or ranges proposed by individuals or institutions

Various individuals and institutions have proposed ICER threshold values or ranges. As noted, in 1982 Kaplan and Bush were first to propose a cost-effectiveness threshold. (Kaplan & Bush 1982) The threshold was proposed for use in the setting of the US health care system and consisted of three levels: cost-effective by current standards if below $20,000 per Well-year; possibly controversial if between $20,000 and $100,000 per Well-year, but justifiable by many then current examples; and questionable (in comparison with other current health care expenditures) if greater than $100,000 per Well-year. (Kaplan & Bush 1982)

Laupacis et al. (1992) published guidelines for the use of cost-effectiveness evidence in the Canadian health care system in 1992. (Laupacis et al. 1992) As for Kaplan and Bush's proposal, Laupacis et al. (1992) proposed a threshold consisting of three levels: cost-effective if less than CAN$20,000 per QALY; moderately cost-effective if between $20,000 and $100,000 per QALY; and unlikely to be cost-effective if in excess of $100,000 per QALY.

Based in part upon the cost-effectiveness of implemented health care programmes, in 1992 Goldman et al. (1992) proposed a cost-effectiveness threshold for use in the US setting. (Goldman et al. 1992) The proposed threshold consisted of four levels: very attractive if below $20,000 per QALY; consistent with implemented programmes if between $20,000 and approximately $40,000 per QALY; in excess of the majority of implemented programmes if between $60,000 and $100,000 per QALY; and unattractive if above $100,000 per QALY. An interpretation of ICERs ranging between $40,000 and $60,000 per QALY was not provided.

Kanis et al. (2002) recommended a threshold value of $60,000 per QALY for the evaluation of osteoporosis treatments when accounting for future costs. When future costs were excluded a corresponding value of $30,000 per QALY was recommended. (Kanis et al. 2005;Kanis & Jonsson 2002)

The Council for Public Health and Health Care (Raad voor de Volksgezondheid en Zorg) is an independent body with the role of advising the Netherlands' government on public health and health care. The Council has stated that it is not entitled to define a cost-effectiveness threshold and that it should be determined through democratic discussion. However, to stimulate debate, in 2006 the Council suggested a value of €80,000 per QALY gained as the maximum acceptable cost-effectiveness ratio. Although the Council considers cost-effectiveness evidence when issuing its recommendations, it is claimed

that no threshold value is in operation. (Cleemput et al. 2008;Raad voor de Volksgezondheid en Zorg 2006)

As noted above (Section 2.6.1.2), on occasion, thresholds benchmarked to GDP per capita have been proposed. Sachs proposed a threshold of 3xGDP per capita per DALY averted for use in developing countries. This threshold was elaborated upon by the WHO-CHOICE, which proposed a three-level threshold (Section 2.6.1.2). (Tan-Torres et al. 2003;WHO 2005) Williams proposed a threshold of 1xGDP per capita per QALY for use by NICE in the UK. (Williams 2004)

2.6.4.1 The $50,000 per QALY cost-effectiveness threshold

Approximately half of all cost-utility studies published up to 2003 used the $50,000 per QALY benchmark value. (Greenberg, Winkelmayer, & Neumann 2006;Neumann et al. 2000) Although suggested that the $50,000 threshold originated in the 1970s or 1980s, a recent study claims that it was first used in 1992 in a study evaluating optimal management strategies for HIV patients. (Freedberg et al. 1992;Grosse 2008) However, it was not until 1996, following the report from the Panel on Cost-Effectiveness in Health and Medicine, that the $50,000 threshold per QALY became routinely used. (Gold et al. 1996;Grosse 2008;Siegel, Weinstein, & Torrance 1996)

While the origins of the $50,000 per QALY threshold are debated, many suggest that the valuation arose from using the cost-effectiveness of haemodialysis for the treatment of ERSD as the benchmark. (Hirth et al. 2000;Laufer 2005) The rationale is that as haemodialysis, with a cost-effectiveness ratio of approximately $50,000 per QALY, was a Medicare benefit, interventions of similar cost-effectiveness should be deemed sufficiently cost-effective. It should be noted, however, that there is uncertainty whether the $50,000 per QALY valuation was ever truly reflective of the cost-effectiveness of haemodialysis. (Bridges, Onukwugha, & Mullins 2010;Grosse 2008;Hirth et al. 2000)

Although the $50,000 per QALY threshold valuation may not have its foundations in society's willingness to pay for a QALY, or with the opportunity cost associated with investing in a new technology, it does provide a "rule of thumb" and some method to compare the cost-effectiveness of various health care technologies.

The $50,000 per QALY threshold has been extensively criticised in the literature. (Braithwaite, Meltzer, King, Jr., Leslie, & Roberts 2008;Bridges, Onukwugha, & Mullins 2010;Evans, Tavakoli, & Crawford 2004;Grosse 2008). If the threshold were benchmarked to the cost-effectiveness of haemodialysis, it would be logical to expect its value to increase in line with inflation and not remain static over time. Indeed, as Hirth et al. (2000) highlight, in 1997 dollars the value would have increased to an approximate value between $74,000 and $95,000. (Hirth et al. 2000) Also, if the threshold is linked to a Medicare benefit, it may be inappropriate to use it across sectors of the health care system other than Medicare. In 2005, Cutler suggested that the threshold should be much higher, proposing a value of $100,000. (Cutler 2005) In 2008, Brathwaite et al. used two approaches to assess the consistency of the $50,000 threshold with resource allocation decisions. (Braithwaite et al. 2008) The lower bound of the threshold was estimated using a comparison of the incremental cost-effectiveness of recent (2003) versus pre-"modern era" (1950) medical care in the United States; the upper bound was estimated using the incremental cost-effectiveness of unsubsidised health insurance versus self-pay for nonelderly adults (ages 21–64) without health insurance. Lower and upper bounds were estimated as $183,000 per life-year and $264,000 per life-year, respectively, notably higher than the $50,000 valuation. Brathwaite et al's suggested value of the upper bound is consistent with that proposed by Ubel et al. In 2003, Ubel et al. argued that thresholds of $50,000 and $100,000 were too low and suggested that medical practice reflects a valuation of a QALY much higher than $100,000. Based upon the median value of their review of behavioural and contingent valuation studies, Ubel et al. (2003) suggest a higher threshold of approximately $265,000. (Ubel et al. 2003)

2.6.5. Empirical work into the value of the threshold

There have been three notable attempts to empirically estimate the value of the cost-effectiveness threshold in the UK. In a series of studies, Martin et al. used a programme budgeting approach to model the link between health care spending and life years saved across various diseases (Martin S, Rice N, & Smith PC 2008a; Martin S, Rice N, & Smith PC 2008a). Appleby et al. (2009) evaluated coverage/disinvestment decisions made in the NHS by local decision makers to estimate the appropriate cost-effectiveness threshold. (Appleby et al. 2009) Baker et al. (2010) attempted to estimate the value of the QALY through the use of willingness-to-pay (WTP) and standard gamble approaches. (Baker et al. 2010)

Using programme budgeting data from Primary Care Trusts (PCTs) in England, Martin et al. performed a series of studies that provide insight into the appropriateness of the cost-effectiveness threshold operated in the UK. The authors used a theoretical model in which decision-makers are required to allocate a fixed budget across health care programmes in order to maximize social welfare while accounting for a health production function for each programme. In the first study, Martin et al. (2008a) used 2004/2005 PCT data to model the link between health care spending and life years saved for care related to cancer and circulatory diseases. The authors estimated the cost of saving a life year in cancer and circulation at approximately £13,000 and £8,000, respectively. (Martin S, Rice N, & Smith PC 2008a) In subsequent studies the same researchers updated the analysis using more recent data (2005/2006 and 2006/2007). In both instances, the findings were similar to those of the previous analysis in terms of the estimated cost of saving a life year in cancer and circulation. (Martin S, Rice N, & Smith PC 2008a; Martin S, Rice N, & Smith PC 2009)

Through the use of structured interviews, Appleby et al. (2009) evaluated coverage decisions made by local decision makers in order to gauge the appropriateness of the cost-effectiveness threshold operated by NICE. The research consisted of interviews

with senior staff from six NHS purchasers and 18 providers together with financial and public health information. Despite estimating the cost-effectiveness of a number of interventions, the researchers could not determine if they were truly the marginal available services in the NHS and thus could not definitively draw conclusions regarding the appropriateness of the existing value of the cost-effectiveness threshold. (Appleby et al. 2009)

The study by Baker et al. (2010) had two objectives: first, to identify characteristics of beneficiaries of health care over which relative weights should be derived and to estimate these relative weights; second, to determine the feasibility of using willingness-to-pay (WTP) and standard gamble approaches to estimate the value of a QALY. (Baker et al. 2010) Internet-based surveys were used for both aspects of the study. With respect to the relative weighting aspect of the study, the authors concluded that additional research is required to explore the methodological differences with respect to age and severity weighting. With respect to the valuation of a QALY, estimates ranged from values within NICE's existing range for the threshold to extremely high values. The authors did, however, state concerns regarding their measurement approach. (Baker et al. 2010)

2.6.6. Cost-effectiveness acceptability curves

As described, in the absence of complete information on the costs and benefits of available health care programmes, the cost-effectiveness threshold is necessary to interpret findings of cost-effectiveness analysis. However, as this chapter has shown, there is much debate with respect to the appropriate threshold value. Cost-effectiveness acceptability curves are proposed as a method for interpreting the findings of cost-effectiveness studies while conveniently evading the question of the value of the cost-effectiveness threshold. Cost-effectiveness acceptability curves present the probability of an intervention being the most cost-effective of those considered across a range of maximum willingness-to-pay values for a unit of health gain (Figure 9), thus avoiding the requirement of a single fixed threshold (Drummond et al. 2005;Fenwick, Claxton, &

Sculpher 2001;Fenwick, O'Brien, & Briggs 2004). Although a graceful way for analysts to avoid the value of a cost-effectiveness threshold, CEACs do not remove the need for decision makers to value a unit of health. (Buxton 2005)

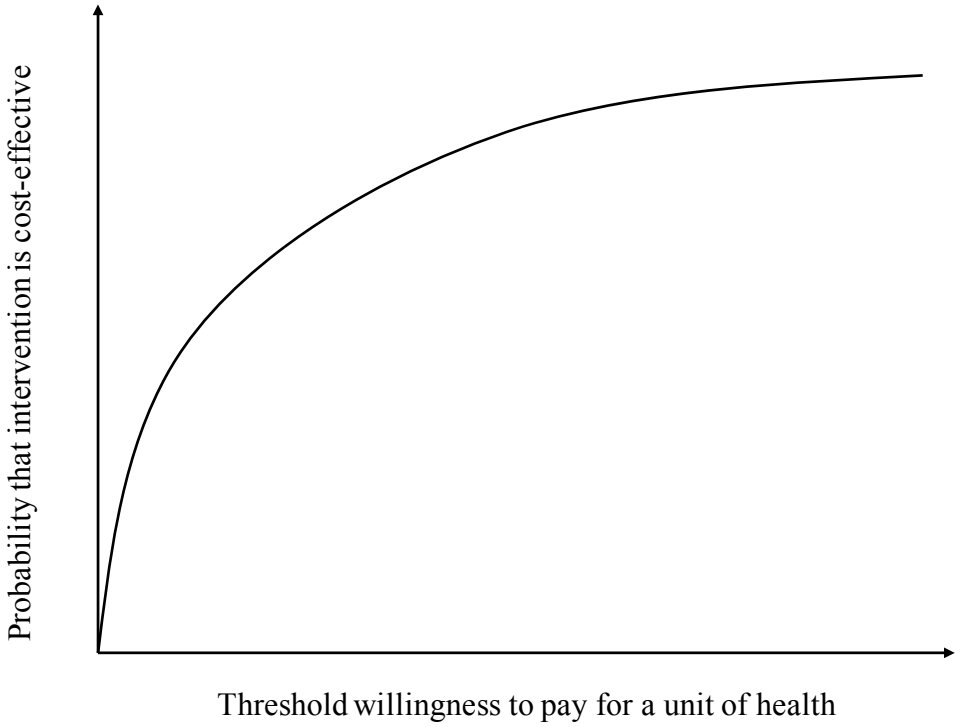


Figure 9. Cost-Effectiveness Acceptability Curve (CEAC)

The presentation of CEACs in published cost-effectiveness analyses is becoming more commonplace. A recent study showed the use of CEACs increasing, with inclusion in 32.6% of studies published in 2006 compared to only 2.1% of studies published in 2001 ($p<0.0001$). (Meckley et al. 2010) The presentation of CEACs is recommended by a number of institutions, including NICE in the UK and the PBAC in Australia. (NICE 2008a;PBAC 2008)

2.7. Chapter summary

There is increasing awareness that resource allocation must be addressed in a systematic rather than intuitive manner. While the reduction of waste in the health care system will allow greater opportunity for investment, this is likely to prove insufficient to curtail the rise in health care spending. (Donaldson et al. 2008;Garner & Littlejohns 2011) Difficult resource allocation choices are inevitable, and decision makers must choose between available interventions.

In this chapter, I have highlighted that markets are insufficient in health care and that economic evaluation offers an approach to the allocation of scarce resources. I present the framework underpinning cost-effectiveness analysis and have shown how cost-effectiveness information can be used to inform coverage decisions for medical technologies. The requirement for a cost-effectiveness decision rule, along with the various approaches to setting its valuation, is illustrated.

In the next chapter I describe various countries with respect to the relationship between health care spending and their relative rankings with respect to key health statistics. Then, I illustrate how economic evaluation is used to inform health care resource allocation in practice. I have chosen countries that illustrate the varying approaches to decision making for health care interventions, some that have embraced cost-effectiveness evidence in their health care system, i.e., the UK, Australia, Canada, and Sweden, and others in which cost-effectiveness evidence has been incorporated into decision making to a much lesser extent, i.e., France, Germany, and the US. Special attention is paid to the US health care system, in which, despite an apparent urgent need to achieve increased value from health care resources, cost-effectiveness evidence is often excluded from review. To gain insight into the resistance to using cost-effectiveness evidence, I review attempts to incorporate cost-effectiveness evidence into decision-making at Medicare and through the Oregon Health Insurance Experiment.

3. Practice

3.1. Introduction

In Chapter 2, I showed that reliance on market forces is insufficient to guide health care resource allocation. Economic evaluation offers an alternative approach and is used in a number of jurisdictions to inform the prioritisation of resources between competing interventions. I reviewed the underlying theoretical frameworks for economic evaluation in health care and illustrated how cost-effectiveness evidence can guide efficient health care resource allocation. I also illustrated the requirement for a cost-effectiveness decision rule, i.e., the cost-effectiveness threshold, described the various approaches for setting its value, and presented examples of cost-effectiveness thresholds used in practice.

In Chapter 3, I build on the theory presented in Chapter 2 and review how economic evaluation is used in practice to inform health care resource allocation across various countries. I review countries that help to illustrate the different approaches taken to using economic evidence, in particular cost-effectiveness evidence, in the prioritisation of health care resources. While some countries, namely the UK, Australia, Canada, and Sweden have embraced the use of economic evidence, other countries, namely France, Germany, and the US are notable for the limited extent that economic evidence is considered. The decision-making processes employed by these countries are described through comparing and contrasting varying approaches. Special attention is paid to the US health care system in which, despite a particularly apparent need to improve the return from health care spending, the use of economic evidence has a limited role.

To provide some background on the featured health care systems and some perspective on the institutions and processes for evaluating health care interventions, I first provide a number of comparative statistics. I describe the health care systems with respect to relative health care spending, abundance of resources, and performance, both in terms of health outcomes and rankings in terms of overall performance.

3.2. Health care system in context

In this section, I compare the health care systems of the UK, Australia, Canada, Sweden, France, Germany, and the US in terms of health care spending, abundance of health care resources, and performance.

3.2.1. Health care spending

While health care spending has increased at a faster rate than economic growth for all Organisation for Economic Co-operation and Development (OECD) countries, there is great variation between countries with respect to health care spending. In terms of percentage of gross domestic product (GDP), the US is by far the highest spender on health care (Figure 10). (Davis, Schoen, & Stremikis 2010;Pearson M 2009) In 2010, the US spent 17.4% of GDP on health care, almost 50% more than France, the next highest spender, and twice as much as Australia. (OECD 2011)

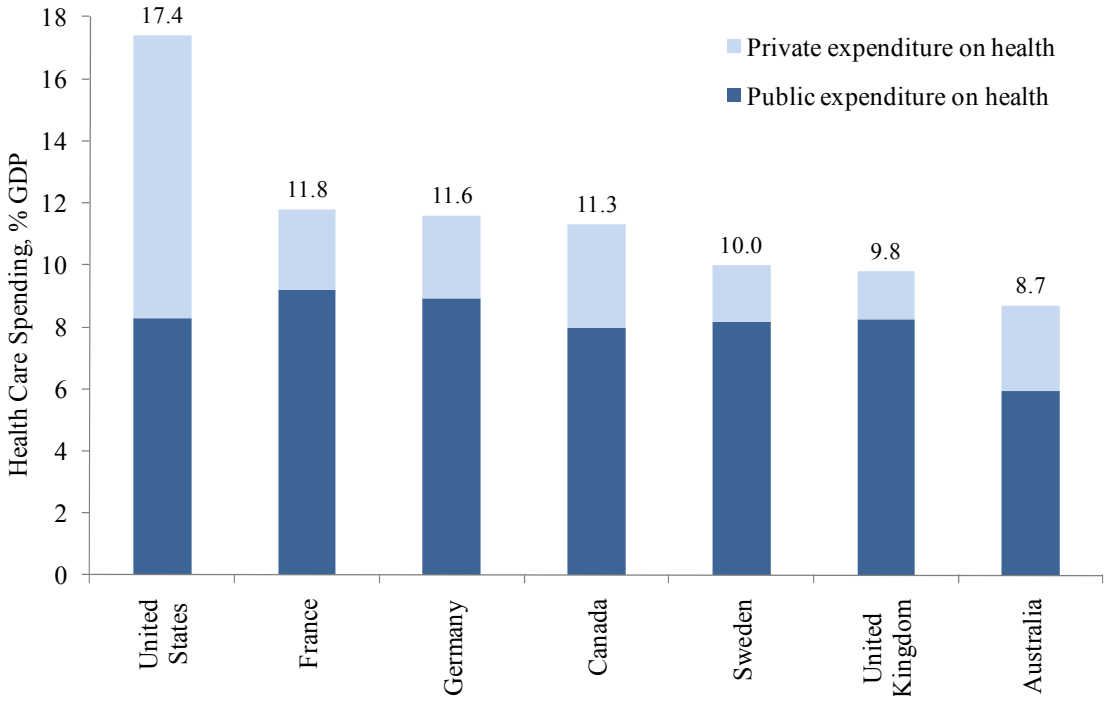


Figure 10. Health expenditure as a share of GDP, 2010 (or most recent year available) (OECD 2011)

In terms of per capita health care spending, the US remains a notable outlier (Figure 11). In 2010, per capita health care spending in the US was $7,960, more than twice per capita spending in France, Sweden, Australia, and the UK, and approximately $3,500 more than the next highest spender, Canada. Although the US has a higher income per capita than other countries, it has been suggested that this does not fully account for relative per capita spending. (Davis, Schoen, & Stremikis 2010)

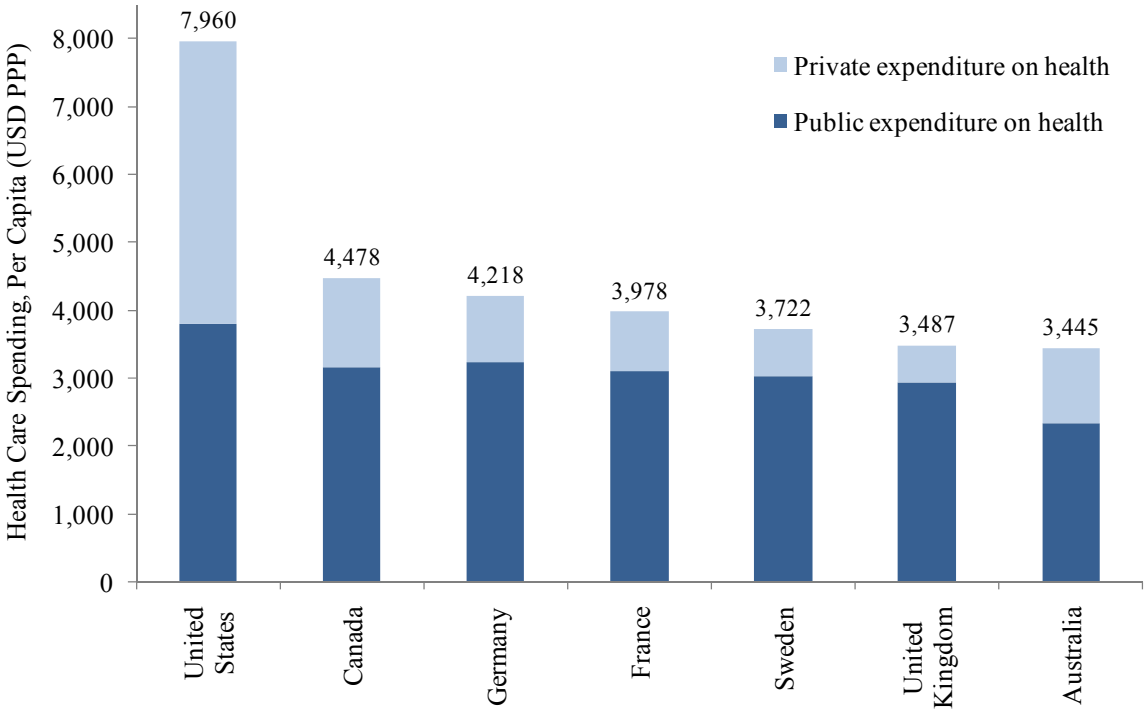


Figure 11. Health expenditure per capita, US$ PPP, 2010 (or most recent year available) (OECD 2011)

There has been much debate as to why US health care spending is much greater than spending in other countries. (Anderson et al. 2003;Neumann 2005;Neumann 2009;Pearson M 2009) Rather than a single explanation, it would appear that there are several contributing factors. There is evidence suggesting that medical procedures are performed much more frequently in the US than elsewhere and that the US pays more for medical procedures than other countries. (Antoniou et al. 2004;Peterson & Burton 2007) Outpatient care is utilised much more frequently in the US than elsewhere and is

estimated to represent the greatest difference in spending between the US and other nations. Indeed, elective interventions are estimated to have accounted for a quarter of the growth in US health spending between 2003 and 2006. (Farrell et al. 2008) Consistent with other developed countries, spending on prescription drugs has increased more rapidly than total health spending. Pharmaceutical spending per capita is, however, higher in the US than in other OECD countries. (Danzon & Furukawa 2003;Pearson M 2009) Although there are fewer physicians per capita in the US than compared to the UK and the OECD average (Section 3.2.2), physicians located in the US, and health care professionals in general, are more often paid wages above what would be predicted by US national income. (Cutler & Ly 2011;Peterson & Burton 2007) Administrative costs are often cited as a significant contributing factor of the overall cost of health care in the US. The cost of health care administration is twice as high in the US than the OECD average and represents 7% of total health care spending. (Cutler & Ly 2011;Pearson M 2009)

3.2.1.1 Future trends for health care spending

Concerns about US health care spending are not only due to the magnitude of current spending, but also with respect to the rate of growth. Globally, there is a trend for growth in health care spending as a percentage of GDP. However, the US has outpaced other countries with the percentage of GDP spent on health care almost doubling between 1980 and 2008 (Figure 12). Currently, more than 17% of GDP is spent on health care, and projections from the congressional budget office (CBO) suggest that by 2050 the percentage of GDP spent on health care will have reached 37%. (CBO 2007) Projected future spending on health care increased the urgency for US health care reform. (CBO 2007;Cutler, Davis, & Stremikis 2009;Orszag & Emanuel 2010;Presidential Executive Office 2009;Sutherland, Fisher, & Skinner 2009)

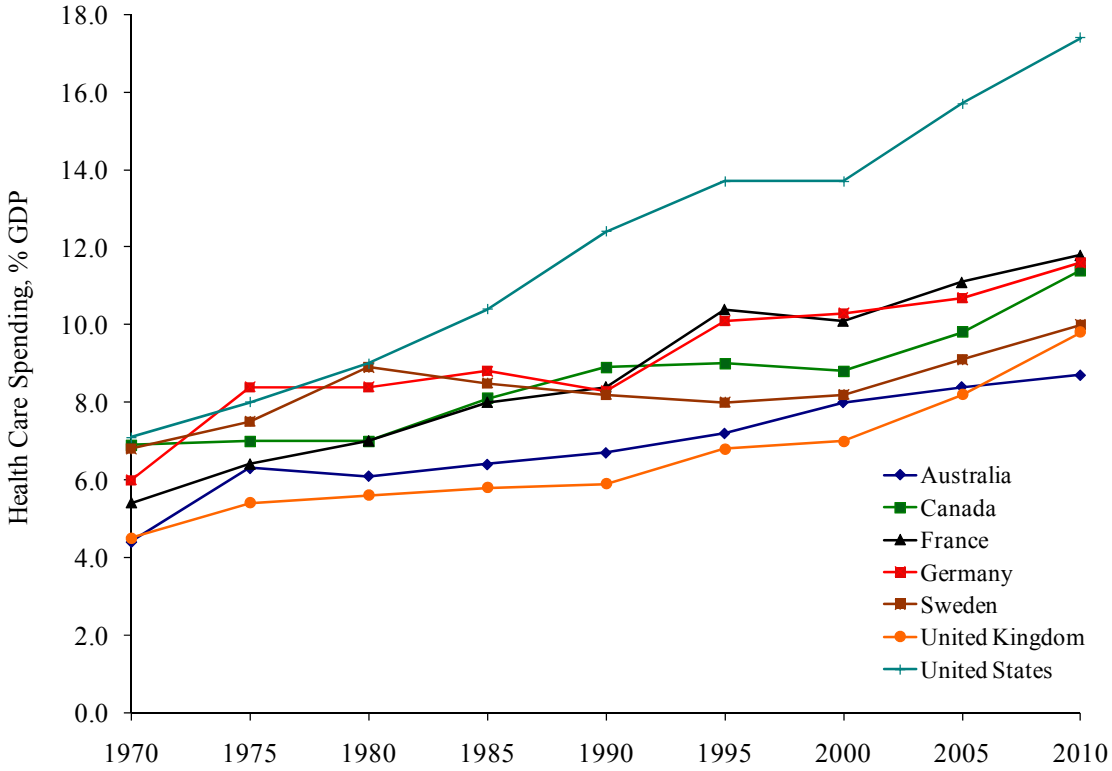


Figure 12. Increase in spending on health care as a percentage of GDP

3.2.2. Abundance of health care resources

Despite relative levels of spending, it is not necessarily the case that health care resources are more abundant in the US than in other countries. Table 6 illustrates the relative abundance of two key health care resources, physicians and hospital beds. Of the countries considered here, the US and Canada equivalently have the fewest physicians per 1,000 of the population. Further, only Sweden has both fewer physician consultations per capita and fewer hospital beds per 1,000 of the population than the US. The US fairs marginally better with respect to acute hospital beds but still lags behind other countries, with Germany, in particular, having twice as many acute hospital beds. (OECD 2011)

Table 6. Abundance of key health care resources

	Australia	Canada	France	Germany	Sweden	UK	US
Physicians per 1,000 population	3.0	2.4	3.3	3.6	3.7	2.7	2.4
Physician consultations per capita	6.5	5.5	6.9	8.2	2.9	5.0	3.9
Acute hospital beds per 1,000 population	3.5	1.8	3.5	5.7	2.0	2.7	2.7
Total hospital beds per 1,000 population	3.8	3.3	6.6	8.2	2.8	3.3	3.1

While abundance of physicians and hospital beds is informative, it does not account for the intensity of patient interactions with the medical system. Diagnostic imaging is one aspect where care delivery is more intense in the US than elsewhere. There are many more CT and MRI scanners per million of the population in the US than in other countries, with scans performed much more frequently as part of routine care. (Cutler & Ly 2011;OECD 2011) Also, the rate of certain surgical procedures is much higher in the US, with revascularisation procedures, knee replacements, and caesarean sections performed more frequently than in other countries. (Ko et al. 2007)

3.2.3. Key health statistics

The OECD produces statistics regarding the performance of health care systems using a variety of metrics. The most recent data was published in June 2011. (OECD 2011)

Table 7. Inter-country comparison of key health statistics – life expectancy and infant mortality

	Australia	Canada	France	Germany	Sweden	UK	US
Life expectancy*	81.6	80.7	81.5	80.3	81.5	80.4	78.2
Infant mortality*	4.3	5.1	3.7	3.5	2.5	4.6	6.5

2010 or most recent year available

3.2.3.1 Life expectancy

Over the latter half of the 20th century many countries achieved significant gains in average life expectancy. From1960 to 2009, average life expectancy increased in OECD countries by 11.2 years, from 68.1 to 79.3 years. In contrast, life expectancy in the US increased by only 8.4 years, from 69.8 to 78.2 years. Notably, average life expectancy in the US has fallen from being 1.7 years longer than the OECD average in 1960 to 1.1 year less than the OECD average in 2009. (OECD 2011) Of the countries discussed above, the US has the shortest average life expectancy, 3.4 years less than average life expectancy in Australia, the country with the highest out of all OECD countries (Table 7).

3.2.4. Infant mortality

A pattern similar to average life expectancy emerges when considering infant mortality. Considering the OECD average, infant mortality rate has declined drastically from a rate of 40.4 deaths per 1,000 live births in 1960 to 4.3 deaths per 1,000 live births in 2009. In the US, infant mortality fell from 26.0 to 6.5 deaths per 1,000 live births (2008 most recent data available), notably higher than for other countries considered here (Table 7).

3.2.5. Global rankings

Often claimed to be the world's best health care system, recent studies have shown that the US health care system ranks unfavourably against others across a variety of criteria. In 2000, the World Health Organization (WHO) published its widely cited rankings of health care systems, and the US health care system placed at number 37. Ranking was based upon 'overall efficiency' with a single composite score calculated from five indicators: health, health quality, responsiveness-level, responsiveness-distribution, and fair-financing. France was ranked as the health care system with the highest efficiency; the UK was ranked 18th (Table 8). (World Health Organization 2000)

Table 8. World Health Organization (WHO) rankings of health care systems

Country	WHO Ranking
France	1
UK	18
Sweden	23
Germany	25
Canada	30
Australia	32
US	37

The Commonwealth Fund is a US-based foundation that promotes a high performing health care system that achieves better access, improved quality, and greater efficiency. (The Commonwealth Fund 2010) Since 2004, the Commonwealth Fund has made four attempts to rank various health care systems. (Davis et al. 2006;Davis et al. 2007;Davis, Schoen, & Stremikis 2010;Hussey et al. 2004) The most recent rankings, published in June 2010, include seven countries: Australia, Canada, New Zealand, the United Kingdom, Germany, the Netherlands, and the US. (Davis, Schoen, & Stremikis 2010) A

ranking is decided upon using the following criteria: quality of care; access; efficiency; equity; and long, healthy, and productive lives. The US ranked last in 2010 in terms of overall rankings, consistent with the 2006 and 2007 study findings (an overall ranking was not presented in the 2004 report). (Davis et al. 2006;Davis et al. 2007) Despite being the most expensive in the world, the comparative analyses by the WHO and Commonwealth Fund suggest that the US health care system underperforms relative to other countries across the majority of performance dimensions. (Davis, Schoen, & Stremikis 2010)

3.3. Using economic evaluation to inform resource allocation

The evaluation of medical technology, commonly referred to as health technology assessment (HTA), is a global practice, as an HTA agency exists in virtually every developed country. HTA is a term used to encompass multiple aspects of decision-making and is defined by the International Network of Agencies for Health Technology Assessment (INAHTA) as, "*a multidisciplinary field of policy analysis, studying the medical, economic, social and ethical implications of development, diffusion and use of health technology*". (INAHTA 2011;Luce et al. 2010)

There is, however, much inter-country variation with respect to HTA activity. A notable source of this variation is the type of evidence included in the assessment. While there is broad consistency between HTA agencies with respect to consideration of safety and efficacy evidence, there is much variation how, and the extent to which, cost-effectiveness evidence is considered. While in some countries cost-effectiveness evidence is a fundamental part of technology assessment, in other countries it plays a lesser role. To illustrate, I have chosen countries that best highlight inter-country differences with respect to decision-making criteria. To this end, the UK, Sweden, Australia, and Canada serve as examples of countries in which cost-effectiveness evidence plays an integral role in decision-making. In contrast, Germany, France, and the US serve as examples of countries in which cost-effectiveness evidence plays less of a role in decision-making.

3.3.1. Countries in which cost-effectiveness evidence plays an integral role in decision-making

In a number of countries, cost-effectiveness evidence plays an instrumental role in coverage and reimbursement decisions, or in recommendations for the efficient use of medical technology. The use of cost-effectiveness evidence in decision-making in the UK, Sweden, Australia, and Canada is presented below.

3.3.1.1 UK and NICE

Health care in the UK is dominated by the National Health Service (NHS) which provides health care to approximately 60 million people. One of the NHS's fundamental principles is universal access to care regardless of ability to pay. (Boyle 2011a;NHS 2011) The NHS is funded through general taxation, although approximately 12% of the population is also covered through private medical insurance. (Boyle 2011b) Health care spending in the UK consumes 9.8% of GDP (2009 data) and the UK health care system was ranked 18th in the WHO's 2000 global ranking of health care systems. (OECD 2011)

April 1999 saw the introduction of the National Institute for Clinical Excellence (NICE). Established as a Special Health Authority, NICE was founded to eradicate the 'postcode lottery', terminology used to describe NHS patients' variable access to medical technology contingent on where they lived. (NICE 2011c) In 2005, the institute merged with the Health Development Agency (HDA) and was renamed the "National Institute for Health and Clinical Excellence". NICE's principal functions are to provide guidance to the NHS with respect to public health, health technologies, and clinical practice. Its most notable function is the technology assessment programme through which NICE is commissioned by the Department of Health (DoH) to evaluate new and existing medical technologies. (NICE 2009a;NICE 2009b;NICE 2011a) NICE is renowned for its open and transparent process, with representatives from the health service, industry, patient advocacy groups and the public providing input. (NICE 2009a;NICE 2009b;NICE 2010a) NICE does not have the authority to restrict access to medical technologies in the NHS; rather, British law dictates that the NHS must provide funding for medical technologies for which NICE issues a positive recommendation. (NICE 2011b;Sorenson et al. 2008) Further, NICE does not have the authority to negotiate, or set, the price of medical technology, though its role is evolving towards a policy of Value Based Pricing (VBP). (DoH 2010) NICE evaluates interventions through one of two processes, single technology appraisal (STA) or multiple technology appraisal (MTA). The STA process

is designed to appraise a single product, device, or other technology with a single indication for which most of the relevant evidence lies with one manufacturer or sponsor. (NICE 2006) The MTA process is designed to appraise single or multiple products, devices, or other technologies with one or more indications. An independent academic group performs the health technology assessment, and additional evidence is sought from selected clinical specialists, NHS-commissioned experts, and patient experts. (NICE 2009a)

Decision-making criteria

NICE is noted for the significant role that cost-effectiveness plays in its recommendations and to the extent to which its methods are in accordance with economic principles. To ensure consistency across appraisals, NICE has adopted the approach of using a 'reference case'. The reference case lays out NICE's requirements for key aspects of their appraisal, e.g., costs should be considered from the perspective of the NHS and the Personal Social Services (PSS), health benefits should be measured using QALYs, costs and health benefits should be discounted at an annual rate of 3.5%, etc. (NICE 2008a) NICE accounts for the opportunity cost of implementing a new technology in the NHS through the use of a cost-effectiveness threshold. To the best of my knowledge, NICE is the only agency that operates an explicit cost-effectiveness threshold. Cost-effectiveness is not, however, the only decision-making criterion. In the Guide to the Methods of Technology Appraisal, NICE states, "*consideration of the cost effectiveness of a technology is a necessary, but is not the sole, basis for decision-making*". (NICE 2008a) NICE have laid out the other factors important in the appraisal and described how social value judgements should be incorporated. (NICE 2008b)

Although all available evidence is considered in the appraisal[i], NICE has a strong preference for head-to-head RCTs. In NICE's appraisals incremental effectiveness data

[i] NICE's reference case states that synthesis of evidence on outcomes should be based on a systematic review, and, when necessary, indirect and mixed treatment comparisons.

are synthesised along with incremental cost data to generate the estimate of cost-effectiveness. Potential budget impact does not determine the outcome of NICE appraisals, yet in the Guide to the Methods of Technology Appraisal, NICE states*, "the [Appraisal] Committee will want to be increasingly more certain of the cost effectiveness of a technology as the impact of the adoption of the technology on NHS resources increases"*. (NICE 2008a) Indicative of the role other factors play in decision-making, NICE does not operate a fixed threshold. Rather cost-effectiveness is considered over a range (see section 2.6.3.2). Below an ICER of £20,000 per QALY gained, the NICE Appraisal Committee's recommendation is normally largely based on cost-effectiveness. Above an ICER of £20,000 per QALY gained, other factors are accounted for, including degree of certainty around the ICER, whether HRQL has been inadequately captured, and the innovative nature of the technology. (NICE 2008a) Above an ICER of £30,000 per QALY gained, the Appraisal Committee needs an increasingly stronger case that the technology is an effective use of NHS resources. As described in Section 2.6.3.2, in addition to the factors stated above, special circumstances have been accounted for in a number of NICE' recommendations. (Rawlins, Barnett, & Stevens 2010)

3.3.1.2 Sweden and TLV/SBU

The Swedish health care system is built around a principle of universal coverage for all members of society. Sweden spends approximately 10% of GDP on health care with around 70% of health care services funded through local government taxes. (Anell 2009;OECD 2011) The vast majority of health care is provided by publicly funded entities with only about 10% provided by privately funded entities.(Anell 2009) Sweden was ranked 23rd in the WHO's 2000 global ranking of health care systems. (OECD 2011;WHO 2000)

Two agencies perform HTA in Sweden, the Dental and Pharmaceutical Benefits Board (Tandvårds- och läkemedelsförmånsverket), or TLV, and the Swedish Council of Technology Assessment in Health Care, or SBU. Established in 2002, the TLV is an

independent authority under the Department of Health and Welfare and is financed through government grants. Before a drug can be included in Sweden's Pharmacy Benefit Scheme (PBS), it must first be approved by the TLV. (TLV 2011) Established in 1987 by the Swedish government, the SBU was charged with evaluating the effectiveness and value of medical technology and providing guidelines for evidence-based medicine to the county councils and medical community. (Jonsson 2009b) Although SBU publications have no direct mandate for influencing drug reimbursement, the TLV may take their recommendation into account. (TLV 2011)

Decision-making criteria

The TLV evaluates the clinical and cost-effectiveness of new drugs. (Anell & Persson 2005) The evaluation is performed from a societal perspective that includes productivity costs and the impact on a patient's family and carers. Three broad criteria are used to evaluate technologies (Anell & Persson 2005): first, human value, i.e., health care is to be provided equally to all members of society; second, need and solidarity, i.e., those with the greatest need for health care receive more resources than others; third, cost-effectiveness, i.e., drug costs must be reasonable from medical, humanitarian, and socioeconomic standpoints. Generally, decisions are made at the product level, i.e., the cost-effectiveness of a product is evaluated across its indications. On occasion, exceptions are made to this policy with coverage decisions made for certain indications or in certain subgroups. (Anell & Persson 2005) Cost-effectiveness analysis is a "*central principle*" of the TLV's evaluation. (LFN - Pharmaceutical Benefits Board 2007) However, as human need and solidarity are also decision-making criteria necessary, trade-offs between them must be made. Consequently, a single fixed cost-per-QALY threshold is not operated. (Anell & Persson 2005;Ramsberg, Odeberg, Engstrom, & Lundin 2004) Although the TLV does not have the authority to negotiate price, if the technology is rejected on the basis of cost-effectiveness, the manufacturer may decide to reapply for reimbursement using a lower price. (LFN - Pharmaceutical Benefits Board 2007)

The SBU's remit is to select medical technologies for review and to consider them from a number of perspectives, including medical, economic, ethical, and social standpoints. (Jonsson 2009b;SBU 2011) The SBU simultaneously evaluates and compares the effectiveness and cost-effectiveness of the alternative technologies. (Jonsson 2010) One of the functions of the SBU's reports is to aid the efficient allocation of health care resources. (O'Donnell et al. 2009)

3.3.1.3 Australia and the PBAC/MSAC

Australians have universal health care coverage through the Australian health care system, the predominant aspect of which is Medicare, the publicly funded insurance programme. (Bulfone, Younie, & Carter 2009a) Health care in Australia is funded through a mixture of public and private financing, with the former accounting for approximately 70% of total funding. Australia spends 8.7% of GDP on health care and was ranked 32nd in the WHO's 2000 global ranking of health care systems. (OECD 2011;WHO 2000)

Two government agencies are responsible for HTA in Australia, the Pharmaceutical Benefits Advisory Committee (PBAC) and the Medical Services Advisory Committee (MSAC). The PBAC is a statutory independent expert committee established under the National Health Act of 1953 and is appointed by the Health minister. (PBAC 2010) The PBAC committee's role is to evaluate drugs and to provide recommendations to the Minister of Health and Ageing regarding their inclusion on the Pharmaceutical Benefits Scheme (PBS), the national formulary that includes drugs and vaccines subsidised by the Australian government. (Lopert 2009;PBS 2011) The PBAC's recommendations fall into one of three categories: unrestricted benefit, restricted benefit, and authority required. Only drugs recommended by the PBAC can be added to PBS; the Health Minister may, however, decide not to list a recommended drug. (Lopert 2009;PBS 2011)

Established in 1998, the role of the MSAC is to advise the Federal Minister for Health and Ageing regarding the strength of evidence relating to new medical technologies and procedures and to recommend under what circumstances they should be used. (MSAC 2010) Requests for the inclusion of devices on the Medicare Benefits Schedule (MBS) are most commonly made by the manufacturer but may also be made by medical organisations, individual physicians, or patients. (Bulfone, Younie, & Carter 2009a;MSAC 2011) The role of the MSAC is to improve health outcomes for patients by ensuring that new and existing medical procedures that attract funding under the MBS are supported by evidence of their safety, clinical effectiveness, and cost-effectiveness.

Decision-making criteria

In 1987, an amendment was made to legislation that required the PBAC to account for the effectiveness and cost of a drug relative to other therapies. (National Health Act 1987) Since then, consideration of cost-effectiveness has been fundamental in PBAC's review. (Bulfone, Younie, & Carter 2009a) Multiple factors are, however, considered relevant to decision-making and include: cost-effectiveness, including estimation of uncertainty; clinical need, including consideration of alternative treatment options; total cost of implementation to the PBS; and affordability of the drug to the patient in the absence of a subsidy. Consistent with multiple decision-making criteria, the PBAC does not operate a fixed cost-effectiveness threshold but considers and weighs a number of relevant factors in deliberations. (Henry, Hill, & Harris 2005;Lopert 2009) Accordingly, although drugs with a lower cost-effectiveness ratio are more likely to be recommended, those with a higher ratio may be recommended if indicated for a life threatening condition, or if a lack of effective alternatives exist. (Lopert 2009)

The MSAC's role is to advise the Minister for Health and Ageing on the strength of evidence relating to the safety, effectiveness, and cost-effectiveness of medical services and technologies and to provide a recommendation as to under what circumstances they

should be covered on the MBS. (MSAC 2011) With respect to the economic evaluation, MSAC requests that a societal perspective be adopted. (Bulfone, Younie, & Carter 2009b) Consistent with the PBAC, the MSAC does not operate an explicit cost-effectiveness threshold. Based upon the strength of the evidence, the MSAC may recommend that the technology should receive public funding, recommend that the technology should not receive public funding, or deem that the evidence is inconclusive.

3.3.1.4 Canada and CADTH

Canada has a national health care system, commonly referred to as Medicare. Canadian residents have 'reasonable access' to 'medically necessary' health-care services independent of their ability to pay. (Health Canada 2011;The Commonwealth Fund 2011a) Canada spends 11.3% of GDP on health care and the WHO ranked the Canadian health care system 30th in their 2000 global health care system rankings. While approximately 70% of health care is publicly funded, approximately two thirds of Canadians have some form of supplementary private insurance often gained through employment based insurance plans. Canada has a decentralised health care system with 13 separate provincial and territorial health insurance plans. (Menon & Stafinski 2009)

Health technology assessment (HTA) is performed at multiple levels throughout the Canadian health care system. While the majority of HTA activity is performed by the Canadian Agency for Drugs and Technologies in Health (CADTH), regions often have their own HTA programs, e.g., Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS) in Quebec. (AETMIS 2011;Menon & Stafinski 2009) Established in 1989, CADTH is a national, independent, not-for-profit organisation funded by Canada's federal, provincial, and territorial ministers of health. (CADTH 2011b) Originally named the Canadian Coordinating Office for Health Technology Assessment (CCOHTA), the agency is charged with providing credible, impartial advice and evidence-based information about the effectiveness and cost-effectiveness of drugs and other health care technologies to Canada's decision makers at the federal, provincial

and territorial level. (CADTH 2011b) In 2002, CCOHTA was given the additional responsibility of the Common Drug Review (CDR), the process of evaluating and recommending drugs for their inclusion on publicly funded federal, provincial, and territorial drug benefit plans. In 2006, CCOHTA changed its name to CADTH to better reflect its roles and responsibilities. (CADTH 2011b)

CADTH has three distinct programs: Health Technology Assessment (HTA), Common Drug Review (CDR), and Canadian Optimal Medication Prescribing and Utilization Service (COMPUS). Through its HTA programs, CADTH evaluates technologies deemed to be of national interest and performs comprehensive reviews of the clinical effectiveness, cost-effectiveness, and broader impact of drugs, drug classes, and health technologies. (CADTH 2011a;INAHTA 2011) The HTA programme's focus is most often on more mature technologies for which there is a larger and higher quality body of evidence available. (CADTH 2011a) In addition to the CADTH's HTA programme, the CDR performs HTAs. The CDR's mandate is to evaluate new drugs, except for anti-cancer agents[ii], before they can be listed on federal, provincial, and territorial drug benefit plans. The CDR submits a report to the Canadian Expert Drug Advisory Committee (CEDAC), which considers the comparative effectiveness and cost-effectiveness of the drug in comparison to the established standard of care. CEDAC makes one of three funding recommendations to participating plans: list without conditions, list with conditions, or do not list. Each plan considers the recommendation and ultimately makes an independent decision as to coverage. (Menon & Stafinski 2009)

Decision making criteria

For each technology assessment, CADTH produces a comprehensive report that includes an evaluation of the technology's clinical effectiveness, cost-effectiveness, and consideration of its various impacts on the health care system, including budget impact,

[ii] The pan-Canadian Oncology Drug Review (pCODR) was recently established (2010) to assess cancer drugs and make recommendations to the provinces and territories to guide their drug funding decisions. Although a number of reviews are in process none have yet been completed.(pCODR 2011)

legal and regulatory issues, and ethical, equity and psychosocial issues. (CADTH 2011a) With respect to clinical effectiveness and efficacy, CADTH typically performs a systematic review of the evidence base. Consideration of all available alternatives must be made, with the recommended reference case including comparison with 'usual care'. (CADTH 2006) The assessment of a technology's cost-effectiveness is performed from the perspective of the publicly funded health care system and typically includes a cost-utility analysis. As the broader impact on both patient health and the health care system is considered, CADTH does not operate a fixed cost-effectiveness threshold. (CADTH 2011a)

Fundamental to each CDR submission is evidence of the product's efficacy, effectiveness, and safety. An appropriate pharmacoeconomic evaluation is required for all submissions to the CDR. (CDR 2010) A cost-effectiveness or cost-utility study is required in the following circumstances: the drug is the first available for a particular indication; the drug is the first in a newly established therapeutic class; the drug has demonstrated differences in safety or efficacy compared to available alternatives in head-to-head randomised controlled trials; or, in the absence of head-to-head trials, the drug's manufacturer assumes that differences exist (manufacturer must provide evidence to support this assertion). Cost-effectiveness or cost-utility analyses must be based upon final outcomes such as life-years, QALYs, important disease specific units (e.g., myocardial infarction, stroke, or fracture), or validated surrogate outcomes. (CDR 2010) A cost-consequence analysis may be considered for products demonstrating benefits in other outcomes, e.g., those that are patient-reported, non-clinical or surrogate. In most cases budget impact analyses (BIAs) are also required.

3.3.2. Countries in which cost-effectiveness evidence plays less of a role in decision-making

Cost-effectiveness is a central component in assessments performed by the HTA agencies described in the preceding sections. Not all countries, however, have HTA agencies that

consider cost-effectiveness evidence to the same extent. Described in the following sections are Germany, France, and the US, three countries in which cost-effectiveness evidence plays a lesser role.

3.3.2.1 Germany and IQWiG

In Germany, universal health care coverage is provided through a multi-payer system. (The Commonwealth Fund 2011b) Germany spends 11% of GDP on health care and the WHO ranked the German health care system 25th in their 2000 global health care system rankings. (OECD 2011;WHO 2000) Germans whose income is below a certain level receive health insurance through the Statutory Health Insurance (SHI) system of private non-profit sickness funds. While the majority of Germans with incomes above the threshold opt into the sickness fund system, some purchase private insurance. (The Commonwealth Fund 2011b)

The Institute for Quality and Efficiency in Health Care (IQWiG) is an independent scientific institute established in July 2004 to provide advice to the Federal Joint Committee, the main decision-making body in German health care. Advice is based upon evidence-based evaluations of the costs and benefits of health technologies and services. (IQWiG 2011a;Nasser & Sawicki 2009;Perleth, Gibis, & Gohlen 2009) The institute is responsible for the scientific evaluation of the benefits and harms, and the quality and efficiency, of health care services. (IQWiG 2011b) The Federal Joint Committee considers IQWiG's evaluations and issues coverage and payment directives. Since January 1st, 2011, all new drugs are subject to assessment with associated medical benefit compared against appropriate therapeutic alternatives. Requests for review topics originate from a combination of government sources, patient advocacy groups, or the Federal commissioner for patient affairs. (The Commonwealth Fund 2011b) The Federal Joint Committee ultimately selects topics to be considered by IQWiG.

Decision making criteria

IQWiG is responsible for the scientific evaluation of the clinical effects as well as the quality and efficiency of health-care services. (Caro et al. 2010) The assessment is a two-step process that includes a clinical assessment, and subsequently, a cost-benefit-assessment. (Fricke & Dauben 2009;Gerber, Stock, & Dintsios 2011)

IQWiG's assessment considers a new technology's medical benefit by evaluating both comparative and non-comparative clinical studies. The assessment is based upon a dossier submitted to IQWiG by the technology's manufacturer that must include all relevant studies and information regarding the medical benefit relative to therapeutic alternatives. According to regulation, IQWiG categorises the medical benefit of a new technology into six categories (Gerber, Stock, & Dintsios 2011):

1. Remarkable additional benefit
2. Considerable additional benefit
3. Minor additional benefit
4. Additional benefit not quantifiable
5. No evidence of additional benefit
6. Less benefit than the comparator

In April 2007, federal law expanded IQWiG's duties by adding a cost-benefit-assessment to the appraisal process. However, under law a technology cannot be excluded from coverage due to its cost. In January 2008[iii], IQWiG published their methods guidance for the submission process. (Caro et al. 2010;IQWiG 2009a) The recommended methodology differs somewhat from the requirements of other HTA agencies. The efficiency frontier, a fundamental aspect of this methodology, is a plot of the incremental costs and benefits of available technologies, with health benefit in terms of "patient

[iii] An updated version was published March 2009(IQWiG 2009b)

relevant health outcomes" presented on the Y-axis and costs presented on the X-axis (Figure 13).

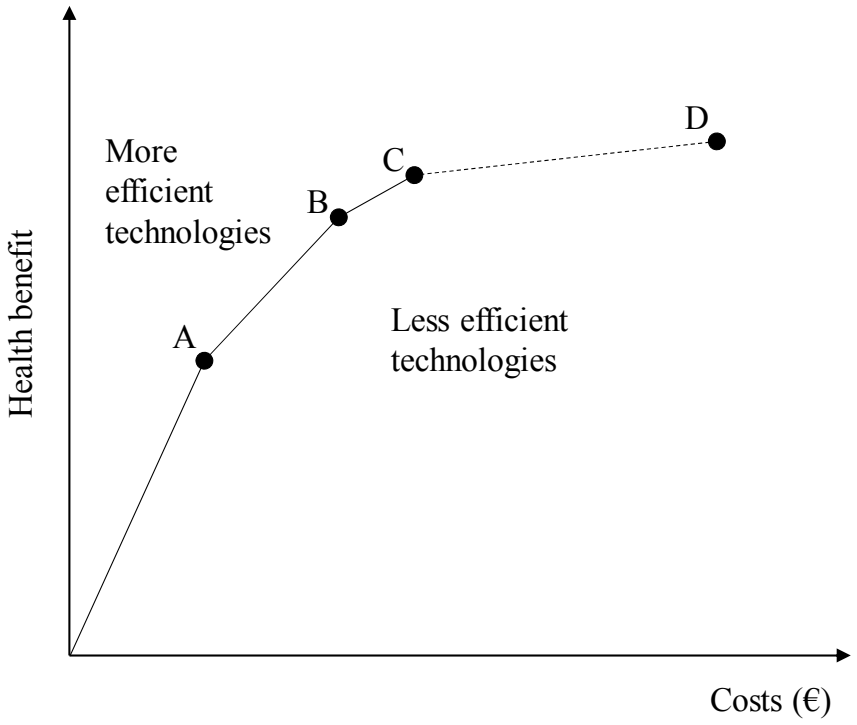


Figure 13. Efficiency frontier as used by IQWiG

New technologies falling below and to the right of the frontier are considered less efficient; technologies falling above and to the left of the frontier are considered more efficient. However, it is expected that rather than falling in either of these categories, a new technology will typically be more effective and more costly than currently available care. The frontier is said to inform whether a new health care programme with a positive ICER represents good value for money by providing the "going rate" for the additional cost per health benefit, i.e., the ICER associated with programme C, the reciprocal of the gradient of the frontier B-C. (Caro et al. 2010) It is suggested that if there are sufficient points on the efficiency frontier, an estimation of the rate at which efficiency has been decreasing as a function of increasing value can be made. This estimate will indicate "what is to be expected" with respect to future increases in the value of a unit of health

benefit, and will provide a basis for assessing the reasonableness of a decrease in efficiency. (Caro et al. 2010) For example, technology D in Figure 1 represents a new technology associated with a positive ICER, i.e., is more effective and more costly than technology C. The ICER associated with technology C represents the "going rate", i.e., the current cost of producing a unit of health gain. Technology D's ICER is greater than programme C's, and thus greater than the "going rate". Further, it is suggested that a willingness to pay approach may inform whether a new technology should be implemented, although the challenges associated with this approach are noted. The decision rule is, however, unclear for technologies that are both more effective and more expensive than their comparator. While the Federal Joint Committee may consider the "going rate" and WTP estimates, the maximum cost per health outcome that would be deemed permissible is uncertain.

IQWiG do not usually consider the QALY. Rather, the agency focuses its assessment of benefit and harm in terms of patient-relevant health outcomes. It is stated that due to not being a "universally accepted method", QALYs will not be used as the outcome metric of choice. (Caro et al. 2009) Although the use of disease specific units to quantify health benefit is appealing, their use will likely be challenging in therapeutic areas in which multiple dimensions of health are affected or for technologies that positively impact multiple therapeutic areas.

An important distinction of IQWiG's methods is that, rather than to serve as a method to allocate resources across diseases, the purpose is to recommend maximum reimbursable prices. IQWiG's methodological approach has been subject to criticism, particularly as opportunity cost does not appear to be accounted for. (Sculpher & Claxton 2010) Similar to other HTA agencies, IQWiG publishes reports, rapid reports and working papers, and invites input from all stakeholders and the general public. (IQWiG 2011a)

3.3.2.2 France and HAS

In 2000, the French health care system was ranked as the world's best by the WHO. (WHO 2000) Health care coverage in France is universal and while it is predominantly government funded, it is possible to purchase supplementary private insurance. (Rochaix & Xerri 2009) France spends approximately 12% of GDP on health care. (OECD 2011) It is said that France's health care system is characterised by "solidarity and universal coverage and responsibility". (de Pouvourville G. 2010;Weill & Banta 2009)

The Autorité de santé (HAS), or French National Authority for Health, is an independent public body established by the French government. (HAS 2011) HAS was created by the National Health Insurance Reform Act of 2004 and was established January 2005, to unite under a single entity a number of activities designed to improve the quality of patient care and guarantee equity within the health care system. (HAS 2011;Weill & Banta 2009) HAS is mandated by law to carry out particular research projects that it reports to the French government and parliament. While HAS's recommendations are advisory, the decision-making bodies, i.e., the union of sickness funds or the Ministry of Health, generally accept its findings. HAS is responsible for a broad range of activities that include: the assessment of drugs, medical devices, medical and surgical procedures, and biological tests; physician certification; the generation of clinical guidelines; and providing information with respect to the coverage of services and reimbursement. (HAS 2011;Weill & Banta 2009)

Through its technology evaluation role, HAS performs two types of technology assessments; single technology assessments (STA) and multiple technology assessments (MTA). (Rochaix & Xerri 2009) Before a new drug, medical device, or procedure can be added to the health insurance benefit list, a mandatory STA is performed. Product manufacturers or professional societies can initiate STAs. HAS assesses a technology or procedure's intrinsic value and its effectiveness relative to competing therapies. (Rochaix & Xerri 2009)

HAS gives an "opinion" on the absolute health benefit, i.e., expected or actual benefit and the relative health benefit, i.e., effectiveness relative to usual care, of the technology or procedure. HAS's opinion is given to the Ministry for Health and Social Security and union of sickness funds, and is used to support coverage, reimbursement, and pricing decisions. Current regulation dictates that "*medicines that neither provide a therapeutic added value nor cost savings*" may not be included on the benefit list. Therefore, technologies or procedures that do not provide additional clinical benefit will only be reimbursed if they are offered at a lower cost. For technologies or procedures judged to provide additional health gain, the pricing committee may grant a higher price. Technologies are reassessed by HAS at five year intervals; procedures are reassessed at variable time intervals. (Rochaix & Xerri 2009)

In contrast to STAs, MTAs generally review an entire class of technologies or procedures. Also, MTAs may take the form of public health guidelines or concern the organisation of care. Although topics for review may be chosen internally by HAS, typically they originate from public agencies or other interested parties. Rather than providing an opinion on certain technologies and procedures, MTAs are designed to provide more high level guidance on coverage policy, health care delivery, or health care organisation.

Decision making criteria

With respect to intrinsic value, HAS considers the severity of the condition treated, the efficacy/safety ratio, and how treatment fits into the current therapeutic strategy. With respect to relative effectiveness, the incremental clinical benefit of the product is categorised on a five-level scale:

I: major improvement (new therapeutic area, reduction of mortality)

II: significant improvement in efficacy and/or reduction of side-effects

III: modest improvement in efficacy and/or reduction of side-effects

IV: minor improvement

V: no improvement.

In 2008, HAS's mission was expanded to include, "*recommendations and medico-economic opinions on the most effective strategies of care, prescription, and disease management*". (Rochaix & Xerri 2009) Although HAS had performed a small number of economic analyses prior to 2008 this legislation signalled a change in direction. To perform this function a new department within HAS was created, the Commission for Economic and Public Health Evaluation (CEESP). CEESP is overseen by an interdisciplinary committee responsible for evaluating the quality and ethics of completed work, providing scoping guidance, and considering potential conflicts of interest. Economic evaluations performed by HAS help illustrate the opportunity costs associated with reimbursement decisions, thus increasing the efficiency of the use of medical technology. For the most part, economic evaluations are performed as part of MTAs rather than STAs. Therefore, the introduction of economic evaluation has not influenced the STA process and so the method for pricing and reimbursing technologies remains principally determined through consideration of clinical efficacy. (de Pouvourville G. 2010) Economic evaluation may impact the price of medical technologies and procedures through a re-examination of treatment classes through MTAs. Typically, economic evaluations consider a whole class of treatments and are used to optimise the overall delivery of care, rather than to evaluate individual medical technologies or procedures. (de Pouvourville G. 2010) As economic evaluations are performed in a within-class basis, disease specific units are often used, e.g., cost of reduction of 1g/L of LDL cholesterol. (de Pouvourville G. 2010)

3.3.2.3 US

The US health care system has been described as fragmented and uncoordinated. (Sullivan et al. 2009) With no single national entity, or set of policies, to guide it, multiple agencies administer health care at the national, state, community, and practice levels. (Shih et al. 2008) In contrast to the countries described above, the majority of Americans (67.5%) obtain health insurance through private providers. (DeNavas-Walt, Proctor, & Smith 2008) Approximately 28% of Americans receive health insurance through government programmes, of which Medicare is the largest. Prior to the Patient Protection and Affordable Care Act (PPACA) legislation in 2010, approximately 15% of the population did not have health insurance coverage.[iv] (PPACA 2010) As described in Section 3.2.1, despite spending considerably more on health care than other countries, the US health care system ranked 37th in the WHO's 2000 global health care system rankings. (WHO 2000)

Consistent with the decentralised nature of the health care system, a number of independent public and private HTA agencies exist in the US rather than a single HTA agency. Notable publicly funded agencies include Medicare and Medicaid, the Agency for Healthcare Research and Quality (AHRQ), the Drug Effectiveness Review Project (DERP) in Oregon, and the HTA programme of the Washington State Medicaid programme, among others. (AHRQ 2011;CMS 2005a;CMS 2011e;Shih, Davis et al. 2008;Washington State Health Care Authority 2010) Other federally funded programmes include the Pharmacy Benefits Management Strategic Healthcare Group at the Department of Veterans Affairs and the Department of Defense PharmacoEconomic Center (PEC) in the Military Health System. (DoD PEC 2010;U.S.Department of Veterans Affairs 2010) The National Institutes of Health, although not having a HTA programme, occasionally perform evidence reviews when developing clinical practice policies. (NIH 2011) Private health care plans often make coverage and reimbursement decisions, although in many cases Pharmacy Benefit Managers (PBMs) are used to

[iv] It is expected that the proportion of uninsured individuals will decrease considerably if the recent health reform legislation it is fully implemented in 2014.

design and administer drug formularies. However, most private organisations do not make information about their HTA programmes readily available. There is likely considerable variation between them. (Sullivan et al. 2009)

Comparative effectiveness research (CER), i.e., the direct comparison of existing health care interventions to determine relative effectiveness, has been advocated as an approach to improve quality of care while helping to arrest rising costs. Initial support for CER came as part of the American Recovery and Reinvestment Act (ARRA) of 2009, in which a provision dedicated $1.1 billion to study CER. (ARRA 2009) The PPACA legislation of 2010 further advocated the use of CER but placed restrictions on how such information should be used. (PPACA 2010) The Patient-Centered Outcomes Research Institute (PCORI) was established as part of the PPACA and has the role of coordinating CER studies, assisting in their funding, and disseminating study findings. (PPACA 2010)

Cost-effectiveness evidence is not typically part of technology assessment in the US. The Centers for Medicare & Medicaid Services (CMS) are the administrators of Medicare, which is the largest payer in the US. CMS states in its Guidance for the Public, Industry and CMS Staff that cost-effectiveness evidence is not a factor CMS considers in making national coverage determinations (NCDs), although recent decisions appear to suggest that cost-effectiveness evidence plays a role in NCDs for preventative care (Table 9).[v] (CMS 2010e) Rather, coverage decisions for medical technologies are made using the 'reasonable and necessary' criterion (Section 3.4.2.1). Both the Department of Defence (DoD) and Department of Veterans Affairs (VA) have internal groups that evaluate cost-effectiveness, but how it is incorporated into decision-making is not described. (DoD PEC 2010;U.S. Department of Veterans Affairs 2010) One of the stated goals of the state of Washington's HTA programme is to make "*State purchased health care more cost effective by paying for medical tools and procedures that are proven to work"*.

[v] N.B. The empirical work presented in chapters 4 through 7 evaluates CMS NCDs from the perspective of cost-effectiveness.

(Washington State Health Care Authority 2010) Again, however, guidance is not given as to how cost-effectiveness evidence is factored into decision-making.

Cost-effectiveness analysis is used to a limited extent in the private health care industry. The Academy of Managed Care Pharmacy (AMCP) has published guidelines that serve as a template for drug companies to submit dossiers to Pharmacy and Therapeutics (P&T) Committees. These guidelines include recommendations regarding cost-effectiveness analysis. (FMCP Format Executive Committee 2010) Wellpoint, one of the largest private health insurance companies, has issued guidelines providing a framework for the submission of cost-effectiveness evidence. (Wellpoint 2010) Also, the Drug Effectiveness Review Project (DERP) is an alliance of 13 states and private organisations that synthesise and judge clinical evidence for drug class reviews. (DERP 2010) Consumers Union (CU), an independent, non-profit organisation, adapts DERP reviews in developing a consumers "Best Buy" guide. Recommendations in the guide are based upon a comparison of a drug to others in the same therapeutic class. The criteria used for these recommendations include relative effectiveness, safety, side-effect profile, convenience, and price. However, as only drug price is considered and not associated costs, CU does not estimate the overall value or cost-effectiveness of drugs. (Consumer Reports 2010)

By and large, the use of cost-effectiveness analysis is not established in the US health care system. Although the examples listed above provide evidence that some decision makers are aware of the benefits of cost-effectiveness evidence, the fragmented nature of the US health care system results in great variability with respect to its use. When decision makers do consider cost-effectiveness evidence, the role that it plays in decision-making is unclear.

3.3.3. Inter-country comparison

There is noticeable variation in decision-making criteria across agencies in the countries described above. While no country relies solely on cost-effectiveness evidence to guide coverage and reimbursement decisions or to make recommendations for the efficient use of medical technology, it plays a more important role in some countries than in others. Cost-effectiveness evidence plays a fundamental role in technology assessment in the UK, Sweden, Australia, and Canada. However, indicative of the fact that multiple criteria are important in decision-making, acceptable cost-effectiveness either exists over an explicit range, e.g., the UK, or no range is given, e.g., Sweden, Australia, and Canada. Germany and France provide examples of countries in which economic considerations play less of a role in decision-making. It is noteworthy that despite Germany and France grounding technology assessment in the clinical evidence, cost-effectiveness evidence still features to a limited extent. In Germany, although IQWiG considers the costs and benefits of medical technology, this information may not be used to deny coverage of a medical technology. Rather, cost-effectiveness information is used on the fringe of the decision-making process to inform the maximum allowable price. In France, while the consideration of cost-effectiveness evidence has recently been added to HAS's mandate, it is considered in only the minority of instances and resistance to its use remains. (de Pouvourville G. 2010)

The US could be considered near the end of the spectrum with respect to its use of cost-effectiveness information. The US health care system is largely decentralised and with multiple entities evaluating medical technology, decision-making is fragmented and uncoordinated. Although some public and private payers use cost-effectiveness information sporadically, the extent to which it informs decision-making is unclear. Further, Medicare, the largest payer in the US, states that cost-effectiveness is not a factor it considers in national coverage determinations.

Why cost-effectiveness evidence plays such a limited role in decision-making in the US is not obvious. In the following sections, I review attempts to incorporate cost-effectiveness evidence into decision-making in the US and consider reasons why resistance to it exists.

3.4. The use of cost-effectiveness evidence in the US health care system

As described in Section 3.2.1, health care spending in the US is greatly in excess of spending in other countries, but the US performs relatively poorly when considering key health metrics. Given the comparatively poor return from spending, it would seem that the US health care system would benefit greatly from the use of cost-effectiveness evidence. However, as described in Section 3.3.2.3, the US is notable for the minimal role that cost-effectiveness evidence plays in decision-making.

There have been, however, a number of prominent attempts to advocate the use of cost-effectiveness analysis and to incorporate cost-effectiveness information into decision-making; these are described below.

3.4.1. The US Panel on Cost-Effectiveness in Health and Medicine

Paradoxically, given the current unwillingness to embrace cost-effectiveness evidence, the US was one of first countries to establish methodological guidelines for conducting cost-effectiveness analysis. In 1993, the US Public Health Service convened the U.S. Panel on Cost-Effectiveness in Health and Medicine. (Gold et al. 1996) The Panel's task was to assess the state of the science of cost-effectiveness analysis and to provide recommendations for the conduct of cost-effectiveness studies. In 1996, the Panel published its recommendations in a book entitled "*Cost-effectiveness in health and medicine*" and in three summary articles. (Gold et al. 1996;Russell et al. 1996;Siegel et al. 1996;Weinstein et al. 1996) Among the Panel's recommendations were the

appropriateness of analytic techniques (cost-effectiveness analysis was proposed as the method of choice), relevant outcome measures (the QALY), the discount rate (3%), and the presentation of incremental ratios. The Panel stated that cost-effectiveness analysis should be useful to various audiences, including insurers, managed care organisations, policy makers, and the general public, among others. The Panel's recommendations proved influential among health economic researchers but had only a limited impact on the proliferation of the use of cost-effectiveness analysis in the US health care system. (Phillips & Chen 2002)

3.4.2. Attempts to incorporate cost-effectiveness analysis into health care

In the following sections, I present two case studies that provide useful insight into the resistance to the use of cost-effectiveness evidence in the US.

3.4.2.1 Cost-effectiveness and Medicare

As highlighted in Section 3.3.2.3, CMS state that cost-effectiveness evidence is not considered in national coverage determinations. (CMS 2010e) There have been, however, attempts to incorporate cost-effectiveness evidence into CMS's technology assessment.

The first attempt to incorporate cost-effectiveness evidence into CMS coverage determinations was in the Health Care Financing Administration's (HCFA) 1989 proposed regulations. (Federal Register 1989b) The HCFA supported this intention by stating, *"We believe the requirement of section 1882(a)(1) that a covered service be 'reasonable' encompasses the authority to consider cost as a factor in Medicare coverage determinations"*. (Federal Register 1989c) It was further reasoned that the systematic assessment of the cost-effectiveness of technologies would "*vastly improve our knowledge base and be a deterrent to coverage of procedures that may be costly, but have little or no impact on improving health outcomes"*. (Federal Register 1989d) The

proposed regulations were scrutinised from a number of sources, e.g., the *New York Times* suggested that it represented *"a fundamental shift… the Federal Government will explicitly weigh cost as a factor in deciding whether Medicare should pay for new medical procedures, devices and drugs for elderly people"*. (Pear 1991) Despite support from the Department of Health and Human Services, opposition from the medical device industry and consumer groups meant that the proposed regulations were not released in final form. (Neumann 2005;Pear 1991)

In the mid-1990s, the HCFA attempted to revive the proposed regulation and publish it as a final rule. Once more, the regulation faced opposition, this time from medical and industry groups. Among those opposing the proposed regulation were the Pharmaceutical Research and Manufacturers Association of America (PhRMA), the American College of Physicians, the American Medical Association (AMA), and various politicians. The opposition ultimately resulted in the HCFA announcing the formal withdrawal of the proposed 1989 regulation in 1999. (Neumann 2005;U.S. Congress 1997)

Today, coverage decisions remain guided by the legislation that created Medicare, which states that *"Medicare coverage is limited to items and services that are reasonable and necessary for the diagnosis or treatment of an illness or injury"*. Cost-effectiveness evidence is effectively excluded from review. (CMS 2010g) In the Guidance for the Public, Industry and CMS Staff states *"Cost effectiveness is not a factor CMS considers in making NCDs. In other words, the cost of a particular technology is not relevant in the determination of whether the technology improves health outcomes or should be covered for the Medicare population through an NCD"*. (CMS 2010e)

3.4.2.2 The Oregon Experiment

The Oregon Experiment is an often-cited example of an attempt to introduce cost-effectiveness into resource allocation decision-making in the US. Oregon's proposed approach to using cost-effectiveness drew much attention and scrutiny from sources inside and outside the US. (Neumann 2005)

Oregon Medicaid Programme

In the late 1980s, the state of Oregon was struggling to finance its state Medicaid programme. In 1987, in response to ongoing budgetary pressures, the Oregon legislature removed major organ transplants as a benefit from the state's Medicaid programme. Later that year, a seven-year-old boy died after not receiving a bone marrow transplant. (Buist 1992) The resultant public outcry prompted a reconsideration of the Medicaid benefit with respect to eligibility and service provision. Attempts to reintroduce bone marrow transplants to the Medicaid benefit were opposed by the president of the state senate, John Kitzhaber. Kitzhaber contended that the resources required to provide bone marrow transplants to a few individuals would be sufficient to provide basic health insurance to many more uninsured individuals. (Fruits, Hillard, & Lewis 2009) Further, it was argued that by restricting the basic services offered as part of the Medicaid programme, it would be possible to expand coverage to all uninsured eligible individuals. To this end, a bill was proposed in 1989 with the ambitious goal of providing health insurance for all Oregon residents. This was to be achieved through two mechanisms, by mandating private employers to provide health insurance to employees and expanding the Medicaid programme to all Oregon residents below the poverty line. Although, attempts to mandate employer health insurance were unsuccessful due to political and business opposition, reforms to the Medicaid programme with the Oregon Health Plan (OHP) were finally implemented in 1994. (Buist 1992;Neumann 2005) The evolution of the OHP from conception to implementation is described below.

Prioritisation of services

The expansion of the Medicaid programme was to be achieved by prioritising services. A novel approach was taken to decide what services should be offered as part of the Medicaid benefit. First, a list of 709 condition/treatment pairs was generated by a state appointed Health Services Commission by ranking technologies in order of net benefit. Through a process that included community meetings, a public survey of quality of life preferences, and consideration of treatment cost, interventions were essentially ranked with respect to their approximate cost-effectiveness. The intention was to systematically produce an objective list of technologies that represented a ranking based upon cost-effectiveness, included input from the community, and was consistent with public preferences. (Fox & Leichter 1991;Neumann 2005;Ubel 2001)

The initial list met fierce opposition and the plan soon became the subject of intense debate. It was claimed that as Medicaid beneficiaries were predominantly poor, young, non-white, and female, the plan discriminated on the basis of class, age, race, and sex. (Brown 1991;Neumann 2005) Others argued that the process used to rank the services was neither open nor fair, and that the poor were not represented in the decision-making process. (Daniels 1991;Jacobs, Marmor, & Oberlander 1999) The list was widely criticised, particularly since much of the ranking appeared to be counterintuitive. For example, in the original list, tooth capping was ranked above surgery for ectopic pregnancy, and splints for temporomandibular joints ranked above appendectomies. (Eddy 1991) Ultimately, the initial list was not submitted to the Health Care Financing Administration (HCFA) for approval due to the strong opposition.

In response to the criticisms, the list was amended. Most notably, the list was rearranged in accordance with expert opinion rather than cost-effectiveness. The amended list was submitted to the HCFA in 1992 but was again rejected, this time on the grounds that it violated the Americans with Disabilities Act (ADA). As quality of life measures did not have input from disabled patients, it was deemed that potential existed for the programme

to discriminate against them. A third and final list that addressed these concerns, excluding the influence of cost-effectiveness evidence, was submitted to the HCFA, and the plan was eventually enacted in 1994. (Buist 1992;Neumann 2005)

Success of the Oregon Health Care Experiment

The OHP achieved its goal of expanding the Medicaid programme and added 100,000 state residents. However, expansion came at a high cost. (Leichter 1999) Expenditures increased by 39%, in contrast to 30% nationally, with the additional cost attributed to the implementation of the new programme. (Bodenheimer 1997) Also, evidence suggests that the use of the list of services actually reduced access to needed services. (Mitchell et al. 2002)

Key lessons from the Oregon Health Plan

Still relevant today, the underlying rationale of the OHP was to provide universal access to health care by prioritising access to health care services. The plan represented a step away from "do everything possible medicine" and shifted the debate away from what populations to cover and towards what benefits to cover. (Bodenheimer 1997;Jacobs, Marmor, & Oberlander 1999;Leichter 1999) After a bold attempt to systematically allocate resources based on cost-effectiveness while involving the community and taking into account public preferences, the OHP was ultimately implemented only after removal of the cost-effectiveness provisions.

Why the inclusion of the cost-effectiveness component ultimately failed has been subject to much analysis. Multiple reasons have been suggested for its failure, with a mixture of technical, political, legal, and ethical factors playing a role. It is argued that the list was technically flawed with the taken approach lacking methodological rigor (Gold et al., 1996) and not an actual reflection of cost-effectiveness. (Tengs et al., 1996) Further, despite attempts to incorporate public input into the list, it is thought that the ranking

failed to capture public preferences. (Ubel et al., 1996) The method used to generate preference weights has been criticised for relying on a rating scale rather than more established techniques such as time trade-off or standard gamble. Others suggest that the principal reason for failure was neither methodological nor legal, rather Americans' ingrained aversion to the imposition of limits and suspicion of governments that impose them. (Neumann 2005) This cultural phenomenon, often referred to as American exceptionalism, is discussed in Section 3.5.1.

Notably, no state Medicaid programme has attempted to implement a similar policy to the OHP. Politically, the explicit use of cost-effectiveness evidence to allocate resources proved unpalatable. The OHP was used by politicians to gain political capital and to serve as an example of what was wrong with the health care system. Importantly, the OHP exposed the limits of publicly applying explicit rationing policies within the United States. (Neumann 2005) It has been suggested that the enduring lesson from Oregon is that the use of cost-effectiveness evidence is unlikely to produce a social or political definition of necessary care in the US. (Hadorn 1991)

3.5. Resistance to cost-effectiveness evidence in the US health care system

Despite comparing unfavourably with respect to key health metrics, the US health care system is substantially more expensive than its international counterparts (Section 3.2.1). (Pearson M 2009) One would expect, therefore, that the US health care system would provide the ideal environment for cost-effectiveness analysis to flourish and be of real benefit to health care decision makers. However, as described above, despite attempts to consider cost-effectiveness evidence in health care decision-making, such evidence is largely excluded from deliberations over the allocation of health care resources.

Resistance to cost-effectiveness evidence, and comparative effective evidence and evidence-based medicine in general, is a notable feature of the US health care system. Chalkidou and Walley 2010 suggest that "*no other developed or developing healthcare system and its users view evidence as suspiciously as US stakeholders, including the medical technology industry and a large proportion of policy makers*". (Chalkidou & Walley 2010;Dhruva et al. 2009) In many respects, the US health care system is unique. The US is the only country in the developed world not to provide universal health care coverage to its citizens, although the recent passing of health care reform legislation should reduce the number of uninsured considerably. Further, the composition of health care financing in the US is different from the majority of other countries, with a much greater proportion coming from private as opposed to public sources. (Davis, Schoen, & Stremikis 2010;World Health Organization 2000) In comparison to other countries' health care systems, the US health care system is fragmented, marked by a mixture of multiple public and private payers. It is unclear, however, why differences in health care financing, or structural differences, would lead to the observed resistance to cost-effectiveness evidence.

It is suggested that the powerful health care lobby is a principal obstruction to the use of cost-effectiveness evidence. (Neumann 2005) The pharmaceutical industry has the

largest lobby in Washington, employing more lobbyists than there are Congressmen. (Angel 2004) It is reported that from January 2005 to June 2006, manufacturers of pharmaceuticals, medical devices, and other health products spent nearly $182 million on federal lobbying. (Ismail 2007)

The competitive nature of the private health insurance market presents a further obstacle for the penetration of cost-effectiveness evidence. A private insurance plan may risk its competitive standing in the marketplace by using cost-effectiveness evidence, as it may be viewed as rationing care. There is, therefore, an understandable reluctance to be the first private insurance plan to openly use cost-effectiveness evidence. (Sullivan et al. 2009)

A study by Bryan et al. (2009) evaluated the acceptability of cost-effectiveness evidence to US decision makers and provides a useful insight into the lack of use of this type of evidence in the US health care system. (Bryan et al. 2009) Through a series of workshops and surveys, Bryan et al. (2009) showed that US decision makers, i.e., regulators, and private and public insurers, broadly support the use of cost-effectiveness evidence as an input into coverage decisions. The researchers did, however, identify major obstacles preventing the greater use of cost-effectiveness evidence, including; litigation fears, concerns of the biased nature of manufacturer funded studies, and the failure of studies to address shorter time horizons of more relevance to decision makers. Notably, despite the broad support for the use of cost-effectiveness evidence as an input in decision-making, the research showed that approximately 40% of decision makers remained uncomfortable with the concept of rationing.

The discomfort with the concept of rationing is, along with other aspects of American culture, suggested to be an obstacle to the inclusion of cost-effectiveness evidence into US health care decision-making. Often referred to as “American Exceptionalism”,

Americans' supposed uniqueness is thought to help explain why the US health care system is different from others. (The Hastings Center 2009)

3.5.1. American Exceptionalism

The term 'American exceptionalism' is used to describe the suggestion that the US is different from other nations because of the uniqueness of its origins, evolution, and institutions. The term is also used in reference to the general American attitude to business (free markets and long work hours), the environment (national attitudes toward energy policy and global warming), consumers (higher birth rates and higher rates of obesity), and politics (maintenance of capital punishment and the right to bear arms) among other aspects of American life. (Neumann 2009;Reiner et al. 2006)

High spending relative to GDP, the lack of universal coverage, and relatively poor health outcomes in the US health system compared to other developed nations are attributed in part to the paradigm of American exceptionalism. (Neumann 2009;Rodwin 1987;The Hastings Center 2009) The reluctance to embrace cost-effectiveness analysis has been viewed as a symptom of American exceptionalism. Americans have a clear dislike for limit setting and appear to overlook the scarcity of health care resources. (Robinson 2001) Also, there is a dislike and mistrust of 'big government' and reluctance to accept bureaucrats making decisions in lieu of patients and physicians. (Neumann 2009;The Hastings Center 2009)

3.5.2. The Patient Protection and Affordable Care Act

Passage of the Patient Protection and Affordable Care Act (PPACA) in March 2010 appears to have further distanced the US health care system from the use of cost-effectiveness evidence. (Chambers & Neumann 2010;Neumann & Weinstein 2010;PPACA 2010) In reference to the PCORI, the legislation states:

"[The PCORI] shall not develop or employ a dollars per quality adjusted life year (or similar measure that discounts the value of a life because of an individual's disability) as a threshold to establish what type of healthcare is cost effective or recommended. The Secretary shall not utilize such an adjusted life year (or such a similar measure) as a threshold to determine coverage, reimbursement, or incentive programs...".

- Patient Protection and Affordable Care Act (PPACA), March 2010

While the language clearly prohibits the use of cost-per QALY thresholds, it may be interpreted as not prohibiting the conduct of cost-utility studies, i.e., cost-per QALY ratios can still be calculated as long as they are not compared with a threshold value. Also, the excerpt is specific to the PCORI, and does not necessarily affect the evidence that is considered by other agencies. Nevertheless, the absolute nature of the language in such a major piece of legislation is noteworthy. Indeed, Neumann and Weinstein (2010) suggest that the language "*suggests a broader ban on the use of cost-utility analyses — and this could have a chilling effect on the field*."

Consequently, it is surprising that cost-effectiveness evidence has been featured in a number of CMS NCDs made after the enactment of the PPACA (Table 9)[vi]. Notably, all NCDs included in Table 9 pertain to preventative care, a type of intervention that appears to have a special relationship with CMS NCDs. What is driving this phenomenon is the Medicare Improvement for Patients and Providers Act (MIPPA) of 2008 (§1861(ddd)(2)). Through the MIPPA legislation, Congress authorised the US Department of Health and Human Services (DHHS) to add preventive services rated 'A' or 'B' by the US Preventives Services Task Force (USPSTF) to Medicare without congressional action. Accordingly, each of the preventative services included in Table 9 are associated with a USPSTF 'A' or 'B' grading. Further, the MIPPA legislation

[vi] The empirical aspect of this thesis considers NCDs made from 1999 through 2007. Therefore, the NCDs presented in Table 9 do not feature in the analysis.

includes language that is used in these cases to explain CMS's consideration of cost-effectiveness. The MIPPA legislation authorises CMS to "*conduct an assessment of the relation between predicted outcomes and the expenditures for [preventative] services*".

Still, it is noteworthy that CMS have reviewed cost-effectiveness evidence in each of these cases. While consistent with the MIPPA legislation, the review of cost-effectiveness evidence seems incongruent with the PPACA. Also, as these preventative interventions are associated with a USPSTF 'A' or 'B' grading, they are required to be covered regardless of CMS's independent review of the evidence base. The relevance of these cases of CMS considering cost-effectiveness information for preventative interventions is discussed in Chapter 8.

Table 9. National Coverage Determinations including cost-effectiveness evidence made after enactment of the PPACA legislation

Title	Date	Cost-effectiveness	National Coverage Determination
Counselling to Prevent Tobacco Use	August 25, 2010	Study identified from the literature (Solberg et al. 2006) $1100 per QALY when excluding savings from smoking-attributable disease prevented. Cost saving when including savings from smoking-attributable disease prevented.	Positive coverage decision for counselling to prevent tobacco use.
Screening for Depression in Adults	October 14, 2011	Study identified from the literature (Simon et al. 2007) Intervention estimated to be dominant, accumulating 61 additional depression free days while accumulating savings of $314 per patient	Positive coverage decision for screening for depression in adults is reasonable and necessary for the prevention or early detection of illness or disability.
Screening and Behavioral Counseling Interventions in Primary Care to Reduce Alcohol Misuse	October 14, 2011	Study identified from the literature (Solberg, Maciosek, & Edwards 2008) Dominant from societal perspective $1755/QALY saved from a health-system perspective	Positive coverage decision for screening and behavioural counselling to reduce alcohol misuse, in primary care settings, is reasonable and necessary for the prevention of early illness or disability
Screening for Sexually Transmitted Infections (STIs) and High-Intensity Behavioural Counselling (HIBC) to prevent STIs	Expected November 2011	AHRQ study (Glass, Nelson, & Villemyer 2005) Screening all women aged 18-31 years more cost-effective than selective screening. For men, standard practice (e.g., history and examination) is more cost saving than enhanced screening strategies.	Proposed positive coverage decision for screening for chlamydia, gonorrhea, syphilis and hepatitis B, as well as high intensity behavioral counseling (HIBC) to prevent STIs.

3.6. Chapter Summary

In this chapter, I put the US health care system into the perspective of others with respect to spending, abundance of health care resources, and key health statistics. The comparator countries were chosen on the basis that they help illustrate different approaches to using cost-effectiveness evidence in coverage and reimbursement decisions or in recommendations for the efficient use of medical technology. The UK, Sweden, Australia, and Canada, while all using different processes, illustrate countries in which cost-effectiveness evidence plays a fundamental role in decision-making. In contrast, Germany and France illustrate countries in which cost-effectiveness evidence, and economic evidence more generally, plays less of a role in decision-making.

In spite of health care spending greatly in excess of spending in other countries, the US health care system performs poorly across a number of key health metrics, including life expectancy and infant mortality. Further, the US health care system consistently ranks poorly in global health care system rankings. Despite an evident need to increase the return from health care spending, I have shown that the US health care system is notable for the limited extent that cost-effectiveness evidence is used to inform coverage and reimbursement decisions. Although in the US some public and private payers use cost-effectiveness evidence sporadically, how, and the extent to which, it informs decision-making is unclear. Notably, Medicare, the largest payer in the US, states that cost-effectiveness is not a factor considered in their coverage decisions.

To provide insight into the resistance to cost-effectiveness analysis in the US health care system, I described in this chapter attempts by Medicare and the state of Oregon's Medicaid programme to incorporate cost-effectiveness evidence into decision-making. Further, I described 'American exceptionalism', a term used to describe the suggestion that the US is different from other nations because of the uniqueness of its origins, evolution, and institutions, and how this may help explain such resistance. Lastly, I highlight the restrictions that the PPACA legislation imposed on the PCORI with respect to the use of cost-per QALY thresholds, and yet the recent trend of cost-effectiveness evidence featuring in decision memos for NCDs pertaining to preventative care.

The remainder of this thesis describes my empirical work, which focuses on CMS NCDs. Given the prominence of Medicare in the US health care system and its stated position on the use of cost-effectiveness evidence, CMS NCDs are a particularly attractive aspect of the US health care system to evaluate.

Chapter 4 forms the foundation for my empirical work. Here, I present some background of the Medicare programme and Medicare coverage policies. Also, I present the research questions and review the relevant literature that helped inform the methods I use for the research and analyses presented in chapters 5 through 7. Finally, I describe the database I created for the analyses presented in the following chapters, including the literature search I performed to identify relevant estimates of the cost-effectiveness of interventions considered in coverage decisions included in NCDs.

4. Introduction to Empirical Work

4.1. Introduction

In Chapter 2, I presented the underlying theory and rationale for the use of economic evaluation to allocate scarce health care resources and illustrated how cost-effectiveness information can guide efficient health care resource allocation. In Chapter 3, I put the US health care system into the perspective of others with respect to spending, abundance of health care resources, and key health care statistics. I also described that despite health care spending greatly exceeding the spending in other countries, the US health care system performs poorly across key health metrics.

In spite of the US's poor return from health care spending, the US is notable for the limited extent that cost-effectiveness evidence is used to inform coverage and reimbursement decisions. This is particularly evident when considering Medicare. The Centers for Medicare and Medicaid Services (CMS), the administrators of Medicare, state that cost-effectiveness is not a factor considered in National Coverage Determinations (NCDs).

This chapter introduces the empirical component of this thesis. The overarching purpose of the empirical work is to study the use of cost-effectiveness evidence in the US health care system. Given its importance in the US health care system, I chose the Medicare programme as the empirical component of this thesis.

This chapter has three components: first, background is given on the Medicare programme and the coverage policies described; second, the research questions that constitute the empirical work are presented; and third, the database that I created for this research is described, including a description of the included variables.

4.2. Introduction to Medicare

The empirical aspect of this thesis focuses on CMS national coverage determinations (NCDs). CMS administers Medicare, the largest health insurance programme in the US. Established in 1965, Medicare provides coverage for US citizens aged 65 years and older, certain people with disabilities under age 65, and people of all ages with end-stage renal disease (ESRD). (CMS 2005a) More than 46 million Americans (almost one sixth of the population) receive health insurance coverage through Medicare. (CMS 2011a) Medicare is a major health care payer and its coverage decisions may have far reaching effects, influencing the coverage policies of other public and private payers. (Neumann 2005)

With an annual cost of upwards of $600 billion, the Medicare programme is a major part of the US economy. (CMS 2011a) It is estimated that one in five dollars used to purchase health services in 2006 came through the Medicare programme, and it finances about one-third of all hospital stays nationally. (The Henry J.Kaiser Family Foundation 2008) As a percentage of GDP, total expenditure on the Medicare programme is projected to increase from 3.5% to 4.6% between 2009 and 2020, spending that will exceed $1 trillion. (CMS 2009a;Medicare 2010) A major driver of these increasing costs is medical technology. (Ginsburg 2004;Ginsburg 2008) Consequently, as exemplified by recent examples, e.g., autologous cellular immunotherapy treatment of metastatic prostate cancer and screening computed tomography colonography (CTC) for colorectal cancer, CMS's rulings regarding medical technology are increasingly scrutinised and debated. (Chambers & Neumann 2011;CMS 2009b;CMS 2011c;Dhruva, Phurrough, Salive, & Redberg 2009;Garg & Ahnen 2010)

4.2.1. Coverage of medical technology in Medicare

In the Social Security Amendments (SSA) that established Medicare, Congress broadly defined the services to be covered by the programme: (Foote 2002;Marmor 1970)

1. Benefit categories covered – e.g., hospital services and physician services;
2. Services with severe limitations – e.g., dental or chiropractic care;

3. Categories excluded – e.g., personal comfort items or cosmetic surgery.

Within these broad categories, CMS adjudicates payment for specific items and services through its coverage processes.[vii]

4.2.2. Medicare coverage policies

The statutory language that established Medicare did not provide an all-inclusive list of items and services to be covered by Medicare. Rather, the legislation provided the criteria to be used to guide the coverage of items and services. The legislation (1862(a)(1)(A) of Social Security Act) states that Medicare may not reimburse *"items and services which are not reasonable and necessary for the diagnosis and treatment of an illness or injury or to improve the functioning of a malformed body member"*. (Social Security Act 1965) Throughout the first two decades of the programme, how this language should be interpreted was not clarified and was generally considered to mean safe, effective, non-investigational, and appropriate. (Neumann 2005) Indeed, text included in the US Federal Register in 1989 stated that the HCFA did not *"think it possible, or advisable, to try to set quantitative standards or develop formula for the applications of those criteria"*. (Federal Register 1989a;Neumann 2005) How to interpret 'reasonable and necessary' has remained unclear. (Foote 2002;Garber 2001;Neumann 2005) However, as noted in Section 3.4.2.1, one clarification that has been made is with respect to the use of cost and cost-effectiveness evidence. The Guidance for the Public, Industry and CMS Staff states *"Cost effectiveness is not a factor CMS considers in making NCDs. In other words, the cost of a particular technology is not relevant in the determination of whether the technology improves health outcomes or should be covered for the Medicare population through an NCD"*. (CMS 2010e) While attempts have been made to incorporate cost-effectiveness evidence into the review process (Section 3.4.2.1), this remains CMS's stated position on the use of cost-effectiveness evidence.

[vii] Most services available in Medicare are not subject to formal coverage policies. Prospective payment policies, i.e., diagnosis-related groups (DRGs) for inpatient care and ambulatory service payment categories (APCs) for outpatient hospital care, facilitate payment of services not formally evaluated by CMS.

As decisions regarding the availability of interventions in the Medicare programme are becoming increasingly contentious, uncertainty regarding CMS's decision-making criteria is increasingly the focus of debate. (Chambers & Neumann 2011;Dhruva et al. 2009;Fox 2010;Garg & Ahnen 2010)

4.2.2.1 Coverage Determinations

Formal coverage determinations for health services are made by the CMS at either the local or the national level. Local coverage policies, or local coverage determinations (LCDs), are made by 14 independent regional Medicare Administrative Contractors (MACs) in the absence of a national coverage policy and represent the majority of Medicare coverage policies. (CMS 2010b;Foote, Halpern, & Wholey 2005) National coverage policies, or National Coverage Determinations (NCD), are binding to all MACs and are reserved for interventions deemed particularly controversial or projected to have a major impact on the Medicare programme. (CMS 2003d) Medicare Administrative Contractors are bound by NCDs. CMS make approximately 15 NCDs each year and since 1999, a total of 171 have been made. For the most part, NCDs are made by CMS's internal coverage group. On occasion, CMS supplements their review with an external technology assessment (TA) and/or consultation with the Medicare Evidence Development and Coverage Advisory Committee (MEDCAC). (CMS 2010f;CMS 2010h)

Within a NCD CMS often evaluates multiple similar technologies, or more commonly, a single technology for multiple indications. Typically, a proposed NCD is subject to a one month comment period before the final NCD is made. The final coverage policy may be coverage without restrictions, coverage with restrictions, non-coverage, or a referral to regional Medicare contractors. On occasion, CMS has utilised a coverage with evidence development (CED) policy, which provides access to technologies while additional evidence is generated to establish whether expanded coverage is warranted. Each NCD is published in the Federal Register, and a decision memorandum, often referred to as a decision memo, is made available on CMS's website. (CMS 2010a;Federal Register 2010) The decision memo is a structured document and is used to communicate CMS's decision. A decision memo presents a brief clinical background of the disease, a review of the history of Medicare's coverage policies for the intervention, a review and

analysis of the relevant scientific and clinical literature, and CMS's reasoning for the ultimate coverage policy.

4.3. Empirical work

Medicare is a fundamental part of the US health care system. The programme is estimated to cost upwards of $600 billion dollars with costs rising at an unsustainable rate. It is suggested that a major cost driver is medical technology. (Ginsburg 2004;Ginsburg 2008)

As stated above, Medicare coverage decisions have far reaching effects and may influence coverage policies of other public and private payers. (Neumann 2005) The criteria CMS use when evaluating medical technology is vague, with decisions guided by the 'reasonable and necessary' criterion. Importantly, one aspect of decision-making that is clear is the use of cost-effectiveness evidence, which CMS state is not a factor considered in making NCDs.

CMS coverage decisions, and in particular NCDs, are an attractive aspect of the US health care system from a research perspective. NCDs are typically made for interventions expected to have the most significant impact on the Medicare programme and thus could be considered CMS's most important coverage decisions. Although much has been written about NCDs, to the best of my knowledge they have not been subject to empirical analysis. (Gillick 2004;Neumann, Rosen, & Weinstein 2005) An advantage of focusing on CMS NCDs is that the publicly available decision memos provide a rich source of information regarding CMS's review of the evidence base and ultimate coverage decision.

Given these factors, the empirical aspect of this thesis focuses on CMS NCDs. The empirical work considered in Chapters 5 through 7 is described below.

4.3.1. Chapter 5

The first piece of empirical work is presented in Chapter 5 and has two research questions.

As highlighted above, in the Guidance for Public, Industry and CMA Staff, the CMS state that cost-effectiveness is not a factor considered in making NCDs. (CMS 2010e) The first objective of this research is to evaluate NCDs to determine if they are consistent with CMS's stated position. The second part of this research is to evaluate the cost-effectiveness of coverage decisions made as part of NCDs. The data are evaluated to determine if there is a difference between the cost-effectiveness of positive and non-coverage decisions.

The objectives for the empirical work presented in Chapter 5 are as follows:

Objective one

- To examine NCD decision memos to determine if they are consistent with CMS's stated position on the use of cost-effectiveness evidence.

Objective two

- To determine if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness.

4.3.2. Chapter 6

The empirical work presented in Chapter 6 builds on that presented in Chapter 5. While the approach in Chapter 5 evaluates whether there is a relationship between coverage decisions and cost-effectiveness, it does not control for other factors and thus is insufficient to determine if cost-effectiveness is independently associated with coverage decisions. The approach taken in Chapter 6 attempts to control for these factors in the analysis. The objective for the empirical work presented in Chapter 6 is as follows:

Objective one

- To determine if cost-effectiveness is an independent predictor of coverage when controlling for other factors that may be considered to influence coverage decisions.

4.3.3. Chapter 7

The third piece of empirical work is presented in Chapter 7. The empirical work in Chapter 5 concerns the evaluation of the cost-effectiveness of coverage decisions in CMS NCDs. The empirical work in Chapter 7 builds on this research to estimate the potential efficiency gains in terms of health gains and cost-savings from a hypothetical reallocation of expenditures between interventions subject to NCDs, using a criterion of cost-effectiveness. Potential benefits in terms of aggregate health gain and cost-savings, along with the effects of using a cost-effectiveness decision rule on the distribution of resources between disease areas and types of interventions, are estimated. The objective for the empirical work presented in Chapter 7 is as follows:

Objective one

- To estimate potential gains in aggregate health achieved from reallocating expenditures between interventions covered as part of NCDs in a manner consistent with a cost-effectiveness decision rule.

Objective three

- To estimate the impact of reallocation on the distribution of expenditures across disease areas (oncology, cardiology, and other) and types of intervention (treatment, diagnostic, and other).

4.4. Choice of a quantitative approach

I decided to use a quantitative approach for the empirical aspect of this thesis. The factors that decision makers consider relevant in decision-making have been evaluated qualitatively by a number of researchers. (Bryan et al. 2009;Gold MR et al. 2007;Fischer KE et al. 2011;Williams I et al. 2008; Al MJ, Feenstra T, & Brouwer WB. 2004) These studies used surveys and focus groups to gain an insight into the importance of various decision-making criteria. Studies of particular relevance to this thesis include Bryan et al. (2009) and Gold MR et al. (2007), both of which evaluated the use of cost-effectiveness evidence in decision-making.

While a qualitative approach would provide useful insights, I chose a quantitative approach for the empirical component of this thesis for two principal reasons. First, unlike for agencies in other countries, CMS's coverage decisions for medical technologies and interventions have not been evaluated in a quantitative manner and I considered this to be a significant gap in the literature. Second, the availability of decision memos for each CMS NCD provided a rich data source amenable to quantitative evaluation.

4.5. Literature review

To help inform the methodological approach taken for the empirical aspect of this thesis, I reviewed the relevant literature. First, I review the literature pertinent to the research objectives in Chapter 5 and Chapter 6. These chapters were considered together in the literature review as, in spite of having different objectives, both concern the retrospective evaluation of coverage decisions. Second, I review the literature pertinent to the research objectives in Chapter 7.

4.5.1. Literature pertaining to Chapter 5 and Chapter 6

The broad objective of the empirical work presented in Chapters 5 and 6 was to evaluate the use of cost-effectiveness evidence in CMS NCDs. A literature search was performed using the PubMed database to identify studies with similar objectives that evaluated how decision makers elsewhere had incorporated cost-effectiveness information into decision-making. The search

criteria included the terms: *"Health technology assessment"*; *"cost-effectiveness"*; *"cost-effectiveness threshold"*; *"regression analysis"; "coverage decisions"; "reimbursement decisions"; "recommendations"; "decision-making criteria"; "decision-making framework"*. Searches were limited to English-language articles only and included studies published before October 15th, 2011.

I included studies that used a quantitative approach to retrospectively evaluate coverage decisions for medical technologies and interventions made by various agencies. I excluded studies that used a qualitative approach to evaluate decision-making criteria or that did not concern the coverage or reimbursement of medical technologies or interventions.

Given that the objective of the literature search was to identify studies that evaluated how cost-effectiveness evidence had been used in coverage decisions, the identified studies were relevant to institutions that use cost-effectiveness in decision-making. The studies evaluated coverage decisions by national agencies in the UK, Australia, Canada, and New Zealand, although one US-based study was also identified. The reference lists of the identified studies were also reviewed for relevant publications. In addition, I became aware of relevant research presented at a scientific conference. This study, Devlin et al. 2010, is also reviewed below.

4.5.1.1 George et al. 2001

As described in Section 3.3.1.3, the Australian government subsidises the price that consumers pay for drugs listed on the national drug formulary, the Pharmaceutical Benefits Schedule (PBS). The Pharmaceutical Benefits Advisory Committee (PBAC) evaluates submissions by manufacturers and recommends drugs to be listed on the PBS. Manufacturers voluntarily submitted economic evaluations between 1991 and 1993, after which point the inclusion of an economic evaluation in their submission has been mandatory, although no guidance as to what constitutes an acceptable level of cost-effectiveness was communicated.

The objective of the study by George et al. (2001) was to generate a league table of drugs considered by the PBAC ranked in order of cost-effectiveness. The league table was used to test

the hypothesis that the PBAC's decisions were consistent with a maxim of economic efficiency.[viii] Further, coverage decisions were explored to determine if they revealed a cost-effectiveness threshold.

Submissions to the PBAC from January 1991 through 1996 were reviewed, with those incorporating ICERs reporting a cost-per life year or cost-per QALY ratios identified. PBAC recommendations were considered dichotomously: 'recommended', and 'recommended at a lower price/'not recommended'. Cost-per life year studies (n=26) and cost-utility studies (n=9) were considered separately.

For recommendations associated with cost-effectiveness studies reporting cost-per life year gained ratios, ICERs for recommended drugs ranged from $AU 5,517 to $AU $75,286; ICERs for drugs rejected or recommended at a lower price ranged from $AU 42,679 to $AU 256,950. For recommendations associated with cost-effectiveness studies reporting cost-per QALY gained ratios, ICERs for recommended drugs ranged from $AU 4,690 to $AU 24,343. Only two drugs associated with cost-effectiveness studies reporting cost-per QALY gained ratios were not recommended. One, with an ICER of $AU 17,937 per QALY gained was recommended at a lower price, and a second, with an ICER of $AU 133,337 per QALY gained was rejected.

A Mann-Whitney test was used to evaluate whether there was a difference between the cost-effectiveness of drugs recommended and those not recommended for listing. It was reported that there was a statistically significant difference between the cost per life-year gained for drugs that were recommended for listing and those that were not ($p=0.0008$). The small number of studies associated with cost-per QALY ratios and a relatively narrow range of ratios prevented the researchers from drawing conclusions for these studies.

The findings were deemed consistent with the hypothesis that PBAC's decisions were consistent with a maxim of economic efficiency. The authors did not identify a fixed threshold value of

[viii] The authors' definition of economic efficiency was not presented.

cost-effectiveness for which the PBAC appeared less likely to recommend a drug for listing, although stated that the PBAC was less likely to recommend a drug if associated with an ICER greater than $AU 76,000 (1998/1999 values) and was unlikely to reject a drug if associated with an ICER less than $AU 42,000.

The authors conclude that while it is clear that economic efficiency plays a large role, it is not the only factor that influences PBAC's recommendations.

4.5.1.2 Towse and Pritchard 2002

Recommendations made by the National Institute for Health and Clinical Excellence (NICE) have been the subject of a number of studies. In 2002, NICE had not clarified their position on the value of the cost-effectiveness threshold. To evaluate whether an implicit cost-effectiveness threshold could be inferred from NICE's recommendations, Towse and Pritchard (2002) reviewed all appraisals featuring a cost-utility study from 1999 through May 2002. (Towse, Pritchard, & Devlin 2002) Also included were cost-effectiveness studies that reported ICERs using cost-per life year gained or cost-per-episode avoided ratios that could be converted into a cost-per QALY estimate using "*eminently reasonable assumptions*". (Towse, Pritchard, & Devlin 2002)

In contrast to George et al. (2001), rather than generating a league table based upon cost-effectiveness, recommendations were categorised using the reported cost-per QALY ratio; <£20,000 per QALY, £20-£30,000 per QALY, and >£30,000 per QALY. These categories were chosen because NICE had previously given some indication in appraisal determinations that the range £20-£30,000 per QALY was significant, and that £30,000 per QALY was approaching the highest acceptable cost-effectiveness. Technologies were further categorised with respect to NICE's recommendation (Table 10):

1. Those in which a cost-per QALY range was given and the technology was accepted;
2. Those in which a cost-per QALY range was given and access to the technology was restricted to a proportion of the patient group;

3. Those in which a cost-per QALY range was given and the technology was rejected.

Table 10. Findings of Towse and Pritchard (2002)

Cost per QALY estimate	Technology accepted	Technology restricted	Technology rejected
<£20,000	15	3	2
£20-£30,000	4	5	1
>£30,000	3	4	4

The authors use a chi-squared test to test a null hypothesis that there was no relationship between the cost-per QALY estimate and whether the technology was accepted, restricted, or rejected. The findings suggested that there was a positive relationship between the cost-per QALY and recommendations ($p<0.05$). The authors report that the findings were sensitive to their assumptions regarding relationship between cost-per QALY and cost-per life year gain. It is clear from Table 10 that there were exceptions to these findings, i.e., that it is not necessarily the case that technologies are accepted if associated with a cost-per QALY ratio less than £20,000, restricted if in the range of £20,000 to £30,000, and rejected if more than £30,000. The authors reported instances when severity of disease and short survival time lead to NICE accepting a technology with an ICER greater than £30,000 per QALY.

4.5.1.3 Dranove et al. 2003

In their study, Dranove et al. (2003) evaluated health maintenance organisation (HMO) formulary adoption decisions. Their objective was to identify economic and organisational characteristics that affect the likelihood of HMOs, including new drugs on their formularies. (Dranove, Hughes, & Shanley 2003)

Data was obtained from a survey of pharmacy directors and drug-specific data taken from an industry trade journal. Respondents reported on seven drugs and reported information with respect to economic and organisational factors, administrative factors, relationship with pharmaceutical companies, and 'other' factors that included HMO size and per member per month pharmacy costs.

Multivariate logistic regression analysis, adjusting for fixed-drug effects and random-HMO effects, was used to estimate models of formulary inclusion. Five models were estimated. The first included only HMO-specific economic factors, models two through four added various administrative factors, and model five included two drug-specific measures. The dependent variable was dichotomous, taking a value of one if the HMO included the drug on the formulary and zero if they did not. Factor analysis was used to limit the number of predictors.

A number of characteristics were identified as affecting formulary adoption. These included non-profit status (for-profits were estimated to have lower adoption rates); incentives facing the pharmacy director (e.g., if the importance of meeting the drug budget was part of the pharmacy director's performance evaluation, the probability of adoption was reduced); the size of the P&T committee (larger P&T committees tended to approve fewer drugs); the make-up of the P&T committee (replacing two medical personnel on the P&T committee with two nonmedical personnel reduced the likelihood of adoption to 50 percent); the relationship between the HMO and the pharmaceutical company (HMOs tended to favour manufacturers whose representatives pay more visits); and member satisfaction (if the relative importance of member satisfaction increased by 0.8 percent, the chances of adoption increased to 78 percent).

Notably, a number of HMO-specific economic factors, including size, drug expenditures, and whether the primary care physicians were at financial risk for drug costs, did not affect the likelihood of adoption.

4.5.1.4 Devlin and Parkin 2004

In the first of a number of similar studies, Devlin and Parkin (2004) evaluated coverage decisions made by NICE in the UK. (Devlin & Parkin 2004) The primary objective of this research was to explore NICE's cost-effectiveness threshold(s) and the trade-offs made between cost-effectiveness and other factors pertinent to decision-making. The methodological approach taken also facilitated an exploration of NICE's preferences and considered the consistency of recommendations.

NICE's recommendations made prior to May 2002 were considered and comprised of 39 technologies and 51 recommendations. Data were abstracted from the publicly available NICE Technology Appraisals. The recommendation was the dependent variable and took the value 0 if NICE recommended the use of the technology, and the value 1 if NICE recommended against the use of the technology. Recommendations originating from the same technology appraisal were considered independent. Independent variables included: the estimated cost-effectiveness reported as cost-per life year or cost-per QALY gained ratios; uncertainty regarding cost-effectiveness; the burden of disease; the availability of alternatives to the technology under review; and specific factors indicated by NICE (e.g., severity of condition, short life expectancy, etc). For the cost-effectiveness variable, cost-per life year gained ratios were included when cost-per QALY gained ratios were unavailable. The authors acknowledge the weakness of this approach, i.e., assuming a one-to-one correspondence between life years gained and QALYs gained, but claimed that this was necessary to ensure sufficient degrees of freedom.

Recommendations were divided into those with associated cost-effectiveness evidence and those without. Those with cost-effectiveness evidence (n=33) were deemed amenable to quantitative analysis and were explored using logistic regression. A binary choice logistic regression model was used. The value of the cost-effectiveness threshold was estimated by calculating the probability of a favourable recommendation for each ICER, while holding other variables constant at their mean value. Four different models were considered. Model 1 was univariate and included only the cost-effectiveness variable. The other three models increasingly added

explanatory variables, including uncertainty regarding cost-effectiveness (Model 2), burden of disease (Model 3), and availability of alternatives (Model 4).

A chi square test was used to determine that none of the four models could be rejected. The authors reported that Model 4 was 'preferred' as it was associated with the highest pseudo R^2 and sensitivity (although specificity was slightly lower than for other models). The results indicated that the likelihood of a positive coverage recommendation decreased as the cost-effectiveness ratio increased. Cost-effectiveness along with other variables better explained NICE's decisions than cost-effectiveness alone. The findings also indicated that the threshold appeared to be somewhat higher than the £20-30,000 per QALY range, NICE's then stated '*range of acceptable cost-effectiveness*'.

4.5.1.5 Dakin et al. 2006

Dakin et al. (2006) built on the previous approach by Devlin and Parkin (2004) and tested an alternative model of decision-making. Rather than modelling decision-making using a binary choice model, this paper used a multinomial dependent variable to better reflect NICE decision-making. The aim of this study was to gain additional insight into the determinants of NICE decisions and trade-offs between them.

NICE Technology Appraisals published up to December 31st, 2003 were evaluated, with data abstracted from 73 appraisals, constituting 94 recommendations (a number of appraisals were subdivided into 2-4 separate recommendations). Recommendations were categorised as "recommended for routine use", "recommended for restricted use", or "not recommended". No ranking was assumed in the dependent variable; the authors believed that the three categories were qualitatively different and "recommended for restricted use" did not represent an intermediate point between the other categories. Various independent variables were considered, including: quantity/quality of clinical evidence; cost-effectiveness; decision date; existence of alternative treatments; budget impact; and intervention type. Variables concerning clinical effectiveness included those pertaining to the number, type, quality, and outcome of reviewed

studies. The primary analysis included only estimates of cost-effectiveness reporting cost-per QALY gained, although a secondary analysis pooled cost-per QALY studies along with cost-per life year studies.

The model was estimated using multinomial logistic regression. Univariate and multivariate regressions were performed. Results showed that high cost-effectiveness ratios increased the likelihood of technologies being rejected rather than recommended for restricted use. Pooling cost-per life year studies along with cost-per QALY studies increased the reported pseudo R^2 compared to the reported value when only cost-per QALY studies were included. The authors suggested that this finding confirms that cost-effectiveness evidence is an important factor in NICE decision-making. With respect to the clinical evidence base, the study showed that the number of RCTs and systematic reviews were statistically significant, i.e., technologies with a larger evidence base were more likely to be recommended for routine use. The results also suggested that pharmaceuticals and technologies evaluated earlier were also less likely to be rejected. Patient group submissions increased the likelihood of a recommendation for routine rather than restricted used.

The authors concluded that the factors affecting the recommendation between routine and restricted use, but not that between routine use and rejection, suggests that that the model was an improvement over the binary-choice model reported by Devlin and Parkin (2004), and that modelling the three outcomes as opposed to a binary choice model more closely reflects NICE decision-making.

4.5.1.6 Harris et al. 2008

Similar to George et al. (2001), Harris et al. (2008) evaluated PBAC recommendations. However, in contrast to George et al. (2001), Harris et al. (2008) considered a range of additional independent variables rather than solely focusing on the cost-effectiveness of recommendations.

The objective of the study was to evaluate the relative influence of factors in PBAC's recommendations. PBAC recommendations from February 1994 through December 2004 were considered (n=858). (Harris et al. 2008) Following exclusion of submissions with insufficient data, the final sample included 103 submissions reporting a cost-per QALY gained ratio and 123 submissions reporting a cost-per life year gained ratio.

In addition to cost-effectiveness, a variety of variables were considered, including; an assessment of the clinical evidence base (clinical importance of treatment effect, precision of clinical evidence, relevance of evidence, etc), severity of the condition (whether condition is associated with premature mortality), availability of alternatives, and the associated budget impact of the technology.

The probability of the PBAC recommending a drug was estimated using a probit multiple regression model. The dependent variable was dichotomous: recommendation and non-recommendation. Two models were estimated, the first including all explanatory variables, the second excluding non-significant groups of variables (determined using a Wald test $p>0.05$). The predictive power of the model was assessed by its pseudo R^2, Hosmer and Lemeshow's goodness of fit test, and the area under the receiver operating characteristic.

The results of the regression when including cost-per QALY studies were presented. It was determined that clinical significance, cost-effectiveness, budget impact, and severity of disease were significant predictors of PBAC's recommendation. In comparison to the average submission, drugs estimated to be clinically significant were associated with an increased probability of coverage of approximately 0.2, and drugs indicated for a life-threatening condition were associated with an increased probability of coverage of approximately 0.4. From the mean reported cost-per QALY ratio of $AU 46,400, an increase of $AU 10,000 corresponded to a decrease in the likelihood of a positive recommendation by 0.06. The authors did not report the results of the analysis with submissions reporting cost-per life year ratios included. However, it was reported that this analysis confirmed the findings of the primary analysis.

The authors concluded that while there was evidence of the probability of a positive recommendation decreasing with higher ICERs, there is no evidence of a fixed cost-effectiveness threshold.

4.5.1.7 Rocchi et al. 2008

As described in Section 3.3.1.4, the Common Drug Review (CDR) is the central review agency for new outpatient medications in Canada. As part of the drug review process, the CDR submits a report to the Canadian Expert Drug Advisory Committee (CEDAC), which considers the drug's effectiveness and cost-effectiveness relative to the standard of care. Rocchi et al. (2008) reviewed CEDAC recommendations published from September 2003 through March 2007 to evaluate the role that cost-effectiveness evidence played in oncology reimbursement decision-making. (Rocchi et al. 2008)

The results of the research were not presented in a league table as per George et al. (2001), nor categorised into categories of cost-effectiveness as per Towse and Pritchard (2002). Rather, for the 25 recommendations with accompanying cost-effectiveness data, drugs were categorised as to whether CEDAC considered the drug to be attractive or unattractive from a cost-effectiveness perspective and whether the drug was listed. For each category the reported range of cost-effectiveness for the respective drugs was listed. The authors did not separate studies reporting cost-per life year gained and cost-per QALY gained ratios in their presentation of study findings.

On four occasions the ICER was considered attractive and the drug listed, with ICERs ranging from dominant to $CAN 71,000 per life year gained. Nine drugs were listed with ICERs deemed unattractive, with ICERs ranging from $CAN 50,000 to $CAN 80,000 per QALY. Twelve drugs were not listed with ICERs deemed to be unattractive, with the ICERs ranging from $CAN 32,000 to $CAN 137,000 per QALY.

Of the drugs considered, oncology medications were recommended for listing with the highest ICERs, with the highest $CAN 80,000 per QALY. Anti-retrovirals were another evaluated subset, with the highest ICER of a recommended drug $52,000 per QALY. Only one drug not classified as an oncology medication or an anti-retroviral was listed with an ICER substantially more than $50,000 per QALY. In this instance it was noted that the ICER was "*in excess of traditional standards*".

The authors conclude that oncology drugs seem to be adopted at the higher thresholds of acceptability than non-oncology drugs. However, as no hypothesis test was performed, it is not possible to conclude with any level of certainty that this was the case.

4.5.1.8 Chim et al. 2010

Chim et al. (2010) also evaluated PBAC recommendations. (Chim et al. 2010) The background to the study was concern that existence of a cost-effectiveness hurdle for the reimbursement of drugs in Australia may limit access to new cancer treatments because of their high cost and modest benefits. The primary objective of this study was to test the hypothesis that ceteris paribus, cancer drugs were less likely be recommended by the PBAC for reimbursement on the PBS than non-cancer drugs.

Public summary documents (PSDs) on major submissions to the PBAC from July 2005 to March 2008 were reviewed (n=227), corresponding to 243 recommendations (on occasion multiple recommendations originated from a PSD). Only drugs used to treat cancer were classified as cancer drugs; those indicated for cancer related nausea and vomiting, or neutropaenia or anaemia were classified as non-cancer drugs. All PSDs reported cost-per QALY information as a range or as a single estimate. When reported as a range, the highest value was used for the analysis.

Cost-effectiveness was included as a categorical variable with the type of model (cost-minimisation analysis or cost-effectiveness analysis/cost-utility analysis) and cost-per QALY ratio (≤$AU 45,000 per QALY; >$AU 45,000 to ≤$AU 75,000 per QALY; >$AU 75,000 per

QALY) accounted for. The estimated annual cost of the drug to the PBS was included. Whether the PBAC accepted the manufacturer's clinical claim was included as a binary variable, as was whether the PBAC accepted the manufacturer's nominated comparator. With respect to predicted utilisation of the drug, the number of patients that would use the drug in a year was considered. Further, submissions were categorised as to their type of application, i.e., whether they pertained to a new drug, new indication, or other.

The PBAC recommendation, i.e., whether the drug was approved or non-approved, was classified as a dichotomous variable by merging instances when a drug was rejected with instances when a drug was partially accepted.

The model was estimated using a binomial logistic regression. Univariate and multivariate analyses were performed. In the multivariate model, type of application, cost-effectiveness, and estimated cost to the PBS were statistically significant. For all submissions, it was reported that the likelihood of a positive recommendation decreased with higher estimates of cost-effectiveness. Submissions for cancer drugs were associated with higher cost-per QALY ratios than non-cancer drugs. However, after adjusting for other factors, there was no statistical difference between cancer and non-cancer with respect to PBAC's recommendation for PBS listing.

4.5.1.9 Devlin et al. 2010

Devlin et al. (2010) was a podium presentation presented at the International Society for Pharmacoeconomics and Outcomes Research's (ISPOR) 13th Annual European Congress in November 2010. (Devlin et al. 2010b) The full study is yet to be published in the peer-reviewed medical literature, and the following information is based upon the material in the presentation. As for Devlin and Parkin (2004) and Dakin et al. (2006) the researchers evaluated NICE Technology Appraisals, although the study time period was not presented.

Study aims were to build upon previous modelling approaches to determine the role of cost-effectiveness and other factors on NICE decisions. Specific research questions were: Has NICE's cost-effectiveness threshold or decision-making process changed over time? What effect does evidence and factors other than cost-effectiveness exert on NICE's decisions?

The researchers used data obtained from HTAInSite, a proprietary database of NICE decisions, supplementing this with data from NICE Technology Appraisal documents when required. Technology appraisals were sub-divided into sub-decisions, with the unit of analysis the individual coverage decision.

Variables considered included cost-effectiveness, patient numbers in RCTs, availability of treatment alternatives, whether the decision pertained to children, whether the submission originated from a patient group, and date of decision. Estimates of technologies that were dominant or dominated were excluded from review, i.e., all technologies included were associated with a positive ICER (more effective and more expensive than comparator). Only studies reporting cost-per QALY ratios were included.

Logistic regression was used to estimate the model, with the dependent variable coded dichotomously as positive or non-coverage. The authors found that cost-effectiveness evidence alone explained the majority of NICE's decisions. Cancer treatments were estimated to have a higher probability of recommendation ceteris paribus. Date of decision was not a significant predictor of coverage, suggesting that the decision-making process had not significantly changed over time. With respect to estimation of the cost-effectiveness threshold (when probability of rejection = 50%), cancer treatments were estimated to be associated with a higher threshold (approximately £50,000 per QALY) than non-cancer treatments (approximately £38,000 per QALY).

4.5.2. Literature pertaining to Chapter 7

The objective of the third piece of empirical work was to hypothetically reallocate expenditures between interventions considered in CMS NCDs in accordance with available cost-effectiveness evidence. A literature search was performed using the PubMed database to identify studies with similar objectives that estimated efficiency gains in terms of aggregate health from a hypothetical reallocation of resources. The search criteria included the terms: *"priority setting"*; *"resource allocation"*; *"cost-effectiveness"*; *"cost-effectiveness threshold"*; *"optimisation"*; *"disinvestment"*. Searches were limited to English-language articles only and included studies published before October 15th, 2011.

I included studies that estimated that consequences of using alternative resource allocation criteria to estimate aggregate health gains across multiple interventions and/or indications. I excluded studies that focused on a single intervention or technology; that did not provide an estimate of aggregate health gain, e.g., burden of illness studies; and cost-effectiveness analyses.

While the search strategy identified a number of studies, only three met my inclusion criteria. The majority of identified studies were cost-effectiveness analyses that evaluated various individual health care technologies or policies. While a body of literature was identified concerning evidence-based priority setting, this literature was largely descriptive in nature and did not provide an empirical framework for the research presented in Chapter 7. (Foglia et al. 2008;Mitton & Donaldson 2003) A number of burden of illness studies featured in the literature search results, including some that quantified health using QALYs, e.g., van Hoek et al. (2011) However, as these studies did not evaluate the consequences of using alternative resource allocation criteria, they did not help inform the empirical framework for the research presented in Chapter 7. Tengs et al. (1995) was an example of a study that was excluded because an estimation of efficiency gain was not provided. In this study the authors gathered information on 587 life-saving interventions across a range of industries. However, as an estimation of aggregate health gain achievable from a reallocation of resources was not presented, this study was not reviewed.

Three studies, Cromwell et al. (1998), Zaric and Brandeau (2001) and Ratcliffe et al. (2005), were identified with similar objectives to those employed in Chapter 7 and are described below.

4.5.2.1 Cromwell et al. (1998)

Cromwell et al. (1998) was a demonstration project set in Australia that used an integer programming approach to allocate resources across acute inpatient services. The objective of the study was to find the mix of services that would maximise health gain from available resources. The authors derived effectiveness data from the Oregon Health Services Commission (measured in QALYs) and resource use data from Australian National Diagnosis Related Groups (AN-DRG). Utilisation data was derived from regional activity data. Over a one-year timeframe, the model estimated potential gains of approximately 353,000 QALYs from the treatment of 45,000 patients.

In the study discussion, Cromwell et al. (2008) suggest that this research makes visible the trade-offs implicit in health policy decision-making, and the opportunity costs in terms of health gain. However, they acknowledge a number of challenges that would need to be overcome before such a model could be implemented in practice. The principal hurdle is the requirement for high quality data regarding the costs and benefits of interventions. Also, for the set of included interventions, there were small differences in the cost-effectiveness of interventions close to the margin, i.e., those immediately above and below the cut-off value of cost-effectiveness. This would prove problematic in practice, as it would be challenging to operate a rigid cost-effectiveness decision rule with little to distinguish between interventions in terms of cost-effectiveness. The authors note that expanding the model beyond acute inpatient services would likely have an inhibitive effect on model complexity.

The study had a number of characteristics that restricted its applicability to the empirical work considered here. For example, Cromwell et al. (1998) did not employ the cost-effectiveness literature; rather, they used separate sources for the evidence regarding cost and effect. Also,

included interventions were limited to forms of acute inpatient care. Finally, costs were limited to a one-year time horizon.

4.5.2.2 Zaric and Brandeau (2001)

The objective of Zaric and Brandeau (2001) was to determine the optimal allocation of resources for HIV prevention and to investigate the impact of alternative patterns of resource allocation on health outcomes. The patient population considered for this research were a hypothetical cohort of 1 million injection drug users (IDUs) and non-IDUs. High prevalence and low prevalence communities were considered, each with differing numbers of IDUs, methadone treatment slots, and HIV prevalence. Three HIV prevention interventions were considered: a needle exchange programme, methadone maintenance treatment, and condom availability programmes. With a hypothetical budget of $1 million, the set of expenditures that maximised aggregate QALYs gained and the number of infections averted was determined. A dynamic epidemic model was used to model the spread of HIV and the flow of IDUs into and out of methadone maintenance over a two-year time horizon. In the low prevalence community, the model resulted in an additional 45.0 QALYs gained, and in the high prevalence community, 7.9 QALYs were gained.

A number of characteristics limited the applicability of this research to the empirical work considered here. Zaric and Brandeau (2001) focused solely on HIV prevention, with only three interventions included. Also, the model was limited to a short time horizon (two years). Cost-effectiveness evidence was not used in Zaric and Brandeau's model, and thus their framework had limited applicability to this research.

4.5.2.3 Ratcliffe et al. (2005)

The literature search identified Ratcliffe et al. (2005) as a potentially relevant study. This study estimated the consequences in terms of costs and aggregate health of using alternative allocation rules for donor liver grafts. A discrete choice experiment was used to generate relative weights for several key factors that might be used to prioritise patients for liver transplantation. These weights were used to develop a "patient-specific index" for patients who received a liver

transplant from centres in England and Wales. The patient-specific index was used to guide resource allocation on the basis of equity. Costs and aggregate health resulting from this allocation were compared to when resources were allocated in order to maximise efficiency, i.e., to maximise health from available resources. The authors used a Wilcoxan signed rank test to show that there was a statistically significant difference in ranks ($p<0.001$) between when patients were ranked in order of efficiency and equity. The authors of the study concluded that the general public's priorities may not be in accordance with a pure efficiency objective and using them to guide resource allocation may lead to an increase in costs and a decrease in aggregate health.

While Ratcliffe et al. (2005) provided a useful insight into the estimation of hypothetical efficiency gains/losses when using alternative resource allocation rules it was of limited applicability to the empirical work considered here. The study focused solely on the allocation of liver grafts and the study's limited scope prevented me from emulating the approach for the empirical work presented in Chapter 7.

4.6. Methodological approach

The studies reviewed above informed the methods used for the empirical work presented in Chapters 5 through 7. Although the objectives of the empirical work, particularly Chapter 5 and Chapter 6, have a degree of similarity with a number of the reviewed studies, due to a number of data limitations it was not possible to emulate the studies precisely. Nevertheless, a number of the reviewed studies provided a framework with which to approach the research objectives stated in Section 4.3.

For example, the broad objective of George et al. (2001), Towse et al. (2002), and Rocchi et al. (2008) was to evaluate aspects of the cost-effectiveness of coverage decisions or recommendations. These studies were particularly relevant to the empirical work presented in Chapter 5 in which one objective is to assess if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness. George et al. (2001) proved particularly useful with respect to the presentation of cost-effectiveness studies and statistical approach.

The remainder of the studies described in Section 4.5.1, i.e., Dranove et al. (2003), Devlin and Parkin (2004), Dakin et al. (2006), Harris et al. (2008), Chim et al. (2010), Devlin et al. (2010), helped provide a framework for the empirical work presented in Chapter 6. These studies used variations on a regression approach to evaluate the influence of cost-effectiveness on coverage decisions or recommendations while controlling for other factors. Further, these studies provided useful insight into what factors are important to decision makers and helped inform the variables I included in the database (Section 4.7.4).

Unfortunately, it was not possible to find studies that proved as helpful to inform the methodological approach for the empirical work presented in Chapter 7. Cromwell et al. (1998) and Zaric and Brandeau (2001) illustrated different approaches to estimating aggregate health gains from alternative resource allocations, but their complexity and data requirements restricted their applicability to the empirical work considered here. In contrast, the narrow scope of

Ratcliffe et al. (2005), i.e., its focus on the allocation of liver transplantations limited its applicability to this research.

4.7. Methodology - Generation of database

This section describes the development of the database used for the empirical work presented in chapters 5, 6 and 7.

4.7.1. Identification of NCDs

As noted above, since 1999 CMS have posted a decision memo for each completed NCD on their website. I included decision memos relevant to NCDs made from 1999 through 2007 in this research.

4.7.2. Decision memoranda

I downloaded and reviewed each decision memo. I included only decision memos relevant to national coverage policies, excluding those pertaining to instances when MACs were granted coverage discretion. I also excluded incomplete memoranda, or those that pertained to minor language changes. For example, the NCD for external counterpulsation (ECP) therapy (CAG-00002N) was opened to clarify the language used to describe the device and thus was not included in the database. Finally, I excluded NCDs pertaining to the coverage of medical technology in clinical trials (coverage with evidence development (CED)), as CED policies were not deemed equivalent to either positive or non-coverage decisions.

4.7.3. Unit of analysis

At the outset of the research, my expectation was that the NCD would be the unit of analysis. However, upon review of the decision memos it became apparent that few NCDs represented a single coverage decision. A multiplicity of decisions arose in a single NCD for several reasons. First, two or more related but distinct technologies were considered. One example of this is the NCD for image guidance for breast biopsy in which both stereotactic and ultrasound image guidance were considered. (CMS 1999a) Second, a single technology was considered for multiple indications, as was the case with the NCD for Positron Emission Tomography (PET) imaging for various forms of cancer. (CMS 2000) Third, the NCD may consider multiple uses of

a technology for a particular indication, such as the NCD for PET which considered technologies for both diagnosis (first use) and staging (second use) of cancer. (CMS 2000) Finally, a "coverage with conditions" decision implicitly gives rise to two or more coverage decisions: a positive decision for the covered indication or population and a negative decision for the non-covered applications. Rarely is a positive coverage decision made without restrictions. In the majority of NCDs, CMS placed restrictions on the eligible patient population. For these NCDs, an entry was made in the database for the population deemed eligible for the medical technology or intervention, with additional entries included for the populations deemed ineligible (Figure 14). An example of this scenario was the NCD for bariatric surgery. In this case, a positive coverage decision was made only for Medicare beneficiaries who have a body-mass index (BMI) >35, at least one obesity-related co-morbidity, and for whom previous obesity treatment proved unsuccessful. (CMS 2006a) Accordingly, a non-coverage decision was included for Medicare beneficiaries who did not meet the stipulated criteria.

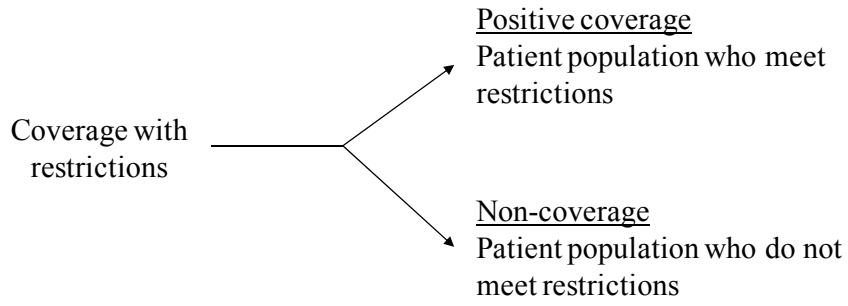


Figure 14. Separate coverage restrictions in a single decision memo

4.7.4. Database

I developed a database in Microsoft Excel and included various data pertaining to each coverage decision. I downloaded one hundred and forty decision memos from the CMS website. (CMS 2010f) Following review, I excluded 37 decision memos based on the exclusion criteria (Section 4.7.2). From the 103 decision memos included in the database, I identified 255 coverage decisions.

The process I used to review each decision memo is illustrated using a worked example for the NCD for deep brain stimulation for Parkinson's disease (Appendix 3). The variables considered in this review and for inclusion in the database are presented below.

4.7.4.1 Coverage decision

I reviewed each decision memo to identify included coverage decisions. I categorised coverage decisions as either positive coverage or non-coverage and thus coded *Coverage decision* as a dichotomous variable. This approach is consistent with Dranove et al. (2003), Devlin and Parkin (2004), and Harris et al. (2008).

As it was often the case that a decision memo was the source of multiple coverage decisions, it was necessary to account for the possibility that these coverage decisions may be related, i.e., there may have been overlap in the reviewed evidence, and that the coverage decisions were made by the same reviewers under the same circumstances. I numbered NCDs chronologically to allow for the 'clustering' of coverage decisions to be accounted for.

I also categorised coverage decisions as implicit and explicit. Explicit coverage decisions were those for which a review of the evidence was presented in the decision memo. Implicit decisions were those for which while it was clear CMS had made a coverage decision a review of the evidence was not presented in the decision memo. An example of this was the NCD for foot care for diabetic patients. (CMS 2001) The NCD included a positive coverage decision for foot care for diabetic patients with peripheral neuropathy with loss of protective sensation (LOPS). For this coverage decision a review of the evidence base was presented in the decision memo. Implicit in the NCD was a non-coverage decision for diabetic patients with peripheral neuropathy but without LOPS, despite the absence of a review of the supporting evidence base. (CMS 2001)

On occasion, coverage decisions were reconsidered at a later date. As CMS revisit a coverage decision only when the body of evidence is sufficient to warrant reconsideration, I considered these to represent unique observations.

I coded *Coverage decision* '1' for positive coverage decisions, and '0' for non-coverage decisions.

4.7.4.2 Quality of evidence

Evaluating the relationship between coverage decisions and the evidence base was an important aspect to this work. *Quality of Evidence* is a categorical variable included in the database that characterises the clinical evidence supporting the coverage decision. This classification is based on an independent review of each decision memo using a grading scale adapted from the United States Preventative Services Task Force (USPSTF) guidelines. The USPSTF classification criteria are presented in Table 11.

Table 11. Grading evidence according to net benefit and quality of evidence (USPSTF Guidelines)

	Magnitude of net benefit, recommendation grade				
Quality of evidence[ix]	**Substantial**	**Moderate**	**Small**	**Zero/ negative**	**Insufficient information**
Good	A	B	C	D	I
Fair	B	B	C	D	I
Poor	I	I	I	I	I

[ix] The USPSTF guidelines used the following criteria :
Good: Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes.
Fair: Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies, generalisability to routine practice, or indirect nature of the evidence on health outcomes.
Poor: Evidence is insufficient to assess the effects on health outcomes because of limited number or power of studies, important flaws in their design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.

The USPSTF classification criteria captures two dimensions of the evidence base, the magnitude of net clinical benefit and evidence quality. For each coverage decision, two trained reviewers from Tufts Medical Center independently reviewed CMS's evaluation of the evidence as presented in the decision memo and assigned a grade using the USPSTF classification criteria (Table 11). The Tufts Medical Center researchers clarified any discrepancies during a consensus meeting. I reviewed the grading assigned to each coverage decision to ensure consistency with the evidence base. On two occasions I spoke with the researchers at Tufts Medical Center to gain clarification before finalising the database. In the database, I classified evidence as "*Good*" when graded as "A" or "B", "*Poor*" when graded as "C", and "*Insufficient*" when graded "I". No evidence was classified with a 'D' grading.

Although combining the magnitude of net clinical benefit and the quality of evidence dimensions is not ideal, evidence grading using the USPSTF classification criteria was the only approach taken by the researchers at Tufts Medical Center. Another critique of this variable is its subjectivity. Clinical studies included within CMS's review were not independently reviewed by the Tufts Medical Center researchers; rather, their grading was based upon the presentation of CMS's review of the evidence in the decision memo. In spite of these limitations, using this approach was deemed defensible and appropriate given the limited resources. Potential approaches that may be taken in the future to improve on this variable and to classify the evidence in an objective manner are discussed in Section 6.4.5.

4.7.4.3 Alternative intervention

Alternative intervention is a dichotomous variable (*Alternative available; No alternative available*) that captures whether or not an alternative intervention was available for the indicated patient population. A number of the studies reviewed above (Section 4.5.1), including Devlin and Parkin (2004), Dakin et al. (2006), and Harris et al. (2008), accounted for the availability of alternatives in the model. Based upon information presented in the decision memo, I made a judgment as to the availability of alternative interventions. I coded the coverage decision as *No alternative available* in situations when it was clear from the decision memo that there were no alternative interventions available. Also, I coded a coverage decision as *No alternative available*

in situations where it was clear from the decision memo that the intervention was permitted only after failure of all available alternatives, because in this usage no further alternatives are available. Therefore, for the purposes of this research I deemed that best supporting care was not an independent treatment option. Abarelix for the treatment of prostate cancer is an example of a coverage decision when no alternative was available. In this case, treatment with abarelix was deemed 'reasonable and necessary' only for patients with advanced symptomatic prostate cancer in whom gonadotropin-releasing hormone (GnRH) agonist therapy is not appropriate, who decline surgical castration, and meet other clinical restrictions. (CMS 2005b)

4.7.4.4 Type of intervention

Various types of intervention are considered through the NCD programme. At the outset, I attempted to categorise the intervention as to its modality, i.e., pharmaceuticals, medical devices, surgeries, and health education and counselling. This categorisation proved unsuitable, as often the intervention did not fit precisely into a single category. Examples are cardiac stents and bariatric surgery, which could be categorised as either surgeries or medical devices. Consequently, I categorised interventions as to their broad indication; those used to treat disease, those used for diagnosis, screening or staging, and those that did not fit into either category, i.e., health education, preventative care, and mobility assistive equipment. Accordingly, I categorised the interventions as *Treatment*, *Diagnostic test*, or *Other*.

4.7.4.5 Coverage requestor

A request for a NCD may originate externally or be internally generated by CMS. An external request for coverage can be made from an individual or entity (e.g., the manufacturers, health plans, providers, etc) who "*identifies an item or service as a potential benefit (or to prevent potential harm) to Medicare beneficiaries*". (CMS 2010e) Alternatively, CMS may internally initiate a NCD in the interest of "*the general health and safety of Medicare beneficiaries*". (CMS 2010e) CMS are prompted to initiate a NCD for an intervention already in use based on several factors, including uncertainty of risks and benefits, the availability of new evidence suggesting a required amendment of existing policies, and non-uniform local coverage policies. (Neumann, Kamae, & Palmer 2008)

I included *Coverage requestor* in the database to account for differences in the origin of the coverage request. *Coverage requestor* is a categorical variable including the following categories: *Manufacturer* (medical device or pharmaceutical company); *Internally generated* (decision to consider the intervention for coverage made by CMS); and *Other* (medical or professional societies and organisations, patient groups, etc). I used information presented in the decision memo and the accompanying 'tracking sheet' to generate *Coverage requestor*. (CMS 2010a)

4.7.4.6 Date of decision

I included the date of the coverage decision in the database. Consideration of the date of decision allowed me to control for unobserved factors that affect the outcome of NCDs that change over time (e.g., composition of decision-making body). I used the date reported in the decision memo. *Date* is a categorical variable including the following categories: *1999-2001*, *2002-2003*, *2004-2005*, and *2006-2007*. I considered a number of alternative ways to code *Date*, including: coding as a continuous variable; coding as a categorical variable using each year as a separate category; and coding as a categorical variable categorising the observations into quartiles. I coded *Date* as a categorical variable because odds ratios for continuous variables are less straightforward to interpret. It coded the variable in groups of years as I thought this to be the most intuitive approach to present the findings. I also investigated the impact of alternative coding on the findings of the analysis (Section 6.3.3).

4.7.4.7 Cost-effectiveness

The inclusion of cost-effectiveness evidence was an important aspect of this research. In the following sections, the development of the cost-effectiveness variable is described.

4.7.4.8 Identification of cost-effectiveness analyses

The first objective for the research presented in Chapter 5 is to examine NCD decision memos to determine if they are consistent with CMS's stated position on the use of cost-effectiveness evidence. Therefore, the first step in identifying relevant estimates of cost-effectiveness was to review each decision memo to determine if a cost-effectiveness analysis had been discussed or cited. (CMS 2010e) In addition, I reviewed accompanying commissioned technology assessments and/or MEDCAC meeting documentation. If no cost-effectiveness analysis was discussed or cited in the decision memo, or conducted as part of a commissioned technology assessment or MEDCAC, I performed a literature search to identify pertinent studies. Primary sources used included the PubMed database, Tufts Medical Center CEA Registry, the Health Economic Evaluations Database (HEED), and the NHS Economic Evaluation Database (NHS EED). (CEA Registry 2010;HEED 2010;NHS-EED 2011;U.S.National Library of Medicine 2011) Secondary sources included internet search engines, conference abstracts, and manufacturers' websites. I performed a literature search for each coverage decision (n=255). Search terms included; generic and brand names of the intervention, all synonyms, and "economic evaluation," "cost-effectiveness," "cost-utility," "cost-minimisation," "decision analysis", "decision model", and "decision analytic model" (See Section 4.7.4.10, for a worked example). Studies were obtained through the Brunel University library. When the electronic copy was not accessible, a paper copy was obtained through the British Library. I included cost-effectiveness studies irrespective of date of publication, although those available at the time of the decision were recorded as such. The search included articles published through December 31st, 2007.

I included cost-effectiveness studies reporting cost-per QALY gained or cost-per life year gained ratios. Also, I included cost-effectiveness studies reporting clinical outcomes measured in disease specific units, such as reduction in blood pressure or decrease in ulcer surface area, when the intervention was dominant, i.e., more effective and less expensive than its comparator, or dominated, i.e., less effective and more expensive than its comparator. I reviewed cost-effectiveness analyses that were a good match to the coverage decision with respect to the intervention, comparator, indication, and patient population.

4.7.4.9 Evaluation of cost-effectiveness studies

I extracted data from each cost-effectiveness study and presented it using a table format (Appendix 3). I chose review criteria based upon the suggested checklist for assessing the quality of decision analytic models reported in Drummond et al. (2005). The criteria were selected to best reflect the aims and objectives of this research, i.e., to select the highest quality cost-effectiveness study most relevant to the coverage decision. To this end, I extracted the following data:

Year of study

- Price year reported in study. If the price year was not reported, the study was reviewed to determine the time period over which data was collected and this was reported accordingly.

Perspective of study

- The perspective of the analysis, e.g., a societal perspective, a health care system perspective, or a patient perspective, was reported.

Comparator

- Each intervention and comparator included in the cost-effectiveness analysis was reported. Only cost-effectiveness analyses that included interventions and comparators relevant to the coverage decision, as determined from review of the decision memo, were included.

Country setting

- The country setting of the cost-effectiveness study was reported.

Study population

- Characteristics of the patient population evaluated in the cost-effectiveness study were reported, e.g., average age, nationality, comorbidities, etc

Incremental cost-effectiveness ratio (ICER)

- ICERs presented in the cost-effectiveness study were reported. As noted above, studies reporting cost-per QALY and cost-per life year ratios were included. Instances when the intervention was estimated to be dominant or dominated were reported accordingly.

Uncertainty associated with reported ICER(s)

- Estimates of uncertainty surrounding the ICER(s) were reported when available. The methodology used for sensitivity analysis, i.e., deterministic or probabilistic, was also reported.

Date of study publication

- The date of study publication was reported. Studies published through December 31st, 2007 were considered.[x]

The purpose of the cost-effectiveness analysis

- The purpose of the cost-effectiveness analysis, i.e., whether the study was prepared for submission to a regulatory body, written for publication, etc, was reported.

Other comments

- Any other aspects of the study deemed important to this research were reported. Examples of reported information included the source of study funding, origin of the clinical evidence, appropriateness of model time horizon, discount rate, etc.

Adjustment of ICER (currency and year)

For cost-effectiveness analyses performed in a country other than the US, I converted the reported ICER into US dollars. I used the purchasing price parity (PPP), available on the OECD website, to convert non-US estimates of cost-effectiveness into US dollars. (OECD 2010) The PPP is a conversion factor that represents how much of a country's currency is needed in that country to buy what $1 would buy in the United States. (World Bank 2010) When necessary, I inflated/deflated the reported ICER to the year in which the NCD was made. I used the health care component of the US consumer price index (CPI) available on the US Bureau of Labor Statistics website for this adjustment. (U.S.Bureau of Labor Statistics 2010b) The CPI is a measure of the average change over time in the prices paid by consumers for a market basket of goods and services. (U.S.Bureau of Labor Statistics 2010a)

[x] December 31st, 2007 was chosen as the cut-off date to ensure consistency between literature searches.

Selection of pertinent study

As noted, I reviewed only cost-effectiveness analyses that were a good match to the coverage decision with respect to intervention, comparator, indication, and patient population. The intervention was clearly described in each decision memo. The background section of decision memos includes description of the disease and available therapeutic management options and was used to evaluate the relevance of the comparator included in the cost-effectiveness study. For some interventions, such as some screening and diagnostic tests, there was no obvious comparator, and so in these cases it was assumed that the appropriate study comparator was no screening/diagnostic test. For others, however, I had to make a decision regarding the appropriate comparator. An example is the NCD for aprepitant for the treatment of chemotherapy-induced nausea. (CMS 2005c) The identified cost-effectiveness analysis compared three treatment strategies; conventional treatment (treatment with a 5-HT3 antagonist and a corticosteroid), conventional treatment plus aprepitant, and conventional treatment with aprepitant added after the onset of chemotherapy-induced nausea and vomiting. (Moore et al. 2007) Review of the decision memo made clear that the CMS considered conventional treatment with a 5-HT3 antagonist and a corticosteroid as the relevant comparator. (CMS 2005c)

I expected that in some cases I would identify multiple relevant estimates of cost-effectiveness. I considered various approaches if this were the case: take an average of the identified estimates of cost-effectiveness; take the median estimate of cost-effectiveness; or, combine the results of studies using meta-analysis. Ultimately, these approaches were not required. In the majority of cases, a single relevant estimate of cost-effectiveness was identified. When multiple estimates of cost-effectiveness were available identified, the best estimate could be selected using the grading system described below.

Study grading

I graded each cost-effectiveness study using an ordered rating scale designed for this research. The scale ranges from "A" to "E," with "A" deemed the most relevant to CMS (Table 12). The objective was to find the study most relevant to the coverage decision. Therefore, I assigned estimates made as part of the NCD or originated from a cost-effectiveness study discussed or

cited in the decision memo an ‘A’ grade. I assigned a cost-effectiveness study submitted to CMS as part of the submission process but not made reference to in the decision memo a ‘B’ grade. I assigned relevant studies set in the US and not reviewed by CMS a ‘C’ grade. Within this grading, studies funded from sources other than the manufacturer were given precedence. In the absence of studies set in the US, studies set in other health systems were considered. I assigned studies set in health systems other than the US and were conducted by a regulatory agency in that country a ‘D’ grade. Finally, I assigned studies set in health systems other than the US and not conducted by regulatory agencies an ‘E’ grade.

When multiple estimates of cost-effectiveness were available, I gave priority to the cost-effectiveness analysis with the best grading. I used the recency of the study to the coverage decision to choose between studies tied on rank.

Table 12. Grading of cost-effectiveness analyses

Ranking	Description
A	Estimate of cost-effectiveness made as part of the NCD. Includes discussion of the cost-effectiveness of the intervention, reference to a cost-effectiveness analysis, or when a cost-effectiveness analysis was commissioned as part of a TA or MEDCAC.
B	Cost-effectiveness analysis submitted to the CMS as part of the submission process.
C	Cost-effectiveness analysis set in the US health care system that the decision memo does not reference.
D	Cost-effectiveness estimate made by a regulatory body in another country.
E	Relevant cost-effectiveness analysis not set in the US health care system.

4.7.4.10 An example of the literature search for cost-effectiveness studies

I searched for a cost-effectiveness study relevant to each coverage decision (n=255). To illustrate the approach taken to identify relevant estimates of cost-effectiveness, the results of the PubMed search strategy used for the NCD for deep brain stimulation (DBS) to treat essential tremor and Parkinson's disease are presented in Table 13.

Table 13. Search strategy and literature search results for the NCD for deep brain stimulation

Search	Search terms	Results
#1	Deep brain stimulation OR DBS OR Globus pallidus deep brain stimulation OR Subthalamic deep brain stimulation OR Thalamic deep brain stimulation	3,416
#2	Economic evaluation OR Cost-effectiveness OR Cost-utility OR Cost-minimisation OR Decision analysis OR Decision model OR Decision analytic model	146,033
#3	#1 AND #2	42
Limits	English, Publication Date through 2007/12/31	

I reviewed titles and abstracts for the 42 results of search #3, and judged whether to include them for further review. I obtained and reviewed three full text studies and abstracted data from them. I performed a search of the additional economic databases, i.e., CEA Registry, NHS EED and HEED, to ensure that all pertinent studies were identified through the PubMed search. This search yielded one additional study. A search of the secondary databases did not yield any additional studies. In total, I reviewed four studies, and data was extracted from them using a data extraction table (data extraction tables are presented in Appendix 3). Of the four studies Tomaszewski and Holloway (2001) was assigned the highest grade and was included in the database.

4.7.4.11 Overview of the cost-effectiveness variable

Using the search strategy described above, I often identified multiple potentially relevant abstracts. As noted above, I reviewed each abstract and the full text article was obtained for those deemed a satisfactory match with the coverage decision in terms of intervention, comparator, and patient population. Of the 103 NCDs, 43 included at least one coverage decision for which an

appropriate cost-effectiveness analysis was identified. In total, I reviewed 87 cost-effectiveness studies pertaining to 64 coverage decisions. Of the 64 cost-effectiveness studies included, 48 pertained to positive coverage decisions and 16 to non-coverage decisions. Of the 191 coverage decisions without an associated estimate of cost-effectiveness, 75 were positive coverage decisions and 116 were non-coverage decisions. An overview of the generation of *Cost-effectiveness* is presented in Figure 15.

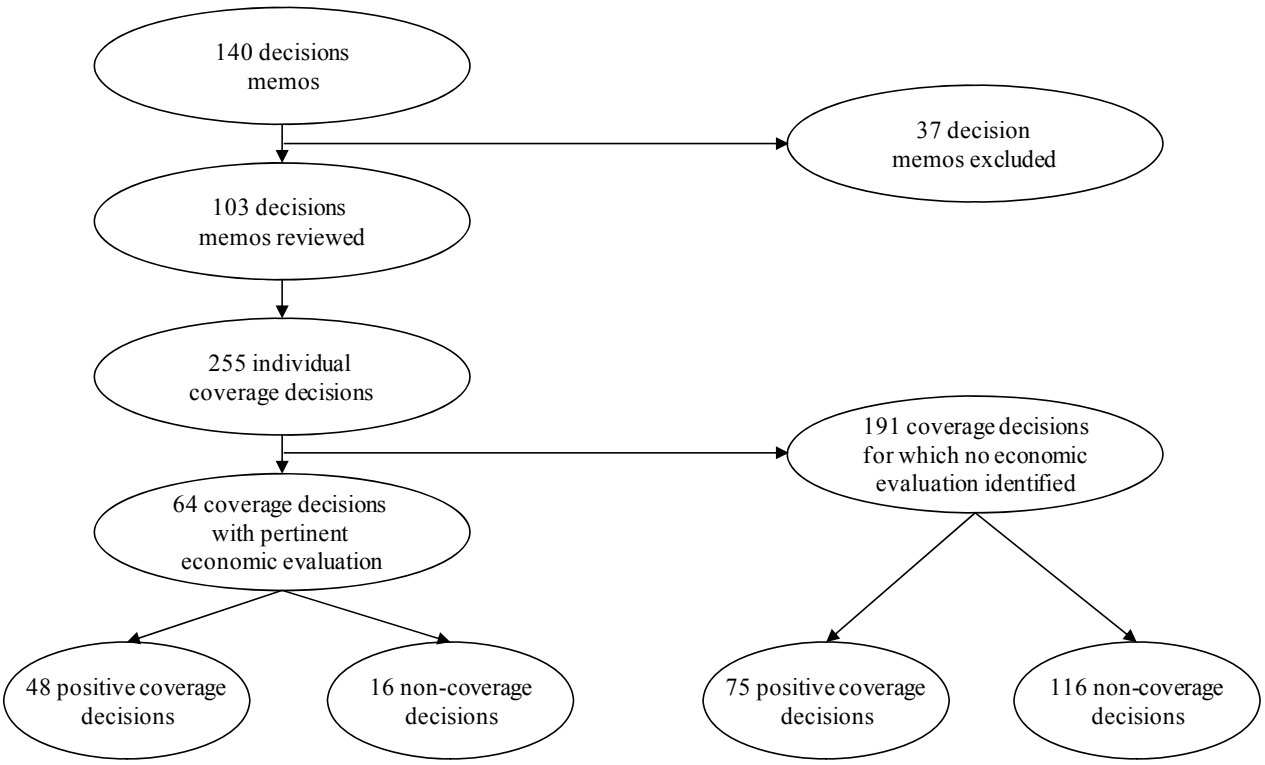


Figure 15. Overview of the generation of *Cost-effectiveness*

For the research presented in Chapter 6, it was necessary to code *Cost-effectiveness* as a categorical variable. While for some studies reviewed above (Section 4.5.1) the researchers included cost-effectiveness evidence in the model as a continuous variable, this meant that dominant and dominated estimates of cost-effectiveness were either dropped from the analysis, e.g., Devlin et al. (2010), or, in the case of dominant estimates, coded as zero, e.g., Dakin et al. (2006). For the research in Chapter 6, I coded the cost-effectiveness data categorically, as it

allowed for the more straightforward interpretation of odds ratios and facilitated the inclusion of coverage decisions associated with dominant or dominated estimates of cost-effectiveness.

4.7.4.12 Additional variables not included in database

I attempted to develop a number of variables that were ultimately were not included in the database. These variables, and reasons for their ultimate exclusion, are discussed below.

Nature of clinical benefit

Interventions subject to NCDs vary with respect to the nature of benefit they provide. Some interventions primarily affect patient survival, others patient quality of life. An attempt was made to capture the nature of benefit that interventions provide in the database. *Nature of benefit* was a categorical variable that included the following categories: *Direct effect on survival; Indirect effect on survival*; *Uncertain effect on survival*; *Quality of life increasing*; and *Not applicable*.

Development of *Nature of benefit* was challenging. Rarely was an intervention solely survival or quality of life increasing, and categorising interventions was complex. Further, for diagnostic and screening tests any positive effect on health is indirect, attributable to treatment as a result of diagnosis of disease. To try to account for this, I included the category *Indirect effect on survival.*

Despite best efforts to include this variable, I deemed that its imprecise nature would prevent meaningful interpretation and I ultimately excluded it from the final database. It is likely that successful inclusion of this variable would require additional input from clinicians or other health care professionals.

Prevalence

NCDs are performed for interventions deemed to have a significant impact on the Medicare programme. Therefore, I attempted to include a variable to capture the potential impact of the

intervention on the Medicare programme. One approach taken was to determine the number of Medicare beneficiaries potentially affected by the coverage decision. *Prevalence* was coded as a continuous variable.

On rare occasions, the prevalence of disease was reported in the decision memo. However, the reported statistics were typically not specific to the precise patient population affected by the coverage decision. For example, in the decision memo for cochlear implantation it states, *"... more than 25 million Americans have hearing loss, including one out of four people older than 65"*. (CMS 2005d) However, the positive coverage decision was made for a population with a number of restrictions, and the reported prevalence statistics do not correlate with the number eligible to receive the intervention. In the vast majority of decision memos, prevalence of disease is not reported, and I performed a search of the medical literature in an attempt to identify estimates. However, in most cases it was not possible to identify an estimate of the number of Medicare beneficiaries affected by the coverage decision. Given the frequency with which estimates could not be identified, *Prevalence* was ultimately not included in the final database.

I used a version of this variable for the research described in Chapter 7. This version relied upon the reporting of ICD-9 codes to estimate the number of eligible beneficiaries in a Medicare claims database. Unfortunately, this information was insufficient to include *Prevalence* here. I found that ICD-9 codes often lacked the precision to identify beneficiaries who met the NCD specifications. Given the imprecise nature of this approach, I ultimately excluded *Prevalence* from the database. I present further details on a version of this variable used for the research presented in Chapter 7 in Section 4.7.5.5.

Budget impact

I used a second approach to attempt to capture the potential impact of the intervention on the Medicare programme by considering the potential budget impact of the coverage decision. *Budget impact* was coded as a continuous variable.

On rare occasions, the potential budget impact of an intervention was reported in the decision memo. However, the reported estimates were typically imprecise. For example, in the decision memo for smoking and tobacco use cessation counseling it states, "*In 1993, smoking cost the Medicare program about $14.2 billion, or approximately 10 percent of Medicare's total budget*." In the vast majority of decision memos, an estimated budget impact is not reported, thus I performed search of the medical literature in an attempt to identify estimates. However, as was the case with *Prevalence*, in most cases it was not possible to identify a relevant estimate of the budget impact associated with the coverage decision. Given the frequency that a relevant estimate of cost-effectiveness could not be identified, *Budget impact* was ultimately not included in the final database.

I used a version of *Budget impact* for the research described in Chapter 7 (4.7.5.2). As for *Prevalence*, I used a Medicare claims database to estimate the number of Medicare beneficiaries receiving an intervention and those eligible for it.

First line technology

To account for instances when CMS recommended the intervention as the standard of care, I attempted to include *First line technology* in the database. I reviewed each decision memo and made a judgement with respect to how CMS recommended the intervention should be incorporated into clinical practice. I coded *First line technology* as a dichotomous categorical variable including the following categories*; First line technology* and *Not first line technology*. *First line technology* differed from *Alternative intervention* as rather than the absolute availability of alternatives, I accounted for how the intervention is prioritised in therapeutic management.

Unfortunately, information provided in decision memos was insufficient for this purpose. While in some cases it was clear that the intervention was not to be used as the first line treatment, i.e., when restrictions on coverage were that the intervention was only permitted after the failure of other interventions, in most cases such information was not presented. Ultimately, despite best

efforts to include this variable, I deemed that its imprecise nature would prevent meaningful interpretation, and it was ultimately excluded from the final database.

4.7.5. Additional variables required for empirical work presented in Chapter 7

The database described above was insufficient for the research presented in Chapter 7. For this research I developed a second smaller database including only coverage decisions with an associated estimate of cost-effectiveness that included the data necessary to perform a hypothetical reallocation of expenditures between interventions subject to NCDs.

For each intervention, the following information was required.

1. Estimate of cost-effectiveness, including incremental cost and incremental effectiveness data;
2. Estimate of intervention and comparator cost in the year following implementation;
3. Estimate of existing utilisation rate (served population), i.e., the utilisation of the intervention within the Medicare population in 2007;
4. Estimate of the size of the total patient population eligible for the intervention in 2007 (to facilitate estimation of the unserved patient population).

4.7.5.1 Cost-effectiveness data

Only interventions with an associated estimate of cost-effectiveness were included. The methods used to identify and evaluate the cost-effectiveness evidence are described in Section 4.7.4.7. In summary, I performed a literature search to identify relevant cost-effectiveness studies for each coverage decision featuring in CMS NCDs from 1999 through 2007. To be included, the cost-effectiveness analysis had to include an intervention matching the coverage decision and also a realistic and relevant comparison intervention. I assessed each cost-effectiveness study using a number of criteria and when multiple studies were available I included the most relevant in the database (Section 4.7.4.9). When necessary, I converted cost-effectiveness estimates to US dollars using the PPP and adjusted them to a 2007 valuation using the health component of the CPI (Section 4.7.4.9). I considered studies reporting cost-per QALY and cost-per life year gained

ratios. Also considered were cost-effectiveness studies reporting the cost-per disease-specific unit when the intervention was estimated to be dominant. Interventions estimated to be dominated, all of which pertained to non-coverage decisions, were excluded from the dataset. I included cost-effectiveness studies irrespective of publication date.

4.7.5.2 Incremental costs

To estimate the consequence of reallocating expenditures between competing interventions in terms of programme costs, incremental cost data was required. I extracted incremental cost data from the cost-effectiveness study when possible. This was possible for the majority of cost-effectiveness studies set in the US. When incremental cost was reported in US dollars other than a 2007 valuation I adjusted the estimate using the health component of the CPI. When the cost-effectiveness study was set in a country other than the US, I performed a search of the medical literature to identify costing studies set in the US that provided the necessary data. Primary sources included the PubMed database and NHS EED. (Centre for Reviews and Dissemination 2011;U.S.National Library of Medicine 2011) If a relevant US-based costing study could not be identified, I converted the incremental cost reported in the cost-effectiveness study into US dollars using the PPP and, when necessary, adjusted it to a 2007 valuation using the health component of the CPI (Section 4.7.4.9).

4.7.5.3 Cost of an intervention in the first year of its use

In addition to incremental costs, estimates of the cost of competing interventions in the year following their first use were required. On occasion, the necessary information was presented in the included cost-effectiveness study and was incorporated accordingly. When not directly reported, it was on occasion possible to calculate the cost from data reported in the cost-effectiveness study, e.g., when cost data was presented graphically. When this was not possible, I estimated the cost of the competing treatments in the year following first use from Medicare and physician reimbursement codes (Section 4.7.5.6). Generally, it was possible to identify the pertinent reimbursement codes from official Medicare documentation, the included cost-effectiveness study, or the manufacturer's website. For interventions subject to non-coverage

decisions, it was possible to either obtain the relevant information from the cost-effectiveness study or from a costing study performed in the US setting.

4.7.5.4 Incremental effectiveness data

Incremental QALY data were necessary to estimate aggregate population health gain/loss from the hypothetical reallocation of expenditures. Quality adjusted life years are an ideal metric for this research as they facilitate a comparison of health gain across interventions and disease areas. The majority of cost-effectiveness studies included in this research included QALYs as the outcome measure, and I extracted the incremental QALY gain accordingly. Consistent with the research presented in Chapters 5 and 6, cost-effectiveness studies reporting cost-per life year gained ratios were considered in the absence of cost-per QALY studies. To maintain consistency with cost-per QALY studies, estimates of incremental life year gains were adjusted with a utility weight, 0.76, for Americans aged 65-69 as reported by Erickson P et al. (1995). (Erickson, Wilson, & Shannon 1995) This adjustment may underestimate incremental QALY gain as it only accounts for years of survival gain, not prior years of treatment when quality of life may have differed between intervention arms. However, I used this approach to facilitate the measurement of aggregate health gains using a single metric.

Cost-effectiveness studies that reported health gain using disease specific units were considered for interventions estimated to be dominant. Although these studies do not provide estimates of incremental QALY gain, they do provide incremental cost data. Separate analyses were performed for the inclusion and exclusion of studies that report health gain using disease-specific units.

4.7.5.5 Utilisation rates

Interventions associated with a positive coverage decision

I estimated utilisation rates from a Medicare claims database obtained for a broader research project performed by researchers at Tufts Medical Center.[xi] The database included both inpatient and outpatient data from a 5% sample of the Medicare population. ICD-9 diagnostic codes, the standard format used to identify illness, injury, or disease, were included in this database. Upon receiving care, Medicare beneficiaries are assigned an ICD-9 diagnostic code that best categorises their ailment. Medicare beneficiaries eligible for an intervention, as defined by the parameters of the NCD, were identified through the reported ICD-9 diagnostic codes. Also included in the 5% Medicare claims database are ICD-9 procedural codes, used for documenting and recording performed medical procedures, and Common Procedural Terminology (CPT) codes, used for physician reimbursement for services performed while working for Medicare, Medicaid, and a majority of private health care payers.

I obtained utilisation rates by estimating the number of beneficiaries that had matching relevant ICD-9 diagnostic codes and CPT or ICD-9 procedural codes. As a 5% Medicare claims database was used, I adjusted the identified utilisation rate by a factor of 20 to estimate of the total number of beneficiaries receiving the intervention.

Separate data files were used for the inpatient and outpatient data, and it was not possible to ensure that beneficiaries did not feature in both datasets. Consequently, it is possible that a beneficiary who received the same therapy on both an inpatient and an outpatient basis in the same year would be counted twice. This double counting is, however, unlikely to be a significant problem. In the majority of cases, interventions are reimbursed exclusively on either an inpatient (e.g., bariatric surgery, transmyocardial revascularisation, ventricular assist devices, etc) or an outpatient basis (e.g., cardiac rehabilitation, PET for various oncology indications, foot care, etc), and the possibility of double counting is minimal.

[xi] See Acknowledgements.

ICD-9 diagnostic codes were used to estimate the size of the eligible patient population. Important in this research was identification of the size of the unserved eligible population, i.e., beneficiaries eligible for the intervention as defined by the NCD parameters but who did not receive the intervention. I calculated the size of the unserved population as the difference between the number of beneficiaries who were a match for both ICD-9 diagnostic codes and CPT/ICD-9 procedural codes, and those who were match solely when considering ICD-9 diagnostic codes (Table 14).

Table 14. Identification of served and unserved populations from Medicare claims

	ICD – 9 Diagnostic codes	
Reimbursement code/ICD procedural code	Match with NCD	Not a match with NCD
Match with NCD	Currently served population	NA
Not a match with NCD	Unserved population	NA

Non-covered interventions

Interventions subject to non-coverage decisions are unavailable to Medicare beneficiaries for the indication defined by the NCD parameters. For interventions subject to non-coverage decisions, the utilisation rate was assumed zero in all cases. In order to include non-covered interventions in the reallocation, it was necessary to estimate the size of the potential eligible patient population if the intervention was offered as a Medicare benefit. As for interventions subject to positive coverage decisions, the size of the eligible patient population was estimated through ICD-9 diagnostic codes. For each non-coverage decision, I reviewed the decision memo to determine the patient population for which the non-coverage decision pertained. This patient population was then categorised using ICD-9 diagnostic codes using an online database. (Centers for Disease Control and Prevention 2011) In turn, I used these diagnostic codes to estimate the size of the potentially eligible patient population through the Medicare 5% claims database.

4.7.5.6 Reimbursement codes

As described above, I used reimbursement codes to estimate current utilisation rate, i.e., the served population, and the cost of the intervention in the first year of its use. The price year was 2007. I describe the sources of these reimbursement codes below.

Physician reimbursement

Common Procedural Terminology codes are developed, published, and licensed by the American Medical Association (AMA). The reimbursement rate associated with each CPT code is available on CMS's website. (CMS 2010c) As reimbursement rate varies by geographic region, the 'neutral' code (Carrier/ Locality code 0000000) was used to maintain consistency in each instance.

Outpatient services

Ambulatory Payment Classifications (APCs) are reimbursement codes used for outpatient services within the Medicare programme. Ambulatory Payment Classifications are a prospective payment system for hospitals and cover all services performed in an outpatient setting, with the exception of physician reimbursement. (MedPAC 2007) This information was obtained for the most part through coding information available on the manufacturer's website. When unavailable through the manufacturer's website, or indeed when the intervention was not associated with a medical technology, APCs were available through published Medicare documentation.

Inpatient services

The Medicare Inpatient Prospective Payment System (IPPS) is a prospective payment system used to reimburse hospitals for inpatient care. This system categorises patients into specific diagnostic categories (MS-DRG) and provides reimbursement accordingly. (MedPAC 2008) Similar to APC codes, this information is most readily accessible through coding information available from manufacturers' websites or through published Medicare documentation.

4.7.5.7 Categorisation of interventions

To facilitate evaluation of the effect of reallocation on the distribution of expenditures, I categorised interventions with respect to disease classification, type, and the size of the associated untreated population. The approach to classifying intervention type is presented in Section 4.7.4.4. Classifications pertaining to disease and size of the associated untreated population are presented below.

Disease classification

- Interventions were categorised as those related to; cardiology, oncology, other (i.e., those unrelated cardiology or oncology).

Size of the associated untreated population

- Interventions were classified as those for which the size of the untreated patient population was large (>1 million beneficiaries), medium (50,000 – 1 million beneficiaries), or small (<50,000 beneficiaries).

4.7.5.8 Software

I used SAS to identify utilisation rates from the Medicare 5% database and Microsoft Excel to perform the hypothetical reallocation of resources.

4.8. Chapter summary

This chapter lays the foundation for the empirical work presented in Chapters 5 through 7. The Medicare programme is described, highlighting its size and importance in the US health care system. The process for coverage of interventions in the Medicare programme is explained, along with the circumstances in which interventions are evaluated through NCDs.

The three pieces of research that constitute the empirical aspect of this thesis are described above, with the objective(s) for each stated. A literature search was performed to identify studies that would help inform the methodological approach for each piece of empirical work. First, a search was performed to identify studies that evaluated the role of cost-effectiveness evidence in coverage and reimbursement decisions, or in recommendations for the efficient use of medical technology. A body of literature was identified and reviewed that evaluated how cost-effectiveness has been used in decision-making by NICE in the UK, the PBAC in Australia, CEDAC in Canada, PHARMAC in New Zealand, and an HMO in the US. Second, a literature search to identify studies that estimated efficiency gains, in terms of aggregate health gains and cost savings, through the hypothetical reallocation of resources was performed. Two studies that estimated hypothetical gains from alternative approaches to resource allocation were reviewed.

A database was developed including the variables required for the empirical work. Variables in the database included cost-effectiveness, quality of supporting clinical evidence, availability of alternative interventions, date of decision, coverage requestor, and type of intervention. The cost-effectiveness variable was primarily generated through literature searches, although on occasion a relevant cost-effectiveness ratio originated from the decision memo. The variable classifying the quality of the supporting clinical evidence was generated through review of the decision memo by two researchers at Tufts Medical Center using the USPSTF guidelines for grading evidence. The remaining variables were generated from review of the decision memos.

Additional variables were required for the third piece of empirical work, including incremental cost and incremental QALY gain data, the cost of the intervention in the first year of its use, the existing utilisation rate, and size of the eligible patient population. For the most part, the incremental cost and incremental effectiveness data originated from the included cost-effectiveness studies. The cost of the intervention and the comparator was captured from a number of sources, including the cost-effectiveness literature, costing studies, or from Medicare reimbursement codes. Utilisation rate and the size of the eligible patient population were estimated from Medicare claims data.

In the next chapter, I describe the first piece of empirical work. The first objective is to evaluate if NCDs are consistent with CMS's stated position on the use of cost-effectiveness evidence, i.e., that cost-effectiveness is not a factor that is considered in NCDs. The second objective is to evaluate the cost-effectiveness of positive coverage and non-coverage decisions and to determine if there is a statistically significant difference between the two groups.

5. Empirical Research: Part 1

5.1. Introduction

In this chapter, I present the first piece of empirical work. In Chapter 2, I described the theory underpinning cost-effectiveness analysis. In Chapter 3, I illustrated how cost-effectiveness evidence is used to inform decision-making in a range of jurisdictions. In some cases, cost-effectiveness analysis has a fundamental role in coverage and reimbursement decisions, or in recommendations for the efficient use of medical technology; in others it plays a lesser role. In particular, despite maintaining levels of health care spending grossly in excess of other countries and faring noticeably worse with respect to health outcomes, cost-effectiveness analysis has not been embraced by the US health care system. This is particularly noticeable when examining the Centers for Medicare and Medicaid Services (CMS), an agency whose stated position is that cost-effectiveness is not a factor considered in national coverage determinations (NCDs).

Chapter 4 provides the foundation for the empirical aspect of this thesis. A background to the Medicare programme and the processes for the coverage of medical technology were summarised. The objectives of the empirical component of this thesis were presented and key studies identified from a literature search that helped inform the methodological approach reviewed. The development of the database necessary for the empirical work is described.

The purpose of this chapter is to examine CMS's use, or lack of use, of cost-effectiveness evidence, and to evaluate the cost-effectiveness of coverage decisions made as part of NCDs. To the best of my knowledge, although the cost-effectiveness of CMS NCDs has been commented upon in the literature, a systematic assessment has not been performed. (Neumann, Rosen, & Weinstein 2005) The research objectives are presented below.

5.2. Objectives and Methods

The empirical work presented in this chapter has two specific objectives:

1. To examine NCD decision memos to determine if they are consistent with CMS's stated position on the use of cost-effectiveness evidence, i.e., that cost-effectiveness is not a factor considered in making NCDs.
2. To determine if there is a difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness.

5.2.1. Objective 1

For the first research objective, I scrutinised the decision memos accompanying NCDs made from 1999 through 2007 (n=140) for any discussion or citation of cost-effectiveness evidence. Also, I reviewed documentation accompanying the NCD, which included that pertaining to associated MEDCAC meetings or external technology assessments. I recorded each instance that cost-effectiveness evidence featured in a decision memo or the accompanying documentation (Section 4.7.4.8).

5.2.2. Objective 2

For the second research objective, the methodological approach was informed by the review of the methods used in the research described in Section 4.5.1. Three studies, George et al. (2001), Towse and Pritchard (2002), and Rocchi et al. (2008), evaluated coverage decisions solely from a cost-effectiveness perspective. An important distinction between the studies and the research described in this chapter is that they evaluated agencies in which cost-effectiveness evidence plays an established role in decision-making. Indeed, rather than evaluating whether there is a difference between positive and non-coverage decisions with respect to cost-effectiveness, the objective of the studies was to identify a threshold value of cost-effectiveness above which a technology is less likely to be covered. Nevertheless, the methodological approaches of the reviewed studies provided a framework with which to perform this research. In particular, George et al. (2001) used a similar methodology to that employed here. Specifically, the study

similarly used tabular presentation of the cost-effectiveness studies as well as a statistical test to evaluate if there was a statistically significant difference between positive and non-coverage decisions with respect to cost-effectiveness.

For coverage decisions made in NCDs from 1999 through 2007, I attempted to identify estimates of cost-effectiveness. As described above, I reviewed each decision memo for discussion or citation of cost-effectiveness evidence. In addition, I performed a literature search for each coverage decision to identify estimates of cost-effectiveness available in the peer-reviewed literature. I considered only cost-effectiveness studies that were a good match with respect to intervention, comparator, indication, and patient population. When I identified multiple relevant cost-effectiveness estimates, the estimate deemed most relevant to the coverage decision was included. A more detailed description of the process used to generate the data used in this research is described in Section 4.7.4.9.

Consistent with the methodological approach taken by George et al. (2001), I performed a Mann Whitney U test to determine if there was a statistically significant difference between the cost-effectiveness of positive coverage decisions and non-coverage decisions. The Mann Whitney U test was deemed the most appropriate statistical test for this analysis. The data take the form of two independent random samples taken from two populations, i.e., positive and non-coverage decisions, and have the necessary characteristics for the Mann Whitney U test, i.e., they are unpaired, categorical[xii] and ordinal. To perform the test, the two samples were pooled and the observations ranked in order of their cost-effectiveness, with ties assigned the average of the next available ranks. The Mann Whitney U statistic (U) can be presented algebraically as follows:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

n_1 = Number of observations available from the first population (positive coverage decisions)

n_2 = Number of observations available from the second population (non-coverage decisions)

[xii] I considered the data to be categorical as the inclusion of observations that were dominant and dominated prevented me from using a numerical variable.

R_1 = Sum of ranks of the observations from the first population

Assuming the null hypothesis that the central locations of the two population distributions are the same, the Mann Whitney U, has mean and variance as presented below:

$$E(U) = \mu_U = \frac{n_1 n_2}{2}$$

$$Var(U) = \sigma^2{}_U = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Further, given the available sample size[xiii] (total coverage decisions, n=64; positive coverage, n=48; non-coverage, n=16), it is possible to assume that the distribution of the random variable is approximated by the normal distribution:

$$Z = \frac{U - \mu_U}{\sigma_U}$$

The decision rule for the Mann Whitney U test is to reject H_0 if:

$$Z = \frac{U - \mu_U}{\sigma_U} < -z_{\alpha/2} \quad \text{or} \quad Z = \frac{U - \mu_U}{\sigma_U} > z_{\alpha/2}$$

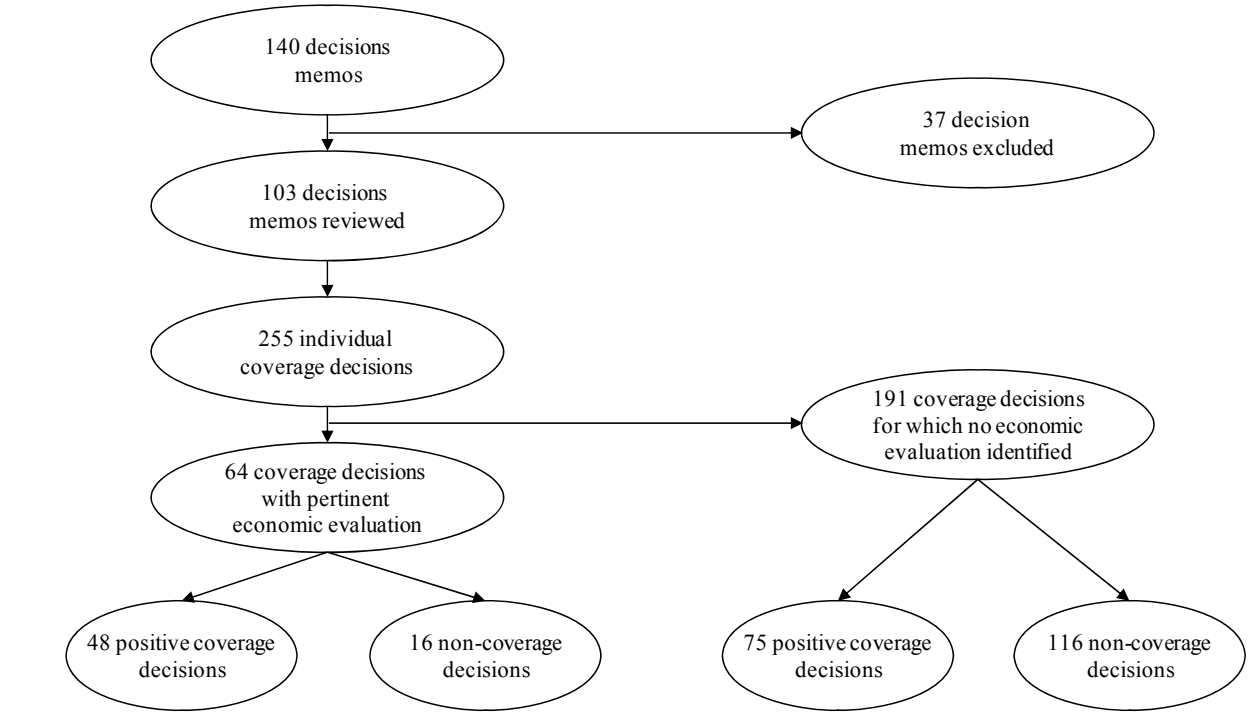
[xiii] Approximation to a normal distribution is appropriate if each sample contains at least 10 observations, i.e., $n_1 \geq 10$ and $n_2 \geq 10$ (Newbold, Carlson, & Thorne 2003)

5.3. Results

In this section, I present the findings of the literature search, illustrating how I arrived at the final sample of cost-effectiveness studies. Also, I describe the distribution of cost-effectiveness studies with respect to study grading (Section 4.7.4.9), and country of study. Next, I present the findings of the research pertinent to the first research objective, i.e., to examine NCD decision memos to determine if the reviewed evidence is consistent with CMS's stated position on the use of cost-effectiveness evidence. Finally, I present the findings of the research pertinent to the second research objective, i.e., to determine if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness.

5.3.1. Literature search findings

An overview of the review process and the final sample of coverage decisions and corresponding cost-effectiveness studies are presented in Figure 16. Thirty-seven of the 140 NCDs made from 1999 through 2007 were excluded based upon the exclusion criteria (Section 4.7.2). From the remaining 103 NCDs, 255 coverage decisions were identified. A relevant estimate of cost-effectiveness was identified for 64 coverage decisions, 48 positive coverage decisions (Table 15), and 16 non-coverage decisions (Table 16). Of the 191 coverage decisions for which no relevant estimate of cost-effectiveness was identified, 75 were positive coverage decisions and 116 non-coverage decisions.



Cost-Effectiveness	Frequency
Dominates	19
<$50k per QALY	12
$50k-$100k per QALY	8
>100k per QALY	9
Dominated	0

Cost-Effectiveness	Frequency
Dominates	3
<$50k per QALY	6
$50k-$100k per QALY	0
>100k per QALY	1
Dominated	6

Figure 16. Review process and final sample of coverage decisions with associated estimate of cost-effectiveness

N.B. QALY = Quality Adjusted Life Year; Life years = Life years gained

Table 15. Positive coverage decisions associated with an estimate of cost-effectiveness

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER (US$)	Outcome measure*	Study country	Reference:
1	Cryosurgery Ablation for Prostate - Primary treatment for clinically localised prostate cancer. (Stages T1-T3)	1999	C	Dominant	Other	USA	Benoit RM et al. (1998)
2	Breast Biopsy - Stereotactic core needle image guidance	1999	C	Dominant	Other	USA	Lee et al. (1997)
3	Breast Biopsy - Ultrasound image guidance	1999	C	Dominant	Other	USA	Liberman L et al. (1998)
4	Diabetic Peripheral Neuropathy with Loss of Protective Sensation - Coverage for diabetic patients who meet specified conditions	2001	E	Dominant	QALY	Sweden	Ragnarson Tennvall G et al. (2001)
5	Positron Emission Tomography - Lung Cancer (non-small cell)	2000	C	Dominant	Other	USA	Valk PE et al. (1996)
6	Positron Emission Tomography - Colorectal Cancer	2000	C	Dominant	Other	USA	Valk PE et al. (1996)
7	Positron Emission Tomography - Melanoma	2000	C	Dominant	Other	USA	Valk PE et al. (1996)
8	Ambulatory blood pressure monitoring - Use in patients with high blood pressure who meet specified criteria	2001	E	Dominant	Other	UK	Aitken (1996)
9	Prothrombin Time (INR) Monitor for Home Anticoagulation Management - Patients with mechanical heart valves that meet specific criteria	2001	E	Dominant	Other	Germany	Völler H et al. (2001)
10	Cardiac rehabilitation programs - Acute Myocardial Infarction	2006	C	Dominant	QALY	USA	Yu C et al. (2004)
11	Cardiac rehabilitation programs - Percutaneous	2006	C	Dominant	QALY	USA	Yu C et al.

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER (US$)	Outcome measure*	Study country	Reference:
	Transluminal Coronary Angioplasty						(2004)
12	Positron Emission Tomography (FDG) for Breast Cancer - Detection of Locoregional Recurrence or Distant Metastasis/Recurrence (Staging and Restaging)	2002	E	Dominant	Other	Canada	Sloka JS et al. (2005)
13	Positron Emission Tomography (FDG) for Myocardial Viability - PET as a primary or initial diagnostic study	2002	E	Dominant	Other	Australia	Miles KA (2001)
14	Intravenous Immune Globulin for Autoimmune Mucocutaneous Blistering Diseases - Pemphigus Vulgaris	2002	C	Dominant	Other	USA	Daoud YJ (2006)
15	Intravenous Immune Globulin for Autoimmune Mucocutaneous Blistering Diseases - Bullous Pemphigoid	2002	C	Dominant	Other	USA	Daoud YJ (2006)
16	Intravenous Immune Globulin for Autoimmune Mucocutaneous Blistering Diseases - Mucous Membrane Pemphigoid	2002	C	Dominant	Other	USA	Daoud YJ (2006)
17	Magnetic Resonance Angiography of the Abdomen and Pelvis - Imaging the renal arteries and the aortoiliac arteries when using MRA is expected to avoid obtaining contrast angiography	2003	C	Dominant	Other	USA	Levy MM et al. (1998)
18	Positron Emission Tomography (N-13 Ammonia) for Myocardial Perfusion - Diagnosis of myocardial perfusion	2003	E	Dominant	Other	Switzerland	Siegrist PT et al. (2007)
19	Smoking & Tobacco Use	2005	A	Dominant	Other	USA	CMS Decision Memo (CAG-

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER (US$)	Outcome measure*	Study country	Reference:
	Cessation Counseling						00241N)
20	Screening Immunoassay Fecal-Occult Blood Test (Hemoccult II FOBT)	2003	A	$1,072	Life years	USA	Report to the Agency for Healthcare Research and Quality (2003)
21	Positron Emission Tomography - Head and Neck Cancers	2000	C	$2,395	QALY	USA	Hollenbeak CS et al. (2001)
22	Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA)	2001	C	$3,079	QALY	USA	Ayas NT et al, (2006)
23	Hyperbaric Oxygen Therapy - Diabetic Wounds of the Lower Extremities that fit specified criteria	2002	C	$5,409	QALY	USA	Guo S et al. (2003)
24	Cochlear implantation - Post lingually hearing impaired patients	2005	C	$10,729	QALY	USA	Francis HW et al. (2002)
25	Cochlear implantation - Pre lingually hearing-impaired patients	2005	C	$10,953	QALY	USA	Francis HW et al. (2002)
26	Bariatric Surgery for the Treatment of Morbid Obesity - Open Roux-en-Y gastric bypass (RYGBP)	2006	E	$12,733	QALY	UK	Clegg A et al. (2003)
27	Bariatric Surgery for the Treatment of Morbid Obesity - Laparoscopic adjustable gastric banding (LAGB)	2006	E	$17,264	QALY	UK	Clegg A et al. (2003)
28	Erythropoiesis Stimulating Agents (ESAs) for non-renal disease indications - Treatment of chemotherapy induced anaemia for patients who meet specified criteria	2007	E	$18,713	QALY	UK	Martin SC et al. (2003)
29	Screening Immunoassay Fecal-Occult Blood Test	2003	A	$21,001	Life years	USA	Report to the Agency for Healthcare

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER (US$)	Outcome measure*	Study country	Reference:
	(iFOBT)						Research and Quality (2003)
30	Autologous Stem Cell Transplantation (AuSCT) for Multiple Myeloma - Treatment of multiple myeloma for patients who meet certain conditions	2000	C	$27,687	Life years	USA	Trippoli S et al. (1998)
31	Implantable Cardioverter Defibrillators (ICDs) - Documented sustained ventricular tachyarrhythmia	2003	C	$36,396	Life years	USA	Mushlin AI et al. (1998)
32	Deep Brain Stimulation for Parkinson's Disease (PD) - PD patients that meet specified criteria	2003	C	$55,826	QALY	USA	Tomaszewski et al. (2001)
33	Microvolt T-wave Alternans - diagnostic testing for patients at risk of sudden cardiac death when the spectral analytic method is used	2007	C	$55,126	QALY	USA	Chan PS et al. (2006)
34	Positron Emission Tomography - Esophageal Cancer	2000	C	$60,544	QALY	USA	Wallace MB et al. (2002)
35	Implantable Defibrillators 2 - NIDCM, documented prior MI, Class II and III heart failure	2005	C	$70,200	QALY	USA	Sanders G et al. (2005)
36	Implantable Cardioverter Defibrillators (ICDs) - Documented familial or inherited conditions with a high risk of ventricular tachyarrhythmias	2003	C	$84,439	Life years	USA	Larsen G et al, (2002)
37	Pancreas Transplants - Patients that meet the specified criteria (type 1 diabetes etc)	2007	C	$90,159	QALY	USA	Kiberd BA et al. (2000)
38	Ultrasound Stimulation for Nonunion Fracture Healing - Tibial	2005	D	$94,848	QALY	Australia	MSAC application 1030 2001)

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER (US$)	Outcome measure*	Study country	Reference:
39	Aprepitant for Chemotherapy-Induced Emesis - For use following specified chemotherapies	2005	C	$97,429	QALY	USA	Moore S et al. (2007)
40	Liver transplantation in patients suffering from hepatitis B	1999	C	$145,749	QALY	USA	Dan YY et al. (2006)
41	Ocular Photodynamic Therapy with Verteporfin for Macular Degeneration - Predominately classic subfoveal CNV lesions	2004	D	$172,770	QALY	UK	Meads et al. (2002)
42	Lung Volume Reduction Surgery - Severe upper lobe emphysema	2003	C	$175,790	QALY	USA	Ramsey et al. (2003)
43	Transmyocardial revascularisation for Severe Angina - Patients with severe angina (stable or unstable), refractory to standard medical therapy.	1998	E	$337,568	QALY	UK	Campbell HE et al. (2001)
44	Lung Volume Reduction Surgery - Non high risk patients suffering from non-upper lobe emphysema with low exercise capacity	2003	C	$343,259	QALY	USA	Ramsey et al. (2003)
45	Ultrasound Stimulation for Nonunion Fracture Healing - Radius	2005	D	$446,384	QALY	Australia	MSAC application 1030 (2001)
46	Ultrasound Stimulation for Nonunion Fracture Healing - Scaphoid	2005	D	$570,379	QALY	Australia	MSAC application 1030 (2001)
47	Insulin Infusion Pump - Type 1 diabetic patients	1999	D	$511,683	QALY	UK	Colquitt et al. (2004)
48	Ventricular Assist Devices as Destination Therapy - Chronic end-stage heart failure patients that meet specified criteria	2003	C	$834,924	QALY	USA	Samson D (2004)

* QALY = Quality Adjusted Life Year; Life years = Life years gained; Other = Study-specific clinical outcome

Table 16. Non-coverage decisions associated with an estimate of cost-effectiveness

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER	Outcome measure*	Study country	Reference:
1	Positron Emission Tomography (FDG) for Breast Cancer - Initial Staging of Axillary Lymph Nodes	2002	E	Dominant	Other	Australia	Miles KA (2001)
2	Warm-Up Wound Therapy aka Noncontact Normothermic Wound Therapy (NNWT)	2002	C	Dominant	QALY	USA	Macario A et al. (2002)
3	Positron Emission Tomography (FDG) for Brain, Cervical, Ovarian, Pancreatic, Small Cell Lung, and Testicular Cancers - Ovarian cancer	2005	C	Dominant	Other	USA	Smith GT et al. (1999)
4	External Counterpulsation (ECP) Therapy	2006	B	$3,126	QALY	USA	(Varricchione 2006)
5	Electrical Bioimpedance for Cardiac Output Monitoring	2006	B	$6,137	QALY	USA	CMS Decision memo – (CAG-00001R2)
6	Bariatric Surgery for the Treatment of Morbid Obesity - BMI of 50 and no comorbidites	2006	C	$11,524	QALY	USA	Craig BM et al. (2002)
7	Lumbar Artificial Disc Replacement	2007	D	$16,957	QALY	Australia	MSAC application 1090 (2000)
8	Acupuncture - Osteoarthritis	2003	E	$17,249	QALY	Germany	Reinhold et al. (2007)
9	Bariatric Surgery for the Treatment of Morbid Obesity - Stated treatments indicated for obesity alone BMI of 40	2006	C	$31,861	QALY	USA	Craig BM et al. (2002)

No.	Technology - Coverage decision	Year	Ranking of evidence	ICER	Outcome measure*	Study country	Reference:
10	Diabetic Peripheral Neuropathy with Loss of Protective Sensation - Coverage for diabetics without loss of protective sensation)	2001	E	$187,472	QALY	Austria	Rauner MS et al. (2005)
11	Percutaneous Transluminal Angioplasty (PTA) of the Carotid Artery Concurrent with Stenting	2001	C	Dominated	Other	USA	Jordan WD et al. (1998)
12	Lung Volume Reduction Surgery - High-risk patients suffering from severe emphysema	2003	C	Dominated	QALY	USA	Ramsey et al. (2003)
13	Lung Volume Reduction Surgery - Non high risk patients suffering from non-upper lobe emphysema with low exercise capacity	2003	C	Dominated	QALY	USA	Ramsey et al. (2003)
14	Implantable Cardioverter Defibrillators (ICDs) - Acute Myocardial Infarction	2003	C	Dominated	QALY	USA	Sanders G et al. (2005)
15	Implantable Cardioverter Defibrillators (ICDs) - Patients who have undergone a coronary artery bypass graft	2003	C	Dominated	QALY	USA	Sanders G et al. (2005)
16	Positron Emission Tomography (FDG) - For Alzheimer's Disease/Dementia	2003	A	Dominated	QALY	USA	Matchar DB et al. (2001)

* QALY = Quality Adjusted Life Year; Life years = life years gained; Other = Study-specific clinical outcome

5.3.2. Characteristics of cost-effectiveness studies

As described in Section 4.7.4.7, I considered cost-effectiveness studies reporting cost-per QALY ratios or cost-per life year gained ratios. In addition, I considered cost-effectiveness studies incorporating disease-specific units when the intervention was estimated to be dominant or dominated. Of the 64 cost-effectiveness studies, 40 (62.5%) reported cost-effectiveness using cost-per QALY ratios, 19 (29.7%) in cost-per disease-specific unit, and five (7.8%) in cost-per life year gained.

Using the criteria described in Table 12 in Section 4.7.4.9, I graded cost-effectiveness studies with respect to relevance. Four cost-effectiveness studies were classified as grade "A" evidence, two as grade "B," 39 as grade "C," six as grade "D," and 14 as grade "E." The distribution with respect to awarded grade was qualitatively similar across positive and non-coverage decisions (Figure 17).

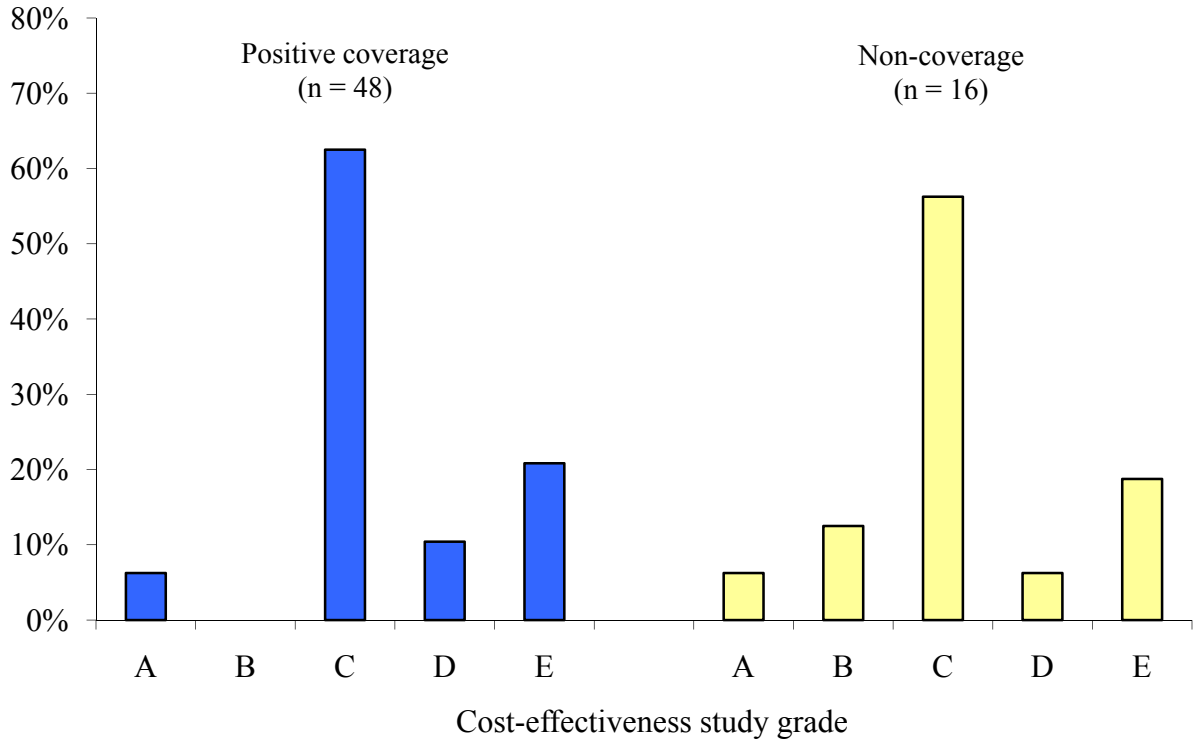


Figure 17. Grading of cost-effectiveness studies

Of studies associated with positive coverage decisions (n=48), 33 (69%) were performed in a US setting; of those associated with non-coverage decisions (n=16), 12 (75%) were performed in a US setting. The distribution with respect to setting was qualitatively similar across positive and non-coverage decisions (Figure 18).

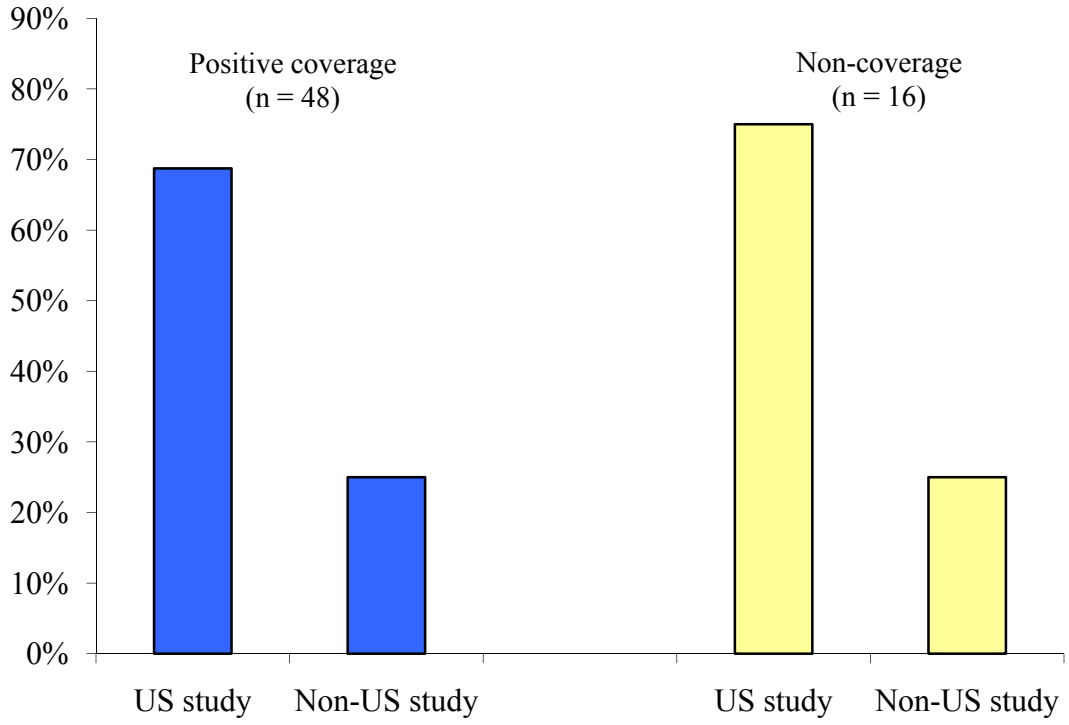


Figure 18. Country of setting of cost-effectiveness studies

5.3.3. Findings – Research Objective 1

The first objective of this research was to examine NCD decision memos to determine if they are consistent with CMS's stated position on the use of cost-effectiveness evidence. For 14 coverage decisions, estimates of cost-effectiveness originated from the decision memo, 12 of which pertained to positive coverage decisions, and two to non-coverage decisions. Of the positive decisions, seven were associated with a cost-effectiveness study that estimated the intervention to be dominant, and five with an ICER less than $30,000 per QALY/life year gained.

Table 17. Decision memos including discussion regarding cost-effectiveness or citation of a cost-effectiveness analysis

No.	Technology - Coverage decision	ICER	Outcome measure*	Reference:
Positive coverage decisions				
1	Cryosurgery Ablation for Prostate - Primary treatment for clinically localised prostate cancer. (Stages T1-T3)	Dominant	Other	Benoit RM et al. (1998)
2	Positron Emission Tomography - Lung Cancer (non-small cell)	Dominant	Other	Valk PE et al. (1996)
3	Positron Emission Tomography - Colorectal Cancer	Dominant	Other	Valk PE et al. (1996)
4	Positron Emission Tomography – Melanoma	Dominant	Other	Valk PE et al. (1996)
5	Cardiac rehabilitation programs - Acute Myocardial Infarction	Dominant	QALY	Yu C et al. (2004)
6	Cardiac rehabilitation programs - Percutaneous Transluminal Coronary Angioplasty	Dominant	QALY	Yu C et al. (2004)
7	Smoking & Tobacco Use Cessation Counseling	Dominant	Other	CMS Decision Memo (CAG-00241N)
8	Screening Immunoassay Fecal-Occult Blood Test	$1,072	Life Years	Report to the Agency for Healthcare Research and Quality (2003)
9	Cochlear implantation - Post lingually hearing impaired patients	$10,729	QALY	Francis HW et al. (2002)
10	Cochlear implantation – Pre lingually hearing-impaired patients	$10,953	QALY	Francis HW et al. (2002)
11	Screening Immunoassay Fecal-Occult Blood Test	$21,001	Life Years	Report to the Agency for Healthcare Research and Quality (2003)
12	Autologous Stem Cell Transplantation (AuSCT) for Multiple Myeloma - Treatment of multiple myeloma for patients who meet certain conditions	$27,687	Life Years	Trippoli S et al. (1998)

No.	Technology - Coverage decision	ICER	Outcome measure*	Reference:
Non-coverage decisions				
1	Positron Emission Tomography (FDG) - Positron Emission Tomography (FDG) for Alzheimer's Disease/Dementia	Dominant	QALY	Matchar DB et al. (2001)
2	Electrical Bioimpedance for Cardiac Output Monitoring	$6,137	QALY	CMS Decision memo (CAG-00001R2)

* QALY = Quality Adjusted Life Year; Life years = Life years gained; Other = Study-specific clinical outcome

5.3.4. Findings – Research Objective 2

The second objective of this research was to determine if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness. An overview of the findings is presented in Figure 19.

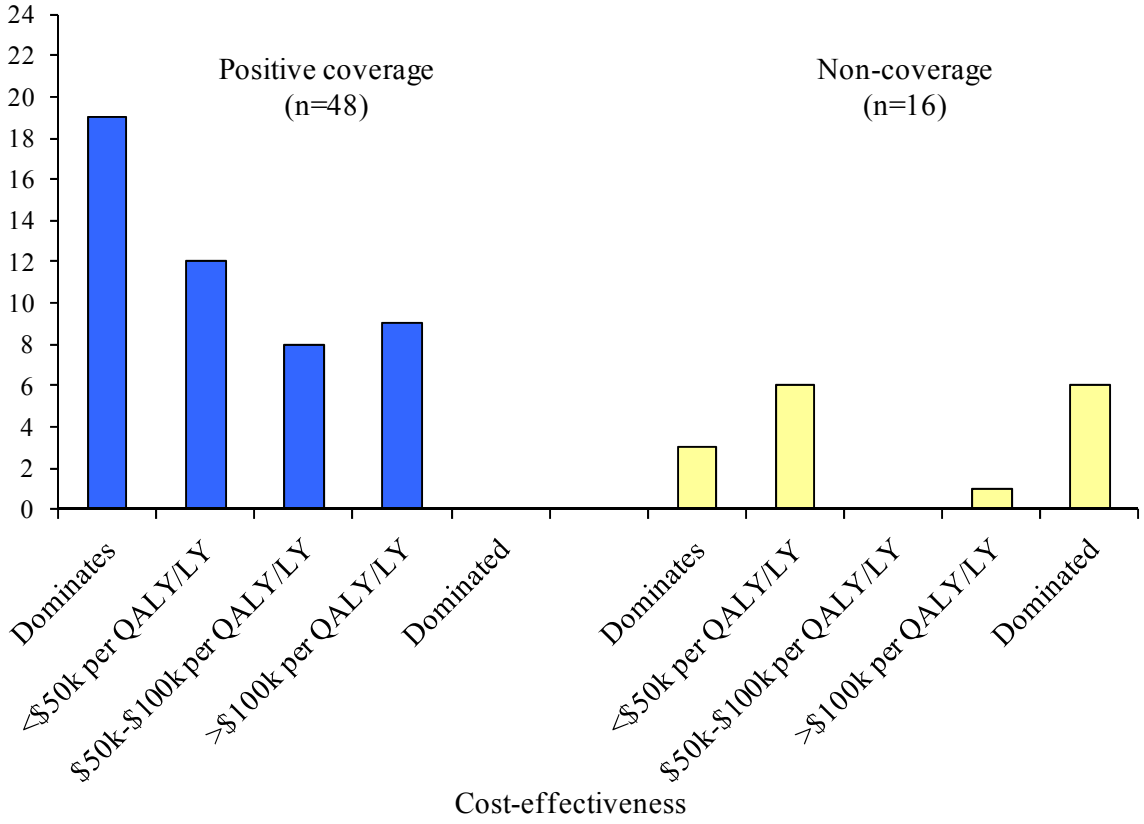


Figure 19. Overview of findings - Cost-effectiveness of coverage decisions

The graphical presentation of the findings is insufficient to determine that there is a statistically significant difference between positive and non-coverage decisions with respect to cost-effectiveness. To evaluate if this was the case, a Mann Whitney U test, as described in Section 5.2.2, was performed. The null hypothesis was defined as there being no difference between the sample of positive coverage decisions and the sample of non-coverage decisions, or, more specifically, that the two population distributions have the same central locations.[xiv] The test showed that there was a statistically significant difference between the cost-effectiveness of interventions subject to positive coverage decisions compared to those subject to non-coverage decisions ($p<0.05$), i.e., the null hypothesis was rejected. This test was repeated when the five studies reporting cost-per life year gained were excluded. The conclusion of the test remained the same, i.e., that there was a statistically significant difference between positive coverage decisions and non-coverage decisions ($p<0.05$) with respect to cost-effectiveness.

[xiv] In formulating the null hypothesis that the central locations of the two population distributions are the same, I assume that apart from any possible differences in central location, the two population distributions are identical (Newbold, Carlson, & Thorne 2003)

5.4. Discussion

Although CMS NCDs have been discussed in the literature, as far as I am aware this is the first empirical study of its kind. (Gillick 2004;Neumann, Rosen, & Weinstein 2005) The research described in this chapter provides important insights into the relationship between CMS NCDs and cost-effectiveness evidence.

5.4.1. Cost-effectiveness evidence featuring in decision memos

Given CMS's stated position on the use of cost-effectiveness evidence, my expectation was that few, if any, decisions memos would include an indication that CMS had considered cost-effectiveness evidence in their review. It was somewhat surprising that for 14 coverage decisions, estimates of cost-effectiveness originated from the decision memo (Table 17). The discussion of cost-effectiveness evidence or citation of a cost-effectiveness study pertained to a positive coverage decision and non-coverage decision in 12 and two of these instances, respectively.

Of the positive coverage decisions, seven were estimated to be dominant. The remaining five had favourable estimates of cost-effectiveness with the highest ICER $27,161 per life year gained. Two decision memos that feature a particularly comprehensive review and discussion of cost-effectiveness evidence refer to screening immunoassay fecal-occult blood test and smoking cessation counselling. (CMS 2003b;CMS 2005e) The decision memo for screening immunoassay fecal-occult blood test included a detailed account of the cost-effectiveness of the intervention. Incremental cost-effectiveness ratios were presented, e.g., "*Hemoccult II® at $4.50 had a cost-effectiveness ratio of $1,071 per life year gained.*" (CMS 2003b)

In the decision memo for smoking and tobacco use cessation counselling, estimates of the cost, cost-effectiveness, and resultant savings associated with the intervention are presented as follows:

> "*Evidence suggests that smoking cessation interventions are highly cost-effective when compared with other medical treatments and prevention programs…. The average annual Medicare cost would be $11.2 million, with a ten-year Medicare cost of $112 million. The*

> *ten-year Medicare savings would be $75 million, with a ten-year non-Medicare savings of $62 million. Over this time, the combined savings to Medicare, state government healthcare programs, third party payers, and to health consumer's out-of-pocket costs, the total savings of the benefit would exceed the costs."* (CMS 2005e)

The two non-coverage decisions for which the decision memo contained an estimate of cost-effectiveness were electrical bioimpedance for cardiac output monitoring (estimated ICER of $6,341 per QALY gained) and PET for Alzheimer's disease/dementia (estimated to be dominant). In the decision memo for electrical bioimpedance for cardiac output monitoring, it is stated that the manufacturer submitted a cost-effectiveness study to CMS. (CMS 2006c) However, CMS excluded this study from their review and in response to its submission CMS stated, "*CMS does not consider cost in making NCDs*". In the decision memo for PET for Alzheimer's disease/dementia, it states that CMS reviewed and made a then unpublished cost-effectiveness study by Silverman et al. available to the Medicare Evidence Development and Coverage Advisory Committee (MEDCAC) convened for the NCD. (CMS 2003a) The now-published study is a reformulated analysis of a previously published cost-effectiveness analysis by the same author. It estimates the use of PET in the early evaluation of dementia patients to increase diagnostic accuracy and reduce the need for nursing home care and unnecessary drug therapy. (Silverman et al. 2002;Silverman et al. 2003) However, these findings are in contrast to a technology assessment (TA) commissioned by CMS for the same NCD, which concluded that using PET in the early evaluation of dementia patients actually decreased aggregate QALYs while increasing costs compared to standard care, i.e., it was a dominated strategy. The findings of the commissioned TA were included in the database in preference to Silverman et al. (2003).

Each decision memo that included discussion of cost-effectiveness evidence or citation of a cost-effectiveness study was re-reviewed to better understand the extent that cost-effectiveness evidence informed the NCD. It was consistently the case that coverage decisions appeared to be guided predominantly by the clinical evidence and that the cost-effectiveness evidence did not appear to have a particularly influential role. Despite this, it is perhaps telling that of the 14 coverage decisions for which the estimate of cost-effectiveness originated in the decision memo,

12 pertained to positive coverage decisions and only two to non-coverage decisions. It is possible that rather than helping inform the coverage decision, the presentation of cost-effectiveness evidence for positive coverage decisions was to justify or support the coverage decision that was based predominantly on clinical evidence.

On occasion, although cost-effectiveness evidence did not explicitly feature in the decision memo, the concept of cost-effectiveness appeared relevant to the coverage decision. In one instance, the NCD for diabetic peripheral neuropathy with loss of protective sensation, CMS made two coverage decisions for the provision of foot care consistent with cost-effectiveness evidence. (CMS 2001) For the first decision, positive coverage for diabetics who suffer from a loss of protective sensation, a cost-effectiveness study estimating foot care to be dominant was identified. (Ragnarson & Apelqvist 2001) For the second decision, non-coverage for diabetics who have not lost protective sensation, a cost-effectiveness study was identified estimating the intervention to be associated with an ICER of approximately $190,000 per QALY. (Rauner, Heidenberger, & Pesendorfer 2005) While the decision memo does not attribute non-coverage directly to cost-effectiveness evidence -- neither cost nor cost-effectiveness is discussed in the decision memo -- given that foot care has some degree of a clinical benefit in this population, evidence of relative value appears to have been considered.

As noted, in CMS's *Guidance for the Public, Industry and CMS Staff* it states that cost-effectiveness is not a factor in NCDs, and in some cases decision memos are consistent with this stated policy. (CMS 2010e) For example, in one decision memo it states, "*From the initial PubMed yield, CMS then excluded abstracts, case reports, review articles, meta-analyses, cost-effectiveness studies…*". (CMS 2010d) On multiple occasions, CMS note that they do not consider cost or cost-effectiveness evidence when making NCDs. (CMS 2002b;CMS 2006b;CMS 2006c;CMS 2006e;CMS 2007) Further, in one case, the NCD for external counterpulsation (ECP) therapy, an intervention's manufacturer submitted cost-effectiveness evidence during the comment period. (CMS 2006d;Varricchione 2006) This evidence, however, did not feature in the decision memo. On other occasions, CMS is inconsistent with respect to their stated policy. For example, as illustrated above for the NCDs for screening immunoassay fecal-occult blood test and

smoking cessation counselling, cost-effectiveness evidence has been explicitly discussed in decision memos and appears to have played some role in coverage determinations. (CMS 2003b;CMS 2005e)

Notably, in the decision memo for screening immunoassay fecal-occult blood test, it states, "*A ratio of $50,000 or less per life saved is often accepted by health economists as indicating that the intervention is 'cost-effective'*". (CMS 2003b) Although this valuation of a cost-effectiveness threshold is said to be acceptable to '*health economists*', and not necessarily CMS, it illustrates CMS's awareness of it. Interestingly, CMS used the same language regarding the cost-effectiveness threshold in a recent NCD (2009) for computed tomography colonography (CTC) for colorectal cancer. (CMS 2009b) Reference to the cost-effectiveness threshold is qualified with the language that the value is often accepted by '*health economists'* as indicating that the intervention is "cost-effective". This is consistent with the decision memo for screening immunoassay fecal-occult blood test.

NCDs made from 1999 through 2007 were considered in this research. As discussed in Chapter 3, the Medicare Improvements for Patients and Providers Act (MIPPA) of 2008 authorises CMS to consider cost-effectiveness evidence for preventative interventions. (MIPPA 2008) In a number of recent NCDs for preventative care, CMS have justified their consideration of cost-effectiveness evidence with the MIPPA legislation. (CMS 2009b;CMS 2009c;CMS 2011d) However, as these NCDs were made after 2007, they were not included in this research. The MIPPA legislation is discussed further in Chapter 3.

5.4.1.1 Other indications of the relevance of cost-effectiveness evidence

The NCD for PET for Alzheimer's disease/dementia was unique among those included in the database as it included discussion of QALYs[xv]. (CMS 2003a) As noted above, as part of this NCD CMS commissioned a technology assessment from the Agency for Healthcare Research and

[xv] As discussed in Section 3.5.2, CMS have considered cost-utility studies in NCDs for preventative interventions in recent years outside the timeframe this research.

Quality (AHRQ). Along with life expectancy and severe dementia-free life expectancy, QALYs were included as an outcome measure in the agency's evaluation. (Matchar et al. 2001) It is notable, however, that, although QALYs were considered, they were not used as part of a cost-effectiveness analysis. Reasons why QALYs were considered in this case are not provided. The NCD for PET for Alzheimer's disease/dementia sets an important precedent, suggesting that CMS values and is willing to use QALYs.

Notably, a cost-effectiveness analysis identified through my literature search was partly funded by CMS.[xvi] Ramsey et al. (2003) conducted a cost-effectiveness analysis of lung-volume-reduction surgery for patients with severe emphysema. (Ramsey et al. 2003) The findings of this study are not, however, presented in the decision memo for lung volume reduction surgery.

5.4.2. Statistically significant difference between positive and non-coverage decisions with respect to cost-effectiveness

Between positive coverage decisions and non-coverage decisions, the results of the Mann Whitney U test described above (Section 5.3.4) show a statistically significant difference with respect to cost-effectiveness, suggesting that interventions subject to positive coverage decisions tend to be associated with more favourable cost-effectiveness evidence than those subject to non-coverage decisions. The results of the Mann Whitney U test remained the same when cost-effectiveness studies reporting cost-per life year gained ratios were included in the dataset. There does not appear, however, to be an upper bound on the value of acceptable cost-effectiveness, i.e., a fixed cost-effectiveness threshold. The highest ICER associated with a non-coverage decision was approximately $190,000 per QALY, for foot care for diabetics who have lost protective sensation in their feet. However, six positive coverage decisions were associated with an ICER greater than this value, with the highest approximately $835,000 per QALY, for ventricular assist devices as destination therapy for chronic end-stage heart failure patients.

[xvi] The cost-effectiveness study performed by the National Emphysema Treatment Trial Research Group was supported by contracts with the National Heart, Lung, and Blood Institute, the Centers for Medicare and Medicaid Services, and the Agency for Healthcare Research and Quality.

It is apparent that CMS covers a number of interventions that do not appear cost-effective by traditional standards. Seventeen interventions subject to positive coverage decisions are associated with an ICER greater than $50,000 per QALY, nine of which are greater than $100,000 per QALY, and three of which are greater than $500,000 per QALY. Notably, often within positive coverage decisions for interventions with high estimates of cost-effectiveness, the language used in the decision memo suggested that the CMS had been aware of the economic implications of the coverage decision. For example, within the decision memos for insulin infusion pumps for diabetic patients (ICER of $558,522) and ventricular assist devices as destination therapy in chronic end-stage heart failure patients (ICER of $820,967), CMS note the high cost of the technology. (CMS 1999b;CMS 2003c)

5.4.3. Challenges and limitations

Identifying cost-effectiveness studies relevant to the coverage decisions was challenging. In contrast to the studies reviewed in Section 4.5.1, estimates of cost–effectiveness were not readily available from the regulatory agency's published documentation, in this case from decision memos. Of the 64 cost-effectiveness studies included in this review, only 14 (22%) originated from the decision memo, with the remainder identified from literature searches.

Of the 103 NCDs included in this research, 43 (42%) included at least one coverage decision for which a relevant cost-effectiveness study was identified. Of the total sample of 255 coverage decisions, a relevant cost-effectiveness estimate was available for 64 (25%). Given the proportion of coverage decisions for which a relevant cost-effectiveness estimate was unavailable, there is a possibility that those associated with a cost-effectiveness estimate are unrepresentative of the overall sample. Of positive coverage decisions (n=123), 48 (39%) were associated with a cost-effectiveness estimate. Of non-coverage decisions (n=132), 16 (12%) were associated with a cost-effectiveness estimate, noticeably less than for positive coverage decisions. This was not entirely unexpected as typically the volume of supporting literature reviewed in the decision memo was much greater for positive coverage decisions than for non-coverage decisions. Also, evidence suggests that there is bias towards the publication of favourable cost-effectiveness estimates that, if CMS NCDs are consistent with the clinical evidence, may increase the

likelihood of a cost-effectiveness study being available for positive coverage decisions. (Bell et al. 2006)

Of the 64 cost-effectiveness studies, 40 (62.5%) reported cost-effectiveness using cost-per QALY gained ratios, 19 (29.7%) using cost-per disease-specific unit, and five (7.8%) using cost-per life year gained ratios. The five cost-effectiveness studies reporting the ICER in terms of cost-per life year reported positive ICERs ranging from $1,072 to $84,439 (Table 15). In some instances, adjusting survival gain with quality of life will decrease the magnitude of the denominator of the ICER equation, causing estimates of cost-effectiveness to be higher when reporting cost-per QALY gained as opposed to a cost-per life year gained ratios. Although not directly comparable, cost-effectiveness studies reporting cost-per life year gained ratios were considered along with those reporting cost-per QALY ratios to maximise the quantity of cost-effectiveness evidence available for this research. For the second research objective, the effect of excluding cost-effectiveness studies reporting cost-per life year gained ratios from the sample was observed. Consistent with the findings when studies reporting cost-per life year gained ratios were included, there was a statistically significant difference ($p<0.05$) between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness.

Although care was taken to ensure that included estimates of cost-effectiveness were high quality and representative of the coverage decision, it was inevitable that there would be some degree of variability among them. For example, while the majority of cost-effectiveness analyses identified were cost-utility analyses, there was variability between them in how the utility weights were calculated. Utilities can be measured either directly (e.g., Standard Gamble [SG] and Time-Trade Off [TTO]) or indirectly (e.g., EuroQol [EQ-5D] and Health Utility Index [HUI]). The method of elicitation will affect their valuation. (Brazier 2008) Research has shown, however, that adjustment of the quality of the utility estimate often does not substantially alter cost-effectiveness estimates. (Chapman et al. 2004) Another source of variation among the included cost-effectiveness analyses was with respect to funding source. Evidence suggests that cost-effectiveness studies funded by the pharmaceutical and medical device industry are more likely to report ICERs below accepted thresholds. (Bell et al. 2006) Although priority was given to non-

industry-funded studies, it was not possible to find a non-industry-funded cost-effectiveness analysis for each coverage decision.

It is possible that, on occasion, CMS considered the intervention to be more or less effective or costly than the inputs used in the corresponding cost-effectiveness analysis. This is illustrated in the example noted above for PET for Alzheimer's disease/dementia. In this instance, while a cost-effectiveness study (Silverman et al. (2003)) estimated the intervention to be dominant, the technology assessment commissioned by CMS estimated the intervention to be dominated. (CMS 2003a;Matchar et al. 2001;Silverman et al. 2003) Nevertheless, in the absence of CMS routinely performing cost-effectiveness studies as part of NCDs, relying on the peer-reviewed literature is a practical and manageable approach.

5.4.4. Next Steps

Future research can build upon the approach used here. I considered NCDs made from 1999 through 2007, and it is possible that updating the research to include subsequent NCDs would increase the number of coverage decisions associated with a relevant estimate of cost-effectiveness. Notably, as discussed in Section 3.5.2, since 2007 CMS have often considered cost-effectiveness evidence in their evaluation of preventative interventions. Expanding the research to include these additional NCDs may allow a comparison between the preventative and non-preventative interventions with respect to cost-effectiveness.

This research was limited to the Medicare programme. It would be valuable to expand the scope to include a broader range of public and private payers. For example, as described in Section 3.3.2.3, the health care programmes in the Department of Defense (DoD) and U.S. Department of Veterans Affairs (VA) state that economic factors are considered when evaluating interventions. A comparison between these agencies with respect to the cost-effectiveness of coverage decisions would prove insightful. As private health care represents the majority of the US health care system, expanding this research to consider private payers would also be of interest. As discussed in Section 3.3.2.3, private payers such as Wellpoint, one of the largest private health insurance

companies, has issued guidelines providing a framework for the submission of cost-effectiveness evidence, and thus may represent a useful starting point for such research.

While NCDs could be considered the most important of CMS's coverage determinations, local coverage policies, or local coverage determinations (LCDs), made in the absence of a national coverage policy by MACs represent the majority of Medicare coverage decisions. (CMS 2010b;Foote, Halpern, & Wholey 2005) To more completely understand the cost-effectiveness of CMS's coverage decisions, this research could be expanded to consider LCDs.

As noted above, one limitation of this research is the relatively small proportion of coverage determinations for which it was possible to identify a relevant cost-effectiveness estimate. One approach to get around the limitations of the cost-effectiveness literature would be to gain input from clinicians, health economists, and other health services researchers. Using an expert panel to make judgements with respect to the cost-effectiveness of interventions for which the available cost-effectiveness literature proved insufficient, while not ideal, would be one potential approach.

Cost-effectiveness is only one economic parameter of potential relevance to CMS. This analysis does not account for the budget impact of the intervention, which may have factored into decisions. It is possible that cost-ineffective interventions are more likely to receive a positive coverage decision if they are associated with a relatively small budget impact. Also, I did not consider reimbursement decisions, which may, in addition to coverage, have an important effect on actual use of a technology, i.e., despite a positive coverage decision, the mode of reimbursement could provide an incentive or disincentive for a physician to prescribe an intervention or for a hospital to offer it. (Neumann, Rosen, & Weinstein 2005)

5.4.5. Policy relevance

Given the current financial difficulties faced by the US health care system, and specifically the Medicare programme, this research is timely and relevant. It provides an insight into the value of many interventions offered by Medicare and the relationship between NCDs and cost-

effectiveness evidence. Importantly, however, this research highlights the lack of knowledge regarding the value of many of the interventions offered by Medicare. Therefore, this research identifies a potential research agenda to better understand the value of interventions offered by the Medicare programme. Highlighted in this research are a number of interventions offered in the Medicare programme that are not cost-effective by traditional standards. Offering these interventions generates relatively little health gain for the expenditure and suggests that resources could provide greater benefits if directed towards alternative interventions.

This research highlights CMS's experience with cost-effectiveness evidence. CMS has helped fund a cost-effectiveness study and has used 'tools' of cost-effectiveness analysis, e.g., QALYs, to inform NCDs. Further, discussion of a cost-effectiveness threshold suggests an awareness of the magnitude of cost-effectiveness ratios that are generally considered to be indicative of value. Therefore, despite CMS's stated position on the use of cost-effectiveness evidence, this research suggests that CMS has considered such evidence worthy of review on occasion.

It is notable that there is a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness, suggesting that interventions subject to positive coverage decisions tend to be associated with more favourable cost-effectiveness evidence. While this research is insufficient to suggest cost-effectiveness evidence is an independent factor in CMS coverage, the finding is encouraging.

5.4.6. Chapter summary

As noted in Section 3.4.2.1, in their "*Guidance for the Public, Industry and CMS Staff*", CMS state that cost-effectiveness is not a factor CMS considers in making NCDs. (CMS 2010e) To the best of my knowledge, the research presented in this chapter is the first systematic attempt to evaluate NCDs from the perspective of cost-effectiveness.

The first objective of this research was to examine NCD decision memos to identify instances when cost-effectiveness evidence was cited or discussed, thus assessing the consistency of CMS's behaviour and its stated position on the use of cost-effectiveness evidence. In 14 instances, the identified estimate of cost-effectiveness associated with a coverage decision originated from the decision memo. It was notable that 12 of these occasions pertained to positive coverage decisions, and the estimate of cost-effectiveness was favourable in each case (maximum ICER of $27,161 per life year gained).

The second objective of this research was to determine if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness. The methodological approach built upon the published studies reviewed in Section 4.5.1. The findings show a statistically significant difference between positive coverage decisions and non-coverage decision with respect to their cost-effectiveness, suggesting that interventions subject to positive coverage decisions tend to be associated with more favourable cost-effectiveness evidence.

While the findings of the research presented in this chapter show that CMS have on occasion considered cost-effectiveness evidence in the review of the evidence base, and that positive coverage decisions tend to be associated with more favourable estimates of cost-effectiveness, the approach taken here is insufficient to determine if cost-effectiveness, or the availability of cost-effectiveness evidence, is an independent factor in CMS NCDs. To evaluate this research question, it is necessary to account for other potentially relevant factors in the NCD process in the analysis. Further, it is necessary to restrict the included cost-effectiveness evidence to only that available at the time of the NCD, i.e., that which CMS may realistically have had the opportunity

to consider. To this end, the research in Chapter 6 presents a logistic regression analysis that includes the independent variables described in Section 4.7.4 to evaluate the factors relevant to CMS when making NCDs.

This research also shows that CMS cover a number of interventions that are not cost-effective by traditional standards. Coverage of cost-ineffective interventions results in a financial burden on the programme, while generating relatively little health gain. Consequently, a reallocation of resources from cost-ineffective interventions to more cost-effective interventions will generate additional health gain for existing levels of expenditure. In an attempt to quantify potential benefits of using cost-effectiveness evidence, a hypothetical reallocation of expenditures between interventions subject to CMS NCDs using a cost-effectiveness decision rule is presented in Chapter 7.

6. Empirical Research: Part 2

6.1. Introduction

In Chapter 4, I provided the foundation for the empirical aspect of this thesis, including a background to the Medicare programme and the processes for the coverage of medical technology in it. Also, I presented the objectives of the empirical work and reviewed the relevant literature that helped inform the methodological approach. Lastly, I described the development of the database and set of variables available for this research.

Chapter 5 constituted the first piece of empirical work. This research had two objectives: to identify instances when cost-effectiveness evidence was cited or discussed in CMS national coverage determinations (NCDs), and to determine if there is a difference between positive and non-coverage decisions with respect to cost-effectiveness. With respect to the first objective, 14 coverage decisions were identified for which cost-effectiveness evidence was cited or discussed in the decision memo, with 12 pertaining to positive coverage decisions and two to non-coverage decisions. Notably, the estimate of cost-effectiveness in each case was favourable (maximum ICER of $27,161 per life year gained). With respect to the second objective, findings show a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to their cost-effectiveness, suggesting that covered interventions tend to be associated with more favourable cost-effectiveness evidence.

The research presented in this chapter builds on Chapter 5 and constitutes my second piece of empirical work. While the findings of the research presented in Chapter 5 show that CMS have on occasion reviewed cost-effectiveness evidence and that positive coverage decisions tend to be associated with more favourable estimates of cost-effectiveness, the methodological approach was insufficient to determine if cost-effectiveness, or the availability of cost-effectiveness evidence, is an independent factor in CMS NCDs.

The objective of the research presented in this chapter is to determine if cost-effectiveness is an independent predictor of coverage, i.e., when controlling for other factors, is cost-effectiveness, or the availability of cost-effectiveness evidence, statistically significantly associated with coverage.

6.2. Objective and Methodology

6.2.1. Objective

The empirical work presented in this chapter has the following objective:

1. To determine if cost-effectiveness is an independent predictor of coverage when controlling for other factors that may be considered to influence coverage decisions.

6.2.2. Methodology

As described in Section 4.5.1, a body of literature exists that describes the evaluation of coverage and reimbursement decisions, or recommendations for the efficient use of medical technology, made by agencies across a number of countries. The methodological approaches taken in conducting these studies helped inform the research presented here.

I included NCDs made from 1999 through 2007 in the dataset. The unit of analysis was the coverage decision, with all coverage decisions made within NCDs considered separately. The variables I considered for inclusion in this research are discussed in Section 4.7.4 and were chosen in an attempt to account for the key aspects of decision-making. The final set of independent variables included; *Quality of evidence*, *Alternative intervention*, *Cost-effectiveness*, *Type of intervention*, *Coverage requestor*, and *Date* (Table 18). Full details of the methodology used to generate the database are presented in Section 0.

Table 18. Variables included in the analysis

Variable	Description	Variable		
		Construction	Definition	% of observations
Dependent variable				
Coverage decision	Outcome of the coverage decision.	Dichotomous variable	Positive coverage	54%
			Non-coverage decision	46%
Independent variables				
Quality of evidence	A review of the supporting clinical evidence as presented in the decision memo performed independently by two reviewers.	Categorical variable – Categorised using USPSTF guidelines (Table 2)	Good	53%
			Poor	10%
			Insufficient	37%
Alternative intervention	The availability of an alternative intervention for the same indication.	Dichotomous variable	Alternative available	83%
			No alternative available	17%
Cost-effectiveness	Estimate of cost-effectiveness for the intervention.	Categorical variable**	No estimate	79%
			Dominates	8%
			ICER <$50k/QALY	6%
			ICER >$50k/QALY	8%
Type of intervention	The broad indication of the intervention.	Categorical variable	Treatment	67%
			Diagnostic test (includes staging/ screening/ monitoring)	28%
			Other (including health education, preventative care, and mobility assistive equipment)	5%
Coverage requestor	The group or individual that requested coverage.	Categorical variable	Manufacturer	32%
			Internally generated	37%
			Other (includes medical/professional society or organization)	41%
Date	Decisions grouped into years	Categorical variable	1999-2001	22.6%
			2002-2003	36.9%
			2004-2005	14.9%
			2006-2007	25.6%

* Percentages may not add to 100% due to rounding

** Percentages represent those cost-effectiveness studies available at the time of the NCD

6.2.3. Exploratory analysis of interaction effects

The independent variables (Table 18) were chosen as I considered them to represent factors likely to have an effect on coverage decisions. Despite the relevance of the variables to various aspects of the coverage decision, it was important to evaluate whether there was any interaction between them that may influence the findings of the model. The term "interaction" describes the instance when the combined effects of two variables are not a sum of their individual effects. Interaction effects can have important implications for the interpretation of the findings of the statistical model as the changing value of one variable will have unpredictable consequences on the result of the model.

For the included independent variables there are various sources of potential interaction. Cost-effectiveness is the aggregate of clinical and cost data. Consequently, I considered it a possibility that interaction would exist between *Cost-effectiveness* and *Quality of evidence*. Similarly, as cost-effectiveness analyses typically compare two or more competing interventions I considered it a possibility that interaction would exist between *Cost-effectiveness* and *Alternative intervention*. Also, evidence suggests that the number of cost-effectiveness studies published each year is increasing (Neumann PJ et al. 2005). Therefore, I considered it a possibility that an interaction effect would exist between *Cost-effectiveness* and *Date.*

For each possible combination of variables I included their interaction term, i.e., the product of the two independent variables, in a multivariate regression including all independent variables. In each instance the interaction term was not significant. Therefore, I deemed that interaction between independent variables did not have an important effect on the results of the model.

6.2.4. Analyses

The model was estimated using binomial logistic regression, regressing the coverage decision against the independent variables (Table 18). I chose logistic regression primarily as a dichotomous decision output was used. I considered a production function to be the conceptual

framework that best represented CMS's decision-making process (Figure 20). As described in Section 4.2.2, NCDs are guided by the 'reasonable and necessary' criterion, with little guidance provided regarding the relative importance of different decision inputs. In a production function model, decision-making inputs are considered concurrently. An alternative approach would be to consider decision-making as a hierarchical process. However, greater knowledge regarding how CMS prioritises different decision inputs would be required to facilitate this approach.

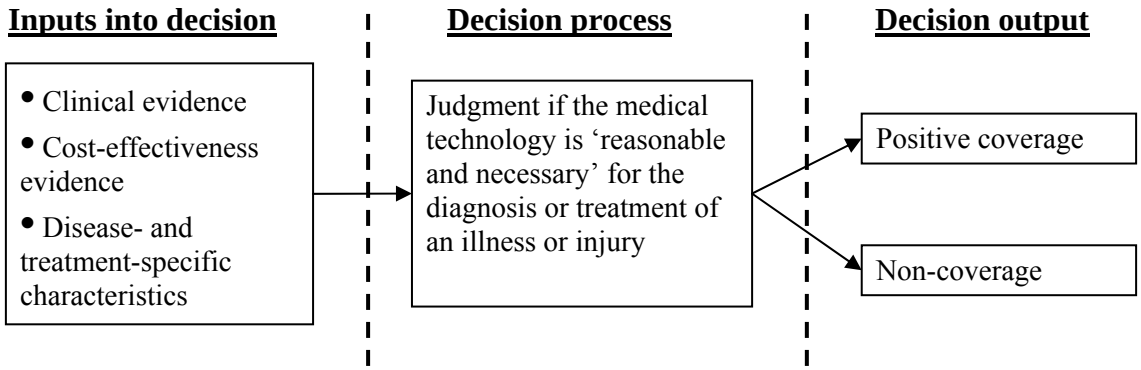


Figure 20. Conceptual framework - Production model of CMS decision-making

For multivariate regressions, I assessed model goodness of fit using a number of statistics. The pseudo R^2 statistic is automatically estimated when performing a multivariate regression. The closer the pseudo R^2 is to 1.0, the better the model fits the data. I also performed the Hosmer-Lemeshow test. This test divides observations into deciles based on predicted probabilities, and then computes a chi-square from observed and expected frequencies. The null hypothesis that there is no difference between observed and predicted values is tested. When the test result is non-significant, the null hypothesis is rejected and the model can be considered to 'fit' the data. The final test I used is the area under the receiver operating characteristic (ROC) curve, or C-statistic. This summary measure of predictive power plots sensitivity versus 1 - specificity. The area under the ROC line is estimated, which ranges between 0.5 and 1.0, with values close to 1.0 representing good predictive power and a high degree of goodness-of-fit.

I performed both univariate and multivariate regression analyses. Following each regression, the predicted probability of a positive coverage decision was computed for each variable. A predicted probability is the estimated probability of a positive coverage decision associated with each category in each independent variable while controlling for the other included variables. In multivariate models, the covariates were fixed at their sample mean values to facilitate the calculation of this statistic.

In the primary set of analyses, univariate and multivariate analyses were performed. First, I performed a univariate analysis that included each independent variable (Table 19). Next, I performed a multivariate analysis that included all independent variables (Table 20). Lastly, I performed a second multivariate analysis that included only those variables with at least one significant category in the initial multivariate analysis (Table 20).

In the primary analyses, I categorised the date of decision variable (*Date*) into groups of years (Table 18). However, it was uncertain how best to code *Date,* and therefore I performed a set of analyses to evaluate the impact of alternative coding approaches on study findings.

I performed a number of analyses to examine the cost-effectiveness variable (*Cost-effectiveness*). In the primary analyses, only cost-effectiveness studies available at the time of the NCD were included. Further, cost-effectiveness studies reporting cost-per life years gained were included along with studies reporting cost-per QALY gained. This approach assumes equivalency between the two metrics. To evaluate if the inclusion of cost-per life year studies affected study findings, I performed univariate and multivariate regressions using a dataset excluding cost-per life year gained studies. Analyses were also performed using a version of *Cost-effectiveness* that included cost-effectiveness data from studies published after the NCD (under the theory that CMS may have had access to unpublished data on cost-effectiveness or could have independently calculated cost-effectiveness in the absence of published studies). Also, I deconstructed the cost-effectiveness variable into the numerator and denominator of the ICER to evaluate to what extent incremental costs or incremental effectiveness was driving the results. Finally, I performed a set

of regressions using a dataset including only coverage decisions with an associated estimate of cost-effectiveness.

A p-value below the 5% level is regarded as statistically significant, and between 5% and 10% is regarded as weakly significant. Analyses were undertaken using Stata SE version 11. (Stata 2009)

6.3. Results

6.3.1. Primary analyses

6.3.1.1 Univariate analyses

In the univariate analyses, *Quality of evidence* ($p<0.01$), *Alternative intervention* ($p<0.01$), *Cost-effectiveness* ($p<0.05$), and *Date* ($p<0.01$) were statistically significantly associated with the coverage decision (Table 19). *Type of intervention* and *Coverage requestor* were non-significant. For *Quality of Evidence*, interventions associated with good quality supporting evidence (*Good*), compared to interventions associated with insufficient evidence (*Insufficient*), were approaching seven times more likely to receive a positive coverage decision (over twice as likely when considering predicted probabilities). Interventions with an available alternative were approximately seven times less likely to be associated with a positive coverage decision (approaching half as likely when considering predicted probabilities) than those without an available alternative. Compared with interventions estimated to be dominant, those for which there was no estimate of cost-effectiveness were approximately six times less likely to be associated with a positive coverage decision (approaching half as likely when considering predicted probabilities); the categories *ICER<$50k/QALY* and *ICER>$50k/QALY* were not statistically significant. Coverage decisions made after 2001 were significantly less likely to be associated with a positive coverage decision than those made prior to that year, with those made in 2006-2007 over 13 times less likely to be associated with a positive coverage decision than those made from 1999 through 2001 (less than half as likely when considering predicted probabilities).

Table 19. Results of univariate logistic regression

Summary statistics				
Number of observations = 195				
Variable	**OR**	**95% CI**		**Predicted probability**
Quality of evidence				
Good	6.678***	3.418	13.049	0.777
Poor	0.686	0.222	2.122	0.263
Insufficient	Reference category			0.342
Joint significance p<0.001				
Alternative intervention				
No	Reference category			0.879
Yes	0.138***	0.046	0.41	0.500
Cost-effectiveness				
No estimate	0.173**	0.038	0.800	0.510
Dominates	Reference category			0.857
ICER<$50k/QALY gained	0.444	0.060	3.285	0.727
ICER>$50k/QALY gained	0.458	0.070	3.017	0.733
Joint significance p=0.035				
Date				
1999-2001	Reference category			0.864
2002-2003	0.176***	0.066	0.469	0.528
2004-2005	0.258**	0.082	0.809	0.621
2006-2007	0.074***	0.026	0.212	0.320
Joint significance p<0.001				
Type of intervention				
Treatment	Reference category			0.557
Diagnostic	0.856	0.453	1.615	0.519
Other	7.151*	0.880	58.076	0.900
Joint significance p=0.149				
Coverage requestor				
Manufacturer	Reference category			0.565
Internally generated	1.106	0.557	2.194	0.589
Other	0.882	0.432	1.800	0.533
Joint significance p=0.813				

*p<0.1, **p<0.05, ***p<0.01

6.3.1.2 Multivariate analyses

The results from the multivariate analysis when including all variables are presented in Table 20. This model had an adjusted pseudo R^2 of 0.312, indicating room for improvement. However, the C statistic is reasonably high (0.86) and the null hypothesis was rejected when using the Hosmer-Lemeshow goodness-of-fit test, indicating that the model fits the data reasonably well. *Quality of evidence* ($p<0.01$), *Alternative intervention* ($p<0.05$), and *Date* ($p<0.01$) were significantly related to the coverage decision. Although the cost-effectiveness variable was not significant ($p=0.143$), the category "*No estimate*" was ($p<0.1$). *Type of intervention* and *Coverage requestor* were non-significant variables. Consistent with the univariate findings, interventions associated with good quality supporting evidence were estimated to be approximately six times more likely to be associated with a positive coverage decision compared to interventions associated with insufficient evidence (approximately twice as likely when considering predicted probabilities). Interventions with an available alternative were seven times less likely to be associated with a positive coverage decision than those without an alternative intervention available (approximately two-thirds as likely when considering predicted probabilities). With respect to *Cost-effectiveness*, interventions not associated with an estimate of cost-effectiveness were almost six times less likely to be associated with a positive coverage decision compared with those estimated to be dominant (approaching half as likely when considering predicted probabilities); the categories *ICER<$50k/QALY* and *ICER>$50k/QALY* were not statistically significant. Coverage decisions made after 2001 were significantly less likely to be associated with a positive coverage decision than those made prior to that year, with those made from 2006 through 2007 approximately 10 times less likely to be associated with a positive coverage decision than those made from 1999 through 2001 (approaching half as likely when considering predicted probabilities).

Table 20. Results of multivariate logistic regression

	Multivariate logistic regression including all variables				Multivariate logistic regression including those variables identified as significant			
	Summary statistics				**Summary statistics**			
	Pseudo R^2 = 0.312				Pseudo R^2 = 0.304			
	Number of observations = 195				Number of observations = 195			
	Area under ROC curve = 0.858				Area under ROC curve = 0.850			
	Hosmer-Lemeshow goodness-of-fit = 0.137				Hosmer-Lemeshow goodness-of-fit test = 0.421			
Variable	**Adj. OR**	**95% CI**		**Predicted probability**	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence								
Good	5.900***	2.602	13.354	0.715	6.040***	2.762	13.209	0.715
Poor	1.218	0.300	4.951	0.424	1.423	0.367	5.522	0.445
Insufficient	Reference category			0.389	Reference category			0.381
	Joint significance p<0.01				Joint significance p<0.01			
Alternative intervention								
No	Reference			0.809	Reference			0.823
Yes	0.147**	0.031	0.686	0.521	0.130***	0.035	0.483	0.515
Cost-effectiveness								
No estimate	0.185*	0.032	1.085	0.529	0.190*	0.035	1.024	0.527
Dominates	Reference category			0.781	Reference category			0.779
ICER<$50k/QALY gained	0.653	0.052	8.159	0.724	0.646	0.055	7.589	0.719
ICER>$50k/QALY gained	0.319	0.035	2.893	0.616	0.375	0.046	3.021	0.637
	Joint significance p=0.143				Joint significance p=0.110			
Date								
1999-2001	Reference category			0.761	Reference category			0.765
2002-2003	0.334*	0.110	1.018	0.578	0.311**	0.103	0.937	0.569
2004-2005	0.324*	0.085	1.228	0.572	0.310*	0.084	1.144	0.569
2006-2007	0.101***	0.024	0.375	0.365	0.109***	0.031	0.383	0.380
	Joint significance p<0.01				Joint significance p<0.01			
Type of intervention								
Treatment	Reference category			0.575				
Diagnostic	0.759	0.328	1.754	0.532				
Other	1.676	0.117	24.014	0.653				
	Joint significance p=0.744							
Coverage requestor								
Manufacturer	Reference category			0.572				
Internally generated	1.156	0.434	3.074	0.594				
Other	0.721	0.270	1.927	0.522				
	Joint significance p=0.610							

*p<0.1, **p<0.05, ***p<0.01

Comparing the results of the univariate and multivariate analyses suggests that there may be some degree of collinearity in the model. In the univariate logistic regression results for *Type of*

intervention, the *Other* category had an odds ratio (OR) of 7.15 (95% CI 0.880 – 58.076). However, in the multivariate analysis the OR dropped to 1.68 (0.117 – 24.014), suggesting that collinearity may be causing the change. To investigate the possibility that two or more of the predictors were non-significant because of collinearity, a number of diagnostic tests were performed. First, various specifications of the model were tested. Dropping individual variables from the model did not produce a noticeable shift in the results, consistent with the presence of collinearity. Coding the variables differently (e.g., coding *Date* as a continuous variable) also did not have a noticeable effect on the results. The Variance Inflation Factor (VIF) was calculated for each coefficient following each multivariate regression. The VIF provides an estimate of how much of the variance of the coefficient estimate is being inflated by multicollinearity. A commonly used rule of thumb is that VIFs≥10 may be a reason for concern and suggest that multicollinearity is a problem. (O'Brien 2007) Following a multivariate regression, the VIFs were calculated and none were approaching a value of 10. Finally, following a multivariate regression the correlations of the estimated coefficients were evaluated. A high correlation (0.8 or higher) between pairs of coefficients indicates that problematic collinearity may exist. (Grewal, Cote, & Baumgartner 2004) No high correlations between any pairs of coefficients were identified. An additional method of determining the presence of multicollinearity is to include different samples in the dataset and observe any large changes in the results of the model. However, given the limited size of the dataset, it was not possible to use this method. The tests described above suggest that despite the rather large difference in the magnitude of the OR for the category *Other* in *Type of intervention* between the univariate and multivariate analyses, multicollinearity is not problematic. Nevertheless, the impact of simply dropping *Type of intervention* from the analysis was evaluated. Compared to the multivariate logistic regression including *Type of intervention*, excluding this variable did not have a notable effect on the results.

When considering a model that included only variables in which at least one category was statistically significant, the model had a pseudo R^2 of 0.304, indicating room for improvement. However, the C statistic is reasonably high (0.850) and the null hypothesis was rejected when using the Hosmer- Lemeshow goodness-of-fit test, indicating that the model fits the data reasonably well. *Quality of evidence* ($p<0.01$), *Alternative intervention* ($p<0.01$), and *Date* ($p<0.01$) were significantly associated with the coverage decision. Although C*ost-effectiveness*

was not significant (p=0.110), the category *No estimate* was (p<0.1). Interventions associated with good quality supporting evidence were six times more likely to receive a positive coverage decision compared with those associated with insufficient evidence (approximately twice as likely when considering predicted probabilities). Compared to interventions with no available alternative, those that had an available alternative were approximately eight times less likely to be associated with a positive coverage decision (approaching half as likely when considering predicted probabilities). Compared with interventions estimated to be dominant, those with no associated estimate of cost-effectiveness were approximately five times less likely to receive a positive coverage decision (approximately two thirds as likely when considering predicted probabilities); the categories *ICER<$50k/QALY* and *ICER>$50k/QALY* were not statistically significant. Each of the categories in *Date* was significant. Interventions considered in more recent time periods were increasingly less likely to be associated with a positive coverage decision. Coverage decisions made from 2006 through 2007 were approximately 10 times less likely to be associated with a positive coverage decision than those made from 1999 through 2001 (half as likely when considering predicted probabilities).

6.3.2. Controlling for multiple coverage decisions from a single decision memo

It was often the case that multiple coverage decisions were made in a single NCD. It may be the case that coverage decisions made in the same NCD are not independent as they are made using similar or related evidence and by the same committee. In order to control for this potential relationship, univariate and multivariate regressions were performed using cluster analysis, i.e., the estimated standard errors were adjusted for within-decision memo clustering of decisions (Table 21 and Table 22). To perform a cluster analysis, observations are assigned to a particular subset in which it there may be some similarity between decisions. In this case, coverage decisions were clustered with respect to the NCD in which they were made. Clustering observations does not have an effect on estimated OR or predicted probabilities, but it does affect the estimated 95% CI and the statistical significance. The findings from the cluster analyses did not vary greatly from the findings of the models when observations were unclustered. A notable change in the univariate analysis was that *Type of intervention* was statistically significant (p<0.01) which was not the case in the unclustered analysis. In the multivariate clustered

analysis, a noticeable change from the unclustered analysis was that in *Date, 2002-2003* ($p=0.143$) and *2004-2005* ($p=0.152$) were non-significant. Also, *Cost-effectiveness* was jointly significant ($p<0.1$), which was not the case in the unclustered analysis.

Table 21. Univariate regression clustering sub-decisions from the same decision memo

Summary statistics				
Number of observations = 195				
Independent variable	**OR**	**95% CI**		**Predicted probability**
Quality of evidence				
Good	6.678***	2.330	19.138	0.777
Poor	0.686	0.134	4.524	0.263
Insufficient	Reference category			0.342
Joint significance p<0.01				
Alternative intervention				
No	Reference category			0.879
Yes	0.138***	0.053	0.357	0.500
Cost-effectiveness				
No estimate	0.173**	0.037	0.843	0.510
Dominates	Reference category			0.857
ICER<$50,000/QALY gained	0.444	0.056	3.546	0.727
ICER>$50,000/QALY gained	0.458	0.064	3.297	0.733
Joint significance p<0.1				
Date				
1999-2001	Reference category			0.864
2002-2003	0.176***	0.053	0.590	0.528
2004-2005	0.258*	0.058	1.158	0.621
2006-2007	0.074***	0.019	0.287	0.320
Joint significance p<0.01				
Type of intervention				
Treatment	Reference category			0.557
Diagnostic	0.856	0.304	2.407	0.519
Other	7.151***	2.759	18.535	0.900
Joint significance p<0.01				
Coverage requestor				
Manufacturer	Reference category			0.565
Internally generated	1.106	0.370	3.300	0.589
Other	0.882	0.274	2.832	0.533
Joint significance p=0.926				

*p<0.1, **p<0.05, ***p<0.01

Table 22. Multivariate regression clustering sub-decisions from the same decision memo

	Multivariate logistic regression including all variables				Multivariate logistic regression including those variables identified as significant			
	Summary statistics				**Summary statistics**			
	Pseudo R^2 = 0.312				Pseudo R^2 = 0.305			
	Number of observations = 195				Number of observations = 195			
	Area under ROC curve = 0.858				Area under ROC curve = 0.850			
	Hosmer-Lemeshow goodness-of-fit test = 0.137				Hosmer-Lemeshow goodness-of-fit test = 0.421			
Independent variable	**Adj. OR**	**95% CI**		**Predicted probability**	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence								
Good	5.900***	1.819	19.102	0.715	6.040***	1.825	19.987	0.715
Poor	1.218	0.182	8.142	0.424	1.423	0.239	8.459	0.445
Insufficient	Reference category			0.389	Reference category			0.381
	Joint significance p<0.05				Joint significance p<0.05			
Alternative intervention								
No	Reference			0.809	Reference category			0.8233
Yes	0.147**	0.036	0.601	0.521	0.130***	0.042	0.404	0.515
Cost-effectiveness								
No estimate	0.185**	0.037	0.933	0.529	0.190*	0.045	0.803	0.527
Dominates	Reference category			0.781	Reference category			0.779
ICER<$50,000/QALY gained	0.653	0.104	4.123	0.724	0.646	0.115	3.631	0.719
ICER>$50,000/QALY gained	0.319	0.029	3.460	0.616	0.375	0.046	3.089	0.637
	Joint significance p<0.1				Joint significance p<0.1			
Date								
1999-2001	Reference category			0.761	Reference category			0.765
2002-2003	0.334	0.077	1.447	0.578	0.311	0.071	1.363	0.569
2004-2005	0.324	0.069	1.514	0.572	0.310	0.073	1.324	0.569
2006-2007	0.101***	0.024	0.434	0.365	0.109***	0.029	0.405	0.380
	Joint significance p=0.021				Joint significance p<0.05			
Type of intervention								
Treatment	Reference category			0.575				
Diagnostic	0.759	0.220	2.612	0.532				
Other	1.676	0.349	8.061	0.653				
	Joint significance p=0.745							
Coverage requestor								
Manufacturer	Reference category			0.572				
Internally generated	1.156	0.358	3.734	0.594				
Other	0.721	0.228	2.284	0.522				
	Joint significance p=0.752							

*p<0.1, **p<0.05, ***p<0.01

6.3.3. Alternative specification of the Date variable

There were a number of alternative available approaches to code *Date*. Initially, I coded *Date* as a continuous variable, using a day as the unit of analysis. Although significant, the OR was so close to 1.0 that it was difficult to interpret in a meaningful way. *Date* was then included as a continuous variable, with a year as the unit of analysis. In this case, the estimated OR was 0.71 ($p<0.01$) for the univariate analysis and 0.70 ($p<0.01$) for the multivariate analysis. However, reporting ORs for continuous variables is not ideal as interpretation is not straightforward. Including *Date* as a categorical variable as opposed to a continuous variable is less restrictive because it allows the effect of a unit change in the variable (e.g., a year) to be not constant across the values of the variable. NCDs made from 1999 through 2007 were included in the analysis. A number of options were available to code this variable categorically. First, an analysis was performed using each year as a separate category. When running a multivariate analysis that included all variables, the years 2003, 2006, and 2007 were significant ($p<0.1$). Although these findings show that in these particular years the likelihood of coverage is less than in 1999, the approach is insufficient to establish a temporal trend in the data. Consequently, to develop the variable, coverage decisions were grouped together with respect to the year, or order, in which they were made. When grouped together into quartiles and using the first quartile as the reference category, the second, third and fourth quartiles were associated with ORs of 0.39 ($p<0.01$), 0.20 ($p<0.01$), and 0.08 ($p<0.01$), respectively, in a multivariate analysis including all variables. Alternatively, when grouped with respect to the year in which the coverage decision was made and using years 1999-2001 as the reference category, the years 2002-2003, 2004-2005, and 2006-2007 were associated with ORs of 0.33 ($p<0.1$), 0.32 ($p<0.1$), and 0.10 ($p<0.01$), respectively, in a multivariate analysis including all variables. Specifying *Date* with groupings of years was chosen for the final model, as I considered these results to be the most intuitive and straightforward to interpret.

6.3.4. Evaluation of the cost-effectiveness variable

The cost-effectiveness variable was evaluated in a numbers of ways. These are described below.

6.3.4.1 Exclusion of cost-per life year studies

In the primary analyses, cost-effectiveness studies reporting cost-per life year gained ratios were included along with those that reported cost-per QALY gained. This approach assumes equivalency between these metrics, which is not correct in the majority of scenarios. To evaluate the effect of including cost-effectiveness studies that report cost-per life year gained ratios, analyses were performed excluding these studies from the dataset. Five studies were excluded, each associated with a positive coverage decision (Table 23).

Excluding cost-effectiveness studies reporting cost-per life year gained ratios did not greatly impact the findings, with the results qualitatively the same as when these studies were included (Table 24 and Table 25).

Table 23. Cost-effectiveness studies reporting cost-per life year gained ratios

Technology - Coverage decision	Year	ICER (US$)	Reference:
Screening Immunoassay Fecal-Occult Blood Test (Hemoccult II FOBT)	2003	$1,072	Report to the Agency for Healthcare Research and Quality (2003)
Screening Immunoassay Fecal-Occult Blood Test (iFOBT)	2003	$21,001	Report to the Agency for Healthcare Research and Quality (2003)
Autologous Stem Cell Transplantation (AuSCT) for Multiple Myeloma - Treatment of multiple myeloma for patients who meet certain conditions	2000	$27,687	Trippoli S et al. (1998)
Implantable Cardioverter Defibrillators (ICDs) - Documented sustained ventricular tachyarrhythmia	2003	$36,396	Mushlin AI et al. (1998)
Implantable Cardioverter Defibrillators (ICDs) - Documented familial or inherited conditions with a high risk of ventricular tachyarrhythmias	2003	$84,439	Larsen G et al. (2002)

Table 24. Univariate regression with cost-per life year studies excluded from the dataset

Summary statistics				
Number of observations = 195				
Variable	**OR**	**95% CI**		**Predicted probability**
Cost-effectiveness				
No estimate	0.184*	0.040	0.850	0.525
Dominates	Reference category			0.857
ICER<$50k/QALY gained	0.222	0.027	1.846	0.571
ICER>$50k/QALY gained	0.417	0.063	2.768	0.714
Joint significance p=0.102				

*p<0.1, **p<0.05, ***p<0.01

Table 25. Multivariate regression with cost-per life year studies excluded from the dataset

	Multivariate logistic regression including all variables				Multivariate logistic regression including those variables identified as significant			
	Summary statistics				**Summary statistics**			
	Pseudo R^2 = 0.306				Pseudo R^2 = 0.298			
	Number of observations = 195				Number of observations = 195			
	Area under ROC curve = 0.852				Area under ROC curve = 0.846			
	Hosmer-Lemeshow goodness-of-fit test =0.109				Hosmer-Lemeshow goodness-of-fit test = 0.116			
Variable	**Adj. OR**	**95% CI**		**Predicted probability**	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence								
Good	6.166***	2.731	13.924	0.721	6.376***	2.929	13.880	0.723
Poor	1.171	0.290	4.722	0.410	1.362	0.354	5.238	0.429
Insufficient	Reference category			0.382	Reference category			0.373
	Joint significance p<0.01				Joint significance p<0.01			
Alternative intervention								
No	Reference category			0.809	Reference category			0.819
Yes	0.149**	0.032	0.689	0.512	0.139***	0.038	0.509	0.516
Cost-effectiveness								
No estimate	0.196*	0.034	1.141	0.538	0.195*	0.036	1.057	0.536
Dominates	Reference category			0.780	Reference category			0.782
ICER<$50k/QALY gained	0.387	0.022	6.780	0.645	0.364	0.022	6.029	0.636
ICER>$50k/QALY gained	0.288	0.031	2.646	0.600	0.330	0.040	2.692	0.621
	Joint significance p=0.284				Joint significance p=0.233			
Date								
1999-2001	Reference category			0.760	Reference category			0.764
2002-2003	0.343*	0.113	1.040	0.580	0.317**	0.105	0.952	0.569
2004-2005	0.313*	0.083	1.185	0.563	0.303*	0.082	0.112	0.561
2006-2007	0.106***	0.029	0.391	0.367	0.116***	0.033	0.406	0.345
	Joint significance p<0.01				Joint significance p<0.01			
Type of intervention								
Treatment	Reference category			0.574				
Diagnostic	0.792	0.346	1.813	0.537				
Other	1.626	0.115	23.002	0.647				
	Joint significance p=0.794							
Coverage requestor								
Manufacturer	Reference category			0.580				
Internally generated	1.097	0.416	2.890	0.594				
Other	0.655	0.248	1.728	0.514				
	Joint significance p=0.526							

*p<0.1, **p<0.05, ***p<0.01

6.3.4.2 Addition of cost-effectiveness data from studies published after the NCD

In the primary set of analyses, the dataset was restricted to include only estimates of cost-effectiveness available at the time of the NCD, i.e., those to which CMS may feasibly have had access. Nineteen estimates of cost-effectiveness were identified that were published following the NCD, and additional analyses were performed including these studies. One hundred and thirty-six (69.7%) coverage decisions were included in the category *No estimate*, 22 (11.3%) in *Dominates*, 16 (8.2%) in *<$50k per QALY*, and 21 (10.8%) in *>$50k per QALY*. Considering the univariate analysis, and including *Dominates* as the reference category, the estimated ORs for *No estimate*, *<$50k per QALY*, and *>$50k per QALY* were 0.13 ($p<0.01$), 0.47 ($p=0.378$), and, 0.67 ($p=0.632$), respectively. These findings are qualitatively similar to those from the univariate analysis, in which case only estimates of cost-effectiveness available at the time of the NCD were included. The findings of the multivariate analysis are presented in Table 26. As for the univariate analysis, with respect to *Cost-effectiveness*, the results of the multivariate regression when including cost-effectiveness evidence published after the NCD were qualitatively similar to when only cost-effectiveness evidence available at the time of the NCD was included. One notable change, however, is that the categories in *Cost-effectiveness* were jointly significant when estimates published following the NCD were included ($p<0.01$). The ORs for the other independent variables included in the multivariate analyses were qualitatively the same as those when only cost-effectiveness evidence available at the time was included.

Table 26. Results of multivariate logistic regression – inclusion of cost-effectiveness studies published following the coverage decision

	Multivariate logistic regression including all variables				Multivariate logistic regression including those variables identified as significant			
	Summary statistics				**Summary statistics**			
	Pseudo R^2 = 0.357				Pseudo R^2 = 0.347			
	Number of observations = 195				Number of observations = 195			
	Area under ROC curve = 0.877				Area under ROC curve = 0.877			
	Hosmer-Lemeshow goodness-of-fit test = 0.239				Hosmer-Lemeshow goodness-of-fit test = 0.049			
Independent variable	**Adj. OR**	**95% CI**		**Predicted probability**	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence								
Good	6.275***	2.633	14.956	0.706	6.275***	2.755	14.292	0.704
Poor	1.169	0.249	5.494	0.419	1.457	0.326	6.516	0.450
Insufficient	Reference category			0.393	Reference category			0.387
	Joint significance p<0.01				Joint significance p<0.01			
Alternative therapy available								
No	Reference category			0.824	Reference category			0.829
Yes	0.107***	0.0267	0.428	0.515	0.103***	0.026	0.408	0.511
Cost-effectiveness								
No estimate	0.106***	0.024	0.463	0.483	0.115***	0.028	0.471	0.482
Dominates	Reference			0.810	Reference			0.805
ICER<$50,000 per QALY gained	0.708	0.088	5.683	0.768	0.701	0.094	5.220	0.760
ICER>$50,000 per QALY gained	0.290	0.041	2.022	0.641	0.362	0.058	2.272	0.665
	Joint significance p<0.01				Joint significance p<0.01			
Date								
1999-2001	Reference category			0.739	Reference category			0.747
2002-2003	0.339*	1.10	1.045	0.568	0.308**	0.101	0.940	0.561
2004-2005	0.378	0.095	1.504	0.587	0.349	0.091	1.341	0.582
2006-2007	0.107***	0.027	0.428	0.380	0.111***	0.029	0.417	0.391
	Joint significance p<0.05				Joint significance p<0.05			
Type of intervention								
Treatment	Reference category			0.575				
Diagnostic	0.771	0.324	1.832	0.538				
Other	1.321	0.082	21.231	0.614				
	Joint significance p=0.818							
Coverage requestor								
Manufacturer	Reference category			0.569				
Internally generated	1.284	0.461	3.581	0.604				
Other	0.672	0.244	1.849	0.512				
	Joint significance p=0.443							

*p<0.1, ** p<0.05, ***p<0.01

6.3.4.3 Subdividing the ICER into incremental cost and incremental benefit

To examine the cost-effectiveness variable, regressions were performed including the numerator (incremental cost) and denominator (incremental effectiveness) as separate variables. The version of the cost-effectiveness variable including estimates of cost-effectiveness published after the NCD and cost-effectiveness studies reporting cost-per life year gained ratios was used for this analysis to maximise the number of observations associated with an estimate of cost-effectiveness.

Incremental cost was coded as a categorical variable using the following categories: *No estimate* (n=138), *Cost-saving* (n=23), *Incremental cost <$5,000* (n=15), and *Incremental cost >$5,000* (n=19). Incremental effect was coded as a categorical variable using the following categories: *No estimate* (n=158), *Negative incremental benefit* (n=4), *0-0.1 QALY gain* (n=13), *0.1-1.0 QALY gain* (n=13), and *>1.0 QALY* gain (n=7). These categorisations were used in both cases to ensure a sufficient number of observations in each category. The findings of the univariate analyses are presented in Table 27.

Table 27. Univariate analysis – Incremental costs and effectiveness

Summary statistics				
Variable	**OR**	**95% CI**		**Predicted probability**
Incremental cost (n=195)				
No estimate	0.119***	0.034	0.419	0.442
Cost-saving	Reference category			0.870
<$5,000	0.600	0.104	3.463	0.800
>$5,000	1.275	0.190	8.545	0.895
Joint significance p<0.01				
Incremental benefit (n=191)				
No estimate	0.180	0.021	1.528	0.519
Negative effect	Dropped from analysis (perfectly predicts model outcome)			
0-0.1 QALYs	0.917	0.682	12.322	0.846
0.1-1.0 QALYs	0.917	0.682	12.322	0.846
>1.0 QALYs	Reference category			0.857
Joint significance p<0.05				

*p<0.1, **p<0.05, ***p<0.01

The categories in *Incremental cost* were jointly significant (p<0.01). The likelihood of a positive coverage decision when there is no estimate of incremental cost is approximately eight times less than when the intervention is estimated to be cost-saving (reference category). Neither *Incremental cost <$5,000* nor *Incremental cost >$5,000* were significant. The categories in *Incremental benefit* were jointly significant (p<0.05). As those interventions associated with negative incremental benefit perfectly predicted a non-coverage decision, *Negative Effect* was dropped from the regression. Compared to the reference category, *>1.0 QALY,* all other categories were associated with an OR of less than 1.0, indicating a decreased likelihood of coverage. No categories in *Incremental effectiveness* were significant.

Incremental cost and *Incremental effectiveness* were included along with all other variables in a multivariate logistic regression (Table 28). The findings of this analysis suggested the presence of multicollinearity. The ORs for a number of the categories in *Incremental cost* and *Incremental effectiveness* were exceptionally high and associated with uncalculated 95% confidence intervals. To evaluate the potential presence of multicollinearity, the VIF was estimated for each coefficient

generated from the multivariate regression. In *Incremental cost,* the VIFs for *<$5,000* and *>$5,000* were 6.44 and 5.45, respectively. In *Incremental effectiveness,* the values for *No estimate* and *0-0.1 QALYs* were 12.64 and 5.44. These relative high estimates, in particular the VIF for *No estimate*, suggested the presence of multicollinearity. Further, I estimated the correlation between coefficients. Although no correlations were greater than 0.8, those between the incremental cost and incremental effectiveness variables ranged up to a value of 6.7. Given the concerns and uncertainty regarding the presence of multicollinearity, separate regressions were performed when including *Incremental cost* and *Incremental effectiveness.*

Table 28. Multivariate regression including incremental cost and incremental effectiveness

	Multivariate logistic regression including all variables			
	Summary statistics			
	Pseudo R^2 = 0.448			
	Number of observations = 191			
	Area under ROC curve = 0.913			
	Hosmer- Lemeshow goodness-of-fit test = 0.442			
Variable	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence				
Good	7.812***	2.922	20.882	0.705
Poor	2.285	0.345	15.144	0.528
Insufficient	Reference category			0.413
	Joint significance p<0.01			
Alternative intervention				
No	Reference category			0.829
Yes	0.079***	0.013	0.503	0.526
Incremental cost				
No estimate	0.068***	0.012	0.377	0.393
Cost-saving	Reference category			0.733
<$5,000	34.951	0.195	6253.629	0.935
>$5,000	3.10×10^7	Not estimated		0.995
	Joint significance p<0.01			
Incremental effectiveness				
No estimate	1.99×10^7	Not estimated		0.601
Negative effect	Dropped from analysis (perfectly predicts model outcome)			
0-0.1 QALYs	298975.100	Not estimated		0.226
0.1-1.0 QALYs	488113.400	Not estimated		0.253
>1.0 QALYs	Reference category			0.087
	Joint significance p=0.480			
Date				
1999-2001	Reference category			0.767
2002-2003	0.297*	0.086	1.021	0.598
2004-2005	0.268*	0.059	1.217	0.582
2006-2007	0.053***	0.010	0.285	0.359
	Joint significance p<0.01			
Type of intervention				
Treatment	Reference category			0.585
Diagnostic	0.744	0.284	1.948	0.550
Other	1.737	0.089	34.038	0.651
	Joint significance p=0.774			
Coverage requestor				
Manufacturer	Reference category			0.575
Internally generated	1.591	0.468	5.408	0.629
Other	0.636	0.206	1.960	0.521
	Joint significance p=0.292			

*p<0.1, **p<0.05, ***p<0.01

A multivariate logistic regression was performed including *Quality of evidence*, *Alternative intervention*, *Date*, *Type of intervention*, *Coverage requestor*, and *Incremental cost*. With *Cost-saving* serving as the reference category in *Incremental cost*, only the category *no estimate* was significant (OR=0.09, $p<0.01$). The categories *<$5,000* and *>$5,000* were not significant (Table 29).

Table 29. Multivariate regression including incremental cost

	Multivariate logistic regression including all variables			
	Summary statistics			
	Pseudo R^2 = 0.390			
	Number of observations = 195			
	Area under ROC curve = 0.893			
	Hosmer- Lemeshow goodness-of-fit test = 0.144			
Variable	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence				
Good	6.061***	2.508	14.648	0.695
Poor	1.111	0.209	5.907	0.425
Insufficient	Reference category			0.409
	Joint significance p<0.01			
Alternative intervention				
No	Reference			0.824
Yes	0.095***	0.018	0.492	0.514
Incremental cost				
No estimate	0.093***	0.021	0.404	0.466
Cost-saving	Reference category			0.809
<$5,000	0.729	0.084	6.294	0.771
>$5,001	1.110	0.103	11.938	0.821
	Joint significance p<0.01			
Date				
1999-2001	Reference category			0.735
2002-2003	0.338**	0.108	1.057	0.572
2004-2005	0.343	0.083	1.411	0.574
2006-2007	0.094***	0.021	0.411	0.377
	Joint significance p<0.05			
Type of intervention				
Treatment	Reference category			0.570
Diagnostic	0.848	0.350	2.059	0.548
Other	1.468	0.089	24.175	0.621
	Joint significance p=0.897			
Coverage requestor				
Manufacturer	Reference category			0.567
Internally generated	1.330	0.463	3.822	0.604
Other	0.689	2.245	1.943	0.517
	Joint significance p=0.450			

*p<0.1, **p<0.05, ***p<0.01

Similarly, a multivariate logistic regression was performed including *Quality of evidence*, *Alternative interventions*, D*ate*, *Type of intervention*, *Coverage requestor*, and *Incremental effectiveness* (Table 30). With *>1.0 QALYs* serving as the reference category, only *No estimate* was significant (OR=0.05, $p<0.05$). As all interventions associated with a deleterious effect on health were non-covered, *Negative effect* perfectly predicted the final coverage decision and was dropped from the regression. The remaining variables, *0-0.1 QALYs*, and *0.1-1.0 QALYs*, were not significant. As *Quality of evidence* takes into account both quality of evidence and degree of clinical benefit (Table 11, Section 4.7.4.2), it was assumed that there may be some collinearity between it and incremental effectiveness. This was evaluated through the calculation of VIFs and the correlation between coefficients. Although no evidence of collinearity was identified, an additional multivariate logistic regression was performed excluding *Quality of evidence*. Consistent with the previous regression, only *No estimate* was significant (OR=0.03, $p<0.01$), with the remaining variables either dropped or not significant (Table 31).

Table 30. Multivariate regression including incremental effectiveness

	Multivariate logistic regression including all variables			
	Summary statistics			
	Pseudo R^2 = 0.3728			
	Number of observations = 191			
	Area under ROC curve = 0.874			
	Hosmer- Lemeshow goodness-of-fit test = 0.231			
Variable	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence				
Good	6.689***	2.766	16.177	0.719
Poor	1.442	0.300	6.923	0.469
Insufficient	Reference category			0.409
	Joint significance p<0.01			
Alternative intervention				
No	Reference category			0.797
Yes	0.141**	0.026	0.758	0.539
Incremental effectiveness				
No estimate	0.053**	0.035	0.787	0.530
Negative effect	Dropped from analysis (perfectly predicts model outcome)			
0-0.1 QALYs	0.261	0.010	6.621	0.748
0.1-1.0 QALYs	0.489	0.019	12.553	0.817
>1.0 QALYs	Reference category			0.881
	Joint significance p<0.05			
Date				
1999-2001	Reference category			0.828
2002-2003	0.202***	0.061	0.669	0.589
2004-2005	0.169**	0.042	0.687	0.559
2006-2007	0.037***	0.008	0.170	0.319
	Joint significance p<0.01			
Type of intervention				
Treatment	Reference category			0.565
Diagnostic	1.096	0.471	2.549	0.578
Other	2.937	0.222	38.841	0.709
	Joint significance p=0.705			
Coverage requestor				
Manufacturer	Reference category			0.555
Internally generated	2.027	0.670	6.133	0.649
Other	0.752	0.269	2.103	0.516
	Joint significance p=0.166			

*p<0.1, **p<0.05, ***p<0.01

Table 31. Multivariate regression including incremental effectiveness and excluding quality of evidence

	Multivariate logistic regression including all variables			
	Summary statistics			
	Pseudo R^2 = 0.296			
	Number of observations = 191			
	Area under ROC curve = 0.844			
	Hosmer- Lemeshow goodness-of-fit test = 0.554			
Variable	**Adj. OR**	**95% CI**		**Predicted probability**
Alternative intervention				
No	Reference			0.846
Yes	0.108***	0.021	0.574	0.534
Incremental effectiveness				
No estimate	0.028***	0.002	0.321	0.518
Negative effect	Dropped (perfectly predicts model outcome)			NA
0-0.1 QALYs	0.152	0.008	2.765	0.776
0.1-1.0 QALYs	0.236	0.013	4.406	0.827
>1.0 QALYs	Reference			0.940
	Joint significance p<0.01			
Date				
1999-2001	Reference			0.867
2002-2003	0.213***	0.071	0.640	0.624
2004-2005	0.190**	0.051	0.701	0.602
2006-2007	0.021***	0.005	0.085	0.240
	Joint significance p<0.01			
Type of intervention				
Treatment	Reference			0.565
Diagnostic	0.906	0.413	1.985	0.549
Other	7.304	0.597	89.382	0.840
	Joint significance p=0.280			
Coverage requestor				
Manufacturer	Reference			0.606
Internally generated	1.000	0.394	2.539	0.606
Other	0.588	0.225	1.5436	0.520
	Joint significance p=0.412			

*p<0.1, **p<0.05, ***p<0.01

6.3.4.4 Including only observations with an associated estimate of cost-effectiveness

As noted above, when including estimates of cost-effectiveness published after the date of the NCD, 59 coverage decisions (approximately 30% of the total sample) were associated with an estimate of cost-effectiveness. Univariate and multivariate logistic regressions were performed using a dataset that included only coverage decisions with an associated estimate of cost-effectiveness (Table 32 and Table 33).

In this restricted dataset, all interventions without an available alternative were associated with non-coverage decisions. Consequently, *Alternative intervention* perfectly predicted the outcome of the model and was dropped from the analysis. Also, in *Type of intervention* the category "*Other*" perfectly predicted coverage and was dropped from the analysis.

In the univariate analysis, *Quality of evidence* was the only significant variable ($p<0.05$). Compared to interventions associated with insufficient evidence, those associated with good quality evidence were approximately seven times more likely to be associated with a positive coverage decision ($p<0.05$).

The findings of the multivariate analysis had a similar pattern to the univariate analysis, with *Quality of evidence* the sole significant variable ($p<0.05$). As for the univariate analysis, compared to interventions associated with insufficient evidence, those associated with good quality evidence were more likely to be associated with a positive coverage decision, albeit with a much greater OR (59.4 vs. 7.2).

Table 32. Univariate logistic regression – Only coverage decisions with associated estimate of cost-effectiveness

Summary statistics				
Number of observations in dataset = 59				
Independent variable	**OR**	**95% CI**		**Predicted probability**
Quality of evidence (n=59)				
Good	7.200**	1.524	34.022	0.923
Poor	0.600	0.066	5.447	0.500
Insufficient	Reference category			0.625
	Joint significance p<0.05			
Alternative intervention (n=51)				
No	Reference category			NA
Yes	Dropped (perfectly predicts model outcome)			NA
Cost-effectiveness (n=59)				
Dominates	Reference category			0.864
ICER<$50,000/QALY gained	0.474	0.090	2.497	0.750
ICER>$50,000/QALY gained	0.671	0.131	3.438	0.810
	Joint significance p=0.678			
Date (n=59)				
1999-2001	Reference category			0.941
2002-2003	0.167	0.018	1.546	0.727
2004-2005	0.563	0.031	10.117	0.900
2006-2007	0.146	0.013	1.658	0.700
	Joint significance p=0.303			
Type of intervention (n=55)				
Treatment	Reference category			0.800
Diagnostic	1.00	0.253	3.949	0.800
Other	Dropped (perfectly predicts model outcome)			NA
Coverage requestor (n=59)				
Manufacturer	Reference category			0.842
Internally generated	0.50	0.106	2.355	0.727
Other	1.50	0.220	10.218	0.889
	Joint significance p=0.411			

*p<0.1, **p<0.05, ***p<0.01

Table 33. Multivariate logistic regression – Only those variables with an associated cost-effectiveness estimate

	Multivariate logistic regression including all variables			
	Summary statistics			
	Pseudo R^2 = 0.367			
	Number of observations = 55			
	Area under ROC curve = 0.812			
	Hosmer-Lemeshow goodness-of-fit test = 0.716			
Independent variable	**Adj. OR**	**95% CI**		**Predicted probability**
Quality of evidence				
Good	59.406**	2.014	1,751.867	0.937
Poor	0.910	0.027	30.679	0.461
Insufficient	Reference category			0.474
	Joint significance p<0.05			
Cost-effectiveness				
Dominates	Reference category			0.898
ICER<$50,000 per QALY gained	0.135	0.003	5.914	0.733
ICER>$50,000 per QALY gained	0.173	0.007	4.242	0.756
	Joint significance p=0.518			
Date				
1999-2001	Reference category			0.953
2002-2003	0.060	0.002	2.031	0.730
2004-2005	0.123	0.003	6.003	0.813
2006-2007	0.057	0.001	2.804	0.723
	Joint significance p=0.444			
Type of intervention				
Treatment	Reference category			0.769
Diagnostic	3.035	0.167	55.184	0.865
Other	Dropped			NA
Coverage requestor				
Manufacturer	Reference category			0.828
Internally generated	0.248	0.023	2.718	0.698
Other	7.346	0.146	370.927	0.934
	Joint significance p=0.224			

*p<0.1, **p<0.05, ***p<0.01

6.4. Discussion

In Chapter 5, I evaluated the cost-effectiveness of coverage decisions made in NCDs from 1999 through 2007. The findings show a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness, suggesting that interventions subject to positive coverage decisions tend to be associated with more favourable cost-effectiveness evidence. However, this approach is insufficient to determine if cost-effectiveness, or the availability of cost-effectiveness evidence, is an independent factor in CMS NCDs. The research presented in this chapter builds on the research in Chapter 6 and evaluates whether cost-effectiveness, or the availability of cost-effectiveness evidence, is an independent predictor of coverage, i.e., is statistically significantly associated with coverage when controlling for other factors that may be considered to influence coverage decisions.

Medicare law dictates that coverage of interventions is limited to those that are 'reasonable and necessary' for the diagnosis or treatment of an illness or injury. (CMS 2010e) Other than the notable exclusion of cost or cost-effectiveness evidence from the decision-making process (Section 4.2.2), CMS has not provided comprehensive guidance as to how this language should be interpreted. Although interpretation of the 'reasonable and necessary' terminology has been discussed qualitatively at some length in the literature, to the best of my knowledge this is the first time that CMS NCDs have been evaluated in an empirical manner. (Foote 2002;Neumann 2005)

6.4.1. Methodology

As described in Section 4.5.1, coverage and reimbursement decisions, or recommendations for the efficient use of medical technology, made by agencies across the globe have been subject to evaluation. Although their research objectives vary, e.g., exploration of a cost-effectiveness threshold vs. the relative likelihood of positive recommendations for cancer vs. non-cancer treatments, essentially the studies share the goal of gaining a better understanding of the role of cost-effectiveness evidence in decision-making. A number of these studies have used variations of logistic regression to evaluate the role of cost-effectiveness evidence while controlling for other factors.

The development of the database used for this research is described in Chapter 4. For the most part, information used to generate the variables originated from the decision memo published for each NCD. This approach was consistent with similar studies that also used publicly available documentation to obtain the necessary data. (Chim et al. 2010;Dakin, Devlin, & Odeyemi 2006;Devlin & Parkin 2004;Harris et al. 2008) Unfortunately, decision memos do not provide the same breadth of information as documentation provided by other decision makers. For example, information regarding budget impact and prevalence of disease are often reported in NICE technology appraisals but are rarely reported in NCD decision memos. Variables considered for inclusion but did not feature in the research are described in Section 4.7.4.12.

The coverage decision was the dependent variable in this analysis, i.e., positive coverage and non-coverage. This approach was taken because multiple coverage decisions often were made in NCDs. Prior to the decision to use a dichotomous dependent variable, various alternative approaches to model CMS coverage decisions were considered. For example, if using a dependent variable that allowed for more than two coverage outcomes, e.g., positive coverage, coverage with restrictions and non-coverage options, potential approaches include multinomial logistic or ordered logistic regression. An advantage of the coverage decision being coded as a dichotomous variable is that it facilitates use of a binomial logistic regression model, modelling the dependent variable as simply positive or non-coverage.

As multiple coverage decisions were made in a single NCD, it may be the case that they are not truly independent. To evaluate whether this had an effect on the findings of the analysis, I performed an additional regression in which I clustered coverage decisions together.

The model was estimated using binomial logistic regression, regressing the coverage decision against the independent variables.

6.4.2. Overview of research findings

Findings show that the quality of supporting clinical evidence, the availability of alternative interventions, the available of cost-effectiveness evidence, and the date of the decision are statistically significantly associated with coverage decisions. Neither *Type of intervention* or *Coverage requestor* was statistically significant in the primary analyses. Findings were broadly consistent across the univariate model, the multivariate model including all variables, and the multivariate model including only variables that had at least one statistically significant category in the multivariate model including all variables. Findings remained qualitatively the same when alternative constructions of *Cost-effectiveness* and *Date* were evaluated. With respect to model fit, although the pseudo-R^2 statistics indicated that there was room for improvement, the C-statistics were reasonable high, indicating that the model fits the data reasonably well (0.85 for the primary multivariate analysis).

6.4.2.1 Quality of the supporting clinical evidence

Within each decision memo, CMS presents a review of the supporting clinical evidence and discusses its strengths and weaknesses. The impression gained from reviewing the decision memos is that CMS are careful to ensure that the ultimate coverage decision is supported by their review of the clinical evidence. The findings of the regression analysis are consistent with this impression, as the quality of evidence is statistically significantly associated with coverage decisions. In the primary multivariate analyses, interventions associated with good quality evidence were estimated to be approximately six times more likely to be associated with a positive coverage decision than interventions associated with insufficient evidence when considering ORs (twice as likely when considering predicted probabilities). Good quality clinical evidence was consistently a significant predictor of the coverage decision across analyses.

This finding is encouraging, illustrating the evidence-based approach taken for NCDs. It is important to note that *Quality of evidence* has a number of limitations (Section 6.4.3), including the subjective nature of the variable along with the fact that the USPSTF guidelines used to grade the evidence combine quality of evidence and net benefit into a single measure. These limitations will be addressed in future work.

6.4.2.2 Availability of alternative interventions

Alternative intervention was a statistically significant variable. In the primary multivariate analyses, interventions with available alternatives were approximately seven times less likely to be to be covered than those for which no alternative was available (approaching two times less likely when considering predicted probabilities). *Alternative intervention* was consistently a statistically significant variable across the various analyses.

A number of the reviewed studies (Section 4.5.1) included a variable similar to *Alternative therapy*. (Dakin, Devlin, & Odeyemi 2006;Devlin et al. 2010a;Devlin & Parkin 2004;Harris et al. 2008) *Alternative therapy* accounts for the importance a decision maker attributes to ensuring a patient population has access to care. In this case, the results for *Alternative therapy* show that CMS have a strong preference for the coverage of interventions for indications for which no alternatives exist. The finding also indicates that CMS do not consider interventions in isolation and do account for the treatment landscape when making NCDs.

6.4.2.3 Cost-effectiveness

The objective of this work was to determine if cost-effectiveness was an independent predictor of coverage. To facilitate the interpretation of categorical variables, typically one of the extreme categories is chosen as the reference category. In the case of *Cost-effectiveness*, *Dominates* was chosen. Therefore, the estimated ORs are interpreted as the likelihood of positive coverage relative to interventions estimated to be both more effective and less costly than their comparator.

In the primary univariate and multivariate regressions, the category *No estimate* was statistically significant, with ORs of approximately 0.18 and 0.19, respectively. This finding indicates that compared to interventions estimated to be dominant, those with no associated estimate of cost-effectiveness are approximately five times less likely to be associated with a positive coverage decision. In the primary analyses, the categories *<$50k per QALY* and *>$50k per QALY* were not statistically significant. While in the multivariate analyses the ORs were less than 1.0, consistent

with a hypothesis that interventions with more favourable estimates of cost-effectiveness are more likely be covered, the lack of statistical significance prevented any conclusions to be drawn.

Under the theory that CMS may have had access to unpublished data on cost-effectiveness or could have independently calculated cost-effectiveness in the absence of published studies, cost-effectiveness data from studies published after the NCD were included in an additional set of analyses. The findings of these analyses were qualitatively similar to those including only estimates of cost-effectiveness available at the time of the NCD.

Cost-effectiveness combines estimates of both the costs and effectiveness of the intervention. It may be the case that any impact of cost-effectiveness on positive coverage decisions demonstrated in the analysis is actually due to the impact of effectiveness on the decision rather than cost or cost-effectiveness. The quality of evidence variable included in the analysis simultaneously accounts for evidence quality and the magnitude of the clinical benefit shown by the intervention. I attempt to control, albeit imperfectly, for the clinical effectiveness of each intervention and analyse the impact of cost-effectiveness conditional on this variable.

The cost-effectiveness variable was further explored by subdividing the ICER into incremental cost and incremental effectiveness components. When considered in univariate analyses, *Incremental cost* and *Incremental effect* were jointly significant (Table 27). However, while *No estimate* was a significant category in *Incremental cost*, *No estimate* was not significant in *Incremental effect*. When including *Incremental cost* and *Incremental effectiveness* together as independent variables in a multivariate regression, the presence of collinearity hindered the interpretation of the findings. Therefore, separate multivariate regressions were performed including *Incremental cost* and *Incremental effectiveness,* and in each case, the variable was significant. For both variables, the category *No estimate* was significant; no estimate of incremental cost was associated with an OR of 0.09 ($p<0.01$), and no estimate of incremental effectiveness was associated with an OR of 0.05 ($p<0.01$). For both variables, all other categories were non-significant. This analysis provided little insight into the association of incremental cost or incremental effectiveness with coverage. As categories other than *No estimate* were not

statistically significant, no relationship between magnitude of incremental cost or incremental effectiveness with coverage was identified.

The majority (approximately 70%) of coverage decisions in the dataset were not associated with an estimate of cost-effectiveness. To further evaluate *Cost-effectiveness*, an additional set of analyses was performed including only coverage decisions with an associated estimate of cost-effectiveness. The small size of this restricted dataset (n=59) led to a number of difficulties. Most notably, *Alternative intervention* was dropped from the model as it perfectly predicted model outcome. Of the remaining variables, *Quality of evidence* was the only significant variable; *Date* and *Cost-effectiveness* were not significant. While this analysis further confirms the significance of *Quality of evidence,* it is not informative with respect to the other variables. This analysis would likely provide a greater insight if the dataset were extended to include more recent coverage decisions, or if additional estimates of cost-effectiveness were available for interventions in the current dataset.

In summary, across the primary analyses interventions for which there were no available estimates of cost-effectiveness were less likely to be associated with a positive coverage decision compared with interventions estimated to be dominant. This is an important finding and suggests that, controlling for other factors, the availability of dominant estimates of cost-effectiveness, or lack thereof, impacts the likelihood of a positive coverage decision.

6.4.2.4 Date of decision

The date of decision variable was significant across the primary analyses. As described in Section 4.7.4.6, various alternative approaches for coding *Date* were considered. Ultimately, in the primary analyses, *Date* was included as a categorical variable, with coverage decisions grouped into the years they were made, i.e., 1999-2001, 2002-2003, 2004-2005, and 2006-2007. Notably, for the multivariate analyses each category was statistically significant, and the reported OR decreased for consecutive groups of years, suggesting that CMS became increasingly restrictive throughout the considered time period. This trend was confirmed when including a

version of *Date* in the multivariate analysis in which coverage decisions were ordered with respect to year and grouped into quartiles (Section 4.7.4.6).

While it is apparent that CMS coverage decisions have become increasingly restrictive, this research does little to explain why this was the case. *Date* was included to control for unobserved factors that affect the outcome of NCDs that change over time. It may be that Medicare's ever greater fiscal challenges – the cost of Medicare more than doubled from $213 billion in 1997 to $431 billion in 2007 – influenced the outcome of NCDs. (CMS 2011a) Alternatively, the changing composition of CMS's coverage team may have influenced the likelihood of coverage.

6.4.3. Limitations and challenges

The principal challenge of this research was obtaining the necessary data. Unfortunately, CMS do not present the same breadth of information in decision memos typically presented by agencies in other countries. For example, unlike NICE's Technology Appraisals, CMS decision memos do not present a budget impact estimate or the number of beneficiaries likely to receive the intervention. This made emulating the methods used in the studies reviewed in Section 4.5.1 difficult, and restricted the number of variables available for analysis. Most notably, the majority of included cost-effectiveness studies were obtained from a literature search and thus, there is inevitably variability between with respect to quality. Also, it may be the case that CMS considered the intervention to be more or less effective or costly than the inputs used in the corresponding cost-effectiveness analysis. Nevertheless, as CMS do not routinely perform cost-effectiveness analysis as part of NCDs, relying on the peer-reviewed literature is a practical and manageable approach.

As described in Section 4.7.4.2, the USPSTF guidelines were used to grade the quality of the supporting clinical evidence for each coverage decision. The USPSTF grading scale accounts for magnitude of net clinical benefit and evidence quality in terms of study design and conduct. Two researchers from Tufts Medical Center independently performed the grading of the clinical evidence considered here. The grading was based upon an evaluation of CMS's review of the

clinical evidence presented in the decision memos. Combining net clinical benefit and evidence quality into a single variable is not ideal, and unfortunately this was the only approach used by the Tufts Medical Center researchers. Further, while the grading was based on independent reviews, *Quality of evidence* is a subjective review of the evidence base. *Quality of evidence* would prove more informative if considered in an objective manner, i.e., the assessment of the evidence base was quantifiable and not based on interpretation. Potential approaches to achieving this goal are described in Section 6.4.5 below.

The significance of the *No estimate* category in *Cost-effectiveness* was notable. However, as only 21% of coverage decisions, or 30% when studies published after the NCD were included, were associated with a cost-effectiveness estimate, the cost-effectiveness of the majority of coverage decisions was unavailable. As described in Section 6.3.4.4, analyses were performed that included only coverage decisions for which an estimate of cost-effectiveness was available. However, the small sample size limited the interpretability of the findings.

In contrast to a number of reviewed studies (Section 4.5.1), *Cost-effectiveness* was coded as a categorical variable. The principal reason for this approach was to include all available cost-effectiveness evidence in the research. Including interventions estimated to be dominant or dominated in the dataset would have proved challenging if *Cost-effectiveness* was coded as a continuous variable. Researchers have gotten around this problem by dropping observations that were dominant or dominated from the dataset. (Devlin et al. 2010) However, this approach was not an option given the available sample size, and the exclusion of relevant data in this way is questionable. Other researchers have gotten around this problem by coding dominant interventions with a zero value. (Dakin, Devlin, & Odeyemi 2006) Again, however, this approach is questionable and requires manipulation of the data. Ultimately, *Cost-effectiveness* was coded as a categorical variable. Unfortunately, I was unable to include as many categories in *Cost-effectiveness* as I would have liked. The categories were chosen to ensure a sufficient number of positive and non-coverage decisions in each category. This necessitated interventions associated with an ICER >$50,000 per QALY to be pooled with interventions estimated to be dominated. This approach is not ideal as it includes interventions that are more effective than their

comparator (ICER >$50,000 per QALY) in the same category as interventions that are less effective than their comparator (dominated). It is expected that as the available sample increases, more categories will be included. As a result, pooling interventions with an ICER of >$50,000 per QALY with those that are dominated can be avoided.

Finally, when evaluating cost-effectiveness evidence, it is important to consider uncertainty in the estimate. (Claxton, 2008) As described in 4.7.4.9, I reported estimates of uncertainty when reviewing the studies. However, given the inconsistent nature of reporting estimates of uncertainty, it proved infeasible to include here.

6.4.4. Policy relevance

As far as I am aware, this is the first study of its kind to quantitatively evaluate CMS NCDs. This study is particularly important given the uncertainty surrounding the interpretation of the 'reasonable and necessary' criterion and, thus, CMS coverage of medical technology. Recent NCDs, e.g., autologous cellular immunotherapy treatment of metastatic prostate cancer, serve to underline the importance of CMS NCDs and the emotion and debate that surrounds them. (Chambers & Neumann 2011)

The findings of the analyses presented above provide insight into CMS's decision-making process and the factors important in CMS decision-making. First, they underscore that CMS has adopted evidence-based medicine, with interventions associated with good quality evidence several times more likely to be covered than those associated with insufficient evidence. Second, the findings highlight the importance of the availability of alternative interventions at the time of NCDs, with interventions with an available alternative much less likely to be covered than those without an alternative available. This may provide an insight into how CMS considers the "necessary" component of the 'reasonable and necessary' criterion. Third, the findings suggest that the availability of cost-effectiveness evidence plays a role in CMS coverage. However, as the categories *<$50,000 per QALY* and *>$50,000 per QALY* were not statistically significant, the findings are insufficient to conclude that CMS NCDs are consistent with cost-effectiveness. If

CMS coverage decisions were consistent with cost-effectiveness, one would expect that these categories would be significant, and the respective ORs would reflect a decreased likelihood that interventions associated with higher ICERs are associated with positive coverage. Nevertheless, when controlling for other factors, the absence of an associated estimate of cost-effectiveness, when compared to instances when a dominant estimate is available, reduces the likelihood of coverage by approximately a factor of five. While preference for the coverage of dominant interventions is intuitive, the finding is contradictory to CMS's stated position, that "*Cost effectiveness is not a factor CMS considers in making NCDs. In other words, the cost of a particular technology is not relevant in the determination of whether the technology improves health outcomes or should be covered for the Medicare population through an NCD.*" (CMS 2010e) Finally, the findings suggest that, when controlling for the other factors in the model, CMS has become more restrictive over time with respect to the coverage of interventions.

Studies like this have the potential to increase the transparency of coverage decisions and increase the accountability of CMS. This study has the potential to help the entire medical community better understand the evidence that CMS considers, thus reducing uncertainty associated with NCDs. The findings of this research go some way to reveal CMS's interpretation of the 'reasonable and necessary' criterion.

6.4.5. Next steps

While the findings of this research provide important insights, there is much room for improvement with respect to the data and methodological approach. Expanding the dataset to include NCDs made after 2007 will increase the sample size and likely improve results. In particular, this may be the case for the cost-effectiveness variable as a number of NCDs have featured cost-effectiveness evidence since 2007 (see Section 3.5.2 for further discussion). Adding a number of variables to the model will also likely prove beneficial.

As noted above, a limitation of the quality of evidence variable is its subjectivity. An objective review of the supporting evidence could be achieved by categorising the evidence base using a

number of criteria pertaining to: study design (e.g., randomised studies, non-randomised study, retrospective study, etc); study outcomes ('hard' endpoints vs. surrogate endpoints); inclusion of active comparators; consistency of findings across studies; patient population (e.g., whether the study included Medicare beneficiaries); country of study (e.g., US-based vs. non-US-based); and recency of study publication.

The availability of alternative interventions proved to be a significant predictor of the final coverage decision. While it was included here as a binary variable, it is possible that the likelihood of a positive coverage decision decreases for interventions with multiple available alternatives. Unfortunately, information presented in the decision memos was insufficient to code the variable in this manner. However, it may be possible to develop the variable along these lines with input from health care practitioners.

As described in Section 4.7.4.12, the inclusion of a number of variables proved difficult, and they were ultimately excluded from this research. Variables pertaining to budget impact and prevalence would be particularly useful to include, as both would give an insight into the impact of the coverage decision on the Medicare programme, something that is currently not accounted for in the model. Again, one approach to help develop these variables would be to gain input from clinicians or other health services researchers.

It has been suggested that social values, e.g., disease severity and equitable access to care, should play a role in health care resource allocation and therefore should be accounted for in coverage decisions. (Dolan et al. 2005) Severity of disease could be included by gaining input from health care practitioners to categorise diseases based upon whether they are life threatening. Another approach would be to use utility weights as a proxy to disease severity.

Also, although challenging to source the information, it would be useful to include the extent of lobbying to CMS in support of a positive coverage decision that occurred throughout the NCD time period. A potential proxy for this would be to count the number of comments submitted to

CMS during the NCD's comment period. Although a somewhat tenuous link to lobbying, this approach would at least account for the amount of public input into the decision.

As described in Section 6.2.3, a 'production function' formed the conceptual framework for this research. This approach was deemed appropriate as CMS provides little guidance on their decision-making criteria. As noted above, alternative frameworks could be considered, including the use of hierarchical models. When decision-making is viewed as a hierarchical process, multi-level models can used to reflect the decision-making process. For example, if hypothesised that clinical evidence is the most important aspect of decision-making, this can be reflected in the model structure. Interviews with staff from the CMS coverage group may offer additional insight into how CMS prioritises different decision-making criteria and thus help inform different modelling approaches.

The objective of a number of the studies reviewed in Section 4.5.1 was to identify an implicit cost-effectiveness threshold from coverage decisions. It is important to note, however, that these studies evaluated decisions by agencies for which cost-effectiveness evidence played an established role. While CMS do not routinely use cost-effectiveness evidence, a similar evaluation could be performed here. However, the lack of a sufficient number of non-coverage decisions associated with high ICERs inhibits such an analysis with the current data. Nevertheless, such an evaluation may be feasible as the sample of cost-effectiveness studies grows.

As described above, I used cluster analysis to control for the fact that coverage decisions made in the same NCD may not be independent because they are made by the same committee and use similar or related evidence (Section 6.3.2). A similar approach could be used to group coverage decisions that are reconsiderations together with the initial NCD. In the current dataset, reconsiderations of previous coverage decisions were considered independent. However, as the evidence reviewed for reconsiderations is likely related to the evidence reviewed for the preceding NCD, it would be interesting to control for this using cluster analysis. It is important to note that reconsiderations do not typically concern an identical coverage decision to the preceding

NCD; rather, they result in coverage determinations for different indications and/or patient populations. Thus, simply including the most recent NCD would exclude a number of unique coverage decisions.

6.5. Chapter summary

In Chapter 5, I identified occasions when CMS had cited or discussed cost-effectiveness evidence in the decision memo and showed a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to their cost-effectiveness. This suggested that covered interventions tend to be associated with more favourable cost-effectiveness evidence. The research presented in this chapter builds on these findings. The objective was to determine if cost-effectiveness is an independent predictor of coverage, i.e., when controlling for other factors, cost-effectiveness, or the availability of cost-effectiveness evidence, is statistically significantly associated with coverage.

A number of variables were included in the model, including; *Quality of evidence*, *Alternative intervention*, *Cost-effectiveness*, *Type of intervention*, *Coverage requestor*, and *Date*. The model was estimated using binomial logistic regression, regressing the coverage decision (positive/non-coverage) against the independent variables. Univariate and multivariate regressions were performed.

Across the primary analyses, the quality of supporting clinical evidence, availability of alternative interventions, and the date of decision were statistically significantly associated with the coverage decision. Although the cost-effectiveness variable was not significant overall, the category *No estimate* was. Key findings are summarised below:

- Interventions associated with good quality supporting evidence were six times more likely to receive a positive coverage decision compared to those associated with insufficient evidence (approximately twice as likely when considering predicted probabilities);
- Compared to interventions with no available alternative, those with an available alternative were approximately eight times less likely to be associated with a positive coverage decision (approaching half as likely when considering predicted probabilities);
- Compared with interventions estimated to be dominant, those with no associated estimate of cost-effectiveness were approximately five times less likely to receive a positive

coverage decision (approximately two thirds as likely when considering predicted probabilities);

- Coverage decisions made in 2006-2007 were approximately 10 times less likely to be associated with a positive coverage decision than those made in 1999-2001 (half as likely when considering predicted probabilities). Interventions considered in more recent time periods were increasingly less likely to be associated with a positive coverage decision.

While the findings are insufficient to conclude that CMS coverage decisions are consistent with cost-effectiveness, they are notable given CMS's stated position on the use of cost-effectiveness evidence. As the available sample increases, it will be interesting to re-evaluate the data to gain a greater insight into the association of cost-effectiveness with coverage decisions made in NCDs. This research provides insight into CMS's interpretation of the 'reasonable and necessary' criterion. The findings suggest that CMS operate an evidence-based coverage policy and that the availability of alternatives is relevant to decision-making. It is interesting that a trend was identified that showed that CMS became more restrictive with respect to coverage over the time period considered. While the research does not provide insight into why this is the case, it may reflect Medicare's ever greater fiscal challenges or the changing composition of CMS's coverage team.

The research presented in Chapter 7 is the final piece of empirical work and builds on the research presented in Chapter 5 and Chapter 6. The findings of the research presented in Chapter 5 illustrate that CMS are covering interventions that are not cost-effective by traditional standards, e.g., nine with ICERs greater than $100,000 per QALY and three with ICERs greater than $500,000 per QALY. Offering these interventions generates relatively little health gain for the expenditure and suggests that resources could provide greater benefits if directed towards alternative interventions. The research in this chapter support this finding and shows that CMS coverage decisions are not entirely consistent with cost-effectiveness. The primary objective of the research in Chapter 7 is to estimate potential gains in aggregate health from a hypothetical reallocation of expenditures between interventions subject to NCDs, using a criterion of cost-effectiveness. Also, the impact of using a cost-effectiveness decision rule to hypothetically

reallocate expenditures between interventions subject to NCDs on the distribution of resources among disease areas and types of intervention is evaluated.

7. Empirical Research: Part 3

7.1. Introduction

In Chapter 4, I provided the foundation for the empirical aspects of this thesis, including a background to the Medicare programme and the processes for the coverage of medical technology. Also, I presented the objectives of the empirical work and reviewed the relevant literature that helped inform my methodological approach. Lastly, I described the development of the database used for this research and the included set of variables.

In Chapters 5 and 6, I evaluated coverage decisions made in CMS national coverage determinations (NCDs). In Chapter 5, I found that on occasion CMS have discussed or cited cost-effectiveness evidence in NCDs. Also, while findings show a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness, suggesting that covered interventions tend to be associated with more favourable cost-effectiveness evidence, interventions have been covered in NCDs that are not cost-effective by traditional standards. Indeed, seventeen covered interventions are associated with an ICER greater than $50,000 per QALY, nine of which greater than $100,000 per QALY and three of which greater than $500,000 per QALY. The research presented in Chapter 6 built on these findings and evaluated whether, when controlling for other factors likely to influence coverage decisions, cost-effectiveness, or the availability of cost-effectiveness evidence, is an independent predictor of coverage outcome. The findings showed that compared with interventions estimated to be dominant, those with no associated cost-effectiveness estimate were approximately five times less likely to be covered. However, as the categories including ICERs with positive values were not statistically significant, it was not possible to conclude that CMS coverage decisions were consistent with cost-effectiveness, i.e., the likelihood of coverage decreased with higher estimates of cost-effectiveness. Other independent variables included those concerning the quality of the supporting clinical evidence, availability of alternative interventions, and the date of decision.

The research presented in this chapter represents the final piece of my empirical work. As described in Chapter 3, the US health care system is under financial strain, with health care costs rising at an unsustainable rate. Spending on Medicare represents a large proportion of total US

health care spending, approximately 5% of GDP, with projections suggesting an increase to 5.9% by 2017. (Orszag & Ellis 2007) It is claimed that Medicare is underperforming, with approximately 30% of administered care either inappropriate or unnecessary. (Bentley et al. 2008;Fisher et al. 2003;Garber, Goldman, & Jena 2007;Orszag 2008)

The primary objective of the research presented in this chapter is to estimate potential gains in aggregate health from a hypothetical reallocation of expenditures between interventions subject to NCDs using a cost-effectiveness decision rule. In addition, the impact of this hypothetical reallocation on the distribution of resources across disease areas and types of intervention is evaluated.

7.2. Objective and Methods

The empirical work presented in this chapter has two specific objectives:

1. To estimate potential gains in aggregate health from a hypothetical reallocation of expenditures using a cost-effectiveness decision rule between interventions subject to NCDs in terms of the net present value of future commitments and expenditures associated with interventions in the year following first use.
2. To estimate the impact of reallocation on the distribution of resources across disease areas and types of intervention.

7.2.1. Sample of interventions

The database used for this research is described in Section 4.7.5. For each coverage decision the following information were required:

1. Estimate of cost-effectiveness, including incremental cost and incremental effectiveness data;
2. Estimate of the cost of the intervention and comparator in the year following first use;
3. Estimate of the existing utilisation rate (served population), i.e., the number of Medicare beneficiaries using the intervention in 2007;
4. Estimate of the number of Medicare beneficiaries eligible for the intervention in 2007.

I reviewed each cost-effectiveness study to extract reported incremental cost and incremental effectiveness data. With respect to incremental effectiveness, incremental QALY gained data was used when available. Incremental life-years gained data were included by adjusting incremental survival gain with a reported utility weight for Americans aged 65-69. (Erickson, Wilson, & Shannon 1995) Cost-effectiveness studies that estimated the intervention to be dominant and reported health outcome using disease-specific units were included. While the inclusion of these studies did not provide incremental health gain data in an appropriate form, the reported estimates of cost-savings were relevant to this work. Separate analyses were performed when including and excluding these studies.

I obtained estimates of the costs associated with the interventions and comparators in the year following first use from the cost-effectiveness study when available. When not reported, I estimated the values from Medicare reimbursement codes (Section 4.7.5.6).

I estimated inpatient and outpatient utilisation rates from a Medicare claims database that included data from a 5% sample of the Medicare population. I used ICD-9 diagnostic codes to identify beneficiaries eligible for the intervention as specified by the parameters in the NCD, and CPT codes and ICD-9 procedural codes to identify beneficiaries who had received the intervention in the dataset. I estimated the existing utilisation rate by identifying beneficiaries who were a match for both ICD-9 codes and CPT codes. I calculated the size of the unserved patient population, i.e., Medicare beneficiaries who were eligible for the intervention but did not receive it, as the difference between the two estimates (Section 4.7.5.5). For non-coverage decisions, I assumed existing utilisation of the intervention to be zero in all cases. I estimated the size of the potentially eligible patient population, i.e., the population for which the non-coverage decision was made, using ICD-9 codes. A more complete description of the database used for this research is presented in Section 4.7.5.

Only interventions for which all the required data were available were included in this research.

7.2.2. Analytic approach

I performed a literature search and review to identify studies with a similar objective to the research considered here (Section 4.5.2). While two studies, Cromwell et al. (1998) and Zaric and Brandeau (2001), appeared to be relevant, neither provided a suitable framework with which to perform this research. Cromwell et al. (1998) used integer programming to allocate resources across acute inpatient services in an Australian setting. The limited scope of this research – it focused solely on inpatient services – facilitated the authors' approach. Unfortunately, the scope of my research, i.e., the inclusion of interventions indicated for a much broader range of indications, prevented me from using a similar framework. Zaric and Brandeau (2001) used a

dynamic model to estimate optimal resource allocation for HIV prevention in a hypothetical cohort of injection drug users (IDUs) and non-IDUs. A dynamic model was used to account for the infectious nature of the disease. A similar approach was not applicable to my research. Also, the focus of the study was narrow, and although the objective of estimating optimal resource allocation was similar to the objective considered here, the limited number of alternative interventions (three) reduced the applicability of the methods to this research.

7.2.2.1 Objective 1

The first research objective was to estimate potential gains in aggregate health from a hypothetical reallocation of expenditures using a cost-effectiveness decision rule between interventions subject to NCDs in terms of the net present value of future commitments and expenditures associated with interventions in the year following first use. To achieve this objective, I ranked the interventions in order of cost-effectiveness and simulated disinvestment/increased investment by adjusting intervention utilisation rates.

It was necessary to make a number of assumptions in this analysis:

1. The comparator included in the cost-effectiveness study was the only true alternative to the intervention, i.e., in all cases, beneficiaries not receiving the intervention would instead receive the study comparator.
2. The following assumptions consistent with a league table approach (Section 2.5.1.1). (Johannesson & Weinstein 1993)
    a. Perfect divisibility, i.e., a health care programme can be partially implemented and still maintain the characteristics of the entire programme.
    b. Constant returns to scale, i.e., costs and effects are proportional to the scale of implementation.
3. When considering the net present value of future commitments, unrestricted finance is available.
4. Supply of organs is not a limited factor in the delivery of transplant related interventions.

Included were all non-dominated interventions with complete data. Reallocations were required to be expenditure-neutral, i.e., no net change in total expenditure was permitted. Also, all eligible beneficiaries must have received either the intervention or the comparator.

A worked example of the reallocation using a simplified scenario is shown below. For simplicity, the worked example includes four interventions (A, B, C, and D) and costs are presented in terms of the net present value of future commitments. Table 34 presents the state of the world prior to the reallocation of expenditures. The ICER (cost-per QALY gained), the existing utilisation level, the total number of eligible beneficiaries, the number of unserved eligible beneficiaries, the incremental cost and QALY gain associated with the intervention, and the aggregate incremental cost and QALY gain across the population, is included for each intervention in Table 34.

Table 34. Worked example – Existing distribution of expenditures across available interventions

Intervention	ICER ($/QALY)	Existing utilisation	Eligible beneficiaries	Unserved eligible beneficiaries	Inc. Cost	Inc. QALYs	Total inc. cost	Total inc. QALYs
A	250,000	100	250	150	25,000	0.10	2,500,000	10
B	80,000	200	350	150	20,000	0.25	4,000,000	50
C	30,000	50	400	350	15,000	0.50	750,000	25
D	25,000	100	200	100	20,000	0.80	2,000,000	80
		450					**9,250,000**	**165**

Step 1 - Decrease utilisation of least cost-effective intervention (Intervention A)

The first step is to disinvest in the least cost-effective intervention available, in this case intervention A (ICER = $250,000 per QALY). This is achieved by decreasing the existing utilisation of the intervention by 50%. In this case the utilisation of intervention A is reduced from 100 to 50. The consequences of this change are shown by the highlighted text in Table 35.

Table 35. Worked example – Step 1

Intervention	ICER ($/QALY)	Existing utilisation	Eligible beneficiaries	Unserved eligible beneficiaries	Inc. Cost	Inc. QALYs	Total inc.cost	Total inc. QALYs
A	**250,000**	**50**	**250**	**200**	**25,000**	**0.10**	**1,250,000**	**5**
B	80,000	200	350	150	20,000	0.25	4,000,000	50
C	30,000	50	400	350	15,000	0.50	750,000	25
D	25,000	100	200	100	20,000	0.80	2,000,000	80
		400					**8,000,000**	**160**

Step 2 - Increase utilisation of most cost-effective intervention (Intervention D)

The second step is to increase investment of the most cost-effective intervention available, in this case intervention D (ICER = $25,000 per QALY). This is achieved by decreasing the size of the unserved patient population by 50%. In this case the number of unserved eligible beneficiaries was decreased from 100 to 50, with utilisation of intervention D increasing from 100 to 150. The consequences of this change are shown by the highlighted text in Table 36.

Table 36. Worked example – Step 2

Intervention	ICER ($/QALY)	Existing utilisation	Eligible beneficiaries	Unserved eligible beneficiaries	Inc. Cost	Inc. QALYs	Total inc.cost	Total inc. QALYs
A	250,000	50	250	200	25,000	0.10	1,250,000	5
B	80,000	200	350	150	20,000	0.25	4,000,000	50
C	30,000	50	400	350	15,000	0.50	750,000	25
D	**25,000**	**150**	**200**	**50**	**20,000**	**0.80**	**3,000,000**	**120**
		450					**9,000,000**	**200**

Step 3 - Decrease utilisation of next least cost-effective intervention (Intervention B)

The third step is to disinvest in the second least cost-effective intervention available, in this case intervention B (ICER = $80,000 per QALY). This is achieved by decreasing the existing utilisation of the intervention by 50%. In this case the utilisation of intervention B is reduced from 200 to 100. The consequences of this change are shown by the highlighted text in Table 37.

Table 37. Worked example – Step 3

Intervention	ICER ($/QALY)	Existing utilisation	Eligible beneficiaries	Unserved eligible beneficiaries	Inc. Cost	Inc. QALYs	Total inc.cost	Total inc. QALYs
A	250,000	50	250	200	25,000	0.10	1,250,000	5
B	**80,000**	**100**	**350**	**250**	**20,000**	**0.25**	**2,000,000**	**25**
C	30,000	50	400	350	15,000	0.50	750,000	25
D	25,000	150	200	50	20,000	0.80	3,000,000	120
		350					**7,000,000**	**175**

Step 4 - Increase utilisation of next most cost-effective intervention (Intervention C) to achieve expenditure neutrality

The fourth and final step is to increase utilisation of the second most cost-effective intervention in order to achieve expenditure neutrality, in this case intervention C (ICER = $30,000 per QALY). Expenditure neutrality was achieved by decreasing the size of the unserved patient population from 350 to 200, a decrease of approximately 43%. Net change in expenditure is zero (total incremental cost of $9,250,000) with a gain in aggregate health of 85 QALYs (Table 38).

Table 38. Worked example – Step 4

Intervention	ICER ($/QALY)	Existing utilisation	Eligible beneficiaries	Unserved eligible beneficiaries	Inc. Cost	Inc. QALYs	Total inc.cost	Total inc. QALYs
A	250,000	50	250	200	25,000	0.10	1,250,000	5
B	80,000	100	350	250	20,000	0.25	2,000,000	25
C	**30,000**	**200**	**400**	**200**	**15,000**	**0.50**	**3,000,000**	**100**
D	25,000	150	200	50	20,000	0.80	3,000,000	120
		500					**9,250,000**	**250**

The worked example illustrates the stepwise approach. The process was continued until there was no further opportunity for reallocation between interventions.

For each hypothetical reallocation, I present the aggregate health gain in terms of QALYs and the number of additional beneficiaries receiving the most effective treatment option, i.e., of the pair

of treatments available, those that received the intervention as opposed to the comparator[xvii]. In the worked example, 50 additional beneficiaries received the most effective available treatment option following the reallocation of expenditures. I also report the ICER of the marginal intervention, i.e., the intervention with the highest ICER for which utilisation was increased. Lastly, I report the average incremental QALY gain per beneficiary affected by the reallocation, i.e., the average incremental QALY gain per beneficiary for whom the reallocation changed the intervention they received.

In the worked example, the reallocation of expenditures is illustrated using the net present value of future commitments. An identical process was used when considering expenditures in the year following first use of the intervention. This more restrictive analysis considered the difference in cost between the intervention and comparator in the first 12 months of use.

Also reported are the findings of an additional analysis when the utilisation of dominant interventions was increased while maintaining the existing utilisation of non-dominant interventions, i.e., those associated with a positive ICER. As dominant interventions are more effective and less costly than their comparator, this hypothetical reallocation results in aggregate health gains and cost-savings.

From a practical standpoint, I concluded it was infeasible to increase utilisation of the intervention to 100%. Despite Medicare covering a particular intervention, patients may be reluctant to receive it. Also, it may often be the case that physicians are reluctant to change their approach to care and will be resistant to offering a different intervention. Cost-effectiveness studies report the cost-effectiveness of an intervention when used for the average patient in a population. However, populations are often likely to be heterogeneous, with interventions of high value for some patients but of low value for others. Therefore, in the base-case analysis a maximum of a 50% shift in patients between competing interventions was allowed, with a range of 10%-90% reported. For example, if 400,000 beneficiaries were eligible for a relatively high value

[xvii] N.B. As noted, an assumption was that each eligible beneficiary received care. If the beneficiary did not receive the intervention, it was assumed they received the comparator.

intervention, and 100,000 received it, it was assumed that further investment could increase utilisation by 50% of the 300,000 eligible beneficiaries not currently receiving it, i.e., 150,000 beneficiaries. If 200,000 beneficiaries received a relatively low value service, I assumed that utilisation could only be reduced by 50% (10% to 90% range), i.e., to 100,000 beneficiaries.

Hypothetical reallocations were performed using Microsoft Excel.

7.2.2.2 Objective 2

The second research objective was to estimate the impact of the hypothetical reallocation of expenditures between interventions on the distribution of expenditures across disease areas and types of intervention. As described in Section 4.7.5.7, interventions were characterised with respect to disease area, magnitude of incremental health gain, type of intervention, size of patient population, cost of intervention, and potential budget impact. To determine the impact of the reallocation on the distribution of expenditures between the aforementioned categories, I made a comparison between the distribution of expenditures prior to and following the hypothetical reallocation.

7.2.3. Datasets

I evaluated two datasets. First, I present the results when using a dataset limited to covered interventions, i.e., those subject to positive coverage decisions (Table 40). Second, I present the results when using a dataset including interventions subject to either positive or non-coverage decisions (Table 41). Results are presented separately when including interventions irrespective of the unit of health gain, and when including only interventions for which an estimate of incremental QALY gain was available.

7.3. Results

Figure 21 shows the process taken to arrive at the final sample of interventions. Only interventions with complete data were included. Thirty-six of the 64 interventions associated with an estimate of cost-effectiveness were included in the final sample (Table 39). Twenty-six of the 28 excluded interventions were excluded due to incomplete data, for which the most common reason was the inability to accurately identify the utilisation rate for the intervention in the indicated patient population. For example, in the NCD for ultrasound stimulation for non-union fracture healing, coverage of the intervention was restricted to patients with non-union bone fractures. (CMS 2005f) However, despite identifying an ICD-9 code for non-union fracture healing (733.82) and for fractures of the relevant bones (tibial [823]; scaphoid [814]; radius [813]), it was not possible to identify beneficiaries with a combination of both codes in the Medicare claims database. For other interventions, the appropriate patient population was unidentifiable in the Medicare claims database, an example being the NCD for intravenous immune globulin for autoimmune mucocutaneous blistering diseases. (CMS 2002a) The appropriate patient population was identified for two of the three indications for which the intervention received a positive coverage decision (bullous pemphigoid and pemphigus vulgaris). For the third indication, pemphigus foliaceus, the Medicare claims database did not include beneficiaries that suffered from this condition (ICD-9 694.5) and thus this diagnosis and treatment combination was excluded from the final dataset.

Two interventions were excluded as they were dominated by another intervention in the sample. The first was breast biopsy using stereotactic core needle image guidance, an intervention estimated to be dominated by ultrasound image guidance for the same indication. (Liberman et al. 1998) The second was screening immunoassay fecal-occult blood test (iFOBT) for colorectal cancer screening. The iFOBT test was extendedly dominated by a second screening immunoassay fecal-occult blood test, Hemoccult II, and was thus excluded from the sample included here. (AHRQ 2003)

Table 39. Interventions eligible for reallocation

Patient Dyad Characteristics		Utilisation			Cost-effectiveness			Costs in year following first use		
Intervention	**Population**	**Eligible patient population**	**Received Tx for diagnosis**	**Eligible but did NOT receive treatment**	**Inc. cost**	**Inc. QALY**	**ICER**	**Cost of intervention**	**Cost of comparator**	**Cost difference in yr1**
Ventricular Assist Devices	Destination Therapy - Chronic end-stage heart failure patients that meet specified criteria	1,474,420	20	1,474,400	$416,545	0.42	$986,630	$331,878	$65,177	$266,701
Transmyocardial revascularisation	Patients with severe angina (stable or unstable), which has been found refractory to standard medical therapy.	143,180	40	143,140	$19,777	0.04	$489,417	$18,123	$4,086	$14,037
Liver transplantation	Patients suffering from hepatitis B	14,320	40	14,280	$150,967	0.74	$204,186	$117,624	$8,558	$109,066
Ocular Photodynamic Therapy with Verteporfin	Macular Degeneration - Predominately classic subfoveal CNV lesions	73,400	1,200	72,200	$14,504	0.03	$195,566	$9,570	$0	$9,570
Lung Volume Reduction Surgery	Severe upper lobe emphysema	109,180	120	109,060	$60,243	0.50	$120,460	$87,905	$28,727	$59,178
Implantable Cardioverter Defibrillators (ICDs)	Patients with documented familial or inherited conditions with a high risk of life-threatening ventricular tachyarrhythmias	1,305,060	28,180	1,276,880	$21,102	0.16*	$99,782	$92,783	$65,846	$26,937
Pancreas transplantation	Pancreas Transplants - Patients that meet the specified criteria (type 1 diabetes etc)	67,920	720	67,200	$198,351	2.20	$90,159	$227,788	$4,218	$223,570
Positron Emission Tomography	Esophageal Cancer	80,400	200	80,200	$5,598	0.07	$81,485	$4,192	$1,438	$2,755
Implantable Cardioverter Defibrillators (ICDs)	NIDCM, documented prior MI, Class II and III heart failure	3,240	0	3,240	$77,113	1.01	$76,244	$37,474	$7,090	$30,384
Deep Brain Stimulation	Parkinson's Disease	727,800	39,860	687,940	$47,121	0.72	$65,970	$53,853	$5,988	$47,864

Patient Dyad Characteristics		Utilisation			Cost-effectiveness			Costs in year following first use		
Intervention	**Population**	**Eligible patient population**	**Received Tx for diagnosis**	**Eligible but did NOT receive treatment**	**Inc. cost**	**Inc. QALY**	**ICER**	**Cost of intervention**	**Cost of comparator**	**Cost difference in yr1**
Implantable Cardioverter Defibrillators (ICDs)	Documented sustained ventricular tachyarrhythmia	959,060	28,040	931,020	$34,375	0.65*	$39,971	$101,310	$73,912	$27,398
Autologous Stem Cell Transplantation (AuSCT)	Patients suffering from Multiple Myeloma	1,600	80	1,520	$83,123	1.69*	$37,275	$2,396	$106	$2,289
Acupuncture	Osteoarthritis	744,860	0	744,860	$536	0.02	$20,383	$97	$0	$97
Lumbar Artificial Disc Replacement	Back pain	140,700	0	140,700	$7,625	0.39	$18,939	$25,986	$16,547	$9,439
Laparoscopic adjustable gastric banding (LAGB) - bariatric surgery	Treatment of Morbid Obesity	5,983,500	6,600	5,976,900	$8,100	0.45	$18,028	$3,366	$142	$3,224
Cochlear implantation	Post lingually hearing impaired patients	32,340	1,120	31,220	$41,520	3.80	$11,653	$26,748	$0	$26,748
Hyperbaric Oxygen Therapy	Hypoxic Wounds and Diabetic Wounds of the Lower Extremities - Diabetic Wounds of the Lower Extremities	1,240,600	43,800	1,196,800	$1,771	0.27	$6,649	$524	$0	$524
Electrical Bioimpedance for Cardiac Output Monitoring	Hypertension	1,429,060	0	1,429,060	$314	0.05	$6,408	$628	$515	$113
External Counterpulsation (ECP) Therapy	Various cardiac conditions	5,018,500	0	5,018,500	$820	0.26	$3,264	$5,343	$0	$5,343
Positron Emission Tomography	Head and Neck Cancers	576,000	800	575,200	$1,425	0.44	$3,224	$6,022	$4,597	$1,425
Screening Immunoassay Fecal-Occult Blood Test - Hemoccult II	Screening for colon cancer	533,200	56,400	476,800	$400	0.13*	$1,318	$5	$0	$5
Ultrasound image guidance	Breast cancer - Breast biopsy	1,986,600	49,600	1,937,000	-$358	NA	Dominates	$613	$972	-$358
Foot care	Diabetic Peripheral Neuropathy with Loss of Protective Sensation	473,600	400	473,200	-$386	0.05	Dominates	$207	$0	$207

Patient Dyad Characteristics		Utilisation			Cost-effectiveness			Costs in year following first use		
Intervention	**Population**	**Eligible patient population**	**Received Tx for diagnosis**	**Eligible but did NOT receive treatment**	**Inc. cost**	**Inc. QALY**	**ICER**	**Cost of intervention**	**Cost of comparator**	**Cost difference in yr1**
Cardiac rehabilitation programs	Acute Myocardial Infarction	200,200	46,400	153,800	-$470	0.60	Dominates	$69	$0	$69
Cardiac rehabilitation programs	Percutaneous Transluminal Coronary Angioplasty	631,400	152,400	479,000	-$470	0.60	Dominates	$69	$0	$69
Positron Emission Tomography (FDG)	Breast Cancer - Initial Staging of Axillary Lymph Nodes	1,257,240	0	1,257,240	$609	NA	Dominates	$901	$0	$901
Positron Emission Tomography	Lung Cancer (non-small cell)	838,400	3,000	835,400	-$698	NA	Dominates	$2,038	$2,736	-$698
Positron Emission Tomography (FDG)	Breast cancer - staging and restaging	1,951,200	2,400	1,948,800	-$759	NA	Dominates	$953	$0	$953
Ambulatory BP monitoring	White coat hypertension	250,800	1,800	249,000	-$915	NA	Dominates	$110	$14	$96
Positron Emission Tomography (FDG)	Colorectal Cancer	605,000	800	604,200	-$892	NA	Dominates	$2,038	$2,929	-$892
Positron Emission Tomography (FDG)	Melanoma	388,600	600	388,000	-$906	NA	Dominates	$2,038	$2,943	-$906
Cryosurgery Ablation	Primary treatment for clinically localised prostate cancer. (Stages T1-T3)	1,388,600	5,000	1,383,600	-$2,189	NA	Dominates	$6,017	$8,206	-$2,189
Positron Emission Tomography (FDG)	Ovarian cancer	230,500	0	230,500	-$3,467	NA	Dominates	$2,956	$0	$2,956
Warm-Up Wound Therapy aka Noncontact Normothermic Wound Therapy NNWT	Stage III and IV ulcers	1,119,120	0	1,119,120	-$14,706	0.12	Dominates	$5,753	$8,431	-$2,678
Intravenous Immune Globulin	Bullous Pemphigoid	8,400	200	8,200	-$157,773	NA	Dominates	$44,613	$105,321	-$60,708
Intravenous Immune Globulin	Pemphigus Vulgaris	3,600	200	3,400	-$217,840	NA	Dominates	$102,656	$165,777	-$63,121

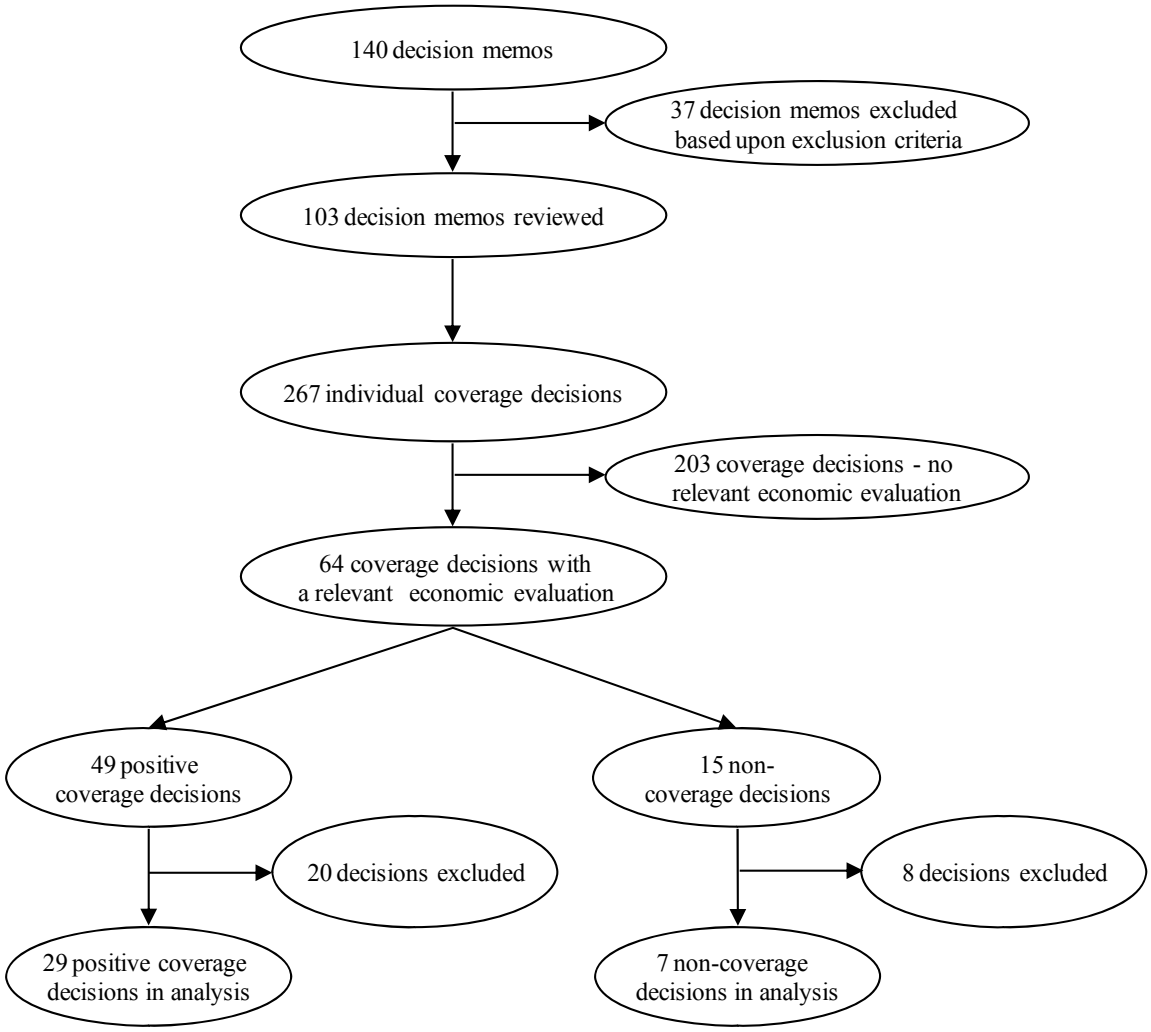


Figure 21. Process of identifying final set of coverage decisions for analysis

7.4. Reallocation including only positive coverage decisions

Prior to reallocation, 470,000 beneficiaries received one of the included interventions, at a cost of approximately $8 billion. Findings of the hypothetical reallocation are reported when adjusting utilisation rates by 50% (range; 10% – 90%).

First, a dataset including only positive coverage decisions was considered (Table 40). When considering the net present value of future commitments and expenditures in the year following first use of the intervention, reallocating expenditures to maximise aggregate health while maintaining a net total expenditure change of zero resulted in approximately an additional 5.85

(1.17 – 10.5) million and 6.13 (1.23 – 11.04) million beneficiaries receiving the most effective available intervention, respectively. This corresponded to gains in aggregate health of approximately 0.79 (0.16 – 1.42) million and 0.92 (0.18 – 1.65) million QALYs gained. Approximately 5.95 (1.19 – 10.72) million and 6.23 (1.25 – 11.22) million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively; corresponding to a per beneficiary gain of 0.13 QALYs and 0.15 QALYs. The ICER of the marginal technology, autologous stem cell transplantation (AuSCT), was $37,275 per life year saved when considering the net present value of future commitments and expenditures in the year following first use of the intervention.

When increasing the utilisation of dominant interventions, while maintaining the existing utilisation rate of interventions with positive ICERs, an additional 4.23 (0.85 – 7.62) million beneficiaries, approximately, received the dominant intervention (Table 40). This approach yielded savings of approximately $4.71 ($0.94 - $8.48) billion and $1.94 ($0.39 – $3.50) billion when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively. Gains in aggregate health of approximately 202,000 (40,000 – 363,000) QALYs were achieved. Approximately 4.23 (0.85 – 7.62) million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, equating to a gain of approximately 0.05 QALYs per beneficiary. Savings of $1,113 and $459 per beneficiary were achieved when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively.

Table 40. Reallocation including only positive coverage decisions

	Net present value of future commitments			Year following first use of intervention		
Reallocation	**Additional beneficiaries receiving care (50% [10-90%])**	**Cost savings (millions) (50% [10-90%])**	**QALY gain (50% [10-90%])**	**Additional beneficiaries receiving care (50% [10-90%])**	**Cost savings (millions) (50% [10-90%])**	**QALY gain (50% [10-90%])**
All interventions – irrespective of unit of health outcome						
Maintaining existing levels of expenditure	5,854,613 (1,170,923 - 10,538,304)	NA	790,392 (158,078 – 1,422,706)	6,133,469 (1,226,694 - 11,040,245)	NA	915,878 (183,176 - 1,648,580)
Increase in utilisation of dominant interventions	4,231,800 (846,360 - 7,617,240)	$4,709 ($942 - $8,476)	201,670 (40,334 - 363,006)	4,231,800 (846,360 - 7,617,240)	$1,943 ($389 - $3,498)	201,670 (40,334 - 363,006)
Only interventions with incremental QALY gain data						
Maintaining existing levels of expenditure	1,637,264 (327,453- 2,947,076)	NA	537,340 (107,468 - 967,211)	1,775,312 (365,802 - 3,292,221)	NA	634,332 (126,866 - 1,141,797)
Increase in utilisation of dominant interventions	553,000 (110,600 - 995,400)	$240 ($48 - $432)	201,670 (40,334 - 363,006)	553,000 (110,600- 995,400)	-$70.86 (-$14.17 – (-$128.56))	201,670 (40,334 - 363,006)

Table 40 also includes the findings of an analysis using a dataset restricted to interventions for which an estimate of incremental QALY gain was available[xviii]. Reallocating expenditures to maximise aggregate health while maintaining a net change in total expenditure of zero resulted in approximately an additional 1.64 million (330,000 – 2.95 million) and 1.78 (0.37 – 3.29) million beneficiaries receiving the most effective available intervention when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively. This corresponded to gains in aggregate health of approximately 0.54 (0.11 – 0.97) million and 0.63 (0.13 – 1.14) million QALYs. Approximately 1.74 (0.35 – 3.12) million and 2.56 (0.31 – 2.80) million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively; corresponding to a per beneficiary QALY gain of approximately 0.31 and 0.41. The ICER of the marginal technology was $11,653 per QALY (cochlear implantation for post lingually hearing impaired patients) and $18,028 per QALY (bariatric surgery [LAGB] for the treatment of morbid obesity) when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively.

When increasing the utilisation of dominant interventions, while maintaining the existing utilisation rates of interventions with positive ICERs, an additional 553,000 (111,000 – 995,000) beneficiaries, approximately, received the dominant intervention (Table 40). Cost-savings were estimated at approximately $240 ($48 - $432) million when considering the net present value of future commitments. When considering expenditures in the year following first use of the interventions, dominant interventions were associated with an increased expenditure of $70.86 ($14.17 - $128.56) million. This was because the included dominant interventions were associated with a positive expenditure in the year following their first use before yielding cost-savings in future years. The aggregate health gain associated with this reallocation was approximately 202,000 (40,000 – 363,000) QALYs. Per beneficiary incremental QALY gain was approximately 0.36 and per beneficiary savings approximately $434 when considering the net

[xviii] N.B. This dataset includes cost-effectiveness studies reporting cost-per life year gained estimates with incremental survival gain adjusted with a utility weight.

present value of future commitments. When considering expenditures in the year following first use of the intervention, there was an additional expenditure of $128 per beneficiary.

7.4.1. Reallocation including positive and non-coverage decisions

A second set of analyses used a dataset that included both positive and non-coverage decisions (Table 41).

Reallocating expenditures to maximise aggregate health while maintaining a net change in total expenditure of zero resulted in approximately an additional 11.12 (2.22 – 20.01) million and 6.73 (1.35 – 12.11) million beneficiaries receiving the most effective available intervention when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively. This corresponded to approximately 1.86 (0.37 – 3.36) million and 580,000 (116,000 – 1.00 million) QALYs gained. Approximately 11.22 (2.24 – 20.19) million and 6.88 (1.38 – 12.38) million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively; corresponding to a per beneficiary incremental QALY gain of approximately 0.17 and 0.08. The ICER of the marginal technology was $18,028 per QALY (bariatric surgery [LAGB] for the treatment of morbid obesity) and $3,264 per QALY (External Counterpulsation (ECP) Therapy) when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively.

When increasing the utilisation of dominant interventions, while maintaining existing utilisation rates of interventions with positive ICERs, an additional 5.54 million (1.11 – 9.96 million) beneficiaries, approximately, received the dominant intervention (Table 41). This approach yielded savings of approximately $12.95 ($2.59 – $23.32) billion and $2.54 ($0.51 – $4.56) billion when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively. Aggregate health gain was approximately 269,000 (54,000 – 484,000) QALYs. Approximately 5.54 (1.12 – 9.96) million

beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, equating to a per beneficiary QALY gain of approximately 0.05 QALYs. Per beneficiary savings of $2,340 and $458 were achieved when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively.

Table 41. Reallocation including positive and non-coverage decisions

	Net present value of future commitments			Year following first use of intervention		
Reallocation	**Additional beneficiaries receiving care (50% [10-90%])**	**Cost savings (millions) (50% [10-90%])**	**QALY gain (50% [10-90%])**	**Additional beneficiaries receiving care (50% [10-90%])**	**Cost savings (millions) (50% [10-90%])**	**QALY gain (50% [10-90%])**
All interventions – irrespective of unit of health outcome						
Maintaining existing levels of expenditure	11,118,104 (2,223,621 - 20,012,587	NA	1,863,736 (372,747 - 3,354,725)	6,729,994 (1,345,999 - 12,113,990)	NA	580,281 (116,056 - 1,044,506)
Increase in utilisation of dominant interventions	5,535,230 (1,107,046 - 9,963,414)	$12,954 ($2,591 - $23,318)	268,817 (53,763 - 483,871)	5,535,230 (1,107,046 - 9,963,414)	$2,535 ($507 - $4,563)	268,817 (53,763 - 483,871)
Including only interventions with QALY data						
Maintaining existing levels of expenditure	6,141,655 (1,228,331 - 11,054,979)	NA	1,614,536 (322,907 - 2,906,165)	2,100,071 (420,014 - 3,780,128)	NA	527,432 (105,486 - 949,377)
Increase in utilisation of dominant interventions	1,112,560 (222,512 - 2,002,608)	$8,469 ($1,694 - $15,244)	268,817 (53,763 - 483,871)	1,112,560 (222,512 - 2,002,608)	$1,428 ($286 - $2,570)	268,817 (53,763 - 483,871)

Table 41 also includes the findings of an analysis of a dataset restricted to interventions for which an estimate of incremental QALY gain was available.[xix] Reallocating expenditures to maximise aggregate health while maintaining a net change in total expenditure of zero resulted in approximately an additional 6.14 (1.23 – 11.05) million and 2.10 (0.42– 3.78) million beneficiaries receiving the most effective available intervention when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively. This corresponded to gains in aggregate health of approximately 1.61 (0.32 – 2.91) million QALYs and 530,000 (105,000 – 949,000) QALYs. Approximately 1.74 (0.35 – 3.12) million and 2.56 (0.31 – 2.80) million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively; corresponding to a per beneficiary incremental QALY gain of approximately 0.31 and 0.41. The ICER of the marginal technology was $11,653 per QALY (cochlear implantation for post lingually hearing impaired patients) and $18,028 per QALY (bariatric surgery [LAGB] for the treatment of morbid obesity) when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively. Approximately 6.24 (1.25 – 11.23) million and 2.20 (0.44 – 3.96) million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, when considering the net present value of future commitments and expenditures in the year following intervention implementation, respectively; corresponding to a per beneficiary QALY gain of approximately 0.26 and 0.24. The ICER of the marginal technology was $18,028 per QALY (bariatric surgery [LAGB] for the treatment of morbid obesity) and $3,264 per QALY (External Counterpulsation (ECP) Therapy) when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively.

When increasing the utilisation of dominant interventions, while maintaining the existing utilisation of interventions associated with a positive ICER, approximately 1.11 (0.22 – 2.00) million additional beneficiaries received care (Table 41). Cost-savings were approximately $8.47 ($1.69- $15.24) billion and $1.43 ($0.29 – $2.57) billion when considering the net present value

[xix] N.B. This dataset includes cost-effectiveness studies reporting cost-per life year gained estimates with incremental survival gain adjusted with a utility weight.

of future commitments and expenditures in the year following first use of the intervention, respectively. Aggregate health gain associated with this reallocation was approximately 269,000 (54,000 – 484,000) QALYs. Per beneficiary incremental QALY gain was approximately 0.24, and per beneficiary savings were approximately $7,600 and $1,300 when considering the net present value of future commitments and expenditures in the year following first use of the intervention, respectively.

7.4.2. Effect of reallocation on distribution of expenditures

Using different criteria to allocate resources will inevitably have an effect on the distribution of expenditures across interventions and patient populations. The effect of using a cost-effectiveness decision rule was evaluated by comparing the existing distribution of expenditures with the distribution following a 50% reallocation. In each case, expenditures were reallocated to maximise aggregate health while maintaining a net change in total expenditure of zero. To include the maximum number of interventions, the dataset that included positive decisions, non-coverage decisions, and dominant interventions irrespective of measure of health outcome, was used.

First, interventions were categorised into the following broad disease classifications: *cardiology, oncology,* and *other* (Section 4.7.5.7). Prior to reallocation, approximately 55% of beneficiaries in the dataset received a *cardiology*-related intervention, approximately 25% *oncology*-related, with the remaining 20% *other* (Table 42). Following reallocation, a greater proportion of beneficiaries received an *oncology*-related intervention (approximately 43%), with fewer receiving a *cardiology*-related intervention (approximately 34%) or *other* (approximately 24%).

Table 42. Effect of expenditure reallocation on distribution across broad disease classifications

Disease area	Prior to reallocation		Following reallocation	
	No. of beneficiaries	Distribution	No. of beneficiaries	Distribution
Cardiology	256,880	54.65%	3,893,420	33.60%
Oncology	118,880	25.29%	4,937,110	42.60%
Other	94,260	20.05%	2,757,594	23.80%
	470,020	100%	11,634,101	100%

Interventions were also categorised with respect to type. Categories included *Treatment, Diagnostic,* and *Other* (Section 4.7.5.7). Prior to reallocation, the majority of beneficiaries received an intervention categorised as *Other* (approximately 42%), with the remaining beneficiaries receiving interventions categorised as *Treatment* (approximately 33%) or *Diagnostic* (approximately 25%) (Table 43). Following reallocation there was an increase in the proportion of beneficiaries receiving an intervention categorised as *Treatment* (approximately 50%) or *Diagnostic* (approximately 44%). The proportion of beneficiaries receiving interventions categorised as *Other* decreased to approximately 7%.

Table 43. Effect of expenditure reallocation on distribution across different types of interventions

Type of intervention	Prior to reallocation		Following reallocation	
	No of beneficiaries	Distribution	No of beneficiaries	Distribution
Treatment	155,220	33.0%	5,754,824	49.7%
Diagnostic	115,600	24.6%	5,081,100	43.8%
Other	199,200	42.4%	752,200	6.5%
	470,020	100.00%	11,588,124	100.0%

N.B. Type of intervention "Other" includes health education, preventative care, and mobility assistive equipment

Interventions were categorised with respect to the size of the eligible patient population. Categories included *Large* (>1 million beneficiaries), *Medium* (50,000 – 1 million beneficiaries), and *Small* (<50,000 beneficiaries) (Section 4.7.5.7). Prior to reallocation the majority of

beneficiaries (approximately 71%) received an intervention categorised as *Medium*, with approximately 29% and 0.3% of beneficiaries receiving an intervention classified as *Large* and *Small,* respectively (Table 44). Following reallocation the majority of beneficiaries received an intervention categorised as *Large* (78%), with 21.8% and 0.2% of beneficiaries receiving an intervention that fell into the categories *Medium* and *Small,* respectively.

Table 44. Effect of expenditure reallocation on distribution across patient populations of different sizes

Size of untreated patient population	Prior to reallocation		Following reallocation	
	No of beneficiaries	Distribution	No of beneficiaries	Distribution
Large	135,600	28.8%	9,034,894	78.0%
Medium	332,780	70.8%	2,530,240	21.8%
Small	1,640	0.3%	22,990	0.2%
	470,020	100.00%	11,588,124	100.00%

The impact of resource reallocation at the patient level with respect to incremental health gains and costs was evaluated. For these analyses, the dataset that included only interventions with an associated estimate of incremental QALY gain was used. First, the impact of expenditure reallocation on the distribution of health gains across Medicare beneficiaries was evaluated. Prior to reallocation the estimated per beneficiary incremental QALY gain was 0.49 QALYs (95% CI 0.53 – 1.05), compared to 0.29 (95% CI 0.18 – 0.75) following reallocation.

The impact of expenditure reallocation on the average incremental cost per beneficiary was evaluated (Table 45). Prior to reallocation the estimated average per beneficiary incremental cost was $7,711 (95% CI -$28,143 – $43,563), compared to $6,728 (95% CI -$12,874 – $13,500) following reallocation.

The impact of expenditure reallocation on the average incremental cost per beneficiary in the year following first use of the intervention was evaluated (Table 45). Prior to reallocation the

estimated average per beneficiary incremental annual cost of interventions was $7,800 (95% CI -$26,700 - $42,200), compared to $1,600 (95% CI -$6,400 - $9,500) following reallocation.

Table 45. Effect of expenditure reallocation on distribution of resources at the patient level

Average impact per beneficiary	Prior to reallocation	Following reallocation
Incremental QALY gain	0.49 QALYs (95% CI 0.53 – 1.05)	0.29 QALYs (95% CI 0.18 – 0.75)
Incremental cost	$7,711 (95% CI -$28,143 - $43,563)	$6,728 (95% CI -$12,874 - $13,500)
Cost in year following first use of intervention	$7,800 (95% CI -$26,700 - $42,200)	$1,600 (95% CI -$6,400 - $9,470)

7.5. Discussion

This research is an ambitious attempt to estimate the impacts of using a cost-effectiveness decision rule to allocate Medicare resources on overall expenditures and health outcomes. To the best of my knowledge, this is the first research of its kind. While frameworks for efficient resource allocation exist, e.g., mathematical programming, inhibitive data requirements reduce their practicality. Studies considered in my literature review were of limited relevance to the methodological approach taken here and were limited in scope, i.e., they focused on a single indication or type of service. (Cromwell et al. 1998;Zaric & Brandeau 2001) This empirical work has two objectives: to estimate potential gains in aggregate health from reallocating expenditures using a cost-effectiveness decision rule in terms of the net present value of future commitments and expenditures in the 12 months following first use of an intervention; and, to estimate the impact of reallocation on the distribution of resources among disease areas and types of intervention.

7.5.1. Summary of findings – Gains in efficiency

The total cost of providing the included interventions in 2007 was estimated to be almost $8 billion, approximately 1.8% of Medicare's budget in that year (approximately $440 billion). Prior to reallocation, approximately 470,000 Medicare beneficiaries received one of the included interventions.

The first set of analyses used a dataset that included interventions subject to positive coverage decisions. When including all dominant interventions, i.e., irrespective of reported unit of health gain, reallocating expenditures (50% reallocation) while maintaining a net change in total expenditure of zero resulted in almost 6 million additional beneficiaries receiving the most effective option, with a corresponding gain in aggregate health of almost 800,000 QALYs. When using a dataset that included only interventions with an associated estimate of incremental QALY gain, findings remained substantial but were smaller in magnitude; 1.6 million additional beneficiaries received the most effective option, with an aggregate health gain of approximately 550,000 QALYs. Substantial gains were also achieved when increasing the utilisation (50%

decrease in unserved eligible population) of dominant interventions while maintaining existing utilisation rates of interventions associated with positive ICERs. Increasing the utilisation of the 12 included interventions estimated to be dominant allowed for approximately 4.2 million additional beneficiaries to receive the most effective option, with a corresponding gain in aggregate health of approximately 200,000 QALYs and savings of $4.8 billion.

When considering expenditures in the year following first use of an intervention, findings were broadly consistent with the findings when considering the net present value of future benefits with respect to the number of additional beneficiaries receiving the most effective option and the corresponding gains in aggregate health. This was because for the interventions considered here, the difference in expenditures between interventions and comparators when considering the 12 months following first use of an intervention were proportionally similar to when considering the net present value of future commitments. However, as the magnitude of the incremental cost in terms of net present value of future commitments was greater than the difference in expenditures between competing interventions in the year following first use of an intervention, cost-savings achieved when increasing utilisation of dominant interventions was greatest when considering the former ($4.8 vs. $1.9 billion).

Interestingly, when considering the dataset that included only interventions with an associated estimate of incremental QALY gain, increasing the utilisation of dominant interventions while maintaining existing utilisation rates for interventions associated with positive ICERs resulted in a positive net expenditure in the year following first use of an intervention ($71 million). This is because, despite being cost-saving when accounting for downstream costs, the included dominant interventions required a positive expenditure in the year following first use of an intervention.[xx]

The pattern of findings was qualitatively similar when including interventions associated with non-coverage decisions in the dataset. While CMS decided not to cover these interventions, they

[xx] These interventions include foot care for diabetic patients suffering from diabetic peripheral neuropathy (DPN) with loss of protective sensation, and cardiac rehabilitation programmes following acute myocardial infarction or percutaneous transluminal coronary angioplasty.

were included here to evaluate the impact of using cost-effectiveness as the only criterion for resource allocation. However, while the pattern of findings was similar, the magnitude of each reported statistic was greater (Table 41). This is because of the relatively favourable cost-effectiveness estimates associated with non-covered interventions. Three were estimated to be dominant, and the highest ICER of the remaining four was approximately $20,000 per QALY (acupuncture for osteoarthritis). Also, as these interventions were associated with large potential eligible patient populations, their inclusion had a large impact on the results of the analysis.

7.5.2. Summary of findings – Effect of reallocation on the distribution of resources

It is inevitable that changing resource allocation criteria will affect the distribution of resources. The impact of using a cost-effectiveness decision rule to reallocate resources on the distribution of expenditures across disease areas, intervention type, and size of eligible patient population was evaluated. The impact of reallocation at the patient level was also evaluated, with the average per patient incremental QALY gain, incremental cost, and cost in the year following first use of intervention considered.

Following reallocation, a greater proportion of resources was directed to oncology-related interventions, and a lesser proportion was directed to those related to cardiology and other diseases (Table 42). This may reflect the fact that many of the included oncology-related interventions were diagnostic imaging and tests rather than chemotherapies, which are generally accepted to be associated with high ICERs. (Greenberg et al. 2010) With respect to intervention type, a greater proportion of resources were directed to treatments and diagnostics as opposed to *'other'* (Table 43) following reallocation. This finding was expected as a number of the dominant interventions in the dataset were diagnostic imaging technologies and tests. When considering size of the eligible patient population, a greater proportion of resources were directed to the most prevalent diseases (greater than one million eligible beneficiaries) following reallocation (Table 44). This is principally due to the fact that dominant interventions and those with favourable estimates of cost-effectiveness were indicated for diseases of high prevalence and largely underutilised.

As the reallocation disproportionately increased the utilisation of some interventions associated with lower than average incremental QALY gains, the average per beneficiary incremental QALY gain decreased from 0.49 to 0.29 QALYs. Similarly, following reallocation the average per beneficiary incremental cost decreased from $7,700 to $6,700 when considering net present value and from $7,800 to $1,600 when considering the year following first use of the intervention.

The limited number of interventions included in the hypothetical reallocation makes it difficult to draw meaningful conclusions from the findings. The research demonstrates, however, that changing resource allocation criteria will likely affect the distribution of expenditures.

7.5.3. Limitations and challenges

Coverage decisions included in this research were identified in NCDs made from 1999 through 2007. As the minority of CMS's coverage decisions are made through the NCD pathway and only coverage decisions associated with a cost-effectiveness estimate are included, this research is limited to a relatively select group of interventions. Further, of the 64 coverage decisions associated with a cost-effectiveness estimate, complete information was available for only 36, with 28 ultimately excluded from the dataset. Due to the relatively small sample of interventions included, it is possible that the included interventions are unrepresentative of those offered in the Medicare programme as a whole. Conclusions that can be drawn from the findings for objective 2 are limited due to the small sample size. However, despite only including 36 coverage decisions, the findings illustrate that reallocating resources is likely to have an effect on the distribution of expenditures across diseases and types of treatment.

The data available for this analysis represented one of the largest challenges of this research. Ideally, the following data would have been available.

Cost data

Ideal cost data for this analysis would be specific to Medicare claims data and have specifications that would facilitate potential legislative action based upon study findings, i.e., meet the standards of the Congressional Budget Office (CBO) and CMS's Office of the Actuary. (CBO 2011;CMS 2011b) Cost data should account for all implications of the intervention and comparator's use and be available for each year through the relevant time horizon, i.e., year 1, year 2, year 3, etc.

Effectiveness data

Similarly, ideal effectiveness data for this research would be specific to the Medicare population, and would be reported in QALYs in order to facilitate comparison across conditions. Effectiveness data should be available for all competing interventions, and accrued QALYs should be reported for each year through the relevant time horizon, i.e., year 1, year 2, year 3, etc. Patient heterogeneity would be accounted for, with data available on the effectiveness of the intervention across the range of patients in the evaluated population.

Patient level information

Available patient-level information would be sufficient to allow accurate identification of beneficiaries that met the specifications of the coverage decision, i.e., ICD-9 codes would be reported comprehensively and would be supplemented with additional patient level data, including BMI, comorbidities, whether patient had failed alternative management approaches, etc.

Epidemiological data

Both prevalence and incidence data would be available, i.e., the total number of existing cases and the number of new cases each year.

If this data were available, the research presented here would have been more straightforward. Unfortunately, it was necessary to source the data from a number of different sources, each with its own limitations.

I relied heavily on the cost-effectiveness literature for this research. The process of identifying and selecting this literature is presented in Section 4.7.4.8. Despite reviewing each study to ensure it was an appropriate match to the coverage decision, it is inevitable that there is variation between the included cost-effectiveness studies with respect to quality, perspective, country setting, framework, approach to estimating utility, etc. This lack of consistency introduced uncertainty into the analysis, limiting the generalisability of the findings. Also, I assumed that the comparator in the cost-effectiveness study was the only available alternative intervention. While I ensured in each case that Medicare deemed the comparator to be appropriate (Section 4.7.4.9), it is likely that multiple alternative interventions were available. Further, it may be that in some cases no treatment is the most clinically appropriate course of action, a management approach not accounted for here.

When possible, cost data were extracted from the included cost-effectiveness study. Unfortunately, this was not possible for cost-effectiveness studies not performed in a US setting, and in these cases estimates were converted to US dollars and inflated to a 2007 valuation (Section 4.7.4.9). The cost-effectiveness studies often proved insufficient as the source of estimates of the cost of interventions in the 12 months following first use. In these cases, estimates were calculated from Medicare and physician reimbursement codes (Section 4.7.5.6).

The cost-effectiveness literature was also the source of the effectiveness data. Similar problems to those discussed for the cost data were seen here, i.e., a US study population was not always included, and there was variability with respect to how the data was calculated. For example, while the majority of studies reported QALYs, the approaches to estimating utility were not consistent. As a result, on four occasions, it was necessary to adjust reported incremental survival gain with a utility weight to gain an estimate of incremental QALY gain. A further limitation was

that I used the published cost-effectiveness study to infer the effectiveness of the intervention and comparator for the average Medicare beneficiary, not accounting for patient heterogeneity.

A fundamental part of this research was identifying both the number of Medicare beneficiaries currently receiving the intervention and the size of the unserved eligible patient population, i.e., beneficiaries eligible for the intervention in accordance with NCD specifications but who did not receive it. I used ICD-9 codes reported in a 5% Medicare claims database to identify these groups of beneficiaries. Although necessary for this research, basing eligibility solely on ICD-9 diagnostic codes was a crude approach. In clinical practice, a number of clinical factors not captured in ICD-9 diagnostic codes help guide management. These may include disease severity, the presence of co-morbidities, patient preference, etc. Also, factors related to disease management may influence coverage decisions, e.g., CMS may grant access to an intervention only after the failure of alternative treatment options. Finally, I relied on ICD-9 codes to be accurately reported in the Medicare claims database. Incomplete reporting of ICD-9 codes would affect the accuracy of the estimated number of unserved beneficiaries and those currently receiving the intervention.

A Medicare claims database including 2007 data was used here. It provided a 'snapshot' of the interventions received by Medicare beneficiaries over the one year period. It does not, however, distinguish between incident and prevalent cases. Without this information, determining whether a beneficiary is truly eligible for a particular intervention is challenging. For example, identifying a Medicare beneficiary with Parkinson's disease is insufficient to determine their eligibility for deep brain stimulation (DBS), as this treatment is reserved for use once pharmaceutical management is no longer effective. For other interventions, this is less problematic. For example, foot care would always be an appropriate intervention for diabetic patients suffering from diabetic peripheral neuropathy with loss of protective sensation.

Due to the crude approach for identifying eligible patients, it was deemed inappropriate to decrease utilisation of relatively low value interventions to zero, or to increase the utilisation of

relatively high value interventions to 100% of the eligible patient population. Therefore, I used a maximum of a 50% change in utilisation to reallocate expenditures; when decreasing expenditure on relatively cost-ineffective interventions, utilisation was decreased by 50%, and when increasing expenditure on relatively cost-effective interventions, the size of the unserved eligible patient population was decreased by 50%. The 50% value is arbitrary and, therefore, a range of 10% through 90% was also reported. As discussed below, the feasibility of such a shift in expenditures was not considered in this work.

7.5.4. Policy significance

This research is timely given the current fiscal challenges facing Medicare, and it asks important questions regarding the programme's efficiency. However, this research proved an ambitious task. Data originated from a number of sources, including a Medicare claims database for utilisation rates and the size of the eligible patient population; the cost-effectiveness literature for estimates of incremental costs and benefits; and Medicare sources for reimbursement data. The disparate nature of the data has implications for the accuracy of the estimates and, by extension, the policy significance of the study. In particular, the included cost data are not up to the standard used by the CBO or the CMS's Office of the Actuary, thus restricting the extent to which the findings presented here could be used for legislation action. Nevertheless, the findings illustrate the broader benefits of using cost-effectiveness evidence, and, rather than an accounting framework, this research should be considered a technical exercise that estimates potential efficiency gains within a feasible range.

Various findings of this research have particular policy relevance. A notable finding is the relatively infrequent use of interventions with the highest ICERs, with interventions associated with ICERs greater than $100,000 per QALY having negligible utilisation rates (Table 39). One potential reason for this may be related to reimbursement, i.e., the mode of reimbursement introduces financial incentives/disincentives that influences physician prescribing of these services. (Neumann, Rosen, & Weinstein 2005)

This research highlights the underutilisation of some dominant interventions, i.e., those that are health increasing and cost-saving. One intervention for which this was the case was cardiac rehabilitation for patients recovering from an acute myocardial infarction or percutaneous transluminal coronary angioplasty. Research suggests that despite documented evidence of the clinical and economic benefits of cardiac rehabilitation, it is an underutilised intervention. Suggested reasons for this underutilisation include a lack of referral by physicians, associated comorbidities, reimbursement factors, and perceived benefits of the intervention, among others. (Daly et al. 2002;Parkosewich 2008;Thomas 2007)

For a number of dominant interventions, positive expenditure was required in the 12 months following first use even though cost-savings were estimated over the time horizon of the cost-effectiveness analysis. This was the case for cardiac rehabilitation and foot care for diabetic patients with neuropathy, two important drivers of study findings (Table 40). This finding is important as it highlights that in some cases initial increased investment is required to yield cost-savings downstream. Policy makers should be aware of interventions for which this is the case and be prepared to prioritise resources accordingly, even if programme cost increases transiently in the year of use. Further, this finding emphasises the importance of accounting for future costs and benefits when evaluating interventions; consideration of interventions over the short-term may not adequately account for the potential positive financial impact on the health care system.

The apparent underutilisation of dominant interventions provides an opportunity for policy makers. Increasing utilisation of dominant interventions would have a real positive impact on the Medicare programme, not only increasing aggregate health, but also generating additional resources that can be invested in other aspects of the Medicare programme.

Few interventions were the principal drivers of efficiency gain in this analysis. Unsurprisingly, efficiency gains were largest for interventions affecting the largest number of Medicare beneficiaries. Low value interventions, i.e., those for which utilisation is reduced in expenditure reallocation, with high initial utilisation will have a large impact on results (e.g., deep brain

stimulation for Parkinson's disease and Implantable Cardioverter Defibrillators (ICDs) for documented ventricular tachyarrhythmia). Conversely, high value interventions that are associated with large unserved eligible patient populations, i.e., interventions with low penetration, have a large impact on results (e.g., cryosurgery for prostate cancer, image guidance for breast biopsy, and cardiac rehabilitation). Though to a lesser extent than utilisation rate, the magnitude of the incremental cost and incremental QALY gain associated with interventions also impacts the results. This finding suggests that interventions for which the largest potential gains are achievable should be targeted. However, it is not necessarily the case that this approach would be most advantageous. It would be unrealistic to expect that policy makers could impose disinvestment of deeply entrenched services without huge resistance. In practice, a more manageable approach may be to target interventions for which it would be more feasible to implement a change in therapeutic management.

Findings are presented in terms of the net present value of future commitments and year following first use of the intervention. The findings presented in terms of net present value of future commitments are illustrative of potential efficiency gains. However, they assume that unrestricted finance is available beyond the available annual budget to pay for interventions that yield cost savings in future years, a somewhat unrealistic assumption. When considering the 12 months following first use of the intervention, the findings are more conservative and account for the existing level of resources.

The findings illustrate that using different criteria to allocate resources will have an effect on the distribution of resources across types of treatment and disease areas. However, given the small number of included interventions, it is difficult to draw conclusions from the findings. The findings do, though, draw into focus the fact that trade-offs between decision-making criteria will have to be made when allocating resources. The objective of this research was to maximise aggregate health, though this may not be consistent with the preferences of the decision maker or society. This research is important in that it provides insight into the maximum amount of health gain achievable, and thus the health-related opportunity cost of using alternative criteria to guide health care resource allocation.

7.5.5. Next steps

Two broad alternative approaches can be taken to move this research forward. First, scope of the research could be expanded to include additional interventions from a broad range of conditions. This approach would increase the complexity of this research and would likely suffer from many of the limitations laid out above in Section 7.5.3. Second, the scope of the research could be reduced and focused on a more select group of interventions from a restricted set of conditions. Decreased model complexity may lend itself to different methodological approaches, e.g., integer programming. These differing approaches would be relevant to two very different research questions. The first approach would be consistent with the objective of the research described here, i.e., to estimate potential efficiency gains from using cost-effectiveness information to allocate resources in the Medicare programme. The second approach would have a narrower objective and would be specific to a single condition or a set of few conditions.

Regardless of the methodological approach, many steps could be taken to improve the model inputs. With respect to the included cost-effectiveness evidence, the inclusion of additional studies would allow for a more comprehensive analysis and one that accounts for a greater proportion of the Medicare programme. Also, with a sufficient sample size it may be possible to restrict cost-effectiveness studies included to those that both evaluate a Medicare population and include Medicare-specific costs and effectiveness data. This, however, would be challenging since few cost-effectiveness studies included here meet these strict requirements.

One of the challenges faced in this research was to identify the relevant patient population from a Medicare claims database. In the absence of additional fields within the Medicare claims data to help characterise patients, one approach to improve the estimation of the size of the eligible patient population would be to seek clinical input. For example, with input from physicians it would be possible to estimate the proportion of patients with a specific ICD-9 code that meet certain additional clinical criteria and thus the specifications of the NCD.

The small sample size prevented categorisation of interventions using as many intervention type or disease area categories as I would have liked. With a larger sample it would be possible to better categorise interventions and understand the implications of resource allocation using a cost-effectiveness decision rule. For example, as society has been shown to have a preference towards the treatment of severe diseases, a variable to capture disease severity would be valuable (Coast 2004;Dolan & Cookson 2000;Drummond et al. 2005;Nord et al. 1995;Ubel 2001) However, as described in Section 4.7.4.12, generating such a variable proved difficult and input from health care practitioners may required for its inclusion. By better understanding society's preferences for resource allocation, the impact of alternative patterns of resource use could be evaluated in light of them. Indeed, research to help evaluate whether the current distribution of expenditures is consistent with Medicare beneficiaries preferences would be valuable.

The current analytic approach is essentially deterministic, one that does not account for parameter uncertainty. An alternative approach would be to use a stochastic process to account for uncertainty in the parameter estimates. Such an approach would require inclusion of additional information, which proved challenging when considering uncertainty in cost-effectiveness estimates, as shown in Section 4.7.4.9. A stochastic process could be further developed to move towards a population model that accounts for disease incidence.

Only interventions subject to NCDs, which comprise a small proportion of all CMS coverage decisions, were included in this research. While NCDs are typically made for interventions deemed to have a significant impact on the Medicare programme, those considered here might not be those for which a reallocation of resources would yield the greatest efficiency gains. Further, it is likely that if the sole purpose of NCDs were to increase programme efficiency, different interventions would be considered. This research would benefit from including a broader range of interventions available in Medicare, not just those evaluated through the NCD pathway. Further, while interventions were included that were not subject to NCDs, i.e., those subject to LCDs, in the larger project conducted at Tufts Medical Center[xxi], the scope of the research did not

[xxi] See Acknowledgements

include the impact of reallocation on the distribution of resources or the cost of the interventions in the year of their first use.

7.6. Chapter summary

The research in chapters 5 and 6 evaluated the relationship between coverage decisions made as part of NCDs and cost-effectiveness evidence. The research presented in this chapter has a different set of objectives, i.e., to estimate potential gains in aggregate health from reallocating expenditures in terms of the net present value of future commitments and expenditures in the first 12 months of use, and to estimate the impact of reallocation on the distribution of resources across disease areas and types of intervention. Using the cost-effectiveness literature, a Medicare claims database, and Medicare reimbursement codes, I developed a dataset with which to perform this research. To reallocate expenditures between interventions, utilisation of relatively cost-ineffective interventions was decreased, and utilisation of more cost-effective interventions increased. The findings estimate that substantial gains in aggregate health are achievable when reallocating expenditures to maximise health while maintaining a net change in total expenditure of zero. Also, simply increasing the utilisation of dominant interventions increases aggregate health gains along with cost-savings. When considering resource distribution, the findings show that allocating expenditures in a manner consistent with the cost-effectiveness evidence changed the distribution of expenditures across diseases and interventions.

While this research highlights that efficiency gains may be achievable, a number of data limitations restrict the generalisability to current policy. Further, the research did not account for the difficulty of changing established therapeutic practices, an aspect that would have a huge bearing on the success of the implementation of a policy consistent with this research.

In the final chapter, I summarise each chapter contained in this thesis and present the key findings of my research. Also, I discuss the limitations of this thesis, its policy relevance, and next steps.

8. Summary and Conclusions

8.1. Context of thesis

While health care affordability is a common challenge faced by developed countries, it is of particular concern in the US. Health care spending in the US is currently greatly in excess of spending in other developed countries and is increasing at an unsustainable rate. While steps have been taken to arrest this trend, the future affordability of health care in the US remains the subject of much debate and concern. Despite substantially higher levels of spending, compared to other developed countries, the US health care system performs worse across a variety of key health metrics.

In many countries, economic evaluation, or more specifically cost-effectiveness analysis, is an approach taken to prioritise scarce health care resources between competing interventions. However, despite the relatively poor return from health care spending, the US has been resistant to using economic evaluation to inform coverage decisions for medical interventions. CMS's stated position is that "*Cost-effectiveness is not a factor CMS considers in making NCDs. In other words, the cost of a particular technology is not relevant in the determination*". (CMS 2010)

This thesis considers the use of cost-effectiveness evidence in the US health care system. The empirical work focuses on CMS national coverage determinations (NCDs). National coverage determinations are reserved for interventions deemed particularly controversial or projected to have a major impact on the Medicare programme. (CMS 2003) The empirical aspect of this thesis has two broad aims; first, to examine the relationship between cost-effectiveness and coverage decisions, and second, to estimate the potential benefits in terms of aggregate health gain of using cost-effectiveness evidence to inform the allocation of expenditures across interventions in the Medicare programme.

8.2. Background

In Chapter 2, I summarised the use of economic evaluation to evaluate health care interventions and showed that it offers an approach to inform efficient resource allocation. I presented the theory underpinning cost-effectiveness analysis and how the approach is consistent with an extra-welfarist framework. Using a worked example I illustrated how, as long as certain assumptions hold, adherence to a cost-effectiveness decision rule will lead to efficient resource allocation across multiple health care programmes. I discussed the limitations and criticisms of cost-effectiveness analysis and how the magnitude of required data limits the use of a league table or mathematical programming approach. In the absence of full information on the costs and benefits of available interventions, a cost-effectiveness threshold is required to interpret cost-effectiveness evidence. Here I presented the various valuations of cost-effectiveness thresholds and various approaches to setting its value.

In chapter 3, I examined the US health care system within the context of health care systems in other developed countries. The US spends approximately 17% of GDP on health care, almost twice the average for OECD countries. Despite this, the US has fewer physicians and hospital beds per capita than many other countries. The US health care system performs poorly compared to health care systems in other developed countries and is associated with lower average life expectancy, higher infant mortality, and worse outcomes across certain diseases. I chose the comparator countries included in Chapter 3 on the basis that they help illustrate different approaches to using cost-effectiveness evidence along with other factors in coverage and reimbursement decisions, or in recommendations for the efficient use of medical technology. The UK, Sweden, Australia, and Canada, while all using different processes, are countries in which cost-effectiveness evidence plays a fundamental role in decision-making. In contrast, Germany and France are countries in which cost-effectiveness evidence, and economic evidence more generally, plays less of a role. Cost-effectiveness analysis is used sparingly in the US health care system. While some public and private payers claim to use cost-effectiveness evidence, the extent to which it informs decision-making is unclear. Notably, CMS states that cost-effectiveness evidence is not relevant to NCDs. To provide insight into why this resistance

exists, I described the failed attempts by Medicare and the state of Oregon's Medicaid programme to incorporate cost-effectiveness evidence into resource allocation decision-making. Here I show that, despite an apparent need to increase the efficiency of the health care system, the US may reside at the end of the spectrum with respect to the extent that cost-effectiveness evidence plays a role in decision-making.

8.3. Introduction to empirical work

In Chapter 4, I presented the research objectives, a review of the relevant literature, and the development of the database used for the empirical aspect of this thesis. With respect to the database, 140 NCDs made from 1999 through 2007 were considered, and the accompanying decision memos reviewed. One hundred and three decision memos met the inclusion criteria and were reviewed to extract relevant information into a database. From the 103 decision memos, 255 unique coverage decisions were identified. For each coverage decision, a literature search was performed to identify relevant cost-effectiveness studies. A relevant cost-effectiveness study was identified for 64 coverage decisions, 48 of which were positive coverage decisions and 16 of which were non-coverage decisions. This sample of coverage decisions with accompanying cost-effectiveness evidence was used for the empirical work presented in Chapter 5.

In addition to cost-effectiveness, the database included a number of variables pertaining to factors likely to have an effect on CMS's coverage decisions. These variables included the quality of the supporting clinical evidence, the availability of alternative interventions, intervention type, origin of the coverage request, and date of the decision. These variables were used for the research presented in Chapter 6.

A second smaller database was developed for the research presented in Chapter 7. This database was restricted to coverage decisions associated with a relevant estimate of cost-effectiveness. The database included: the incremental cost and incremental effectiveness associated with the intervention; the cost of the intervention and comparator in the first year of its use; the existing utilisation rate of the intervention; and the size of the eligible patient population.

8.4. Key findings of empirical work

8.4.1. Empirical Research: Part 1

The research presented in Chapter 5 had two objectives. First, to examine NCD decision memos to determine if they are consistent with CMS's stated position on the use of cost-effectiveness evidence, i.e., that cost-effectiveness is not a factor considered in NCDs. Second, to determine if there is a difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness.

With respect to the first objective, I found that in the majority of cases CMS's actions were consistent with their stated policy, i.e., that cost-effectiveness evidence did not feature in decision memos. In some cases, however, CMS's actions were inconsistent with their stated policy; for 14 coverage decisions, cost-effectiveness evidence was cited or discussed in the decision memo. Interestingly, 12 of these coverage decisions were positive, with a favourable cost-effectiveness ratio in each case (maximum ICER of $27,161 per life year gained). On one occasion, the decision memo for screening immunoassay fecal-occult blood test, a reference was made to the often-cited $50,000 per life year cost-effectiveness threshold. On another occasion, for the NCD regarding PET for Alzheimer's disease/dementia, the decision memo included discussion of QALYs. While not discussed with respect to cost-effectiveness evidence, as a discussion of the intervention's cost did not feature in the decision memo, it is notable that QALYS were used in this instance.

With respect to the second objective, I used a Mann Whitney U test to determine that there was a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness ($p<0.05$). This finding suggests that interventions subject to positive coverage decisions tend to be more cost-effective than those subject to non-coverage decisions. However, the research showed that CMS cover a number of interventions that do not appear cost-effective by traditional standards; seventeen were associated with an

ICER greater than $50,000 per QALY, nine with an ICER greater than $100,000 per QALY, and three with an ICER greater than $500,000 per QALY.

8.4.2. Empirical Research: Part 2

The findings of the research presented in Chapter 5 showed a statistically significant difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness. However, the approach taken was insufficient to confirm that cost-effectiveness, or the availability of cost-effectiveness evidence, is independently associated with coverage. Therefore, the objective of the research presented in Chapter 6 was to determine if cost-effectiveness is an independent predictor of coverage when controlling for other factors likely to have an effect on CMS's coverage decisions.

I used a binomial logistic regression for this analysis. Independent variables included: quality of the supporting clinical evidence, availability of alternative interventions, intervention type, origin of the request for coverage, and the date of the decision. The dependent variable was positive coverage or non-coverage. The findings of the multivariate logistic regression showed that the following variables were either statistically significant or included at least one statistically significant category: quality of the supporting clinical evidence, availability of alternatives, the date of the decision, and cost-effectiveness.

For *Cost-effectiveness*, compared to interventions estimated to be dominant, those with no associated estimate of cost-effectiveness were approximately five times less likely to be covered (approximately two thirds less likely when considering predicted probabilities). This is an important finding, showing that the availability of cost-effectiveness evidence had an independent effect on the coverage decision.

With respect to the quality of the supporting clinical evidence, interventions associated with good quality supporting evidence were six times more likely to be covered than those associated

with insufficient evidence (approximately twice as likely when considering predicted probabilities). With respect to the availability of alternative interventions, compared to interventions with no available alternative, those with an available alternative were approximately eight times less likely to be covered (approaching half as likely when considering predicted probabilities). Finally, with respect to the date of decision, coverage decisions made in 2006-2007 were approximately 10 times less likely to be covered than those made in 1999-2001 (half as likely when considering predicted probabilities). Further, interventions subject to coverage decisions in more recent time periods were increasingly less likely to be covered.

8.4.3. Empirical Research: Part 3

The findings of the research presented in Chapter 5 showed that CMS cover interventions that are not cost-effective by traditional standards. Coverage of cost-ineffective interventions generates relatively little health gain for the expenditure and suggests that existing resources could provide greater benefits if directed towards alternative more cost-effective interventions. The research presented in Chapter 7 had two objectives. First, to estimate potential gains in aggregate health from a hypothetical reallocation of expenditures using a cost-effectiveness decision rule. Second, to estimate the impact of reallocation on the distribution of expenditures across disease areas (oncology, cardiology, and other) and types of intervention (treatment, diagnostic, and other).

With respect to the first objective, reallocating expenditures, while maintaining a net change in total expenditure of zero, resulted in approximately 6 million additional beneficiaries receiving the most effective option. This corresponded to a health gain of approximately 800,000 QALYs. Approximately 6 million beneficiaries were affected by the reallocation, i.e., the reallocation changed the intervention they received, and the average per beneficiary gained approximately 0.13 QALYs. The ICER of the marginal intervention, i.e., the intervention with the highest ICER for which utilisation was increased, was approximately $37,000 per life year saved (autologous stem cell transplantation). Substantial gains in aggregate health were also estimated when increasing the utilisation of dominant interventions while maintaining existing utilisation

rates for interventions with positive ICERs. Increasing the use of dominant interventions resulted in approximately 4.2 million additional beneficiaries receiving the most effective treatment option, corresponding to gains in aggregate health of approximately 200,000 QALYs and savings of $4.8 billion. Approximately 4.2 million beneficiaries were affected by the reallocation, with an estimated average per beneficiary gain in health of 0.05 QALYs and per beneficiary cost savings of $1,113.

With respect to the second objective, the reallocation resulted in a greater proportion of expenditures directed to oncology-related interventions, and a lesser proportion to cardiology and other diseases. The reallocation also resulted in a greater proportion of expenditures directed to treatments and diagnostics, with less directed towards those categorised as other (e.g., health education, preventative care).

The research highlighted that dominant interventions are often underutilised and that substantial gains in aggregate health and cost-savings are achievable. While a number of interventions deemed cost-ineffective by traditional standards are covered, they appear to be utilised less frequently than more cost-effective interventions.

8.5. Limitations

This thesis has a number of limitations. First, I highlight limitations specific to the cost-effectiveness evidence used for each piece of empirical work. Second, I discuss the key remaining limitations for each piece of empirical work. Third, I discuss some of the broader limitations of this research.

8.5.1. Cost-effectiveness evidence

Unlike agencies that conduct cost-effectiveness analyses internally, or require cost-effectiveness studies to be included in submissions to them, CMS do not perform, or require, a cost-effectiveness analysis as part of their review. While on occasion the included cost-effectiveness study originated in the decision memo, I predominantly relied on published estimates of cost-effectiveness for much of this research. Therefore, there was inevitable variation between studies across a range of aspects including: country setting, perspective, funding status, etc.

While the majority of studies reported ICERs in terms of cost-per QALY gained, I also included studies reporting cost-per life year gained ratios, and those reporting cost-effectiveness ratios using disease specific units when the intervention was estimated to be dominant or dominated. While not directly comparable, cost-effectiveness studies reporting cost-per life year gained ratios were pooled with those reporting cost-per QALY ratios in the primary analyses in each piece of empirical work. However, for the empirical work presented in Chapters 5 and 6, I reported findings of the analyses when cost-effectiveness studies reporting cost-per life year gained ratios were excluded from the dataset (Section 5.3.4 and Section 6.3.4.1). For the empirical work presented in Chapter 7, for studies reporting cost-per life year gained, I adjusted survival gain with a utility weight to gain an, albeit imperfect, estimate of incremental QALY gain (Section 4.7.5.4).

I did not account for uncertainty associated with the included cost-effectiveness estimates in the empirical aspects of this thesis. Uncertainty was not consistently reported in the included cost-

effectiveness studies and, although it was included in the review of the cost-effectiveness studies, it was ultimately excluded from the dataset (Section 4.7.4.9). However, to more completely understand the influence of cost-effectiveness evidence on coverage decisions, it would be necessary to account for uncertainty.

8.5.2. Empirical work presented in Chapter 5

A minority of the coverage decisions featured in this research, approximately 25% (64 of 255), were associated with a cost-effectiveness estimate. This raises the question of whether coverage decisions with an associated estimate of cost-effectiveness are truly representative of the total sample. The fact that a large proportion of coverage decisions were not associated with cost-effectiveness evidence, and thus unaccounted for in the empirical work presented in Chapter 5, is a limitation of the research.

Also, in the instances when CMS did discuss or cite cost-effectiveness evidence in the decision memo, it was not possible to infer how the evidence influenced coverage decisions.

8.5.3. Empirical work presented in Chapter 6

The research presented in Chapter 6 was limited by data availability. As noted in Section 4.7.4.12, it was not possible to include a number of potentially relevant variables in the analysis. For example, including variables that accounted for the potential budget impact of the intervention and the prevalence of disease proved impractical. Disease severity is another factor suggested to be of importance in decision-making. (Dolan, 2005) However, accounting for disease severity proved difficult and was ultimately not accounted for in this research.

As for Chapter 5, the cost-effectiveness variable was a source of a number of limitations of this research. In the primary analysis, only 21% of coverage decisions were associated with evidence of cost-effectiveness at the time of the decision, though this increased to 30% when including

studies published after the NCD. This was problematic when constructing the cost-effectiveness variable as I was restricted to the number of categories that I could include, i.e., for positive ICERs I was restricted to two categories, <$50,000 per QALY, and >$50,000 per QALY. Further, I was required to pool cost-effectiveness studies that estimated the intervention to be dominated with studies that estimated the intervention to have an ICER of >$50,000 per QALY. This is not an ideal approach, as interventions estimated to be dominated are less effective than their comparator, unlike those associated with positive ICERs.

The variable capturing the quality of the supporting clinical evidence was limited in two key ways. First, the USPSTF guidelines for grading evidence were used as the approach to characterise the evidence. This approach is not ideal, as it requires that incremental benefit and the evidence quality to be accounted for simultaneously. Second, the variable is subjective in nature. The variable originated from a review of the evidence as presented in the decision memo by two researchers from Tufts Medical Center. As the researchers relied on CMS's presentation of the evidence base, rather than independently reviewing the individual studies, the variable is essentially subjective. A potential approach to improving this variable is presented in Section 6.4.5.

As discussed in Chapter 6, there is a possibility that a number of the independent variables in the model may interact, i.e., that the combined effects of two independent variables are not a sum of their individual effects. The presence of interaction effects make interpretation of model findings challenging, as a change in one variable will have unpredictable consequences on the result of the model. While it was deemed that there was no significant interaction between the independent variables in the current model it will be important to retest future models as the dataset is extended and the volume of available data increases.

8.5.4. Empirical work presented in Chapter 7

I had to make a number of necessary assumptions in the empirical work presented in Chapter 7. First, I assumed that every beneficiary included in the reallocation received some form of intervention. Second, if the beneficiary did not receive the intervention under review, it was assumed that they received the comparator. Third, when considering the net present value of future commitments, it was necessary to assume the availability of unrestricted finance.

I included cost data from a variety of sources, including: the cost-effectiveness studies, costing studies, and Medicare reimbursement codes. This variation in the source of the cost information is a limiting factor in this research. The cost-effectiveness studies were also used as the source for the estimate of incremental health gain. A consequence of this approach is that I inferred that the reported estimate of incremental health gain was accurate for the average Medicare beneficiary who received it. The cost-effectiveness literature was insufficient to allow me to account for patient heterogeneity in the analysis.

I relied on ICD-9 codes reported in the Medicare claims database to identify beneficiaries eligible for the intervention. This is a crude approach as ICD-9 codes lacked the required precision to account for all clinical factors specified in NCDs. On occasion ICD-9 codes were insufficient to identify Medicare beneficiaries eligible for an intervention, and the respective coverage decision was excluded from the analysis. The Medicare claims database also prevented me from distinguishing between incident and prevalent cases. This information is often necessary to determine whether a beneficiary is eligible for an intervention, e.g., when the management approach immediately following diagnosis differs from the long-term management of the condition.

I did not account for the feasibility of implementing changes in patient care, i.e., changing the intervention received by the Medicare beneficiary. It is likely that implementing a change will be challenging for interventions deeply entrenched in clinical practice. I accounted for this by

implementing a change in utilisation of 50% (range 10-90%), rather than adjusting utilisation rates to 0% or 100%. This change was, however, arbitrary.

8.5.5. Broader limitations

The empirical aspect of this thesis pertains to NCDs made from 1999 through 2007. Updating the research to include NCDs made through 2011 would be useful, particularly given that for a number of recent NCDs CMS have cited or discussed cost-effectiveness evidence in the decision memo (Section 3.5.2).

Although NCDs are arguably the most important of CMS coverage decisions, a minority of interventions are evaluated through this process. Rather, the majority of coverage decisions are made by Medicare Administrative Contractors (MACS), often referred to as local coverage determinations (LCDs). By excluding LCDs from this research, it is possible that the included coverage decisions are unrepresentative of CMS coverage decisions in general.

The most evident limitation of this thesis is that it focuses solely on the Medicare programme. While Medicare is a dominant component of the US health care system, other public and private insurance bodies also play a major role. To gain a more complete understanding of the role of cost-effectiveness evidence in the US health care system, this research could be broadened to include other payers.

8.6. Contributions of thesis

In the US, health care is an emotive political issue. This thesis is timely and is relevant to the ongoing debate regarding the cost, sustainability, and efficiency of the Medicare programme. Despite a real need to improve the value achieved from health care spending, the US remains largely opposed to the use of cost-effectiveness evidence in coverage decisions for medical technologies and interventions. This thesis makes a number of important contributions and provides important evidence that fills significant gaps in the literature.

The research presented in Chapter 5 is the first to systematically evaluate the cost-effectiveness of CMS's coverage decisions, and to evaluate the consistency of CMS's actions with its stated policy on the use of cost-effectiveness evidence. I highlighted that, contrary to CMS's stated position on the use of cost-effectiveness evidence, CMS has cited or discussed cost-effectiveness evidence on a number of occasions. This research provides insight into the value of interventions offered in the Medicare programme, and shows that CMS cover a number of interventions that are not cost-effective by traditional standards. Despite this, through this research I have established that there is a difference between positive coverage decisions and non-coverage decisions with respect to cost-effectiveness, suggesting that covered interventions tend to be more cost-effective than non-covered interventions. It was possible to identify a cost-effectiveness study for approximately 25% of coverage decisions and, therefore, this research illuminates the lack of knowledge with respect to the cost-effectiveness of the majority of interventions offered in the Medicare programme.

The empirical work presented in Chapter 6 is the first to use regression analysis to evaluate coverage decisions made in CMS NCDs, providing an empirical insight into CMS's coverage of medical interventions. Given the lack of a formal interpretation of CMS's reasonable and necessary coverage criteria, and the existing uncertainty regarding it, this research is an important contribution. (Foote 2002;Neumann, Rosen, & Weinstein 2005) Encouragingly, the research suggests that CMS operate evidence-based decision-making, as coverage decisions appear to be broadly consistent with the supporting clinical evidence base. I have also shown

that the availability of alternative interventions appears to be an important factor in coverage decisions and that value appears to play a role.

Insight into CMS's decision-making criteria is important, as NCDs are increasingly subject to debate and scrutiny. The findings of the empirical work in Chapter 6 have the potential to reduce the uncertainty associated with CMS NCDs and provide a framework with which to understand CMS's reasonable and necessary criteria. The research presented in Chapter 6 provides a starting point for a future research agenda to better understand the evidence required for an intervention to be a covered benefit in the Medicare programme.

While it is apparent that the Medicare programme has faced increasingly difficult fiscal challenges, no study has shown the impact this has had on coverage policy. The research in Chapter 6 suggests that, when controlling for factors relevant in the decision making process, CMS has become increasingly restrictive with respect to the coverage of medical interventions over the considered time period, i.e., 1999 through 2007.

It is claimed that Medicare is underperforming, with approximately 30% of administered care either inappropriate or unnecessary. (Bentley et al. 2008;Fisher et al. 2003;Garber, Goldman, & Jena 2007;Orszag 2008) The research presented in Chapter 7 is the first to estimate the potential efficiency gains achievable from a hypothetical reallocation of resources in accordance with available cost-effectiveness evidence. The findings show that substantial gains in aggregate health are potentially achievable.

The research in Chapter 7 provides an insight into the relationship between the existing utilisation of interventions in the Medicare programme and their cost-effectiveness. For example, the findings highlight the underutilisation of a number of interventions that are estimated to be dominant. These interventions represent obvious targets for gains in aggregate health and cost-savings. In addition, the research shows that although CMS cover a number of

interventions that are not cost-effective by traditional standards, these interventions are used infrequently, with few patients receiving them.

The research in Chapter 7 also illustrates that using different resource allocation criteria will impact the distribution of resources across types of technology and disease, i.e., that implementation of a cost-effectiveness rule for resource allocation may result in resources being reallocated from one type of technology or disease area to another. This research draws attention to the opportunity cost of coverage decisions and the necessary trade-offs that must be made when allocating scarce health care resources.

A common finding across all empirical aspects of this thesis is the often-inadequate nature of the available evidence to perform this type of research. While a number of coverage decisions were associated with estimates of cost-effectiveness, there was much variability between cost-effectiveness studies with respect to quality, methods, perspective, etc. This thesis serves to reinforce the need for the standardisation of cost-effectiveness studies to be relevant to US decision makers' practice. Further, the utilisation and cost data used in Chapter 7 illustrate the difficulties of using existing evidence to estimate the consequences of reallocating resources in terms of aggregate health gain.

8.7. Policy implications

This thesis has a number of important policy implications. These are presented below with respect to each piece of empirical work.

8.7.1. Empirical Research: Part 1

The findings of the research presented in Chapter 5 should be highly relevant to policy makers. First, this research provides an insight into the value of a proportion of interventions evaluated through CMS NCDs. It shows that CMS is covering interventions that are not cost-effective by traditional standards. As noted above, offering these interventions generates relatively little health gain for the expenditure and suggests that resources could provide greater benefits if directed towards alternative interventions. Second, it highlights the lack of knowledge regarding the value of many of the interventions offered by Medicare. Without this knowledge, Medicare is limited to the extent it can make rational coverage decisions and account for value in coverage policy. Third, in the majority of occasions, CMS have appeared not to consider relevant cost-effectiveness studies available at the time of the NCD.

8.7.2. Empirical Research: Part 2

The findings of the empirical work presented in Chapter 6 underscore that CMS has adopted evidence-based medicine in NCDs and illustrates a level of consistency in their decision-making. It is noteworthy that the availability of cost-effectiveness evidence is associated with coverage, in contradiction to CMS's stated position. The findings go some way to reveal CMS's interpretation of the 'reasonable and necessary' criterion. For example, the availability of alternative interventions is associated with coverage, which perhaps provides insight into what CMS considers necessary care. This research has the potential to help the entire medical community better understand CMS's evidence requirements, thus reducing uncertainty associated with CMS NCDs. The study increases the transparency of coverage decisions, increasing CMS's accountability. The findings suggest that CMS has become more restrictive with respect to coverage over time, perhaps reflecting Medicare's increasing fiscal challenges.

8.7.3. Empirical Research: Part 3

The research presented in Chapter 7 is particularly policy-relevant given the fiscal challenges currently faced by the Medicare programme. As described above, CMS covers some very cost-ineffective interventions, and thus there is opportunity for more efficient resource use. The research illustrates the potential benefits in terms of aggregate health gains of using cost-effectiveness evidence to inform resource allocation. Indeed, substantial aggregate health gains are achievable from reallocating resources within the current level of expenditure. Further, the research identifies the underutilisation of a number of dominant interventions. This finding is concerning and suggests efficiency gains are readily achievable. Lastly, the research highlights that the use of different resource allocation criteria will affect the distribution of expenditures across types of intervention and disease areas.

8.8. Next Steps

8.8.1. Research scope

As noted above, a limitation of this research is that it focuses solely on the Medicare programme. One potential starting point for future work would be to expand the scope of this research to included additional public and private payers. In terms of public payers, the DoD or the VA may provide an appropriate place to start. Both payers state that cost-effectiveness is considered in their decision-making, although their respective approach to its incorporation in the decision-making process is unclear (Section 3.3.2.3). An interesting study, given their different positions in the use of cost-effectiveness evidence, would be to compare coverage decisions made by CMS, the DoD, and the VA. Unfortunately, the DoD and VA do not provide similar documentation to CMS's decision memos and obtaining the necessary data for this research may be challenging.

The majority of Americans (67.5%) obtain health insurance through private providers. To gain a complete understanding of the relationship between cost-effectiveness evidence and coverage of interventions in the US health care system, it would be necessary to include private payers in future research. As noted above for the DoD and the VA, this research would likely be limited by the absence of documentation providing details for each coverage/tiering decision. While proprietary drug formularies are often publicly available, documentation supporting the coverage/tiering decision is typically unavailable.

Also, as noted above, a limitation of this research is that it is limited to only NCDs. Coverage decisions made by Medicare Administrative Contractors (MACs), local coverage determinations (LCDs) are the majority of coverage decisions made by Medicare. Including LCDs in this research would provide a broader insight into the coverage of interventions in the Medicare programme. The principal challenge of researching LCDs concerns the volume of policies. In a study to assess variation in coverage across regional Medicare contractors, Foote et al. (2005) reviewed 5,213 individual coverage policies. (Foote, Halpern, & Wholey 2005) The frequent

lack of supporting documentation for LCDs further complicates matters, a fact that would limit replication of the research presented here.

8.8.2. Improving existing variables

Cost-effectiveness

One overarching means to enhance the empirical aspect of this thesis is to update it to include NCDs made through 2011. This would increase the number of coverage decisions and provide a more up-to-date assessment. As the number of NCDs with associated cost-effectiveness evidence increases, it may be possible to stratify the data by type of intervention. This may prove informative when considering preventative care, a subset of NCDs for which cost-effectiveness has been considered with some regularity in recent years (Section 3.5.2).

A larger sample of coverage decisions would provide an opportunity to develop the cost-effectiveness variable. For the existing research, I was restricted to the number of categories I could include in the cost-effectiveness variable. With a larger sample, I could increase the number of categories, capturing cost-effectiveness with more precision.

For many coverage decisions, the medical literature proved insufficient to identify relevant estimates of cost-effectiveness. Although the limitations of the medical literature unavoidable, one approach to supplement the literature searches would be to gain expert opinion into the cost-effectiveness of the coverage decisions here.

Quality of supporting clinical evidence

The variable used to account for the quality of the supporting clinical evidence within the empirical work presented in Chapter 6 was limited due to its reliance on the USPSTF evidence grading criteria, and, as a result, its subjectivity. The quality of the supporting clinical evidence could be captured more accurately using an objective review of the supporting evidence. This

could be achieved by categorising the evidence base using a number of objective criteria, including: study design (e.g., randomised studies, non-randomised study, retrospective study, etc); study outcomes ('hard' endpoints vs. surrogate endpoints); inclusion of active comparators; consistency of findings across studies; patient population (e.g., whether the study included Medicare beneficiaries); country of study (e.g., US-based vs. non-US-based); and recency of study publication.

Availability of alternative therapies

The variable accounting for the availability of alternative interventions is currently coded as a binary variable (alternatives available/no alternatives available). Accounting for the number of available alternatives and coding the variable either continuously or categorically are potential approaches for developing this variable.

Utilisation rate

For the empirical work presented in Chapter 7, using ICD-9 diagnostic codes to estimate existing utilisation rate and size of the eligible patient population proved challenging. Unfortunately, as ICD-9 codes were insufficient to identify eligible beneficiaries that met the specifications of the NCD, I had to exclude a number of interventions from the analysis. Input from health care practitioners would be one approach to supplement the ICD-9 codes and provide estimates of the proportion of beneficiaries with a broad diagnosis that would meet the specifications of a NCD.

8.8.3. Potential additional variables

Adding a number of variables to the database would be useful. As NCDs are typically made for interventions expected to have a significant impact on the Medicare programme, the inclusion of variables that characterise the budget impact and prevalence associated with an intervention/disease would be useful. While the decision memos do not report this information, and the medical literature proved insufficient for these variables, it may be possible to include a

version of these variables in the database with input from health care practitioners or health services researchers.

A variable that captures the nature of disease would be useful to include in the database. Such variables could potentially account for disease severity or nature of clinical benefit. While these variables proved difficult to include in this research (Section 4.7.4.12), input from health care practitioners may again be beneficial here.

The empirical work presented in Chapter 7 found that the least cost-effective interventions were often underutilised. I speculated that this might be related to how the intervention is reimbursed. One option to account for reimbursement would be include physician reimbursement rate as a variable in the dataset.

Lastly, it is often the case that NCDs are controversial. There is often a great deal of scrutiny on CMS with respect to NCDs, and it may be that political factors influence the outcome. Although difficult to include, a variable that accounted for lobbying may prove insightful. A potential proxy for this might be a tally of comments submitted to CMS during the NCD's comment period. Although a somewhat tenuous link to lobbying, this approach would at least account for the amount of public input into the decision.

8.8.4. Alternative conceptual framework

It would be valuable to engage members of the Medicare coverage group in future research. I chose to use a 'production function' for the conceptual framework used in the research described in Chapter 6. With insight from the decision makers, it may be possible to develop this research to account for the relative importance that they place on different aspects of the evidence base. For example, if it was the case that the quality of the clinical evidence was the dominant aspect of the evidence base, and evidence of value was considered secondarily, the regression model could be structured according using a hierarchical modelling framework.

8.8.5. Expanding scope of reallocation work

The empirical research presented in Chapter 7 could be developed using two general approaches. First, the scope of the research could be broadened to include additional interventions available in the Medicare programme, i.e., those covered through LCDs.[xxii] While this approach would include a greater proportion of interventions covered in the Medicare programme, the same challenges faced in the research presented in Chapter 7 would also likely be faced here. Nevertheless, estimated efficiency gains would be more reflective of the benefits of using cost-effectiveness evidence in the Medicare programme as a whole. The alternative approach would be to narrow the scope and focus the research on few interventions and conditions. Research efforts could be focused on aspects of the Medicare programme in which data of sufficient quality was available. However, the narrow scope of the work would make generalisations to the wider benefits of using cost-effectiveness analysis in the Medicare programme more difficult.

The empirical work presented in Chapter 7 evaluated the impact of a hypothetical reallocation of expenditures using a cost-effectiveness rule on the distribution of expenditures between patient populations and types of intervention. With a better understanding of society's preferences, it would be possible to determine if the current distribution of expenditures is consistent with society's preferences, and further, if using cost-effectiveness evidence moves us closer or further away from being in accordance with them.

[xxii] While the broader research project conducted at Tufts Medicare Canter did include interventions covered through LCDs, its scope was more limited compared to the empirical work in Chapter 7.

8.9. Moving towards value based policies – opportunities and challenges

In the preceding sections, I have focused on the empirical aspects of this thesis. In the remainder of this chapter, I will discuss the opportunities and challenges Medicare face in moving towards value-based policies.

Medicare is an important component of the US health care system and is fundamental to its, and the country's, financial future. (Chernew, Baicker, & Hsu 2010) CMS's role is to administer the Medicare programme and thus is responsible for its financial stability. While the cost of Medicare has steadily increased, it is inevitable that a point will be reached when Congress will be unwilling to borrow or raise taxes to continue to fund the programme. Medical technology is a major contributor to the increase in costs (Ginsburg 2004;Ginsburg 2008). In 2008, the Medicare programme's "Triple Aims" were announced: 1) improve the individual experience of care; 2) improve the health of populations; and 3) reduce per-capita costs of care for populations. (Berwick, Nolan, & Whittington 2008) These aims reflect the rationale for using cost-effectiveness analysis, i.e., to improve the health care quality while controlling costs.

This thesis illustrates that using cost-effectiveness evidence to inform coverage decisions offers a potential approach for improving the quality of care that Medicare beneficiaries receive while curtailing the unsustainable growth in programme cost.

8.9.1. Moving forward – Challenges, opportunities, and recommendations

If CMS are to move towards a more value based coverage policy, there are a number of important hurdles to overcome. In the following sections, I discuss some of the challenges facing Medicare and suggest how to overcome them.

8.9.1.1 Restricted authoritative capacity

CMS's role as Medicare's administrator is limited both by its legal authority and a prevailing practice that inhibits its flexibility with respect to the coverage of medical technology. This is illustrated by the recent NCD for sipuleucel-T (Provenge®), a vaccine-based treatment indicated for advanced prostate cancer approved by the FDA in 2010. (CMS 2011a;FDA 2010) While sipuleucel-T is associated with estimated survival gains of 4.1 months compared to placebo, the cost of the treatment is notably high, $93,000 for a course of three treatments. Prostate cancer is a prevalent disease in the Medicare population, and upon sipuleucel-T's approval CMS were faced with the challenge of how to pay for an intervention that could potentially have huge implications for the cost of the programme. Ultimately, after much debate and comment from stakeholders, CMS covered sipuleucel-T in accordance with the approved FDA indication. This NCD serves as a useful case study of the limited flexibility that CMS has with respect to coverage. Without the authority to negotiate on price, and with cost-effectiveness evidence effectively excluded from consideration, CMS's only option is to closely scrutinise the clinical evidence base. The NCD for sipuleucel-T shows that without the authority to consider cost-effectiveness evidence, CMS have few options but to cover interventions that offer marginal incremental health benefits, irrespective of their cost.

This lack of authority was exacerbated by Medicare recently losing its authority to use the long-standing "least costly alternative" (LCA) policy. Essentially a cost-minimisation strategy, the LCA policy allowed Medicare to pay the rate of the lowest cost alternative in situations where there was no evidence of clinical superiority between two products. In December, 2009, Medicare lost a legal challenge on the grounds that Congress establishes payment policy and does not give Medicare explicit authority for LCA. (Hays v. Sebelius, 2009)

CMS will inevitable face future difficult coverage decisions, particularly as highly expensive cancer treatments hit the marketplace. Given its limitations, CMS face a significant challenge to balance the provision of new and expensive interventions while administering Medicare in a

fiscally responsible manner. Payment reform should be accelerated to grant CMS the authority and flexibility to make coverage decisions consistent with these goals.

8.9.1.2 Resistance to the use of cost-effectiveness evidence

Resistance to the use of cost-effectiveness evidence in the US health care system is well established. While the phenomenon of American exceptionalism, i.e., the general resistance to limit setting, is inherent throughout US society, it is maybe most visible in the political realm of health care. With health care a contentious political issue, as exemplified by the debate surrounding, and resistance to, proposed changes in the recent health reform legislation, moving health care towards more value-based policies is likely to be hugely challenging. Indeed, when considering the Medicare programme and the Oregon Health Plan, opposition from politicians contributed to previous failure to introduce cost-effectiveness considerations into decision-making (Section 3.4.2).

Cost-effectiveness analysis remains a difficult political sell. Concerns that cost-effectiveness evidence may adversely affect the revenue streams of manufacturers, providers, insurers and health care professions are likely to hinder its future use. Without politic support for its use, any change regarding the use of cost-effectiveness evidence remains unlikely. However, at a time when the cost of health care is considered one of the most significant threats to the US's fiscal wellbeing, perhaps the current environment is one in which there exists a political willingness to foster a change in health care. Once there is a willingness on the part of politicians to accept that the health care system cannot offer all beneficial services regardless of cost, a debate can begin as to how best to implement the changes.

8.9.1.3 Acceptance of limitations on health care

In Chapter 2, I showed that to use cost-effectiveness evidence, there either must be a budget constraint or a cost-effectiveness threshold in operation; it is notable that CMS do not have either of these. CMS do not have a fixed annual budget; rather, the cost of the programme increases

with the rising cost of care provision. The lack of a fixed budget is important as it has meant that CMS have not had to face considerations of opportunity cost in their coverage decisions.

Setting an annual budget, or limiting spending growth, would help promote debate regarding the limits of the Medicare programme and the relative value of the services it provides. Starting the debate as to how CMS should allocate scarce Medicare resources between services would perhaps provide a path to an open discussion among stakeholders with respect to relative value, opportunity cost, and cost-effectiveness.

8.9.1.4 Ambiguity of decision-making criteria

While the empirical work presented in Chapter 6 provides some insight into CMS's decision-making criteria, in the absence of a definition of how the 'reasonable and necessary' coverage criteria should be interpreted, much ambiguity remains. Though there may be some benefit in CMS shielding decision-making criteria from public scrutiny if, as a result, outcomes are achieved that would not have been through a transparent process, it is unclear how this would be the case with respect to NCDs.

Without explicit decision-making criteria, CMS maintains a degree of flexibility in their coverage decisions. This flexibility, while beneficial to CMS, leads to uncertainty for manufacturers and the medical community as to the coverage of, and access to, medical technology, and prevents parties effectively negotiating with the CMS. Further, a lack of a clear decision rule risks inequitable and inconsistent coverage decisions.

Having clear coverage criteria would be beneficial. It would decrease the uncertainty regarding CMS coverage among stakeholders and promote more consistent decision-making. For manufacturers, explicit coverage criteria would help inform the evidence required to support positive coverage decisions, helping the design of clinical development programmes.

It would also have the effect of making transparent the trade-offs CMS make in their decision-making and providing insight into how value factors into decision-making. Currently, by virtue of the ambiguity of decision-making criteria CMS maintain the illusion that rationing does not occur in the Medicare programme. However, this is evidently not the case. Medicare does make rationing decisions, but it does so in a closeted manner, concealed behind coverage policies that are supposedly based solely on clinical evidence. (Fox 2010) However, as the research presented in Chapter 7 suggests, not explicitly considering cost-effectiveness evidence comes at a cost, with resource allocation less efficient than would otherwise be achievable. Further, because of the veiled nature of coverage decisions, it is impossible to determine their consistency with societal preferences for the coverage of medical technology.

8.9.1.5 Learn from international experiences

This thesis illustrates the benefits of using cost-effectiveness evidence to inform coverage of medical technology. However, for the foreseeable future, it seems unlikely that cost-effectiveness evidence will become as deeply integrated into the US health system as it has in other countries, e.g., the UK, Sweden, Australia, and Canada (3.3.1). However, as described in Chapter 3, international experience shows that the introduction of cost-effectiveness evidence does not require a fundamental change in a health care system. It is notable that it is in recent times economic evidence has been considered by institutions in Germany and France, and it has been incorporated in a manner that complements existing approaches to technology evaluation, with coverage decisions remaining principally grounded in the clinical evidence base.

Also shown in Chapter 3 is that in no institution is cost-effectiveness the sole decision-making criterion. While NICE operates an explicit threshold, the threshold exists over a range to allow for a range of other factors to be accounted for in decision-making (Section 2.6.3.2).

If there was the political will to consider cost-effectiveness evidence, CMS could learn from the various models used in other countries. Experience illustrates that it is possible to include cost-effectiveness evidence in coverage decisions in a transparent manner that would complement existing processes and institutional structures.

8.9.1.6 Comparative effectiveness

While great resistance to the use of cost-effectiveness evidence remains, comparative effectiveness evidence has recently been put under the spotlight as an approach to increase efficiency in the US health care system. Recently, the PPACA legislation created the Patient Centered Outcomes Research Institute (PCORI) to coordinate comparative effectiveness research (CER) studies, assist in CER study funding, and to disseminate study findings. (PPACA 2010) While CER may prove sufficient to increase the value of care in some instances, i.e., identifying dominated interventions, the approach does account for the value of health gain, necessary when allocating scarce resources. A proposed payment model incorporating CER would require Medicare to pay equally for interventions that provide comparable health outcomes, with higher payments set for interventions that have been demonstrated to provide superior health benefits. (Pearson & Bach 2010) However, how to value additional health gain, and thus how to set prices, is unclear, and the approach does not satisfactorily circumvent the rationale for cost-effectiveness analysis. While not met with the same degree of resistance as cost-effectiveness evidence, not all stakeholders have embraced comparative effectiveness evidence, with opponents suggesting a negative impact on innovation and patient care. (Carrier, Pham, & Rich 2010;Vernon & Goldberg 2011)

Certainly, comparative effectiveness evidence is a step towards the consideration of value in decision-making. Considering incremental benefits of competing treatments better informs decisions between competing interventions. However, without consideration of cost, how to interpret the value of the incremental benefit is challenging.

8.10. Conclusions

This thesis is timely and its findings have a number of important policy implications. CMS NCDs are reserved for interventions deemed particularly controversial or projected to have a major impact on the Medicare programme and offer a valuable insight into CMS coverage of interventions. This thesis aimed to evaluate the current use of cost-effectiveness analysis in Medicare, and to estimate the potential value of using it in terms of aggregate health gains.

Despite their stated position on the use of cost-effectiveness evidence the empirical work highlighted that CMS have on occasion discussed or cited cost-effectiveness evidence in decision memos. While I identified instances when CMS have covered interventions not cost-effective by traditional standards, I found that interventions subject to positive coverage decisions tend to be associated with more favourable cost-effectiveness evidence.

The second piece of empirical work evaluated the role of cost-effectiveness evidence in CMS NCDs while controlling for other factors that are likely to have an effect on Medicare coverage decisions. I determined that compared to interventions associated with cost-effectiveness evidence that estimated the intervention to be dominant; those with no associated estimate of cost-effectiveness were approximately five times less likely to be covered. This finding again contradicts CMS's stated position on the use of cost-effectiveness evidence. The findings also illustrated that CMS operate use an evidence based approach for NCDs and that the availability of alternative interventions has a significant effect on the likelihood of coverage. Lastly, the research showed that CMS have become more restrictive over time with respect to the coverage of medical interventions.

The findings from the third empirical aspect of this thesis serve to highlight the potential benefits of considering value in Medicare coverage policy, suggesting that substantial aggregate health gains are achievable from using cost-effectiveness evidence to guide resource use.

While this thesis has shown that cost-effectiveness evidence has been discussed or cited on occasion, it is clear that it is not used, or acted upon, with regularity and, accordingly, the Medicare programme could be more efficient. This is borne out by the research presented in Chapter 7 in which I estimated that substantial gains could be achieved from using cost-effectiveness evidence.

Rationing is an unavoidable reality in the Medicare programme. CMS maintain, however, that cost-effectiveness is not a factor in decision-making, therefore suggesting that interventions with positive benefits are paid for, regardless of costs. This closeted approach to rationing has proved insufficient and maintaining the current position on the use of cost-effectiveness has come at a cost in terms of efficiency. Moving forward, CMS will have to decide if it is prepared to continue to trade-off the use of cost-effectiveness, and the associated gains in aggregate health, for the illusion that health care is unrationed.

While politically difficult for the US government to set explicit limits on access to health care interventions, discouraging the use of cost-effectiveness evidence in Medicare is unfortunate. Only if full information of the relationship between the costs and benefits of interventions is available can a health care system be expected to work efficiently. (Neumann & Weinstein 2010)

It seems certain that the Medicare will suffer from increasing fiscal difficulties. It is inevitable that CMS will have to act to manage the financial sustainability of the programme as new expensive interventions become available. What is uncertain is the approach CMS will take. Irrespective of the taken approach, the goal will be the same, i.e., to improve programme value. Despite the resistance to the use of cost-effectiveness evidence in the US health care system, as shown in this thesis, cost-effectiveness analysis offers one approach to inform efficient resource allocation.

If the political will to use cost-effectiveness evidence to inform resource allocation existed, CMS could learn from the experiences of other countries. While to emulate the processes used in countries in which cost-effectiveness evidence plays a fundamental role in decision-making, e.g., the UK, Sweden, Australia, and Canada (Section 3.3.1), would require radical changes in the US health care system, lessons from Germany and France prove that cost-effectiveness can be incorporated without the need for radical overhauls. As I described in Chapter 3, Germany and France are countries that have in recent times incorporated cost-effectiveness evidence into aspects of decision-making, while ensuring that existing process remained essentially unaltered. In Germany and France decision-making remains grounded in the clinical and comparative evidence with economic evidence incorporated in a manner to complement existing approaches to technology assessment.

The fiscal challenges facing the Medicare programme, and the US health care system in general, are unlikely to diminish in the foreseeable future. CMS face the challenge of balancing the provision of new and expensive interventions, while administering Medicare in a fiscally responsible manner. This thesis shows that cost-effectiveness analysis is an approach that can achieve value based coverage policy. What is required is the political will to make bold and likely hugely unpopular steps to shift coverage policy to one that is explicitly evidence and value based.

9. References

ACCESS ECONOMICS PTY LIMITED 2004, The Economic Impact and Cost of Vision Loss in Australia. Viewed July 17th, 2011 <www.cera.org.au/uploads/CERA_clearinsight_overview.pdf>

AETMIS. Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS) - Agency Background. Viewed July 17th, 2011 <http://www.aetmis.gouv.qc.ca/site/en_historique.phtml>

AHRQ 2003, Report to the Agency for Healthcare Research and Quality. A Comparison of the Cost-Effectiveness of Fecal Occult Blood Tests with Different Test Characteristics in the Context of Annual Screening in the Medicare Population.

AHRQ. Agency for Healthcare Research and Quality (AHRQ). Technology Assessment program. Viewed July 2nd, 2011 <http://www.ahrq.gov/clinic/techix.htm>

AITKEN, L., ADDISON, C. 1996. "The cost-effectiveness of ambulatory blood pressure monitoring". Prof Nurse, vol. 12, no. 3, pp. 198-202

AL, M. J., FEENSTRA, T. L., & HOUT, B. A. 2005, "Optimal allocation of resources over health care programmes: dealing with decreasing marginal utility and uncertainty", Health Econ., vol. 14, no. 7, pp. 655-667.

AL MJ, FEENSTRA T, & BROUWER WB. 2004, "Decision makers' views on health care objectives and budget constraints: results from a pilot study." Health Policy vol. 70, no. 1, pp. 33-48

AMENT, A. & BALTUSSEN, R. 1997, "The interpretation of results of economic evaluation: explicating the value of health", Health Econ., vol. 6, no. 6, pp. 625-635.

ANDERSON, G. F., REINHARDT, U. E., HUSSEY, P. S., & PETROSYAN, V. 2003, "It's the prices, stupid: why the United States is so different from other countries", Health Aff.(Millwood.), vol. 22, no. 3, pp. 89-105.

ANELL, A. 2009, The Swedish Health Care System, 2009, The Commonwealth Fund.

ANELL, A. & PERSSON, U. 2005, "Reimbursement and clinical guidance for pharmaceuticals in Sweden: do health-economic evaluations support decision making?", Eur J Health Econ, vol. 6, no. 3, pp. 274-279.

ANGEL, M. 2004, "Buying Influence - How the Industry Makes Sure It Gets Its Way," in The Truth About the Drug Companies: How They Deceive Us and What to Do About It, M. Angell, ed., New York, pp. 193-216.

ANTONIOU, J., MARTINEAU, P. A., FILION, K. B., HAIDER, S., ZUKOR, D. J., HUK, O. L., PILOTE, L., & EISENBERG, M. J. 2004, "In-hospital cost of total hip arthroplasty in Canada and the United States", J.Bone Joint Surg.Am., vol. 86-A, no. 11, pp. 2435-2439.

APPLEBY J, DEVLIN N, PARKIN D, BUXTON M, CHALKIDOU K. 2009, "Searching for cost effectiveness thresholds in the NHS." Health Policy vol. 91, no. 3, pp. 239-45

ARRA. American Recovery and Reinvestment Act (ARRA). 2009.

AYAS, N.T., FITZGERALD, J.M., FLEETHAM, J.A., WHITE, D.P., SCHULZER, M., RYAN, C.F., GHAELI, R., MERCER, G.W., COOPER, P., TAN, M.C., MARRA, C.A. 2006. "Cost-effectiveness of continuous positive airway pressure therapy for moderate to severe obstructive sleep apnea/hypopnea". Arch Intern Med., vol. 166, pp. 977-984

BAKER R, BATEMAN I, DONALDSON C, JONES-LEE M, LANCSAR E, LOOMES G, MASON H, ODEJAR M, PINTO PRADES JL, ROBINSON A, RYAN M, SHACKLEY P, SMITH R, SUGDEN R, WILDMAN J; SVQ RESEARCH TEAM. 2010, "Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project." Health Technol Assess vol. 14, no. 27, pp. 1-162

BALA, M. V., ZARKIN, G. A., & MAUSKOPF, J. A. 2002, "Conditions for the near equivalence of cost-effectiveness and cost-benefit analyses", Value.Health, vol. 5, no. 4, pp. 338-346.

BALTUSSEN, R., KNAI, C., & SHARAN, M. 2004, "Iron fortification and iron supplementation are cost-effective interventions to reduce iron deficiency in four subregions of the world", J Nutr., vol. 134, no. 10, pp. 2678-2684.

BARTON, P., BRYAN, S., & ROBINSON, S. 2004, "Modelling in the economic evaluation of health care: selecting the appropriate approach", J Health Serv.Res.Policy, vol. 9, no. 2, pp. 110-118.

BELL, C. M., URBACH, D. R., RAY, J. G., BAYOUMI, A., ROSEN, A. B., GREENBERG, D., & NEUMANN, P. J. 2006, "Bias in published cost effectiveness studies: systematic review", BMJ, vol. 332, no. 7543, pp. 699-703.

BENOIT, R.M., COHEN, J.K., MILLER, R.J. 1998. "Comparison of the hospital costs for radical prostatectomy and cryosurgical ablation of the prostate", Urology. vol. 52, no. 5, pp. 820-824

BENTLEY, T. G., EFFROS, R. M., PALAR, K., & KEELER, E. B. 2008, "Waste in the U.S. Health care system: a conceptual framework", Milbank Q., vol. 86, no. 4, pp. 629-659.

BERWICK, D. M., NOLAN, T. W., & WHITTINGTON, J. 2008, "The triple aim: care, health, and cost", Health Aff.(Millwood.), vol. 27, no. 3, pp. 759-769.

BIRCH, S. & DONALDSON, C. 1987, "Applications of cost-benefit analysis to health care. Departures from welfare economic theory", J.Health Econ., vol. 6, no. 3, pp. 211-225.

BIRCH, S. & DONALDSON, C. 2003, "Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism?", Soc.Sci.Med., vol. 56, no. 5, pp. 1121-1133.

BIRCH, S. & GAFNI, A. 1992, "Cost effectiveness/utility analyses. Do current decision rules lead us to where we want to be?", J.Health Econ., vol. 11, no. 3, pp. 279-296.

BIRCH, S. & GAFNI, A. 2004, "The 'NICE' approach to technology assessment: an economics perspective", Health Care Manag.Sci., vol. 7, no. 1, pp. 35-41.

BIRCH, S. & GAFNI, A. 2006a, "Information created to evade reality (ICER): things we should not look to for answers", Pharmacoeconomics., vol. 24, no. 11, pp. 1121-1131.

BIRCH, S. & GAFNI, A. 2006b, "The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold", J Health Serv.Res.Policy, vol. 11, no. 1, pp. 46-51.

BLACK, W. C. 1990, "The CE plane: a graphic representation of cost-effectiveness", Med. Decis. Making, vol. 10, no. 3, pp. 212-214.

BODENHEIMER, T. 1997, "The Oregon Health Plan--lessons for the nation. First of two parts", N.Engl.J.Med., vol. 337, no. 9, pp. 651-655.

BOYLE, S. 2011, "United kingdom (England) health system review", Health Syst.Transit., vol. 13, no. 1, pp. 1-483.

BRAITHWAITE, R. S., MELTZER, D. O., KING, J. T., JR., LESLIE, D., & ROBERTS, M. S. 2008, "What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule?", Med.Care, vol. 46, no. 4, pp. 349-356.

BRANNON, I. 2005, "What Is a Life Worth?", Regulation no. Winter, pp. 60-63.

BRAZIER, J. 2008, "Valuing health States for use in cost-effectiveness analysis", Pharmacoeconomics., vol. 26, no. 9, pp. 769-779.

BRIDGES, J. F., ONUKWUGHA, E., & MULLINS, C. D. 2010, "Healthcare rationing by proxy: cost-effectiveness analysis and the misuse of the $50,000 threshold in the US", Pharmacoeconomics., vol. 28, no. 3, pp. 175-184.

BRIGGS, A., CLAXTON, K., & SCULPHER, M. 2006, Decision Modelling for Health Economic Evaluation, 1 edn, Oxford University Press.

BRIGGS, A. & FENN, P. 1997, "Trying to do better than average: a commentary on 'statistical inference for cost-effectiveness ratios'", Health Econ., vol. 6, no. 5, pp. 491-495.

BROOK, R. H. 2009, "Assessing the appropriateness of care--its time has come", JAMA, vol. 302, no. 9, pp. 997-998.

BROOK, R. H., CHASSIN, M. R., FINK, A., SOLOMON, D. H., KOSECOFF, J., & PARK, R. E. 1986, "A method for the detailed assessment of the appropriateness of medical technologies", Int J Technol.Assess.Health Care, vol. 2, no. 1, pp. 53-63.

BROUWER, W. B., CULYER, A. J., VAN EXEL, N. J., & RUTTEN, F. F. 2008, "Welfarism vs. extra-welfarism", J Health Econ, vol. 27, no. 2, pp. 325-338.

BROUWER, W. B. & KOOPMANSCHAP, M. A. 2000, "On the economic foundations of CEA. Ladies and gentlemen, take your positions!", J Health Econ, vol. 19, no. 4, pp. 439-459.

BROWN, L. D. 1991, "The national politics of Oregon's rationing plan", Health Aff.(Millwood.), vol. 10, no. 2, pp. 28-51.

BRYAN, S., WILLIAMS, I., & MCIVER, S. 2007, "Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals", Health Econ, vol. 16, no. 2, pp. 179-193.

BRYAN S, SOFAER S, SIEGELBERG T, & GOLD M. 2009, “Has the time come for cost-effectiveness analysis in US health care?” Health Econ Policy Law, vol. 4, no. 4, pp. 425-43

BUIST, A. 1992, The Oregon Experiment: Combining Expert Opinion and Community Values to Set Health Care Priorities. HERG Research Report, Health Economics Research Group, Brunel University, 12.

BULFONE, L., YOUNIE, S., & CARTER, R. 2009A, "Health technology assessment: reflections from the Antipodes", Value Health, vol. 12 Suppl 2, p. S28-S38.

BULFONE, L., YOUNIE, S., & CARTER, R. Health technology assessment: reflections from the Antipodes - Supplementary material. Value Health . 2009b. 2011b.

BUXTON, M. J. 2005, "How much are health-care systems prepared to pay to produce a QALY?", Eur.J Health Econ, vol. 6, no. 4, pp. 285-287.

CADTH 2006, Guidelines for the Economic Evaluation of Health Technologies - Canadian Agency for Drugs and Technologies in Health (CADTH), CADTH, Ottawa, Canada.

CADTH. About Health Technology Assessments. http://www.cadth.ca/en/products/health-technology-assessment/health-technology-assessments . 2011a. 7-15-2011a.

CADTH. Canadian Agency for Drugs and Technologies in Health (CADTH): CCOHTA to CADTH: Our History. http://www.cadth.ca/index.php/en/cadth/corporate-profile/history . 2011b. 4-20-2011b.

CAMPBELL, H.E., TAIT, S., BUXTON, M.J., SHARPLES, L.D., CAINE, N., SCHOFIELD, P.M., WALLWORK, J.. 2001. "A UK trial-based cost-utility analysis of transmyocardial laser revascularization compared to continued medical therapy for treatment of refractory angina pectoris." Eur J Cardiothorac Surg., vol, 20, no. 2, pp. 312-318.

CARO, J. J., MOLLER, J., & GETSIOS, D. 2010, "Discrete Event Simulation: The Preferred Technique for Health Economic Evaluations?", Value.Health.

CARO, J. J., NORD, E., SIEBERT, U., MCGUIRE, A., MCGREGOR, M., HENRY, D., DE, P. G., ATELLA, V., & KOLOMINSKY-RABAS, P. 2010, "The efficiency frontier approach to economic evaluation of health-care interventions", Health Econ.

CARRIER, E., PHAM, H. H., & RICH, E. 2010, Comparative Effectiveness Research and Innovation: Policy Options to Foster Medical Advances, National Institute for Health Care Reform.

CBO 2007, The Long-Term Outlook for Health Care Spending - Congress of the United States. Congressional Budget Office (CBO).

CBO 2011, Congressional Budget Office.

CDR. Common Drug Review - Submission Guidelines for Manufacturers. Viewed September 10th, 2011 <http://cadth.ca/media/cdr/process/CDR_Submission_Guidelines.pdf>

CEA Registry 2010, Tufts Medical Center Cost-Effectiveness Analysis (CEA) Registry. Viewed October 1st, 2011 <https://research.tufts-nemc.org/cear4/SearchingtheCEARegistry/SearchtheCEARegistry.aspx>

Centers for Disease Control and Prevention. International Classification of Diseases, Ninth Revision (ICD-9). Viewed September 10th, 2011 <http://www.cdc.gov/nchs/icd/icd9.htm>

Centre for Reviews and Dissemination. NHS Economic Evaluation Database (NHS-EED). Viewed September 10th, 2011 <http://www.crd.york.ac.uk/crdweb/>

CHALKIDOU, K. & WALLEY, T. 2010, "Using comparative effectiveness research to inform policy and practice in the UK HHS: past, present and future", Pharmacoeconomics., vol. 28, no. 10, pp. 799-811.

CHAMBERS, J. D. & NEUMANN, P. J. 2010, "US healthcare reform: implications for health economics and outcomes research", Expert.Rev.Pharmacoecon.Outcomes.Res., vol. 10, no. 3, pp. 215-216.

CHAMBERS, J. D. & NEUMANN, P. J. 2011, "Listening to Provenge--what a costly cancer treatment says about future Medicare policy", N.Engl.J Med, vol. 364, no. 18, pp. 1687-1689.

CHAN, P.S., STEIN, K., CHOW, T., FENDRICK, M., BIGGER, J.T., VIJAN, S. 2006. “Cost-Effectiveness of a Microvolt T-Wave Alternans Screening Strategy for Implantable Cardioverter-Defibrillator Placement in the MADIT-II–Eligible Population”. J Am Coll Cardiol., vol. 48, no. 1, pp. 112-121

CHAPMAN, R. H., BERGER, M., WEINSTEIN, M. C., WEEKS, J. C., GOLDIE, S., & NEUMANN, P. J. 2004, "When does quality-adjusting life-years matter in cost-effectiveness analysis?", Health Econ, vol. 13, no. 5, pp. 429-436.

CHIM, L., KELLY, P. J., SALKELD, G., & STOCKLER, M. R. 2010, "Are cancer drugs less likely to be recommended for listing by the Pharmaceutical Benefits Advisory Committee in Australia?", Pharmacoeconomics., vol. 28, no. 6, pp. 463-475.

CLAXTON, K., WALKER, S., PALMER, S., & SCULPHER, M. 2010, CHE Research Paper 54: Appropriate Perspectives for Health Care Decisions, The University of York. Centre for Health Economics (CHE).

CLAXTON, K., BRIGGS, A., BUXTON, M. J., CULYER, A. J., MCCABE, C., WALKER, S., & SCULPHER, M. J. 2008, "Value based pricing for NHS drugs: an opportunity not to be missed?", BMJ, vol. 336, no. 7638, pp. 251-254.

CLEEMPUT, I., NEYT, M., THIRY, N., DE LAET, C., & LEYS, M. 2008, Threshold values for cost-effectiveness in health care Health Technology Assessment (HTA)., Belgian Health Care Knowledge Centre (KCE)., Brussels, 100C.

CLEGG, A., COLQUITT, J., SIDHU, M., ROYLE, P., WALKER, A. 2003. “Clinical and cost-effectiveness of surgery for morbid obesity: a systematic review and economic evaluation”., Int J Obes Relat Metab Disord vol. 27, no. 10, pp. 1167-77

CLEMENT, F. M., HARRIS, A., LI, J. J., YONG, K., LEE, K. M., & MANNS, B. J. 2009, "Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada", JAMA, vol. 302, no. 13, pp. 1437-1443.

CMS 1999a, Decision Memo for Breast Biopsy (CAG-00040N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-

memo.aspx?NCAId=13&ver=9&NcaName=Breast+Biopsy&NCDId=272&ncdver=1&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 1999b, Decision Memo for Insulin Infusion Pump (CAG-00041N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=40&ver=8&NcaName=Insulin+Infusion+Pump&NCDId=223&ncdver=2&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2000, Decision Memo for Positron Emission Tomography (FDG) (CAG-00065N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=85&fromdb=true Accessed on August 20th, 2011

CMS 2001, Decision Memo for Diabetic Peripheral Neuropathy with Loss of Protective Sensation (LOPS) (CAG-00059N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=22&ver=10&NCDId=171&ncdver=1&NcaName=Diabetic+Peripheral+Neuropathy+with+Loss+of+Protective+Sensation+(LOPS)&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2002a, Decision Memo for Intravenous Immune Globulin for Autoimmune Mucocutaneous Blistering Diseases (CAG-00109N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=43&ver=8&NcaName=Intravenous+Immune+Globulin+for+Autoimmune+Mucocutaneous+Blistering+Diseases&NCDId=158&ncdver=1&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2002b, Decision Memo for Warm-Up Wound Therapy® a/k/a Noncontact Normothermic Wound Therapy (NNWT) (CAG-00114N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=80&fromdb=true Accessed on August 20th, 2011

CMS 2003a, Decision Memo for Positron Emission Tomography (FDG) for Alzheimer's Disease/Dementia (CAG-00088N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=64&ver=10&NcaName=Positron+Emission+Tomography+(FDG)+for+Alzheimer%2527s+Disease%2FDementia&NCDId=288&ncdver=3&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2003b, Decision Memo for Screening Immunoassay Fecal-Occult Blood Test (CAG-00180N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=87&ver=13&NcaName=Screening+Immunoassay+Fecal-

Occult+Blood+Test&TAId=20&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2003c, Decision Memo for Ventricular Assist Devices as Destination Therapy - Chronic end-stage heart failure patients that meet specified criteria (CAG-00119N). Available: http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=79&ver=8&NcaName=Ventricular+Assist+Devices+as+Destination+Therapy&fromdb=true&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2003d, Notice: Medicare Program; Revised Process for Making Medicare National Coverage Determinations; and GAO,Medicare: Divided Authority 68.

CMS 2005a, Centers for Medicare and Medicaid Services (CMS). Medicare Program - General Information. Available: https://www.cms.gov/MedicareGenInfo/ Accessed on August 20th, 2011

CMS 2005b, Decision Memo for Abarelix for the Treatment of Prostate Cancer (CAG-00238N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=129&ver=35&NcaName=Abarelix+for+the+Treatment+of+Prostate+Cancer&NCDId=306&ncdver=1&IsPopup=y&bc=AAAAAAAAIBAA& Accessed on August 20th, 2011

CMS 2005c, Decision Memo for Aprepitant for Chemotherapy-Induced Emesis (CAG-00248N). Available: http://www.cmms.hhs.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=133&ver=15&NcaName=Aprepitant+for+Chemotherapy-Induced+Emesis&bc=AiAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2005d, Decision Memo for Cochlear Implantation (CAG-00107N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=134&ver=14&NcaName=Cochlear+Implantation&bc=BEAAAAAAEAAA&&fromdb=true Accessed on August 20th, 2011

CMS 2005e, Decision Memo for Smoking & Tobacco Use Cessation Counseling (CAG-00241N). Available: http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=130&ver=22&NcaName=Smoking+%26+Tobacco+Use+Cessation+Counseling&NCDId=308&ncdver=1&CoverageSelection=National&KeyWord=smoking+cessation&KeyWordLookUp=Title&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2005f, Decision Memo for Ultrasound Stimulation for Nonunion Fracture Healing (CAG-00022R). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=135&ver=22&NcaName=Ultrasound+Stimulation+for+Nonunion+Fracture

+Healing&NCDId=65&ncdver=2&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2006a, Decision Memo for Bariatric Surgery for the Treatment of Morbid Obesity (CAG-00250R). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=160&ver=34&NcaName=Bariatric+Surgery+for+the+Treatment+of+Morbid+Obesity&MEDCACId=26&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2006b, Decision Memo for Cardiac Catheterization Performed In Other Than A Hospital Setting (CAG-00166N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=15&ver=17&NcaName=Cardiac+Catheterization+Performed+In+Other+Than+A+Hospital+Setting&TAId=28&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2006c, Decision memo for Electrical Bioimpedance for Cardiac Output Monitoring (CAG-00001R2). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=179&ver=23&NcaName=Electrical+Bioimpedance+for+Cardiac+Output+Monitoring&NCDId=267&ncdver=3&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2006d, Decision Memo for External Counterpulsation (ECP) Therapy (CAG-00002R2). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=162&ver=22&NcaName=External+Counterpulsation+(ECP)+Therapy&NCDId=97&ncdver=2&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2006e, Decision Memo for Infrared Therapy Devices (CAG-00291N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=176&ver=22&NcaName=Infrared+Therapy+Devices&bc=BEAAAAAAEAAA&&fromdb=true Accessed on August 20th, 2011

CMS 2007, Decision Memo for Percutaneous Transluminal Angioplasty (PTA) of the Carotid Artery Concurrent with Stenting (CAG-00085R3). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=194&ver=18&NcaName=Percutaneous+Transluminal+Angioplasty+(PTA)+of+the+Carotid+Artery+Concurrent+with+Stenting&NCDId=201&ncdver=9&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2009a, 2009 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance trust Funds. Available: https://www.cms.gov/reportstrustfunds/downloads/tr2009.pdf Accessed on August 20th, 2011

CMS 2009b, Decision memo for Screening Computed Tomography Colonography (CTC) for Colorectal Cancer (CAG-00396N). Available: http://ww.cms.hhs.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=220&ver=19&NcaName=Screening+Computed+Tomography+Colonography+(CTC)+for+Colorectal+Cancer&TAId=58&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2009c, Decision memo for Screening for Human Immunodeficiency Virus (HIV) infection (CAG-00409N). Available: http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=229&ver=19&NcaName=Screening+for+the+Human+Immunodeficiency+Virus+(HIV)+Infection&bc=BEAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2010a, Centers for Medicaid and Medicare Services (CMS) - Medicare Coverage Database. Available: https://www.cms.gov/center/coverage.asp Accessed on August 20th, 2011

CMS 2010b, Centers for Medicare and Medicaid Services (CMS). Available: http://www.cms.gov/ Accessed on August 20th, 2011

CMS 2010c, Centers for Medicare and Medicaid Services: Physician Fee Schedule Search. Available https://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx Accessed on August 20th, 2011

CMS 2010d, Decision Memo for Magnetic Resonance Angiography of the Abdomen and Pelvis (CAG-00142N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=51&ver=5&NCDId=178&ncdver=2&NcaName=Magnetic+Resonance+Angiography+of+the+Abdomen+and+Pelvis&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2010e, Guidance for the Public, Industry and CMS Staff. Factors CMS Considers in Opening a National Coverage Determination. Available: https://www.cms.gov/medicare-coverage-database/details/medicare-coverage-document-details.aspx?MCDId=6&McdName=Factors+CMS+Considers+in+Opening+a+National+Coverage+Determination&mcdtypename=Guidance+Documents&MCDIndexType=1&bc=BAAIAAAAAAAA& Accessed on August 20th, 2011

CMS 2010f, Medicare Coverage Database - Indexes. Available: https://www.cms.gov/medicare-coverage-database/indexes/national-and-local-indexes.aspx Accessed on August 20th, 2011

CMS 2010g, Medicare Coverage Determination Process - Overview. Available: https://www.cms.gov/determinationprocess/ Accessed on August 20th, 2011

CMS 2010h, Medicare Evidence Development & Coverage Advisory Committee (MEDCAC) Meetings. Available: https://www.cms.gov/FACA/02_MEDCAC.asp Accessed on August 20th, 2011

CMS 2011a, Centers for Medicare and Medicaid Services (CMS). National Health Expenditure Data. Available: https://www.cms.gov/nationalhealthexpenddata/ Accessed on August 20th, 2011

CMS 2011b, CMS's Office of the Actuary. Available: https://www.cms.gov/cmsleadership/09_office_oact.asp Accessed on August 20th, 2011

CMS 2011c, Decision Memo for autologous cellular immunotherapy treatment of metastatic prostate cancer (CAG-00422N). Available: http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=247&ver=12&NcaName=Autologous+Cellular+Immunotherapy+Treatment+of+Metastatic+Prostate+Cancer&TimeFrame=7&DocType=All&bc=AgAAYAAAIAAA Accessed on August 20th, 2011

CMS 2011d, Decision memo for Counseling to Prevent Tobacco Use (CAG-00420N). Available: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=242&ver=12&NcaName=Counseling+to+Prevent+Tobacco+Use&NCDId=342&ncdver=1&IsPopup=y&bc=AAAAAAAAIAAA& Accessed on August 20th, 2011

CMS 2011e, Medicaid - General Information - Overview. Available: http://www.cms.gov/MedicaidGenInfo/ Accessed on August 20th, 2011

COAST, J. 2001, "Who wants to know if their care is rationed? Views of citizens and service informants", Health Expect., vol. 4, no. 4, pp. 243-252.

COAST, J. 2004, "Is economic evaluation in touch with society's health values?", BMJ, vol. 329, no. 7476, pp. 1233-1236.

COLQUITT, J.L., GREEN, C., SIDHU, M.K., HARTWELL, D., WAUGH, N. 2004. “Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes”. Health Technol Assess., vol. 8, no. 43, pp. 1-171

CONSUMER REPORTS. Consumer Reports. Best Buy drugs. Viewed September 10th, 2011 <http://www.consumerreports.org/health/about/best-buy-drugs.htm>

CRAIG, B.M., TSENG, D.S. 2002. “Cost-Effectiveness of Gastric bypass for Severe Obesity”. Am J Med., vol. 113, pp. 491-498

CROMWELL, D. A., VINEY, R., HALSALL, J., & HINDLE, D. 1998, "Linking measures of health gain to explicit priority setting by an area health service in Australia", Soc.Sci.Med, vol. 47, no. 12, pp. 2067-2074.

CULYER, A. 1989, "Commodities, characteristics of commodities, characteristics of people, utilities, and the quality of life," in Quality of life. Perspectives and policies, S. Baldwin, ed., Routledge, London, pp. 9-27.

CULYER, A., MCCABE, C., BRIGGS, A., CLAXTON, K., BUXTON, M., AKEHURST, R., SCULPHER, M., & BRAZIER, J. 2007, "Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence", J.Health Serv.Res.Policy, vol. 12, no. 1, pp. 56-58.

CUTLER, D. M. 2005, Your Money or Your Life: Strong Medicine for America's Health Care System Oxford University Press, USA.

CUTLER, D. M., DAVIS, K., & STREMIKIS, K. 2009, Why Health Reform Will Bend the Cost Curve - The Commonwealth Fund.

CUTLER, D. M. & LY, D. P. 2011, "The (paper) work of medicine: understanding international medical costs", J Econ Perspect., vol. 25, no. 2, pp. 3-25.

CUTLER, D. M., RICHARDSON, E., KEELER, T. E., & STAIGER, D. 1997, "Measuring the Health of the U.S. Population", Brookings Papers on Economic Activity.Microeconomics, vol. 1997, pp. 217-282.

DAKIN, H. A., DEVLIN, N. J., & ODEYEMI, I. A. 2006, ""Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making", Health Policy, vol. 77, no. 3, pp. 352-367.

DALY, J., SINDONE, A. P., THOMPSON, D. R., HANCOCK, K., CHANG, E., &

DAVIDSON, P. 2002, "Barriers to participation in and adherence to cardiac rehabilitation programs: a critical literature review", Prog.Cardiovasc.Nurs., vol. 17, no. 1, pp. 8-17.

DAN, Y.Y., WAI, C.T., YEOH, K.G., LIM, S.G. 2006. “Prophylactic strategies for hepatitis B patients undergoing liver transplant: a cost-effectiveness analysis”. Liver Transpl., vol. 12, no. 5, pp. 736-746

DANIELS, N. 1991, "Is the Oregon rationing plan fair?", JAMA, vol. 265, no. 17, pp. 2232-2235.

DANIELS, N. 2000, "Accountability for reasonableness", BMJ, vol. 321, no. 7272, pp. 1300-1301.

DANIELS, N. 2008, Just health: Meeting health needs fairly, 1 edn, Cambridge University Press, Cambridge.

DANZON, P. M. & FURUKAWA, M. F. 2003, "Prices and availability of pharmaceuticals: evidence from nine countries", Health Aff.(Millwood.), vol. Suppl Web Exclusives, pp. W3-36.

DAOUD, Y.J., AMIN, K.G. 2006. “Comparison of cost of immune globulin intravenous therapy to conventional immunosuppressive therapy in treating patients with autoimmune mucocutaneous blistering diseases”. Int Immunopharmacol., vol. 6, no. 4, pp. 600-606

DAVIS, K., SCHOEN, C., SCHOENBAUM, S. C., AUDET, A. J., DOTY, M. M., HOLMGREN, A. L., & KRISS, J. L. 2006, Mirror, Mirror on the Wall: An Update on the Quality of American Health Care Through the Patient's Lens 12.

DAVIS, K., SCHOEN, C., SCHOENBAUM, S. C., DOTY, M. M., HOLMGREN, A. L., KRISS, J. L., & SHEA, K. K. 2007, Mirror, Mirror on the Wall: An International Update on the Comparative Performance of American Health Care 59.

DAVIS, K., SCHOEN, C., & STREMIKIS, K. 2010, Mirror, Mirror on the Wall. How the performance of the US Health Care System Compares Internationally. 2010 Update. The Commonwealth Fund.

DE POUVOURVILLE G. 2010, "A French approach to cost-effectiveness analysis?", Eur J Health Econ, vol. 11, no. 6, pp. 521-523.

DELAUNE, J. & EVERETT, W. 2008, New England Health Care Institute (NEHI); Waste and Inefficiency in the U.S. Health Care System: Clinical Care: A Comprehensive Analysis in Support of System-Wide Improvements.

DENAVAS-WALT, C., PROCTOR, B. D., & SMITH, J. C. 2008, Income, Poverty, and Health Insurance Coverage in the United States: 2007 - U.S.Department of Commerce. Economics and Statistics Administration. U.S. Census Bureau.

Department for Transport 2009, The Accidents Sub-Objective. Transport Analysis Guidance (TAG) Unit 3.4.1.

DERP. Drug Effectiveness Review Project (DERP). Viewed September 10th, 2011 <http://www.ohsu.edu/ohsuedu/research/policycenter/DERP/>

DEVLIN, N., DAKIN, H. A., RICE, N., PARKIN, D., & O'NEILL, P. "NICE's cost effectiveness threshold revisited: new evidence on the influence of cost effectiveness and other factors on NICE decisions".

DEVLIN, N. & SUSSEX, J. 2011, Incorporating Multiple Criteria in HTA. Methods and Processes.

DEVLIN, N. & PARKIN, D. 2004, "Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis", Health Econ, vol. 13, no. 5, pp. 437-452.

DHANASIRI, S. K. & PULIYEL, J. M. 2007, "Regulating vaccines: can health-economics tools be used profitably?", Indian Pediatr., vol. 44, no. 1, pp. 11-14.

DHRUVA, S. S., PHURROUGH, S. E., SALIVE, M. E., & REDBERG, R. F. 2009, "CMS's landmark decision on CT colonography--examining the relevant data", N.Engl.J.Med., vol. 360, no. 26, pp. 2699-2701.

DIENER, A., O'BRIEN, B., & GAFNI, A. 1998, "Health care contingent valuation studies: a review and classification of the literature", Health Econ, vol. 7, no. 4, pp. 313-326.

DoD PEC. The Department of Defense (DoD) PharmacoEconomic Center (PEC). Viewed September 10th, 2011 <http://www.pec.ha.osd.mil/mission.htm>

DoH. A new value-based approach to the pricing of branded medicines - a consultation. Medicines, Pharmacy & Industry Group, Department of Health (DoH). Viewed September 10th, 2011 <http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_122793.pdf . 12-16-2010. 4-20-2011>

DOLAN, P. & COOKSON, R. 2000, "A qualitative study of the extent to which health gain matters when choosing between groups of patients", Health Policy, vol. 51, no. 1, pp. 19-30.

DOLAN, P., SHAW, R., TSUCHIYA, A., & WILLIAMS, A. 2005, "QALY maximisation and people's preferences: a methodological review of the literature", Health Econ, vol. 14, no. 2, pp. 197-208.

DONALDSON, C. 2008, Two cheers for the NHS! Why do we have an NHS? An economist's perspective, National Health Executive.

DONALDSON, C., GERARD, K., MITTON, C., JAN, S., & WISEMAN, V. 2005, Economics of health care financing: the visible hand, 2 edn, Palgrave, Macmillan, London.

DONALDSON, C. 1998, "The (near) equivalence of cost-effectiveness and cost-benefit analyses. Fact or fallacy?", Pharmacoeconomics., vol. 13, no. 4, pp. 389-396.

DONALDSON, C., BATE, A., BRAMBLEBY, P., & WALDNER, H. 2008, "Moving forward on rationing: an economic view", BMJ, vol. 337, p. a1872.

DONALDSON, C., BIRCH, S., & GAFNI, A. 2002, "The distribution problem in economic evaluation: income and the valuation of costs and consequences of health care programmes", Health Econ, vol. 11, no. 1, pp. 55-70.

DRANOVE, D., HUGHES, E. F., & SHANLEY, M. 2003, "Determinants of HMO formulary adoption decisions", Health Serv.Res., vol. 38, no. 1 Pt 1, pp. 169-190.

DRUMMOND, M. F., SCULPHER, M. J., TORRANCE, G. W., O'BRIEN, B. J., & STODDART, G. L. 2005, Methods for the Economic Evaluation of Health Care Programmes, 3 edn, Oxford University Press.

DUINTJER TEBBENS, R. J., THOMPSON, K. M., HUNINK, M. G., MAZZUCHI, T. A., LEWANDOWSKI, D., KUROWICKA, D., & COOKE, R. M. 2008, "Uncertainty and sensitivity analyses of a dynamic economic evaluation model for vaccination programs", Med Decis.Making, vol. 28, no. 2, pp. 182-200.

EARNSHAW, S. R. & DENNETT, S. L. 2003, "Integer/linear mathematical programming models: a tool for allocating healthcare resources", Pharmacoeconomics., vol. 21, no. 12, pp. 839-851.

EDDY, D. M. 1991, "Oregon's methods. Did cost-effectiveness analysis fail?", JAMA, vol. 266, no. 15, pp. 2135-2141.

EICHLER, H. G., KONG, S. X., GERTH, W. C., MAVROS, P., & JONSSON, B. 2004, "Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge?", Value.Health, vol. 7, no. 5, pp. 518-528.

ELBASHA, E. H. & MESSONNIER, M. L. 2004, "Cost-effectiveness analysis and health care resource allocation: decision rules under variable returns to scale", Health Econ., vol. 13, no. 1, pp. 21-35.

ELSHAUG, A. G., HILLER, J. E., TUNIS, S. R., & MOSS, J. R. 2007, "Challenges in Australian policy processes for disinvestment from existing, ineffective health care practices", Aust.New Zealand.Health Policy, vol. 4, p. 23.

ELSHAUG, A. G., WATT, A. M., MOSS, J. R., & HILLER, J. E. 2009, Policy Perspectives on the Obsolescence of Health Technologies in Canada. Canadian Agency for Drugs and Technologies in Health (CADTH).

EPSTEIN, D. M., CHALABI, Z., CLAXTON, K., & SCULPHER, M. 2007, "Efficiency, equity, and budgetary policies: informing decisions using mathematical programming", Med Decis.Making, vol. 27, no. 2, pp. 128-137.

ERICKSON, P., WILSON, R., & SHANNON, I. 1995, "Years of healthy life", Healthy.People.2000.Stat.Notes no. 7, pp. 1-15.

EVANS, A. W. 2007, Fatal Train Accidents on Britain's Main Line Railways: End of 2006 Analysis.

EVANS, C., TAVAKOLI, M., & CRAWFORD, B. 2004, "Use of quality adjusted life years and life years gained as benchmarks in economic evaluations: a critical appraisal", Health Care Manag.Sci., vol. 7, no. 1, pp. 43-49.

FARRELL, D., JENSEN, E., KOCHER, B., LOVEGROVE, N., MELHEM, F., MENDONCA, L., & PARISH, B. 2008, McKinsey Global Institute. Accounting for the cost of US health care: A new look at why Americans spend more, McKinsey Global Institute.

FDA. Food and Drug Administration. Summary summary basis for regulatory action: sipuleucel-T (Provenge). Viewed September 10th, 2011 <http://www.fda.gov/downloads/BiologicsBloodVaccines/CellularGeneTherapyProducts/ApprovedProducts/UCM213114.pdf>

Federal Register 1989a, Federal Register - Medicare Program; Criteria and Procedures for Making Medical Services Coverage Decisions That Relate to Health Care Technology 54 Fed. Reg. 4302.

Federal Register 2010, Federal Register. November 21st, 2011 <http://www.gpoaccess.gov/fr/>

FENWICK, E., CLAXTON, K., & SCULPHER, M. 2001, "Representing uncertainty: the role of cost-effectiveness acceptability curves", Health Econ, vol. 10, no. 8, pp. 779-787.

FENWICK, E., O'BRIEN, B. J., & BRIGGS, A. 2004, "Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions", Health Econ, vol. 13, no. 5, pp. 405-415.

FISCHER KE, LEIDL R, & ROGOWSKI WH. 2011, “A structured tool to analyse coverage decisions: development and feasibility test in the field of cancer screening and prevention.” Health Policy, vol. 101, no. 3, pp. 290-9

FISHER, E. S., WENNBERG, D. E., STUKEL, T. A., GOTTLIEB, D. J., LUCAS, F. L., & PINDER, E. L. 2003, "The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care", Ann Intern Med, vol. 138, no. 4, pp. 273-287.

FITCH, K., BERNSTEIN, S. J., AGUILAR, M. S., BURNAND, B., LACALLE, J. R., LAZARO, P., VAN HET LOO, M., MCDONNELL, J., VADER, J., & KAHAN, J. P. 2001, The RAND/UCLA Appropriateness Method User's Manual.

FMCP Format Executive Committee 2010, The AMCP Format for Formulary Submissions: Version 3.0.

FOGLIA, M. B., PEARLMAN, R. A., BOTTRELL, M. M., ALTEMOSE, J. A., & FOX, E. 2008, "Priority setting and the ethics of resource allocation within VA healthcare facilities: results of a survey", Organ Ethic., vol. 4, no. 2, pp. 83-96.

FOLLAND, S., GOODMAN, A. C., & STANO, M. 2003, The Economics of Health and Health Care, 4 edn, Prentice Hall.

FOOTE, S. B. 2002, "Why Medicare cannot promulgate a national coverage rule: a case of regula mortis", J.Health Polit.Policy Law, vol. 27, no. 5, pp. 707-730.

FOOTE, S. B., HALPERN, R., & WHOLEY, D. R. 2005, "Variation in Medicare's local coverage policies: content analysis of local medical review policies", Am.J Manag.Care, vol. 11, no. 3, pp. 181-187.

FOX, D. M. & LEICHTER, H. M. 1991, "Rationing care in Oregon: the new accountability", Health Aff.(Millwood.), vol. 10, no. 2, pp. 7-27.

FOX, J. 2010, "The Hidden Role of Cost: Medicare Decisions, Transparency and Public Trust", University of Cincinnati Law Review, vol. 79, pp. 1-51.

FOX-RUSHBY, J. A. & HANSON, K. 2001, "Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis", Health Policy Plan., vol. 16, no. 3, pp. 326-331.

FRANCIS, H.W., CHEE, N., YEAGLE, J., CHENG, A., NIPARKO, J.K. 2002. “Impact of cochlear implants on the functional health status of older adults”., Laryngoscope vol. 112, no. 8 Pt 1, pp. 1482-1488

FREEDBERG, K. A., TOSTESON, A. N., COTTON, D. J., & GOLDMAN, L. 1992, "Optimal management strategies for HIV-infected patients who present with cough or dyspnea: a cost-effective analysis", J Gen.Intern.Med., vol. 7, no. 3, pp. 261-272.

FRICKE, F. U. & DAUBEN, H. P. 2009, "Health technology assessment: a perspective from Germany", Value Health, vol. 12 Suppl 2, p. S20-S27.

FROEHLICH, F., BURNAND, B., PACHE, I., VADER, J. P., FRIED, M.,

SCHNEIDER, C., KOSECOFF, J., KOLODNY, M., DUBOIS, R. W., BROOK, R. H., & GONVERS, J. J. 1997, "Overuse of upper gastrointestinal endoscopy in a country with open-access endoscopy: a prospective study in primary care", Gastrointest.Endosc., vol. 45, no. 1, pp. 13-19.

FRUITS, E., HILLARD, A., & LEWIS, L. 2009, The Oregon Health Plan. A "Bold Experiment" That Failed, Cascade Policy Institute. Viewed October 15th, 2011 < http://cascadepolicy.org/wp-content/uploads/2010/09/Oregon_Health_Plan_-_The_Bold_Experiment_That_Failed.pdf>

GAFNI, A. 2006, "Economic Evaluation of Health-care Programmes: Is CEA Better than CBA?", Environmental & Resource Economics, vol. 34, pp. 407-418.

GAFNI, A. & BIRCH, S. 2003, "NICE methodological guidelines and decision making in the National Health Service in England and Wales", Pharmacoeconomics., vol. 21, no. 3, pp. 149-157.

GAFNI, A. & BIRCH, S. 2006, "Incremental cost-effectiveness ratios (ICERs): the silence of the lambda", Soc.Sci.Med., vol. 62, no. 9, pp. 2091-2100.

GARBER, A., GOLDMAN, D. P., & JENA, A. B. 2007, "The promise of health care cost containment", Health Aff.(Millwood.), vol. 26, no. 6, pp. 1545-1547.

GARBER, A. M. 2001, "Evidence-based coverage policy", Health Aff.(Millwood.), vol. 20, no. 5, pp. 62-82.

GARBER, A. M. & PHELPS, C. E. 1997, "Economic foundations of cost-effectiveness analysis", J Health Econ, vol. 16, no. 1, pp. 1-31.

GARBER, A. M., WEINSTEIN, M. C., TORRANCE, G. W., & KAMLET, M. S. 1996, "Theoretical foundations of cost-effectiveness analysis," in Cost-Effectiveness in Health and Medicine, M. R. Gold et al., eds., Oxford University Press, New York.

GARG, S. & AHNEN, D. J. 2010, "Is computed tomographic colonography being held to a higher standard?", Ann.Intern.Med., vol. 152, no. 3, pp. 178-181.

GARNER, S. & LITTLEJOHNS, P. 2011, "Disinvestment from low value clinical interventions: NICEly done?", BMJ, vol. 343, p. d4519.

GEORGE, B., HARRIS, A., & MITCHELL, A. 2001, "Cost-effectiveness analysis and the consistency of decision making: evidence from pharmaceutical reimbursement in australia (1991 to 1996)", Pharmacoeconomics., vol. 19, no. 11, pp. 1103-1109.

GERBER, A., STOCK, S., & DINTSIOS, C. M. 2011, "Reflections on the Changing Face of German Pharmaceutical Policy: How Far is Germany from Value-Based Pricing?", Pharmacoeconomics., vol. 29, no. 7, pp. 549-553.

GILLICK, M. R. 2004, "Medicare coverage for technological innovations--time for new criteria?", N.Engl.J Med., vol. 350, no. 21, pp. 2199-2203.

GINSBURG, P. B. 2004, "Controlling health care costs", N.Engl.J Med, vol. 351, no. 16, pp. 1591-1593.

GINSBURG, P. B. 2008, High and rising health care costs: Demystifying U.S. health care spending.

GLASS, N., NELSON, H. D., & VILLEMYER, K. 2005, Screening for Gonorrhea: Update of the Evidence for the U.S. Preventive Services Task Force. Agency for Healthcare Research and Quality, U.S. Rockville, MD, 05-0579-B.

GOLD MR, FRANKS P, SIEGELBERG T, & SOFAER S. 2007, “Does providing cost-effectiveness information change coverage priorities for citizens acting as social decision makers?” Health Policy, vol. 83, no. 1, pp. 65-72

GOLD, M. & BRYAN, S. 2007, "A response to Birch and Gafni--some reasons to be cheerful about NICE", Health Econ Policy Law, vol. 2, no. Pt 2, pp. 209-216.

GOLD, M. R., SIEGEL, J. E., RUSSELL, L. B., & WEINSTEIN, M. C. 1996, Cost-effectiveness in health and medicine Oxford University Press, New York.

GOLDMAN, L., GORDON, D. J., RIFKIND, B. M., HULLEY, S. B., DETSKY, A. S., GOODMAN, D. W., KINOSIAN, B., & WEINSTEIN, M. C. 1992, "Cost and health implications of cholesterol lowering", Circulation, vol. 85, no. 5, pp. 1960-1968.

GREENBERG, D., EARLE, C., FANG, C. H., ELDAR-LISSAI, A., & NEUMANN, P. J. 2010, "When is cancer care cost-effective? A systematic overview of cost-utility analyses in oncology", J Natl.Cancer Inst., vol. 102, no. 2, pp. 82-88.

GREENBERG, D., WINKELMAYER, W. C., & NEUMANN, P. J. Prevailing judgments about society's willingness to pay for a QALY: do they vary by country? Have they changed over time? Value Health A-60. 2006.

GREWAL, R., COTE, J. A., & BAUMGARTNER, H. 2004, "Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing", Marketing Science, vol. 23, no. 4, pp. 519-529.

GROSSE, S. D. 2008, "Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold", Expert.Rev.Pharmacoecon.Outcomes.Res., vol. 8, no. 2, pp. 165-178.

GUO, S., COUNTE, M.A., GILLESPIE, K.N., SCHMITZ, H. 2003. “Cost-effectiveness of adjunctive hyperbaric oxygen in the treatment of diabetic ulcers”., Int J Technol Assess Health Care. vol. 19, no. 4, pp. 731-737

HADORN, D. C. 1991, "Setting health care priorities in Oregon. Cost-effectiveness meets the rule of rescue", JAMA, vol. 265, no. 17, pp. 2218-2225.

HARRIS, A. H., HILL, S. R., CHIN, G., LI, J. J., & WALKOM, E. 2008, "The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994-2004", Med.Decis.Making, vol. 28, no. 5, pp. 713-722.

HAS. The Haute Autorité de santé (HAS) - About HAS. Viewed August 20th, 2011 <http://www.has-sante.fr/portail/jcms/c_5443/english>

HAYS V. SEBELIUS, No. 08-5508. 2009.

Health Canada. Health Canada - Health Care System. Viewed August 20th, 2011 <http://www.hc-sc.gc.ca/hcs-sss/index-eng.php#skipall>

HEED 2010, Health Economic Evaluation Database (HEED). Viewed August 25th, 2011 < http://onlinelibrary.wiley.com/book/10.1002/9780470510933>

HEMINGWAY, H., CHEN, R., JUNGHANS, C., TIMMIS, A., ELDRIDGE, S., BLACK, N., SHEKELLE, P., & FEDER, G. 2008, "Appropriateness criteria for coronary angiography in angina: reliability and validity", Ann.Intern.Med, vol. 149, no. 4, pp. 221-231.

HENRY, D. A., HILL, S. R., & HARRIS, A. 2005, "Drug prices and value for money: the Australian Pharmaceutical Benefits Scheme", JAMA, vol. 294, no. 20, pp. 2630-2632.

HIRTH, R. A., CHERNEW, M. E., MILLER, E., FENDRICK, A. M., & WEISSERT, W. G. 2000, "Willingness to pay for a quality-adjusted life year: in search of a standard", Med.Decis.Making, vol. 20, no. 3, pp. 332-342.

HOFFMAN, K. & JACKSON, S. 2003, A review of the evidence for the effectiveness and costs of interventions preventing the burden of noncommunicable diseases: How can health systems respond?

HOLLENBEAK, C.S., LOWE, V.J., STACK, B.C. 2001. “The cost-effectiveness of fluorodeoxyglucose 18-F positron emission tomography in the N0 neck”., Cancer vol. 92, no. 9, pp. 2341-2348.

HOLM, S. 1998, "The second phase of priority setting. Goodbye to the simple solutions: the second phase of priority setting in health care", BMJ, vol. 317, no. 7164, pp. 1000-1002.

HUSSEY, P. S., ANDERSON, G. F., OSBORN, R., FEEK, C., MCLAUGHLIN, V., MILLAR, J., & EPSTEIN, A. 2004, "How does the quality of care compare in five countries?", Health Aff.(Millwood.), vol. 23, no. 3, pp. 89-99.

INAHTA. International Network of Agencies for Health Technology Assessment (INAHTA) - Members - CADTH. Viewed August 20th, 2011 <http://www.inahta.org/Members/CADTH/>

IQWiG. General Methods for evaluating the relation between cost and benefit - version 1.0 - IQWiG. Viewed August 20th, 2011 <https://www.iqwig.de/download/General_Methods_for_the_Assessment_of_the_Relation_of_Benefits_to_Costs.pdf >

IQWiG 2009b, Institute for Quality and Efficiency in Health Care (IQWiG): General Methods for the Assessment of the Relation of Benefits to Costs. Version 1.0.

IQWiG. Institute for Quality and Efficiency in Health Care (IQWiG). Viewed August 20th, 2011 <https://www.iqwig.de/index.2.en.html>

IQWiG. Responsibilities and objectives of IQWiG. Institute for Quality and Efficiency in Health Care (IQWiG). Viewed August 20th, 2011 <https://www.iqwig.de/about-us.21.en.html>

ISMAIL, M. A. Spending on Lobbying Thrives: Drug and health products industries invest $182 million to influence legislation - The Center for Public Integrity. Viewed August 20th, 2011 <http://projects.publicintegrity.org/rx//report.aspx?aid=823>

ISPOR. International Society For Pharmacoeconomics and Outcomes Research - Global Health Care Systems Roadmaps. Viewed August 20th, 2011 <http://www.ispor.org/htaroadmaps/default.asp>

JACOBS, L., MARMOR, T., & OBERLANDER, J. 1999, "The Oregon Health Plan and the political paradox of rationing: what advocates and critics have claimed and what Oregon did", J.Health Polit.Policy Law, vol. 24, no. 1, pp. 161-180.

JOHANNESSON, M. 1995, "The relationship between cost-effectiveness analysis and cost-benefit analysis", Soc.Sci.Med., vol. 41, no. 4, pp. 483-489.

JOHANNESSON, M. & O'CONOR, R. M. 1997, "Cost-utility analysis from a societal perspective", Health Policy, vol. 39, no. 3, pp. 241-253.

JOHANNESSON, M. & WEINSTEIN, M. C. 1993, "On the decision rules of cost-effectiveness analysis", J.Health Econ., vol. 12, no. 4, pp. 459-467.

JONSSON, B. 2009A, "Ten arguments for a societal perspective in the economic evaluation of medical innovations", Eur.J Health Econ, vol. 10, no. 4, pp. 357-359.

JONSSON, E. 2009B, "History of health technology assessment in Sweden", Int J Technol.Assess.Health Care, vol. 25 Suppl 1, pp. 42-52.

JONSSON, E. 2010, "Health technology assessment and comparative effectiveness in Sweden", Value Health, vol. 13 Suppl 1, p. S6-S7.

JORDAN, W.D., ROYE, G.D., FISHER, W.S., REDDEN, D., MCDOWELL, H.A. 1998. “A cost comparison of balloon angioplasty and stenting versus endarterectomy for the treatment of carotid artery stenosis”. J Vasc Surg., vol. 27, no. 1, pp. 16-22

KANIS, J. A., BORGSTROM, F., ZETHRAEUS, N., JOHNELL, O., ODEN, A., & JONSSON, B. 2005, "Intervention thresholds for osteoporosis in the UK", Bone, vol. 36, no. 1, pp. 22-32.

KANIS, J. A. & JONSSON, B. 2002, "Economic evaluation of interventions for osteoporosis", Osteoporos.Int., vol. 13, no. 10, pp. 765-767.

KAPLAN, R. & BUSH, J. 1982, "Health-related quality of life measurement for evaluation research and policy analysis", Health Psychology, vol. 1, no. 1, pp. 61-80.

KARLSSON, G. & JOHANNESSON, M. 1996, "The decision rules of cost-effectiveness analysis", Pharmacoeconomics., vol. 9, no. 2, pp. 113-120.

KEENAN, P. S., NEUMANN, P. J., & PHILLIPS, K. A. 2006, "Biotechnology and Medicare's new technology policy: lessons from three case studies", Health Aff.(Millwood.), vol. 25, no. 5, pp. 1260-1269.

KIBERD, B.A., LARSON, T. 2000. “Estimating the benefits of solitary pancreas transplantation in nonuremic patients with type 1 diabetes mellitus: a theoretical analysis”. Transplantation, vol. 70, pp. 1121–1127

KLEINMAN, L. C., KOSECOFF, J., DUBOIS, R. W., & BROOK, R. H. 1994, "The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States", JAMA, vol. 271, no. 16, pp. 1250-1255.

KO, D. T., KRUMHOLZ, H. M., WANG, Y., FOODY, J. M., MASOUDI, F. A., HAVRANEK, E. P., YOU, J. J., ALTER, D. A., STUKEL, T. A., NEWMAN, A. M., & TU, J. V. 2007, "Regional differences in process of care and outcomes for older acute myocardial infarction patients in the United States and Ontario, Canada", Circulation, vol. 115, no. 2, pp. 196-203.

LARSEN, G., HALLSTROM, A., MCANULTY, J., PINSKI, S., OLARTE, A., SULLIVAN, S., BRODSKY, M., POWELL, J., MARCHANT, C., JENNINGS, C., AKIYAMA, T.; AVID INVESTIGATORS. 2002. “Cost-effectiveness of the implantable cardioverter-defibrillator versus antiarrhythmic drugs in survivors of serious ventricular tachyarrhythmias: results of the Antiarrhythmics Versus Implantable Defibrillators (AVID) economic analysis substudy”. Circulation, vol. 105, no. 17, pp. 2049-2057

LAUFER, F. 2005, "Thresholds in cost-effectiveness analysis--more of the story", Value Health, vol. 8, no. 1, pp. 86-87.

LAUPACIS, A. 2002, "Inclusion of drugs in provincial drug benefit programs: who is making these decisions, and are they the right ones?", CMAJ., vol. 166, no. 1, pp. 44-47.

LAUPACIS, A., FEENY, D., DETSKY, A. S., & TUGWELL, P. X. 1992, "How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations", CMAJ., vol. 146, no. 4, pp. 473-481.

LEE, C.H., EGGLIN, T.K., PHILPOTTS, L., MAINIERO, M.B., TOCINO I. 1997. “Cost-Effectiveness of stereotactic core needle biopsy: Analysis by means of mammographic findings”. Radiology, vol. 202, pp.849-854

LEICHTER, H. M. 1999, "Oregon's bold experiment: whatever happened to rationing?", J.Health Polit.Policy Law, vol. 24, no. 1, pp. 147-160.

LEVY, M.M., BAUM, R.A., CARPENTER, J.P. 1998. “Endovascular surgery based solely on noninvasive preprocedural imaging”. J Vasc Surg., vol. 28, no. 6, pp. 995-1003

LFN - Pharmaceutical Benefits Board 2007, The Swedish Pharmaceutical Reimbursement System.

LIBERMAN, L., FENG, T. L., DERSHAW, D. D., MORRIS, E. A., & ABRAMSON, A. F. 1998, "US-guided core breast biopsy: use and cost-effectiveness", Radiology, vol. 208, no. 3, pp. 717-723.

LOOMES, G. 2010, "Valuing life years and QALYs: 'transferability' and 'convertibility' of values across the public sector," in Cost-Effectiveness Thresholds: Economic and ethical issues, A. Towse, C. Pritchard, & N. Devlin, eds..

LOPERT, R. 2009, Evidence-Based Decision-Making Within Australia's Pharmaceutical Benefits Scheme.

LORD, J., LAKING, G., & FISCHER, A. 2004, "Health care resource allocation: is the threshold rule good enough?", J.Health Serv.Res.Policy, vol. 9, no. 4, pp. 237-245.

LUCE, B. R., DRUMMOND, M., JONSSON, B., NEUMANN, P. J., SCHWARTZ, J. S., SIEBERT, U., & SULLIVAN, S. D. 2010, "EBM, HTA, and CER: clearing the confusion", Milbank Q., vol. 88, no. 2, pp. 256-276.

MARMOR, T. 1970, The Politics of Medicare Chicago: Aldine.

MACARIO, A., DEXTER, F. 2002. “Is noncontact normothermic wound therapy cost effective for the treatment of stages 3 and 4 pressure ulcers”. Wounds, vol. 14, no. 3, pp. 93-106

MARTIN, S.C., GAGNON, D.D., ZHANG, L., BOKEMEYER, C., VAN MARWIJK, KOOY, M., VAN HOUT, B. 2003. “Cost utility analysis of survival with epoetin-alfa versus placebo in stage IV breast cancer”. Pharmacoeconomics, vol. 21, no. 16, pp. 1153-1169

MARTIN S., RICE N., SMITH P.C. 2008a. Does health care spending improve health outcomes? Evidence from English programme budgeting data. J Health Econ, vol. 27, pp. 826-42

MARTIN S., RICE N., SMITH P.C. 2008b. The link between health care spending and health outcomes for the new English primary care trusts. York: Centre for Health Economics, 2008. Research paper no. 42 Viewed January 30th 2012 <http://www.york.ac.uk/media/che/documents/papers/researchpapers/rp42_health_care_spending_and_health_outcomes_english_pcts.pdf>

MARTIN S., RICE N., SMITH P.C. 2009. The link between health spending and health outcomes for the new English primary care trusts. London: The Health Foundation. Viewed January 30th 2012 <http://www.health.org.uk/publications/programme-budgeting-report/>

MATCHAR, D. B., KULASINGAM, S. L., MCCRORY, D. C., PATWARDHAN, M. B., RUTSCHMANN, O. T., SAMSA, G. P., & SCHMECHEL, D. E. 2001, Use of Positron Emission Tomography and other neuroimaging techniques in the diagnosis and management of Alzheimer's disease and dementia, AHRQ.

MCCABE, C., CLAXTON, K., & CULYER, A. J. 2008, "The NICE cost-effectiveness threshold: what it is and what that means", Pharmacoeconomics., vol. 26, no. 9, pp. 733-744.

MEADS, C., SALAS, C., ROBERTS, T., MOORE, D., FRY-SMITH, A., HYDE, C. 2002. “Clinical effectiveness and cost utility of photodynamic therapy for wet age-related macular degeneration - NHS R&D HTA Programme”. Viewed February 1, 2008 <http://www.nice.org.uk/nicemedia/pdf/maculardegenerationassessmentreport.pdf>.

MECKLEY, L. M., GREENBERG, D., COHEN, J. T., & NEUMANN, P. J. 2010, "The adoption of cost-effectiveness acceptability curves in cost-utility analyses", Med.Decis.Making, vol. 30, no. 3, pp. 314-319.

Medicare 2010, Medicare.gov. The official US government site for Medicare. Viewed August 20th, 2011 < http://www.medicare.gov/default.aspx>

MedPAC 2007, Outpatient hospital services payment system. Payment basics. Medicare Payment Advisory Commission (MedPAC). Viewed August 20th, 2011. <http://www.medpac.gov/documents/MedPAC_Payment_Basics_07_OPD.pdf>

MedPAC 2008, Hospital acute inpatient service payment system. Payment basics. Medicare Payment Advisory Commission (MedPAC). Revised October 2008. Viewed August 20th, 2011. <http://www.medpac.gov/documents/MedPAC_Payment_Basics_07_hospital.pdf>

MELTZER, D. 1997, "Accounting for future costs in medical cost-effectiveness analysis", J Health Econ, vol. 16, no. 1, pp. 33-64.

MENON, D. & STAFINSKI, T. 2009, "Health technology assessment in Canada: 20 years strong?", Value Health, vol. 12 Suppl 2, p. S14-S19.

MILES, K.A. 2001. "An approach to demonstrating cost-effectiveness of diagnostic imaging modalities in Australia illustrated by positron emission tomography". Australas Radiol., vol. 45, no. 1, pp. 9-18

MIPPA. Medicare Improvements for Patients and Providers Act (MIPPA) of 2008. 1-3-2008.

MITCHELL, J. B., HABER, S. G., KHATUTSKY, G., & DONOGHUE, S. 2002, "Impact of the Oregon Health Plan on access and satisfaction of adults with low income", Health Serv.Res., vol. 37, no. 1, pp. 11-31.

MITTON, C. & DONALDSON, C. 2003, "Resource allocation in health care: health economics and beyond", Health Care Anal., vol. 11, no. 3, pp. 245-257.

MOORE, S., TUMEH, J., WOJTANOWSKI, S., & FLOWERS, C. 2007, "Cost-effectiveness of aprepitant for the prevention of chemotherapy-induced nausea and vomiting associated with highly emetogenic chemotherapy", Value Health, vol. 10, no. 1, pp. 23-31.

MSAC. "Artificial intervertebral disc replacement (Total disc arthroplasty)" MSAC application 1090, Assessment report, ISSN 1443-7120. Viewed August 20th, 2011. <http://www.health.gov.au/internet/msac/publishing.nsf/Content/2CDBC3816FDE8D20CA2575AD0082FD8E/$File/1090%20-%20Artificial%20intervertebral%20disc%20replacement%20Report.pdf>.

MSAC. 2001. "Low intensity ultrasound (LIUS) treatment for acceleration of bone fracture healing – ExogenTM bone growth stimulator". MSAC Application 1030, ISSN 1443-7120. Viewed September 21th, 2011 <http://www.health.gov.au/internet/msac/publishing.nsf/Content/1E0803AD53E44E39CA2575AD0082FD91/$File/1030%20-%20Exogen%20bone%20growth%20stimulator%20Report.pdf>.

MSAC. Medical Services Advisory Committee (MSAC). Viewed September 15th, 2011 <http://www.msac.gov.au/>

MSAC. Medical Services Advisory Committee (MSAC) - About Us. http://www.msac.gov.au/internet/msac/publishing.nsf/Content/about-us-lp-1 . 2011. 7-15-2011.

MUSHLIN, A.I,. HALL, W.J., ZWANZIGER, J., GAJARY, E., ANDREWS, M., MARRON, R., ZOU, K.H., MOSS, A.J. 1998. The cost-effectiveness of automatic implantable cardiac defibrillators: results from MADIT. Circulation, vol. 97, no. 21, pp. 2129-2135

NASSER, M. & SAWICKI, P. T. 2009, Institute for Quality and Efficiency in Health Care: Germany, The Commonwealth Fund 57.

National Health Act. National Health Act 1953, S101 - Functions of Pharmaceutical Benefits Advisory Committee. 1987.

National Institute for Health and Clinical Excellence (NICE) 2001, Guide to the Technology Appraisal Process.

National Institute for Health and Clinical Excellence (NICE) 2008, Guide to the Methods of Technology Appraisal, NICE, London.

NEUMANN, P. J. 2005, Using Cost-Effectiveness Analysis to Improve Health Care: Opportunities and Barriers Oxford University Press.

NEUMANN, P. J. 2009, OHE Briefing: American Exceptionalism and American Health

Care: Implications for the US Debate on Cost-Effectiveness Analysis.

NEUMANN, P. J. & GREENBERG, D. 2009, "Is the United States ready for QALYs?", Health Aff.(Millwood.), vol. 28, no. 5, pp. 1366-1371.

NEUMANN, P. J., KAMAE, M. S., & PALMER, J. A. 2008, "Medicare's national coverage decisions for technologies, 1999-2007", Health Aff.(Millwood.), vol. 27, no. 6, pp. 1620-1631.

NEUMANN, P. J., ROSEN, A. B., & WEINSTEIN, M. C. 2005, "Medicare and cost-effectiveness analysis", N.Engl.J Med., vol. 353, no. 14, pp. 1516-1522.

NEUMANN, P. J., SANDBERG, E. A., BELL, C. M., STONE, P. W., & CHAPMAN, R. H. 2000, "Are pharmaceuticals cost-effective? A review of the evidence", Health Aff.(Millwood.), vol. 19, no. 2, pp. 92-109.

NEUMANN, P. J. & WEINSTEIN, M. C. 2010, "Legislating against use of cost-effectiveness information", N.Engl.J Med, vol. 363, no. 16, pp. 1495-1497.

NEUMANN PJ, GREENBERG D, OLCHANSKI NV, STONE PW, AND ROSEN AB. 2005, “Growth and quality of the cost-utility literature, 1976-2001.” Value Health vol. 8, no. 1, pp. 3-9

NEWBOLD, P., CARLSON, W. L., & THORNE, B. 2003, "Nonparametric Statistics," in Statistics for Business and Economics, 5 edn, P. J. Boardman, ed., Prentice Hall, pp. 531-555.

NHS. About the NHS - Overview. Viewed August 28th, 2011 http://www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx

NHS-EED. NHS Economic Evaluation Database (NHS-EED) - Centre for Reviews and Dissemination (CRD). Viewed August 18th, 2011 http://www.crd.york.ac.uk/crdweb/

NICE 2006, Guide to the single technology appraisal (STA) process - National Institute for Health and Clinical Excellence (NICE).

NICE 2008a, Guide to the Methods of Technology Appraisal - National Institute for Health and Clinical Excellence (NICE), NICE, London.

NICE 2008b, Social Value Judgements - Principles for the development of NICE guidance (Second Edition).

NICE 2009a, Guide to the multiple technology appraisal process - National Institute for Health and Clinical Excellence.

NICE 2009b, Guide to the single technology appraisal process - National Institute for Health and Clinical Excellence.

NICE 2009c, National institute for Health and Clinical Excellence. Appraising treatments which may extend life, at the end of life London.

NICE. National Institute for Health and Clinical Excellence - How we work. Viewed on September 5th, 2011 <http://www.nice.org.uk/aboutnice/howwework/how_we_work.jsp>

NICE. NICE 'do not do' recommendations database. Viewed September 25th, 2011. <http://www.nice.org.uk/usingguidance/donotdorecommendations/search.jsp>

NICE 2010c, Social Value of a QALY. Available:

http://www.nice.org.uk/aboutnice/howwework/researchanddevelopment/SocialValueOfAQALY.jsp Accessed on August 20th, 2011

NICE. National Institute for Health and Clinical Excellence (NICE) - About NICE. Viewed October 2oth, 2011 <http://www.nice.org.uk/aboutnice/>

NICE. NICE 10 years on: past, present and future. Viewed August 9th, 2011 <http://www.nice.org.uk/newsroom/features/NICETenYearsOn.jsp>

NICE. The National Institute for Health and Clinical Excellence (NICE) - "Who we are". Viewed August 9th, 2011 <http://www.nice.org.uk/aboutnice/whoweare/who_we_are.jsp>

NIH. National Institutes of Health. Viewed August 9th, 2011 <http://www.nih.gov/>

NORD, E., RICHARDSON, J., STREET, A., KUHSE, H., & SINGER, P. 1995, "Maximizing health benefits vs egalitarianism: an Australian survey of health issues", Soc.Sci.Med., vol. 41, no. 10, pp. 1429-1437.

NUIJTEN, M. J. & RUTTEN, F. 2002, "Combining a budgetary-impact analysis and a cost-effectiveness analysis using decision-analytic modelling techniques", Pharmacoeconomics., vol. 20, no. 12, pp. 855-867.

O'BRIEN, B. J., GERTSEN, K., WILLAN, A. R., & FAULKNER, L. A. 2002, "Is there a kink in consumers' threshold value for cost-effectiveness in health care?", Health Econ., vol. 11, no. 2, pp. 175-180.

O'BRIEN, R. M. 2007, "A Caution Regarding Rules of Thumb for Variance Inflation Factors", Quality and Quantity, vol. 41, no. 5, pp. 673-690.

O'DONNELL, J. C., PHAM, S. V., PASHOS, C. L., MILLER, D. W., & SMITH, M. D. 2009, "Health technology assessment: lessons learned from around the world--an overview", Value Health, vol. 12 Suppl 2, p. S1-S5.

OECD 2010, Organisation for Economic Co-operation and Development. Purchasing Power Parity (PPP) statistics.

OECD. OECD Health Data 2011 - Frequently Requested Data. Viewed June 30th, 2011 <http://www.oecd.org/document/16/0,3343,en_2649_34631_2085200_1_1_1_1,00.html>

ORSZAG, P. R. 2008, The Overuse, Underuse, and Misuse Health Care. Statement of Dr. Peter R. Orszag before the Committee on Finance, United States Senate. Washington, DC: Congressional Budget Office.

ORSZAG, P. R. & ELLIS, P. 2007, "Addressing rising health care costs--a view from the Congressional Budget Office", N.Engl.J Med., vol. 357, no. 19, pp. 1885-1887.

ORSZAG, P. R. & EMANUEL, E. J. 2010, "Health care reform and cost control", N.Engl.J Med., vol. 363, no. 7, pp. 601-603.

PALMER, S. & TORGERSON, D. J. 1999, "Economic notes: definitions of efficiency", BMJ, vol. 318, no. 7191, p. 1136.

PARKOSEWICH, J. A. 2008, "Cardiac rehabilitation barriers and opportunities among women with cardiovascular disease", Cardiol.Rev., vol. 16, no. 1, pp. 36-52.

PBAC 2008, Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3).

PBAC. Pharmaceutical Benefits Advisory Committee (PBAC). Viewed June 15th, 2011 <http://www.health.gov.au/internet/main/publishing.nsf/content/health-pbs-general-listing-committee3.htm> .

PBS. Pharmaceutical Benefits Scheme (PBS) - Pharmaceutical Benefits Advisory Committee (PBAC). Viewed June 15th, 2011 <http://www.pbs.gov.au/info/industry/listing/participants/pbac>

pCODR. The pan-Canadian Oncology Drug Review (pCODR). Viewed June 15th, 2011 <http://www.pcodr.ca/portal/server.pt/community/home_page/548>

Pear, R. Medicare to weigh cost as a factor in reimbursement. New York Times . 4-21-1991.. Viewed March 3rd, 2011

PEARSON M 2009, Disparities in health expenditure across OECD countries: Why does the United States spend so much more than other countries? Written statement to Senate Special Committee on Aging.

PEARSON, S. D. & BACH, P. B. 2010, "How Medicare could use comparative effectiveness research in deciding on new coverage and reimbursement", Health Aff.(Millwood.), vol. 29, no. 10, pp. 1796-1804.

PERLETH, M., GIBIS, B., & GOHLEN, B. 2009, "A short history of health technology assessment in Germany", Int J Technol.Assess.Health Care, vol. 25 Suppl 1, pp. 112-119.

PETERSON, C. & BURTON, R. 2007, US Health Care Spending: Comparison with other OECD Countries.

PHELPS, C. E. & MUSHLIN, A. I. 1991, "On the (near) equivalence of cost-effectiveness and cost-benefit analyses", Int.J Technol.Assess.Health Care, vol. 7, no. 1, pp. 12-21.

PHILIPSON, T. & SOARES, R. 2001, Human Capital, Longevity, and Economic Growth: A Quantitative Assessment of Full Income Measures, World Bank, Washington DC.

PHILLIPS, K. A. & CHEN, J. L. 2002, "Impact of the U.S. panel on cost-effectiveness in health and medicine", Am.J Prev.Med, vol. 22, no. 2, pp. 98-105.

PPACA. Patient Protection and Affordable Care Act (PPACA). H.R. 3590.

Presidential Executive Office 2009, The Economic Case for Health Care Reform: Update. Executive Office of the President. Council of Economic Advisers.

QUINET, E. 2000, Economic Evaluation of Road Traffic Safety Measures. Conclusions of Round Table 117, Paris, 26-27 October 2000.

Raad voor de Volksgezondheid en Zorg 2006, Raad voor de Volksgezondheid en Zorg. Zinnige en duurzame zorg. Zoetermeer.

RAFTERY, J. P. 2008, "Paying for costly pharmaceuticals: regulation of new drugs in Australia, England and New Zealand", Med.J Aust., vol. 188, no. 1, pp. 26-28.

RAGNARSON, T. G. & APELQVIST, J. 2001, "Prevention of diabetes-related foot ulcers and amputations: a cost-utility analysis based on Markov model simulations", Diabetologia, vol. 44, no. 11, pp. 2077-2087.

RAMSBERG, J., ODEBERG, S., ENGSTROM, A., & LUNDIN, D. 2004, "Examining the quality of health economic analyses submitted to the Pharmaceutical Benefits Board in Sweden. The first year", Eur J Health Econ, vol. 5, no. 4, pp. 351-356.

RAMSEY, S. D., BERRY, K., ETZIONI, R., KAPLAN, R. M., SULLIVAN, S. D., & WOOD, D. E. 2003, "Cost effectiveness of lung-volume-reduction surgery for patients with severe emphysema", N.Engl.J Med., vol. 348, no. 21, pp. 2092-2102.

RATCLIFFE J, BUXTON M, YOUNG T, LONGWORTH L. 2005, “Determining priority for liver transplantation: a comparison of cost per QALY and discrete choice experiment-generated public preferences.” Appl Health Econ Health Policy vol. 4, no. 4, pp. 249-55

RAUNER, M. S., HEIDENBERGER, K., & PESENDORFER, E. M. 2005, "Model-based evaluation of diabetic foot prevention strategies in Austria", Health Care Manag.Sci., vol. 8, no. 4, pp. 253-265.

RAWLINS, M., BARNETT, D., & STEVENS, A. 2010, "Pharmacoeconomics: NICE's approach to decision-making", Br J Clin.Pharmacol., vol. 70, no. 3, pp. 346-349.

RAWLINS, M. D. & CULYER, A. J. 2004, "National Institute for Clinical Excellence and its value judgments", BMJ, vol. 329, no. 7459, pp. 224-227.

REINER, D. M., CURRY, T. E., DE FIGUEIREDO, M. A., HERZOG, H. J., ANSOLABEHERE, S. D., ITAOKA, K., JOHNSSON, F., & ODENBERGER, M. 2006, "American exceptionalism? Similarities and differences in national attitudes toward energy policy and global warming", Environ.Sci.Technol., vol. 40, no. 7, pp. 2093-2098.

REINHOLD, T., WITT, C.M., JENA, S., BRINKHAUS, B., WILLICH, S.N. 2008. “Quality of life and cost-effectiveness of acupuncture treatment in patients with osteoarthritis pain”. Eur J Health Econ. vol. 9, no. 3, pp. 209-219.

ROBINSON, J. C. 2001, "The end of managed care", JAMA, vol. 285, no. 20, pp. 2622-2628.

ROCCHI, A., MENON, D., VERMA, S., & MILLER, E. 2008, "The role of economic evidence in Canadian oncology reimbursement decision-making: to lambda and beyond", Value.Health, vol. 11, no. 4, pp. 771-783.

ROCHAIX, L. & XERRI, B. 2009, National Authority for Health: France - Commonwealth Fund - Issue Brief.

RODWIN, V. G. 1987, "American exceptionalism in the health sector: the advantages of "backwardness" in learning from abroad", Med.Care Rev., vol. 44, no. 1, pp. 119-154.

RUSSELL, L. B., GOLD, M. R., SIEGEL, J. E., DANIELS, N., & WEINSTEIN, M. C. 1996, "The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine", JAMA, vol. 276, no. 14, pp. 1172-1177.

RYAN, M., SCOTT, D. A., REEVES, C., BATE, A., VAN TEIJLINGEN, E. R., RUSSELL, E. M., NAPPER, M., & ROBB, C. M. 2001, "Eliciting public preferences for healthcare: a systematic review of techniques", Health Technol.Assess., vol. 5, no. 5, pp. 1-186.

SACHS, J. D. 2001, World Health Organisation (WHO): Macroeconomics and Health: Investing in Health for Economic Development Geneva.

SAMSON, D. 2004. “Special report: cost-effectiveness of left-ventricular assist devices as destination therapy for end-stage heart failure”. Chicago: Blue Cross and Blue Shield Association. Technology Evaluation Center Assessment Program Report, vol. 19, no. 2, pp. 1-36

SANDERS, G.D., HLATKY, M.A., OWENS, D.K. 2005. “Cost-effectiveness of implantable cardioverter-defibrillators”. N Engl J Med., vol. 353, no. 14, pp. 1471-80.

SBU. SBU – Swedish Council on Health Technology Assessment. Viewed August 6th, 2011 <http://www.sbu.se/en/>

SCULPHER, M. & CLAXTON, K. 2010, "Sins of omission and obfuscation: IQWIG's guidelines on economic evaluation methods", Health Econ.

SIEGRIST, P.T., HUSMANN, L., KNABENHANS, M., GAEMPERLI, O., VALENTA, I., HOEFFLINGHAUS, T., SCHEFFEL, H., STOLZMANN, P., ALKADHI, H., KAUFMANN, P.A. 2008. “(13)N-ammonia myocardial perfusion imaging with a PET/CT scanner: impact on

clinical decision making and cost-effectiveness". Eur J Nucl Med Mol Imaging., vol. 35, no. 5, pp. 889-895

SENDI, P., AL, M. J., GAFNI, A., & BIRCH, S. 2003, "Optimizing a portfolio of health care programs in the presence of uncertainty and constrained resources", Soc.Sci.Med., vol. 57, no. 11, pp. 2207-2215.

SENDI, P., GAFNI, A., & BIRCH, S. 2002, "Opportunity costs and uncertainty in the economic evaluation of health care interventions", Health Econ., vol. 11, no. 1, pp. 23-31.

SENDI, P. P. & BRIGGS, A. H. 2001, "Affordability and cost-effectiveness: decision-making on the cost-effectiveness plane", Health Econ, vol. 10, no. 7, pp. 675-680.

SHIH, A., DAVIS, S., SCHOENBAUM, A., GAUTHIER, R., NUZUM, R., & MCCARTHY, D. 2008, Organizing the U.S. Health Care Delivery System for High Performance, The Commonwealth Fund.

SHOGREN, J. F. & STANO, M. 2002, "Skill and the Value of Life", Journal of Political Economy, vol. 110.

SIEGEL, J. E., WEINSTEIN, M. C., RUSSELL, L. B., & GOLD, M. R. 1996, "Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine", JAMA, vol. 276, no. 16, pp. 1339-1341.

SIEGEL, J. E., WEINSTEIN, M. C., & TORRANCE, G. W. 1996, "Reporting cost-effectiveness studies and results," in Cost-effectiveness in Health and Medicine, M. R. Gold et al., eds., Oxford University Press, New York.

SILVERMAN, D. H., GAMBHIR, S. S., HUANG, H. W., SCHWIMMER, J., KIM, S., SMALL, G. W., CHODOSH, J., CZERNIN, J., & PHELPS, M. E. 2002, "Evaluating early dementia with and without assessment of regional cerebral metabolism by PET: a comparison of predicted costs and benefits", J Nucl.Med, vol. 43, no. 2, pp. 253-266.

SILVERMAN, D. H., TRUONG, C. T., KIM, S. K., CHANG, C. Y., CHEN, W., KOWELL, A. P., CUMMINGS, J. L., CZERNIN, J., SMALL, G. W., & PHELPS, M. E. 2003, "Prognostic value of regional cerebral metabolism in patients undergoing dementia evaluation: comparison to a quantifying parameter of subsequent cognitive performance and to prognostic assessment without PET", Mol.Genet.Metab, vol. 80, no. 3, pp. 350-355.

SIMON, G. E., KATON, W. J., LIN, E. H., RUTTER, C., MANNING, W. G., VON, K. M., CIECHANOWSKI, P., LUDMAN, E. J., & YOUNG, B. A. 2007, "Cost-effectiveness of systematic depression treatment among people with diabetes mellitus", Arch.Gen.Psychiatry, vol. 64, no. 1, pp. 65-72.

SLOKA, J.S., HOLLETT, P.D., MATHEWS, M. 2005. “Cost-effectiveness of positron emission tomography in breast cancer. Mol Imaging Biol., vol. 7, no. 5, pp. 351-360

SMITH, R. 1991, "Where is the wisdom...?", BMJ, vol. 303, no. 6806, pp. 798-799.

SMITH, G.T., HUBNER, K.F., MCDONALD, T., THIE, J.A. 1999. “Cost Analysis of FDG PET for Managing Patients with Ovarian Cancer”. Clin Positron Imaging., vol. 2, no. 2, pp. 63-70

Social Security Act. 1862(a)(1)(A) of Social Security Act. 1965.

SOLBERG, L. I., MACIOSEK, M. V., & EDWARDS, N. M. 2008, "Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness", Am.J Prev.Med, vol. 34, no. 2, pp. 143-152.

SOLBERG, L. I., MACIOSEK, M. V., EDWARDS, N. M., KHANCHANDANI, H. S., & GOODMAN, M. J. 2006, "Repeated tobacco-use screening and intervention in clinical practice: health impact and cost effectiveness", Am.J Prev.Med, vol. 31, no. 1, pp. 62-71.

SORENSON, C., DRUMMOND, M., KANAVOS, P., & MCGUIRE, A. 2008, National Institute for Health and Clinical Excellence (NICE): How does it work and what are the implications for the U.S.?

Stata 2009, Stata Statistical Software: Release 11, StataCorp LP; College Station, TX.

STINNETT, A. A. & PALTIEL, A. D. 1996, "Mathematical programming for the efficient allocation of health care resources", J.Health Econ., vol. 15, no. 5, pp. 641-653.

SUGDEN, R. & WILLIAMS, A. H. 1979, The Principles of Practical Cost Benefit Analysis Oxford University Press, Oxford.

SULLIVAN, S. D., WATKINS, J., SWEET, B., & RAMSEY, S. D. 2009, "Health technology assessment in health-care decisions in the United States", Value Health, vol. 12 Suppl 2, p. S39-S44.

SUTHERLAND, J. M., FISHER, E. S., & SKINNER, J. S. 2009, "Getting past denial--the high cost of health care in the United States", N.Engl.J Med., vol. 361, no. 13, pp. 1227-1230.

TAN-TORRES, E. T., BALTUSSEN, R., ADAM, T., HUTUBESSY, R., ACHARYA, A., EVANS, D. B., & MURRAY, C. J. 2003, Making choices in health: WHO guide to cost-effectiveness analysis, World Health Organization, Geneva.

TENGS, T. O., ADAMS, M. E., PLISKIN, J. S., SAFRAN, D. G., SIEGEL, J. E., WEINSTEIN, M. C., & GRAHAM, J. D. 1995, "Five-hundred life-saving interventions and their cost-effectiveness", Risk Anal., vol. 15, no. 3, pp. 369-390.

The Commonwealth Fund. Viewed August 20th, 2011 <http://www.commonwealthfund.org/>

The Commonwealth Fund 2011a, International Profiles of Health Care Systems - The Canadian Health Care System, 2009.

The Commonwealth Fund 2011b, International Profiles of Health Care Systems - The German Health Care System, 2009.

The Hastings Center 2009, Connecting American Values with Health Reform. Viewed August 17th, 2011 < www.thehastingscenter.org/Publications/Detail.aspx?id=3528>

The Henry J.Kaiser Family Foundation. Financing Medicare: an issue Brief. Viewed December 29th, 2010 <http://www.kff.org/medicare/upload/7731.pdf>

THOMAS, R. J. 2007, "Cardiac rehabilitation/secondary prevention programs: a raft for the rapids: why have we missed the boat?", Circulation, vol. 116, no. 15, pp. 1644-1646.

TLV. The Dental and Pharmaceutical Benefits Agency, TLV. Viewed August 17th, 2011 <http://www.tlv.se/in-english-old/in-english/>

TOMASZEWSKI, K.J., HOLLOWAY, R.G. “Deep brain stimulation in the treatment of Parkinson’s disease: A cost-effectiveness analysis”. Neurology, vol. 57, pp. 663–671

TOPEL, R. & MURPHY, K. 1997, "Unemployment and Nonemployment", American Economic Review, vol. 87, pp. 295-300.

TORRANCE, G. W. 1986, "Measurement of health state utilities for economic appraisal", J Health Econ, vol. 5, no. 1, pp. 1-30.

TOWSE, A. & PRITCHARD, C. 2002, Cost-Effectiveness Thresholds: Economic and Ethical Issues King's Fund; Office of Health Economics, London.

TOWSE, A., PRITCHARD, C., & DEVLIN, N. 2002, "Does NICE have a threshold? An external view," in Cost-Effectiveness Thresholds. Economic and Ethical Issues, A.

Towse, C. Pritchard, & N. Devlin, eds., King's Fund and Office of Health Economics, London.

TOWSE, A. 2009, "Should NICE's threshold range for cost per QALY be raised? Yes", BMJ, vol. 338, p. b181.

TRIPPOLI, S., MESSORI, A., BECAGLI, P., ALTERINI, R., TENDI, E. 1998. “Treatments for newly diagnosed multiple myeloma: analysis of survival data and cost-effectiveness evaluation”. Oncol Rep., vol. 5, pp. 1475-1482

TRUEMAN, P., DRUMMOND, M., & HUTTON, J. 2001, "Developing guidance for budget impact analysis", Pharmacoeconomics., vol. 19, no. 6, pp. 609-621.

TSUCHIYA, A. & WILLIAMS, A. 2010, "Welfare economics and economic evaluation," in Economic evaluation in health care. Merging theory with practice, M. Drummond & A. McGuire, eds., Oxford University Press, Oxford, pp. 22-45.

U.S.Bureau of Labor Statistics 2010a, U.S. Department of Labor. Bureau of Labor Statistics. Viewed July 2nd, 2011 < http://www.bls.gov/>

U.S.Bureau of Labor Statistics 2010b, U.S. Department of Labor. Bureau of Labor Statistics. Inflation calculator. Viewed July 2nd, 2011 <http://www.bls.gov/data/inflation_calculator.htm>

U.S.Congress 1997, Issues Relating to Medicare's Coverage Policy - U.S.Congress,House of Representatives,Committee on Ways and Means,Subcommittee on Health.

U.S.Department of Veterans Affairs. U.S. Department of Veterans Affairs. Health Economics Resource Center. Viewed July 2nd, 2011 <http://www.herc.research.va.gov/home/default.asp>

U.S.National Library of Medicine. PubMed: U.S. National Library of Medicine National Institutes of Health. http://www.ncbi.nlm.nih.gov/pubmed/ . 2011. 2-1-2008.

UBEL, P. A. 2001, Pricing Life: Why It's Time for Health Care Rationing MIT Press.

UBEL, P. A. & GOOLD, S. 1997, "Recognizing bedside rationing: clear cases and tough calls", Ann.Intern.Med., vol. 126, no. 1, pp. 74-80.

UBEL, P. A., HIRTH, R. A., CHERNEW, M. E., & FENDRICK, A. M. 2003, "What is the price of life and why doesn't it increase at the rate of inflation?", Arch.Intern.Med., vol. 163, no. 14, pp. 1637-1641.

VALK, P.E., POUNDS, T.R., TESAR, R.D., HOPKINS, D.M., HASEMAN, M.K. 1996. Cost-effectiveness of PET imaging in clinical oncology”. Nucl Med Biol., vol. 23, pp. 737-743

VAN HOEK, A. J., UNDERWOOD, A., JIT, M., MILLER, E., & EDMUNDS, W. J. 2011, "The impact of pandemic influenza H1N1 on health-related quality of life: a prospective population-based study", PLoS.One., vol. 6, no. 3, p. e17030.

VARRICCHIONE, T. R. 2006, Comments on Proposed Decision Memo External Counterpulsation (ECP) Therapy (CAG-00002R2), posted December 20, 2005.

VERNON, J. & GOLDBERG, R. 2011, Shorter Lives, Less Prosperity: The Impact of Comparative Effectiveness Research on Health and Wealth Center for Medicine in the Public Interest (CMPI).

VISCUSI, W. K. & ALDY, J. E. 2003, "The value of a statistical life: A critical review of market estimates around the world", Journal of Risk and Uncertainty, vol. 27, no. 1, pp. 5-76.

VÖLLER, H., DOVIFAT, C., GLATZ, J. 2001. “HOME management of anticoagulation”. Eur Heart J. (supplement), vol. 3, pp. Q44–Q49

WALLACE, M.B., NIETERT, P.J., EARLE, C., KRASNA, M.J., HAWES, R.H., HOFFMAN, B.J., REED, C.E. 2002. “An analysis of multiple staging management strategies for carcinoma of the esophagus: computed tomography, endoscopic ultrasound, positron emission tomography, and thoracoscopy/laparoscopy”. Ann Thorac Surg., vol. 74, no. 4, pp. 1026-1032

Washington State Health Care Authority. Health Technology Assessment Program (HTA). Washington State. Viewed August 6th, 2011 <http://www.hta.hca.wa.gov/>

WEILL, C. & BANTA, D. 2009, "Development of health technology assessment in France", Int J Technol.Assess.Health Care, vol. 25 Suppl 1, pp. 108-111.

WEINSTEIN, M. C. 1996, "From cost-effectiveness ratios to resource allocation: Where to draw the line?," in Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies, F. Sloan, ed., Cambridge University Press, New York.

WEINSTEIN, M. C. & FEINBERG, H. V. 1980, Clinical Decision Analysis Sauders, Philadelphia.

WEINSTEIN, M. C. & MANNING, W. G., JR. 1997, "Theoretical issues in cost-effectiveness analysis", J Health Econ, vol. 16, no. 1, pp. 121-128.

WEINSTEIN, M. C., SIEGEL, J. E., GOLD, M. R., KAMLET, M. S., & RUSSELL, L. B. 1996, "Recommendations of the Panel on Cost-effectiveness in Health and Medicine", JAMA, vol. 276, no. 15, pp. 1253-1258.

WEINSTEIN, M. C. & STASON, W. B. 1977, "Foundations of cost-effectiveness analysis for health and medical practices", N.Engl.J Med., vol. 296, no. 13, pp. 716-721.

WEINSTEIN, M. C. & ZECKHAUSER, R. J. 1973, "Critical ratios and efficient allocation", J Public Econ, vol. 2, pp. 147-157.

WELLPOINT 2010, Wellpoint Health Technology Assessment Guidelines. Drug Submission Guidelines for Re-Evaluation of Products, Indications, and Formulations. National Pharmacy and Therapeutics Committee.

WHITE, K. L. 1995, "Evidence-based medicine", Lancet, vol. 346, no. 8978, pp. 837-838.

WHO 2000, The World health report 2000: health systems: improving performance. World Health Organization, World Health Organization. <http://www.who.int/whr/2000/en/>

WHO 2005, World Health Organization - CHOosing Interventions that are Cost Effective (WHO-CHOICE) - Tables of Costs and Prices used in WHO-CHOICE Analysis.

WILLIAMS, A. 2004, What could be nicer than NICE?, Office of Health Economics.

WILLIAMS, A. H. 1983, The economic role of 'health indicators'. Measuring the social benefits of medicine. Proceedings of a meeting held at Brunel University, Office of Health Economics, London.

WILLIAMS I, MCIVER S, MOORE D, & BRYAN S. 2008, "The use of economic evaluations in NHS decision-making: a review and empirical investigation." Health Technol Assess vol. 12, no. 7, pp. 1-175

World Bank 2010, World Bank Glossary. AViewed August 6thm, 2011 <http://www.worldbank.org/depweb/english/modules/glossary.html>

World Health Organization 2000, The World health report 2000: health systems: improving performance., World Health Organization.

YU, C.M., LAU, C.P., CHAU, J., MCGHEE, S., KONG, S.L., CHEUNG, B.M., LI, L.S. 2004. "A short course of cardiac rehabilitation program is highly cost effective in improving long-term quality of life in patients with recent myocardial infarction or percutaneous coronary intervention". Arch Phys Med Rehabil., vol. 84, pp. 1915-1922

ZARIC, G. S. & BRANDEAU, M. L. 2001, "Optimal investment in a portfolio of HIV prevention programs", Med Decis.Making, vol. 21, no. 5, pp. 391-408.

10. Appendices

10.1. Appendix 1 - Assigning monetary values to life – worked examples

10.1.1. Compensating wage method (revealed preference)

The compensating wage, or revealed preference, method can be used to estimate the value of a statistical life (VSL) can be estimated by multiplying the wage differential by the inverse of the difference in probability of death or injury. (Brannon 2005;Drummond, Sculpher, Torrance, O'Brien, & Stoddart 2005b) A worked example is presented below:

	Wage	Risk of death
Job A	20,000	1/10,000
Job B	20,500	2,10,000

- VSL = Wage differential*Risk differential

- VSL = 500*10,000

- VSL = 5,000,000

10.1.2. Contingent valuation

Contingent valuation is another approach to estimating the VSL. A worked example of CV is presented below (Drummond, Sculpher, Torrance, O'Brien, & Stoddart 2005b):

- Current risk of a motorcyclist being killed in an accident = 50 in 100,000

- Risk of a motorcyclist being killed in an accident with new safety feature = 25 in 100,000

- Reduction in risk (dR) = 25 in 100,000

- Maximum willing to pay for safety feature (dV) = £100

- Implied value of life = dV/dR

= £100/25 x 10^{-5}

= £400,000

10.2. Appendix 2 – Email correspondence

Email received from Prof. Christopher J.L. Murray (received January 29th, 2007).

Dear James,
The result is from an analysis by Jeff Sacks. It follows form a basic utility maximization model where healthy life years are effectively the integrand for U(c). A log utility function yields something close to 3 if I remember. More concave utility functions would yield a higher multiplier. In fact, many plausible utility functions would argue that the multiplier of gdp per capita would increase as consumption increases. I am not sure if the maths were ever published but they should be easy to replicate.

Regards
Chris Murray

Email received from Prof. Jeffrey D. Sachs (received January 29th, 2007).

Dear James,

The standard of a DALY threshold at 3 times per capita income appears informally in several mentions in published articles. The common US threshold of around $135,000 per DALY is an example. There is no deep-deep theory, but there are relevant articles by Alan Garber (and a co-author, whose name I don't recall for the moment), and value-of-life articles by Chicago economists, explaining why the benefit of an extra life year is equal in fact to three components: the direct effect, a curvature effect (more years to smooth the income), and a leisure effect. Again, as I'm away from my office, I don't have references at hand.

I would suggest that you also look at empirical cutoff points that are used in the U.S., U.K., and perhaps other high-income countries. When I did that a few years ago, the 3x income level was roughly right.

Please let me know what else you find, and I can take this up in more detail when I return from Africa.

Best regards,
Jeffrey Sachs

10.3. Appendix 3 - Review of NCD for Deep Brain Stimulation

Comments

One cost-effectiveness study set in the US was published prior to the decision memo.

Number of decision/sub decisions available from this memo

Two decisions were identified from the decision memo: first, positive coverage of DBS for Parkinson's Disease (PD) patients that meet specified criteria; second non-coverage of DBS for PD patients that do not meet specified criteria

Intervention:

Deep Brain Stimulation (DBS) for Essential Tremor and Parkinson's Disease

Coverage criteria/decision:

Medicare will cover unilateral or bilateral thalamic VIM DBS for the treatment of essential tremor (ET) and/or Parkinsonian tremor and unilateral or bilateral STN or GPi DBS for the treatment of Parkinson's disease only under the following conditions:

1. Medicare will only consider DBS devices to be reasonable and necessary if they are Food and Drug Administration (FDA) approved devices for DBS or devices used in accordance with FDA approved protocols governing Category B Investigational Device Exemption (IDE) DBS clinical trials.
2. For thalamic VIM DBS to be considered reasonable and necessary, patients must meet all of the following criteria:
    a. Diagnosis of essential tremor (ET) based on postural or kinetic tremors of hand(s) without other neurologic signs, or diagnosis of idiopathic PD (presence of at least 2 cardinal PD features (tremor, rigidity or bradykinesia) which is of a tremor dominant form
    b. Marked disabling tremor of at least level 3 or 4 on the Fahn-Tolosa-Marin Clinical Tremor Rating Scale (or equivalent scale) in the extremity intended for treatment, causing significant limitation in daily activities despite optimal medical therapy.
    c. Willingness and ability to cooperate during conscious operative procedure, as well as during post-surgical evaluations, adjustments of medications and stimulator settings.

3. For STN or GPi DBS to be considered reasonable and necessary, patients must meet all of the following criteria:
    a. Diagnosis of PD based on the presence of at least 2 cardinal PD features (tremor, rigidity or bradykinesia).
    b. Advanced idiopathic PD as determined by the use of Hoehn and Yahr stage or Unified Parkinson's Disease Rating Scale (UPDRS) part III motor subscale.
    c. L-dopa responsive with clearly defined "on" periods.
    d. Persistent disabling Parkinson's symptoms or drug side effects (e.g., dyskinesias, motor fluctuations, or disabling "off" periods) despite optimal medical therapy.
    e. Willingness and ability to cooperate during conscious operative procedure, as well as during post-surgical evaluations, adjustments of medications and stimulator settings.

DBS is not reasonable and necessary and is not covered for ET or PD patients with any of the following:

1. Non-idiopathic Parkinson's disease or "Parkinson's Plus" syndromes.
2. Cognitive impairment, dementia or depression which would be worsened by or would interfere with the patient's ability to benefit from DBS.
3. Current psychosis, alcohol abuse or other drug abuse.
4. Structural lesions such as basal ganglionic stroke, tumor or vascular malformation as etiology of the movement disorder.
5. Previous movement disorder surgery within the affected basal ganglion.
6. Significant medical, surgical, neurologic or orthopedic co-morbidities contraindicating DBS surgery or stimulation.

Patients who undergo DBS implantation should not be exposed to diathermy (deep heat treatment including shortwave diathermy, microwave diathermy and ultrasound diathermy) or any type of MRI which may adversely affect the DBS system or adversely affect the brain around the implanted electrodes.

DBS should be performed with extreme caution in patients with cardiac pacemakers or other electronically controlled implants which may adversely affect or be affected by the DBS system.

For DBS lead implantation to be considered reasonable and necessary, providers and facilities must meet all of the following criteria:

1. Neurosurgeons must: (a) be properly trained in the procedure; (b) have experience with the surgical management of movement disorders, including DBS therapy; and (c) have experience performing stereotactic neurosurgical procedures.
2. Operative teams must have training and experience with DBS systems, including knowledge of anatomical and neurophysiological characteristics for localizing the targeted nucleus, surgical and/or implantation techniques for the DBS system, and operational and functional characteristics of the device.
3. Physicians specializing in movement disorders must be involved in both patient selection and post-procedure care.
4. Hospital medical centers must have: (a) brain imaging equipment (MRI and/or CT) for pre-operative stereotactic localization and targeting of the surgical site(s); (b) operating rooms with all necessary equipment for stereotactic surgery; and (c) support services necessary for care of patients undergoing this procedure and any potential complications arising intraoperatively or postoperatively.

Since long-term safety, effectiveness and optimal targeting for DBS have not been established, CMS will review the appropriateness of Medicare coverage as pertinent new evidence becomes available. This review will include clinical follow-up and targeting information from the ongoing, randomized VA/NINDS Cooperative Trial comparing best medical therapy with DBS of the STN and GPi for PD, as well as longer term clinical results from mandatory annual progress reports and final report to the FDA of Medtronic's bilateral DBS PMA post-approval study.

Is there an alternative treatment available?

Yes. The following are excerpts from the decision memo:

"Pharmacotherapy with propanolol (a beta-adrenergic blocker) and primidone (an anticonvulsant medication) are first line agents in the treatment of ET and may improve function by reducing the severity of tremor. However, certain patients do not adequately respond to or cannot tolerate these medications".

"Treatments for PD include those which alleviate symptoms (symptomatic therapy), slow the loss of nerve cells (neuroprotective), and increase and/or improve cell function (restorative). Currently, symptomatic therapy - with medications, lesioning surgery or DBS - is the only available treatment for patients with PD".

"L-dopa is the oldest and most potent symptomatic drug treatment and remains the gold standard for relieving the symptoms of PD".

"Dopamine agonists (such as bromocriptine, pergolide, pramipexole and ropinirole), which directly stimulate dopamine receptors but are not as effective as L-dopa, are also used as an initial form of therapy in order to delay the need for L-dopa and its associated long-term adverse effects".

Type of treatment benefit

As this treatment does not increase life expectancy it is determined that it has an *"Increase in quality of life".*

Is this an explicit decision in the decision memo?

Yes

Prevalence in Medicare population

Not known. However, the following is an extract from the decision memo, *"Parkinson's Disease (PD) effects up to 1 million Americans".*

Budget impact of this technology in the Medicare population

Not known.

First line?

No. The following text is extracted from the decision memo:

"For patients who become unresponsive to pharmacological treatments and/or have intolerable drug side effects, lesioning surgeries and DBS may be helpful for carefully selected patients".

Economic evaluations (4):

The study by Tomaszewski KJ and Holloway RG (2001) was ultimately included in the database. It was assigned the highest grade of the four cost-effectiveness studies review (a 'C' grade).

The following studies were identified from the search strategy and reviewed.

1. Tomaszewski KJ, Holloway RG. Deep brain stimulation in the treatment of Parkinson's disease: a cost-effectiveness analysis. Neurology. 2001 Aug 28;57(4):663-71.
2. Charles PD, Padaliya BB, Newman WJ, Gill CE, Covington CD, Fang JY, So SA, Tramontana MG, Konrad PE, Davis TL. Deep brain stimulation of the subthalamic nucleus reduces antiparkinsonian medication costs. Parkinsonism Relat Disord. 2004 Dec;10(8):475-9.
3. Meissner W, Schreiter D, Volkmann J, Trottenberg T, Schneider GH, Sturm V, Deuschl G, Kupsch A. Deep brain stimulation in late stage Parkinson's disease: a retrospective cost analysis in Germany. J Neurol. 2005 Feb;252(2):218-23.
4. Valldeoriola F, Morsi O, Tolosa E, Rumià J, Martí MJ, Martínez-Martín P. Prospective comparative study on cost-effectiveness of subthalamic stimulation and best medical treatment in advanced Parkinson's disease. Mov Disord. 2007 Nov 15;22(15):2183-91.

Appraisal of Tomaszewski KJ and Holloway RG, 2001

Assessment criteria:	Comment:
Year of study	Price year 2000
Perspective of study	Societal
Comparator	Best medical management
Country setting	USA
Study population	Patients aged 50 years or older who are in the later stages of the disease (Hoehn and Yahr stage between 3 and 5) with intractable motor fluctuations.
Incremental cost-effectiveness ratio (ICER)	Base case analysis - $49,000
Uncertainty associated with reported ICER(s)	As part of a sensitivity analysis the authors varied the efficacy in the treatment of DBS in terms of QALYs gained. This varied from DBS being dominated to $27,147
The year study was published	2001
The purpose of the cost-effectiveness analysis	The purpose of this economic evaluation appears to be for publication only
Other comments	The model time horizon is the remaining life expectancy of the patient. This economic evaluation was available at the time that the decision was made. The economic evaluation did use some theoretical input values; indeed, the author states the requirement for additional randomized controlled trials.

Appraisal of Charles DP et al, 2004

Assessment criteria:	Comment:
Year of study	Price year 2002
Perspective of study	Payer
Comparator	Standard care – no direct comparator was used. As this was a cost study it demonstrated how the cost of the technology was offset in subsequent years
Country setting	USA
Study population	US population, mean age of patients 57 years if age.
Incremental cost-effectiveness ratio (ICER)	Cost saving – only in relation to pharmacological treatment?
Uncertainty associated with reported ICER(s)	No estimate of uncertainty was presented
The year study was published	2004
The purpose of the cost-effectiveness analysis	The purpose of the economic evaluation was for publication only
Other comments	Small sample size The economic evaluation uses hypothetical increases in the cost of pharmacological treatment. Cost savings are in relation to pharmacological treatment only and does not take into account the cost of the procedure

Appraisal of Meissner et al. 2005

Assessment criteria:	Comment:
Year of study	Electrodes were implanted between May 1997 and December 2000 – No price year us given
Perspective of study	Payer – Article states that no patients returned to work following the procedure
Comparator	Standard care
Country setting	Germany
Study population	German population - 58.6±1.0 years and mean disease duration was 16.0±0.7 years.
Incremental cost-effectiveness ratio (ICER)	DBS was dominant after the 1st year
Uncertainty associated with reported ICER(s)	No estimate of uncertainty was presented
The year study was published	2005
The purpose of the cost-effectiveness analysis	The purpose of the evaluation appears to be for publication only
Other comments	The costs were assessed for one year before and two years after implantation of deep brain stimulators focusing on the charges for drug treatment, in-patient hospital care and outpatient care. All calculated costs are indicated in euros.

Appraisal of Valldeoriola F et al, 2007

Assessment criteria:	Comment:
Year of study	Price year not stated. Appears that study was completed in 2006.
Perspective of study	Appears to be societal. Authors state that only 'Direct costs' were included and these were divided into two categories: a. Direct medical costs, related to costs for goods and services used in the prevention, diagnosis, treatment, and rehabilitation of the illness (for example, costs for medical visits, hospitalization, and pharmaceuticals). b. Direct nonmedical costs, generally assumed by the patient, including expenses related to the disease (for example, transportation, social services, adaptation of accommodation and any kind of special equipment, facilities or orthopedic material).
Comparator	STN-DBS and best medical management in patients with advanced PD
Country setting	Spain
Study population	Yes. Mean age of patients in STN-DBS grp 59.9 (SD 6.8), mean age of patients in BMT group 63.8 (SD 6.4)
Incremental cost-effectiveness ratio (ICER)	34,389€/QALY
Uncertainty associated with reported ICER(s)	Only basic sensitivity analyses were conducted. These included: Excluding the BMT patient group patient who had a prolonged hospitalization from the analysis - incremental cost per QALY was of 44,078€ (X1.3). Excluding patients treated with continuous apomorphine infusion, (an expensive therapy) - 62,148€ per QALY (X1.8).
The year study was published	2007
The purpose of the cost-effectiveness analysis	The purpose of the evaluation appears to be for publication only
Other comments	Open, prospective, longitudinal study EQ-5D used to estimate utility in the clinical trial. Study performed over a period of one year.