Economic issues in cost-effectiveness analysis alongside multinational clinical trials

Oliverio Rivero-Arias

Merton College
University of Oxford
Trinity Term
2010

D.Phil. Thesis

To my family

Acknowledgements

Writing this thesis has been, at times, a long, difficult and solitary journey. However, several people have offered both encouragement and useful advice over the last three years, making this journey a pleasant one. I am deeply grateful to my supervisor Prof. Alastair Gray for his unconditional patience and guidance throughout this experience. A special "thank you" is given to the team from the ISAT coordinating centre for always answering my queries promptly with a happy smile. I am indebted to Dr Molyneux and Mr Richard Kerr, the lead investigators of the ISAT study, for access to the data used in this thesis.

I would like to express my gratitude to James Buchanan, Dr. Helen Campbell, Helen Dakin, Dr. Ramón Luengo-Fernandez, Dr. Mara Violato, and Dr. Sarah Wordsworth, from the Health Economics Research Centre (HERC) at the University of Oxford, for spending time reading chapters and providing valuable comments that have helped to improve the overall content of this research. I am grateful to Juan Manuel Ramos-Goñi for his technical advice in Visual Basic and for his programming tips generally.

Financial support for this research was provided by a three-year doctoral research fellowship from the National Institute of Health Research, which is gratefully acknowledged. Additional funding to attend international conferences was provided by HERC and Merton College.

I thank my friend Jose Leal for constructive and active discussions over the years on economic evaluation using patient-level data versus modelling studies. I do not think we have reached an agreement yet but I feel that day will happen soon.

This thesis is dedicated entirely to my family.

Publications and conference presentations from this thesis

Publications

Rivero-Arias O and Gray A. The multinational nature of cost-effectiveness analyses alongside multinational clinical trials. Value in Health 2010; 13(1):34-41

Wolstenholme J, Rivero-Arias O, Gray A, Molyneux A, Kerr R, Yarnold J, Sneade M. Treatment pathways, resource use and costs of endovascular coiling versus surgical clipping following aSAH. Stroke 2008; 39(1):111-119.

Conference presentations

Rivero-Arias O and Gray A. The multinational characteristics of cost-effectiveness analysis alongside multinational clinical trials. 7th International Health Economics Association Conference (iHEA), July 2009, Beijing.

Rivero-Arias O, Wolstenholme J, Gray A, Tok MY. Building an economic model using data from a large multinational randomised controlled trial to extrapolate the results from the within trial analysis. 6th International Health Economics Association Conference (iHEA), July 2007, Copenhagen.

Rivero-Arias O, Wolstenholme J, Gray A. A comparison of the resource use, costs and health outcomes of endovascular coiling versus surgical clipping of subarachnoid aneurysms in the UK: Results from the ISAT study. 6th European Conference on Health Economics (ECHE), July 2006, Budapest.

Abstract

Economic issues in cost-effectiveness analysis alongside multinational clinical trials

Oliverio Rivero-Arias

Merton College, D.Phil. Thesis, Trinity Term 2010

This thesis is about conducting cost-effectiveness analysis when patient-level cost and outcome data are available in more than one country as part of a clinical trial. Economic evaluation alongside a multinational clinical trial poses interesting challenges to analysts. Health care systems are context-specific and as a result, estimates of incremental costs and effects can be influenced by the allocation and costs of resources within each country in the trial. This thesis uses the economic theory of production as a model to explain the behaviour of a potential provider of health care in each country participating in a hypothetical simulated multinational clinical trial. It is shown that the substitution effect of input factors measured by the elasticity of substitution plays a role in the estimation of country-specific incremental costs. Because heterogeneity in incremental costs and outcomes across countries may exist, this thesis explores different tests for heterogeneity and concludes that researchers should explore heterogeneity of treatment effect in the main categories of resource use in addition to costs and outcomes. The thesis recognises the limited power of these tests but suggests that they can be a potentially useful tool for selecting the appropriate method to generate country-specific results. The advantages of multilevel models over standard regression analysis for analysing and presenting the results from these studies are recognised within the thesis. However, analyses also demonstrate that the validity of these models depends crucially upon the number of countries and the sample size within each country. These are aspects that receive little attention in current published studies, and therefore this thesis calls for improvements to be made to the design stage of multinational clinical trials. The multinational nature of published cost-effectiveness analyses alongside multinational clinical trials is also reviewed and it is shown that the reporting of frequencies and country contribution in these studies is currently very poor. As a result, it is recommended that future guidelines should emphasise the importance of reporting the multinational nature of the study.

Table of contents

Acknowledgements ..... **iii**
Publications and conference presentations from this thesis ..... **iv**
Abstract ..... **v**
Table of contents ..... **vi**
Tables ..... **xi**
Figures ..... **xiv**
Common abbreviations used in this thesis ..... **xvi**

Chapter 1:
Introduction **1**
1.1 Background 1
1.2 Structure and content of the thesis 10

Chapter 2:
Economic evaluation alongside multinational clinical trials **12**
2.1 Introduction 12
2.2 Definition of economic evaluation 12
2.3 Measuring outcomes in economic evaluation 16
2.4 The measurement of costs in economic evaluation 20
2.5 The importance of time horizon and the discount rate 22
2.6 Handling uncertainty in economic evaluation of health care interventions 23
2.7 The net benefit framework 27
2.8 The role of multinational clinical trials in economic evaluation 28
2.9 Summary and conclusions 29

Chapter 3:
Theoretical issues relating to the production function of a health care technology in multinational clinical trials **30**
3.1 Introduction 30
3.2 Technology 31
3.2.1 Inputs and outputs 31
3.2.2 Technological constraints 32
3.2.3 The marginal product and the technical rate of substitution 35
3.2.4 Additional examples of isoquants 37
3.2.5 Parametric representation of a technology 39

3.2.6 The elasticity of substitution 40
3.2.7 The constant elasticity of substitution (CES) production function 43
3.3 Costs 45
3.3.1 Cost minimisation 45
3.3.2 Deriving the cost function and expansion paths for the producer 48
3.4 Measures of efficiency in health care 50
3.5 Theory of production in the economic evaluation of multinational clinical trials . 52
3.6 Conclusion 56

Chapter 4:
A structured review on cost-effectiveness analyses alongside multinational clinical trials 58
4.1 Introduction 58
4.2 Methods 60
4.2.1 Inclusion/exclusion criteria 60
4.2.2 Search Strategy 61
4.2.3 Data extraction 63
4.2.4 Presentation of results of the review 64
4.3 Results 65
4.3.1 Summary descriptive information of the overall review 65
4.3.2 Summary of results from the methodological evidence review 67
4.3.3 Summary of results from the applied evidence review 67
4.3.4 Detailed results of the methodological evidence review 76
4.3.5 Detailed results of the applied evidence review 93
4.4 Discussion 98
4.4.1 Applied review 98
4.4.2 Methodological review 100
4.4.3 Limitations of the review 101
4.4.4 Summary 102

Chapter 5:
The Medical Research Council (MRC) International Subarachnoid Aneurysm Trial (ISAT) case study 103
5.1 Introduction 103
5.2 An overview of the ISAT clinical investigation study 105
5.2.1 Aim of the study 105
5.2.2 Primary and secondary outcomes 109
5.2.3 Statistical analysis 109

5.2.4 Clinical results of the study 110
5.2.5 Conclusion of the ISAT study 111
5.2.6 Implications of the ISAT study 111
5.3 Data available to conduct the economic analysis of the ISAT study 112
5.3.1 Resource use data collected 112
5.3.2 Unit cost data collection 113
5.3.3 Unit cost data collected in the United Kingdom only 113
5.3.4 Country-specific unit cost data collection 114
5.3.5 Unavailable unit costs (market-basket method) 115
5.3.6 Intervention costs 118
5.3.7 Quality-adjusted life years (QALYs) 119
5.3.8 Follow-up 122
5.3.9 Descriptive statistics and data analysis 122
5.4 Results 123
5.4.1 Resource use 123
5.4.2 EQ-5D results 135
5.4.3 Quality-adjusted life years (QALYs) 146
5.5 Summary and conclusions 146

Chapter 6:
Statistical methods to identify heterogeneity of treatment effect in multinational cost-effectiveness analyses 149

6.1 Introduction 149
6.2 Statistical methods 151
6.2.1 Treatment effect on health outcomes, resource use, costs and net benefits 151
6.2.2 Formal statistical tests to detect country-by-treatment interactions 152
6.2.3 Standard regression analysis to detect country-by-treatment interactions 155
6.2.4 Multilevel models to model variation of treatment effect across countries 158
6.2.5 Presenting treatment effect results 165
6.2.6 Software implementation 166
6.3 Application of the statistical methods: an example using the ISAT dataset 167
6.3.1 Summary descriptive from the ISAT dataset 167
6.3.2 Formal statistical tests for heterogeneity 174
6.3.3 Standard regression analysis 178
6.3.4 Multilevel models 191
6.3.5 Summary of results from the test for heterogeneity across methods 197
6.3.6 Presenting treatment effect results 198
6.4 Summary and conclusions 201

Chapter 7:
Economic issues in multinational cost-effectiveness analyses – a comparison of costing methods and a simulation study 206

7.1 Introduction 206
7.2 A comparison of multinational costing approaches using ISAT data 208
7.2.1 Results of the comparison between multinational costing approaches 209
7.2.2 Summary of the results from the comparison of multinational costing approaches 219
7.3 The simulation study 219
7.3.1 Background 219
7.3.2 Multinational costing approaches 220
7.3.3 The CES production function 221
7.3.4 Selection of number of countries 223
7.3.5 The cost-minimisation problem 224
7.3.6 The simulation process 226
7.3.7 Results 230
7.4 Summary and conclusions 238

Chapter 8:
Economic analyses of endovascular coiling versus neurosurgical clipping in patients with aneurysmal subarachnoid haemorrhage – evidence from the International Subarachnoid Aneurysm Trial (ISAT) 243

8.1 Introduction 243
8.2 Methods 244
8.2.1 Summary of the ISAT study 244
8.2.2 Unit cost collection 245
8.2.3 Follow-up/perspective 246
8.2.4 Statistical methods 246
8.3 Results 249
8.3.1 Overview of results from previous chapters 249
8.3.2 Country-specific incremental cost and QALYs 250
8.4 Summary and conclusions 256

Chapter 9: Final summary and conclusions 260

References 270

Appendix 1: Properties of a health care production function 295

Appendix 2: Keyword proforma used in the structured review 297

Appendix 3: Full list of references included in the structured review ... 299

Appendix 4: Detailed results of the cost-effectiveness analysis performed in multi-country costing studies (11 studies) 308

Appendix 5: Detailed results of the cost-effectiveness analysis reporting country-specific ICERs (10 studies) 311

Appendix 6: Covering letter and unit cost survey to collect country-specific unit costs 316

Appendix 7: Calculation of residuals and country-specific shrinkage estimates in multilevel models (MLMs)..... 320

Appendix 8: Algebra to calculate $x1$ and $x2$ in the cost-minimisation problem in the simulation study..... 323

Tables

Table 1.1: Terminology for methodological approaches to multinational economic evaluations (Adapted from Table I of an article from Reed and colleagues (Reed et al. 2005a)) 4

Table 2.1: Types of economic evaluation 16

Table 4.1: Number of articles indentified during the main search strategy in MEDLINE, EMBASE and NHS-HEED 65

Table 4.2: Description of countries contributing data and number of studies to which data were contributed using the World Bank classification category 69

Table 4.3: Descriptive statistics of the multinational nature of the studies included in this review 71

Table 4.4: Unit costs information used in the costing analysis in the studies included in the review 72

Table 4.5: Unit cost collection approach used in the studies included in the review 72

Table 4.6: Health outcome data collected and type of effectiveness analysis conducted in the studies 73

Table 4.7: Ways in which cost-effectiveness data were analysed in the studies using Reed et al 2005 (Reed et al. 2005a) classification 74

Table 4.8: Presentation of country-specific incremental cost-effectiveness ratio (ICER) results in the applied studies 75

Table 4.9: Handling uncertainty for the incremental cost-effectiveness ratio (ICER) estimates in the applied studies 75

Table 5.1: Centres and distribution of patients across countries participating in ISAT 106

Table 5.2: Baseline characteristics whole sample 107

Table 5.3: Selected baseline characteristics by country 108

Table 5.4: Clinical outcome at 2 months and 1 year (Molyneux et al. 2005) 110

Table 5.5: Relative medical cost indexes of the study countries 116

Table 5.6: Country-specific unit costs data collected, $US PPP 2008 (bold figures are values imputed using the market-basket method) 118

Table 5.7: Coefficients for the estimation of the EQ-5D index values 121

Table 5.8: Distribution of theatre staffing present during the procedures across centres in the UK 123

Table 5.9: Distribution of theatre staffing present during the procedures in the UK, Sweden, Finland and Germany 124

Table 5.10: Resource use associated with the first episode of care for participants from all countries in the study (n=2,143) 126

Table 5.11: Resource use associated with the first episode of care for UK participants (n=1,644) 127

Table 5.12: Resource use associated with the first episode of care for Swedish participants (n=123) 128

Table 5.13: Resource use associated with the first episode of care for German participants (n=118) 129

Table 5.14: Resource use associated with the first episode of care for Canadian participants (n=89) 130

Table 5.15: Resource use associated with the first episode of care for Finnish participants (n=63) 131

Table 5.16: Risk of readmission for subsequent operational procedures at each follow-up point and country 133

Table 5.17: Risk of having an angiogram after original discharged at each follow-up point and country 134

Table 5.18: Distribution (n, %) of EQ-5D responses across the dimensions for all participants at 2 months, 1 year and 5 years follow-up 136

Table 5.19: Distribution (n, %) of EQ-5D responses across the dimensions for UK participants at 2 months, 1 year and 5 years follow-up 137

Table 5.20: Distribution (n, %) of EQ-5D responses across the dimensions for Swedish participants at 2 months, 1 year and 5 years follow-up 138

Table 5.21: Distribution (n, %) of EQ-5D responses across the dimensions for German participants at 2 months, 1 year and 5 years follow-up 139

Table 5.22: Distribution (n, %) of EQ-5D responses across the dimensions for Canadian participants at 2 months, 1 year and 5 years follow-up 140

Table 5.23: Distribution (n, %) of EQ-5D responses across the dimensions for Finnish participants at 2 months, 1 year and 5 years follow-up 141

Table 5.24: Utility values obtained from the EQ-5D at each follow-up point and using country-specific EQ-5D value sets for participating countries 143

Table 5.25: EQ-5D mean index values comparison using country-specific value set compared to the UK value set in Germany, Canada and Finland at 2 months and 2 years follow-up 145

Table 5.26: Utility values obtained from the EQ-5D at each follow-up point and using the UK value set for the pooled sample and multiple imputation for missing values. 146

Table 6.1: Sample statistics for the selected resource use using all countries and participating countries data 169

Table 6.2: Sample statistics for quality-adjusted years (QALYs) and total cost per patient at five years follow-up using all countries and participating countries data 171

Table 6.3: Sample statistics for net monetary benefits using country-specific unit costs and different thresholds of willingness to pay (λ) in all countries and participating countries data 173

Table 6.4: Quantitative and qualitative tests to detect country-by-treatment interactions on selected resource use 175

Table 6.5: Quantitative and qualitative tests to detect country-by-treatment interactions quality-adjusted life years and total cost per patient at 5 years follow-up 176

Table 6.6: Quantitative and qualitative tests to detect country-by-treatment interactions of net benefits on the monetary scale using country-specific unit costs and different thresholds of willingness to pay 177

Table 6.7: Country-by-treatment interaction test using regression analysis on selected resource use 186

Table 6.8: Country-by-treatment interaction test using regression analysis for quality-adjusted life years and total cost per patient at 5 years follow-up 188

Table 6.9: Country-by-treatment interaction test using regression analysis on net benefits on the monetary scale using country-specific unit costs and different thresholds of willingness to pay 190

Table 6.10: Modelling variation in treatment effect using a multilevel random slope model on selected resource use 193

Table 6.11: Modelling variation in treatment effect using a multilevel random slope model for quality-adjusted life years and total cost per patient at 5 years follow-up 194

Table 6.12: Modelling variation in treatment effect using a multilevel random slope model on the net benefit statistic 196

Table 6.13: Summary results of test for heterogeneity using formal statistical tests, regression analysis and multilevel models 197

Table 6.14: Akaike information criterion results from the standard regression analysis and the multilevel models 199

Table 7.1: Cost per patient for selected resource use using a one-country costing strategy (UK unit costs reported in $US 2008 PPP) in all countries and participating countries data 211

Table 7.2: Cost per patient for selected resource use using a multicountry costing strategy in all countries and participating countries data 212

Table 7.3: Distribution of resource use category contribution to the total cost per patient by costing approach and trial arm 214

Table 7.4: Mean difference between costing approaches for selected resource use categories by trial arm 216

Table 7.5: Total cost per patient at five years follow-up a one country (UK unit costs) and a multicountry costing strategy in all countries and participating countries data 217

Table 7.6: Country-specific unit cost estimates used in the simulation study (unit costs reported in $US 2008 PPP) 227

Table 7.7: Total cost per patient using only the resource use categories ITU and ward days to treat a patient with aSAH 229

Table 7.8: Absolute difference between incremental costs at elasticity of substitution equals to 2 and elasticity of substitution equals 0.1 (difference reported in $US 2008 PPP) 237

Table 8.1: Incremental costs and QALYs between trial arm across countries, and associated incremental cost-effectiveness ratio (ICER) using different multinational costing approaches 252

Table 8.2: Incremental costs and QALYs between trial arm across countries, and associated incremental cost-effectiveness ratio (ICER) 252

Table 9.1: Multinational characteristics to consider when developing the economic evaluation protocol alongside a multinational clinical trial 267

Figures

Figure 2.1: The cost-effectiveness plane (Black 1990) 14

Figure 2.2: Quality-adjusted life years gained from an intervention (adapted from (Gray et al. 2010)) 19

Figure 2.3: Uncertainty on the cost-effectiveness plane and the calculation of confidence interval for the incremental cost-effectiveness ratio 25

Figure 2.4: Uncertainty on the cost-effectiveness plane and the calculation of confidence interval for the incremental cost-effectiveness ratio 26

Figure 2.5: Cost-effectiveness acceptability curve 27

Figure 3.1: A production set and a production function 33

Figure 3.2: An isoquant indicating the degree of substitution between physician and nurses 34

Figure 3.3: Fixed proportions or Leontief technology isoquants between physician and nurses 38

Figure 3.4: Perfect substitutes isoquants between experienced and training nurses 39

Figure 3.5: Parametric approximation of an isoquant 40

Figure 3.6: The CES production function depending on rho 44

Figure 3.7: The cost-minimisation problem for the health care decision maker 47

Figure 3.8: The cost-minimisation problem for three levels of output between doctors and nurses 49

Figure 3.9: The derived total cost function 50

Figure 3.10: Allocative efficiency in the two-input world 51

Figure 3.11: Allocative and technical efficiency in the two input world 52

Figure 3.12: Scenario with the same isoquant in each country but different relative prices between countries 54

Figure 3.13: Scenario with the different isoquants and similar relative prices between countries 55

Figure 3.14: Scenario with different isoquant and relative prices between countries 56

Figure 4.1: Schematic diagram of overall identification and review process 66

Figure 4.2: Frequency of multinational economic evaluations published from 1996 to 2007 67

Figure 4.3: Description of methodological approaches used in the studies by year of publication 74

Figure 5.1: Neurosurgical clipping (a) and endovascular coiling (b) diagrams 104

Figure 5.2: Kaplan-Meier cumulative mortality to 7 years (Molyneux et al. 2005) 111

Figure 6.1: Histogram of duration of operation (in minutes) by randomisation group using all countries data 179

Figure 6.2: Histogram of number of angiograms associated to the first episode of care by randomisation group using all countries data 179

Figure 6.3: Histogram of number of days in intensive therapy unit (ITU) by randomisation group using all countries data 180

Figure 6.4: Histogram of number of days at ward facility by randomisation group using all countries data 180

Figure 6.5: Histogram of number of readmissions at 5 years follow-up by randomisation group using all countries data 181

Figure 6.6: Histogram of quality-adjusted life years (QALYs) at 5 years follow-up by randomisation group using all countries data 181

Figure 6.7: Histograms of total cost per patient using UK and country-specific unit-costs at 5 years follow-up by randomisation group using all countries data 182

Figure 6.8: Histograms of net monetary benefits using different values of willingness to pay and country-specific unit costs by randomisation group using all countries data 182

Figure 6.9: Comparison of country-specific incremental QALY estimates from observed data, ordinary least square estimation and multilevel model 200

Figure 6.10: Comparison of country-specific incremental cost estimates from observed data, ordinary least square estimation and multilevel model 201

Figure 7.1: Proportional change in total cost of endovascular compared with neurosurgery by costing approach 218

Figure 7.2: Relationship between ITU and ward days in each country using a CES production for different values of elasticity of substitution 231

Figure 7.3: Mean cost difference between endovascular and neurosurgery for different values of elasticity of substitution by country (Cv on unit cost 0.2) 234

Figure 7.4: Mean cost difference between endovascular and neurosurgery for different values of elasticity of substitution by country (Cv on unit cost 0.5) 235

Figure 7.5: Mean cost difference between endovascular and neurosurgery for different values of elasticity of substitution by country (Cv on unit cost 0.8) 236

Figure 8.1: 95%, 50% and 5% confidence ellipse of difference in costs and QALYs in the cost-effectiveness plane in each country and costing approach 253

Figure 8.2: Cost-effectiveness acceptability curves in each country indicating the probability of endovascular being cost-effective using a one-country costing approach 255

Figure 8.3: Cost-effectiveness acceptability curves in each country indicating the probability of endovascular being cost-effective using a multicountry costing approach 255

Common abbreviations used in this thesis

CEAC	Cost effectiveness acceptability curve
CEP	Cost-effectiveness plane
CI	Confidence interval
GNI	Gross National Income
ICER	Incremental cost-effectiveness ratio
ISAT	International Subarachnoid Aneurysm Trial
ITU	Intensive therapy unit
MLMs	Multilevel models
NHS	National Health Service
QALYs	Quality-adjusted life years
RCTs	Randomised controlled trials
UK	United Kingdom
VPC	Variance partitioning coefficient

Chapter 1: Introduction

1.1 Background

The demand for health care has rapidly expanded over the last three decades. This has greatly contributed to the increase in health care expenditure during the same period in most OECD countries (Organisation for Economic Co-Operation and Development 2009). As a result, governments worldwide are under constant pressure to invest in new health technologies that represent good value for money for society given the limited resources available. In the absence of a market that would provide the appropriate signals to decision makers to select among health care alternatives, a systematic framework to compare the costs and benefits of competing interventions is needed. This framework should ensure that the health of the population is maximised subject to the limited resources available and this is the aim of economic evaluation of health care technologies.

Economic evaluation has become a valuable methodology to inform health care resource allocation in a large number of developed and developing countries (National Institute for Health and Clinical Excellence 2008; World Health Organisation 2003). As a result, being able to inform decision makers in a particular location using evidence from economic evaluations conducted in another location has long been of interest to health economists (Drummond et al. 1992; O'Brien 1997).

Multinational clinical trials enable costs and outcomes data to be collected in more than one setting, improving at least in theory the generalisability of the cost-effectiveness results to different jurisdictions[1]. The amount of information from a location-specific economic evaluation that could be transferred into another setting depends on the generalisability of

[1] In this research thesis, the terms "location", "jurisdiction", and "setting" are used interchangeably.

the results of the study. Defining generalisability is not an easy task but one of the most useful definitions of generalisability of economic evaluations is presented by Sculpher and colleagues in an extensive report on this issue:

"... generalisability of economic evaluation is the extent to which the results of a study based on a measurement in a particular patient population and/or a specific context hold true for another population and/or different context" (Sculpher et al. 2004b).

Generalizability is often called transferability, but they are not strictly the same concept. Although researchers often use them interchangeably, transferability is the potential of a set of results obtained from a given setting to be valid in other locations.

Factors that influence the generalisability of cost-effectiveness results across locations were identified in a primer article by Drummond and colleagues (Drummond et al. 1992). The authors classified these factors into two broad categories: those related to the health technology under evaluation and those related to the health care system. The first category includes characteristics such as population demography and the epidemiology of the disease, the general and relative price levels of the technology, and the distribution of health care resources available to implement the technology. The second category includes those health care system specific factors such us clinical practice and conventions, incentives to professionals, institutions and patients, and the financing of the health care system. The topic of generalisability in economic evaluation was recently reviewed systematically in the report by Sculpher and colleagues and extensive research on each of these categories was identified (Sculpher et al. 2004b). Similar results were reported in another literature review identifying factors influencing the geographic transferability of economic evaluation data across jurisdictions (Goeree et al. 2007). More recently, Antonanzas and colleagues

developed an index to measure the degree of transferability of results of cost-effectiveness analysis across locations. The authors calculated this index in a sample of 27 Spanish studies in infectious diseases and concluded that the level of transferability across Spanish regions was very low (Antonanzas et al. 2009).

In recent years economic evaluations alongside multinational clinical trials have been suggested as a potential mechanism to improve generalisability of studies. Because individual patient-level data on effects and resource use are available in more than one location, the analyst is potentially able to use local and external information to inform decision makers in their own setting. In the area of oncology and cardiovascular disease, for example, multinational trials have been designed and performed routinely over the last decade (Hilbrich and Sleight 2006; Mano et al. 2006). Conducting randomised controlled trials (RCTs) on a multinational basis is not a novel approach and several authors have noted the benefits associated with these study designs (Camps 2004; de Silva 1997). Expanding clinical trials to regions outside local boundaries increases the recruitment rate and the sample size available for analyses. Consequently, the time needed to collect data, and to be able to analyse and present the results of the clinical trial can be significantly reduced. Because patients are enrolled from different country populations, generalisability of results is often improved. In addition, the expertise and knowledge from experienced research centres in a particular country can be transferred to those centres in other countries with less research experience. The pharmaceutical industry has found additional important advantages in running their studies across different countries (Anon. 2002; Chaikin et al. 2000; Smith 2002): detailed information on unit costs, resource use and effects are collected across countries, providing valuable data needed to address pricing, marketing authorisation, strategic planning and reimbursement in different countries.

In multinational trials, patients from several countries contribute to the overall sample size of the study. This special characteristic introduces methodological challenges for health economists but also the need for specific and clear guidelines to report and present the results from these analyses (Pang 2002; Torti, Jr. et al. 2006). In a workshop in 2005 entitled "Conducting economic evaluations alongside multinational clinical trials", experts in this area gathered to discuss a possible consensus on the methodology and reporting of multinational economic evaluations (Reed et al. 2005a). Generalisability, transparency, statistical power and economic implications related to traditional economic theory were the main methodological concerns identified. The additional burden to a study of collecting unit costs across countries was considered to be the main practical aspect when performing these evaluations. Although the available analytical strategies were discussed during the meeting, no clear consensus was reached on the appropriate method to apply to data. Instead, it was agreed that future applied work as well as new methodological developments were needed to clarify the correct approach to be adopted in future studies. An interesting and useful output from this workshop was the classification of methodological approaches to multinational economic evaluations presented in Table 1.1.

Table 1.1: Terminology for methodological approaches to multinational economic evaluations (Adapted from Table I of an article from Reed and colleagues (Reed et al. 2005a))

Source of clinical effectiveness data	Source of resource use data	Type of analysis	Costing methodology	Classification
All participating countries	All countries	Fully pooled	→ Multicountry	→ Fully pooled with multicountry costing
			→ One country	→ Fully pooled with one-country costing
All participating countries	One country or subset of countries	Partially split	→ Multicountry*	→ Partially split with multicountry costing
			→ One country	→ Partially split with one-country costing
One country or subset of countries	One country or subset of countries	Fully split	→ Multicountry*	→ Fully split with multicountry costing
			→ One country	→ Fully split with one-country costing

*Multicountry costing is feasible only if resource use data are available from more than one country.

This terminology provides an interesting attempt to classify economic analysis alongside multinational clinical trials depending on: 1) the country's contribution to cost and outcomes data and 2) the costing method used in the study. A fully pooled analysis is based on effectiveness and resource use data from patients in all countries. A partially split analysis is based on estimates of effectiveness data from all countries but resource use are based on a single country (or subset of countries). In a fully split analysis, estimates of effectiveness and resource use are based on the same group of patients in a single country (or subset of countries). A multi-country costing approach combines country-specific unit cost estimates with country-specific resource use whereas a one-country costing combines unit cost estimates from one country with resource use data from all countries, a subgroup of countries or a single country. Combining the type of analysis and the costing methodology results in the classification presented in the last column of Table 1.1 (fully pooled with multicountry costing, fully pooled with one-country costing and so on).

This terminology can now be used to classify past and future studies. To date little is known about the multinational nature of applied cost-effectiveness studies in multinational trials. For example, the average number of countries participating in these studies and the average sample contribution from each country has not been reported. To address this gap in the literature, a structured review on cost-effectiveness analysis alongside multinational clinical trials was conducted as part of this research thesis. The classification from Reed and colleagues is followed when presenting the results from this review.

The workshop "Conducting economic evaluations alongside multinational clinical trials" clearly identified the two main challenges to health economists involved in international clinical trials (Reed et al. 2005a). On the one side, the problems associated with collecting reliable country-specific unit cost weights to use in the analysis, and on the other side the

methodology to present and report cost-effectiveness results. Therefore, and as a response to these challenges, health economists have suggested new methods to deal with these issues.

Collecting unit cost data from several countries can be a complex procedure if the number of countries involved in the trial is large. Therefore, it is important to develop appropriate unit cost forms to collect these data. Goeree and colleagues[2] provided some insights on the number of countries and the number of hospitals in each country needed to obtain reliable cost estimates in a multinational clinical trial (Goeree et al. 1999). It was suggested that unit cost data should be collected from as many hospitals as possible within and across countries. However, this is not always feasible and increasing the number of hospitals may increase the amount of missing unit cost data, complicating the cost analysis of the study. Consequently, methods for imputing missing unit cost data have been discussed in the literature on multinational health economic evaluation (Glick et al. 2003; Grieve et al. 2010; Schulman et al. 1998).

The use of regression analysis has monopolised the discussion on how to report cost-effectiveness in multinational trials. Standard regression analysis to control for the multinational nature of the dataset using dummies for each country in the trial was the first suggestion by health economists (Koopmanschap et al. 2001; Willke et al. 1998a). More recently, the use of multilevel models (MLMs) which assume a natural hierarchy in the dataset has generated special interest in the literature to report cost-effectiveness analysis alongside multinational trials (Grieve et al. 2005; Manca et al. 2005b). Patients enrolled in a national clinical trial are recruited normally through centres (e.g. hospitals or general practices) within a country. When a multinational trial is conducted, for example to accelerate the recruitment rate, patients are recruited from centres in different countries. A

[2] All the methodological references included in this introductory chapter will be described in depth in subsequent chapters. Therefore, only a brief summary of the aims and results is provided here.

dataset of this type is said to be hierarchical in nature because patients are nested within centres and centres nested within countries. In statistical terminology, a multilevel structure is defined with patients assigned to level 1, centres to level 2 and countries to level 3. A MLM assumes that participants in each level (patients, centres and countries) are a sample from a population of patients, centres and countries. Therefore, a MLM allows for variability in all levels and can be estimated directly unlike standard regression models which only estimate variability at level 1. Furthermore, centre or country-specific covariates can be included in these models to explain the variability across centres or countries.

MLM inferences rely on the exchangeability assumption to be valid. This assumption states that the inference parameter of interest in each country, for example the treatment effect, is drawn randomly from an unknown parametric population distribution (Spiegelhalter et al. 2003). In a clinical trial setting, the treatment effect in each country becomes a random variable following that distribution. This implies that the treatment effects across countries are similar but not identical. However, the parameters of this population distribution are always unknown *a priori* and for this reason the justification of assuming exchangeability is based on our knowledge on the topic. If we have reasons to believe that the treatment effect across countries does not follow the same parametric distribution because we have *a priori* evidence of systematic differences across countries, the exchangeability assumption may not be justified (Spiegelhalter et al. 2003).

When the assumption of exchangeability is reasonable, and the treatment effect across countries follows the same population distribution, then MLMs provide a useful mechanism where countries can borrow strengths from each other to improve their inferences. For instance, a country with a relatively small sample size in a clinical trial could in principle use information from all countries to improve the inferences for its own setting. This

"borrowing-sharing" process is possible through the use of shrinkage estimates. Shrinkage estimates are calculated using a statistic that combines the variability occurring at all levels (e.g. patients, centres and countries) and estimated as part of a MLM. In the case of a multinational clinical trial, this may be the variability across patients within each country and the variability of treatment effect across countries.

The choice of standard regression or a multilevel framework to report the results from these studies is still a judgment that needs to be made by the analyst. Little is known about applied cost-effectiveness analysis alongside multinational clinical trials and more experience is required using both approaches to truly understand when to implement each method appropriately. This thesis provides new evidence and applies both methods in a real case study using a large multinational clinical trial in the area of subarachnoid haemorrhage.

In clinical trials we normally observe some variability between patients that we can explain using patients characteristics such us age or gender. In multinational clinical trials this variability may be different because the characteristics of patients in a particular trial are different across countries. This heterogeneity is likely to influence treatment effect on resource and outcomes. It is not clear how multilevel models and regression analysis deal with heterogeneity of treatment effect on costs and outcomes. Chapter 6 formally evaluates the implications of heterogeneity in the analysis of cost-effectiveness data in this area.

Although the health economics literature includes several suggested methods to deal with multinational outcomes and cost data, there are other areas that have received little or no attention so far (Reed et al. 2005a). For example, economic theory suggests that a producer of health care combines inputs that achieve a certain level of desired health output and minimise costs. This process is likely to be different across settings and as a result, relative

prices, levels of inputs and overall costs of providing a particular health technology are different across countries. Nevertheless, the research on how these differences affect the analysis in cost-effectiveness alongside clinical trials is still limited.

Raikou and colleagues explored some of these issues using a simulation of a multicentre study (Raikou et al. 2000). The authors evaluated the use of average or centre-specific unit costs to calculate the cost per patient in a multicentre cost-effectiveness analysis framework. It was shown that these two approaches were likely to produce different results if unit costs vary greatly across centres within a country and there was substitution between factor inputs. In a multinational setting, these differences may be even more marked, as centres are distributed across countries and unit costs may vary much more across countries than within them. Economic theory suggests that there is an underlying production function for health care in each centre that relates volumes of resource use and the costs of these resources to achieve a certain level of output. Applying a one-country costing method fails to recognise this. To date, there is no formal comparison between one-country and multicountry costing methods to analyse costs in cost-effectiveness analysis alongside multinational clinical trials.

The aim of this thesis is to investigate these economic issues, in particular, the use of different costing approaches and the impact of differences in the substitution of factor inputs across countries in the analysis of multinational clinical trials. In addition, this thesis reviews the statistical methods available to analyse and report these studies and applies them to a real case study of subarachnoid haemorrhage patients, providing new applied evidence in the area. The thesis also reviews current published economic evaluations alongside multinational clinical trials and describes in depth the multinational nature of such studies.

1.2 Structure and content of the thesis

Chapter 2 provides an overview of the economic evaluation paradigm and the different components of the framework in health care decision making. Chapter 3 then introduces the theoretical economic foundations that will be used later in the thesis. The chapter explains the production function of health care technologies and establishes a link between the theoretical issues and cost-effectiveness analysis alongside multinational clinical trials.

Chapter 4 continues with a structured review that identifies current applied economic evaluations literature and describes in depth the multinational nature of such studies. A brief overview of the methodological approaches to analyse and report this type of study is also included in the chapter.

Chapter 5 presents the case study used in this thesis to explore the economic and statistical issues around cost-effectiveness analyses alongside multinational clinical trials: the International Subarachnoid Aneurysm Trial (ISAT). The chapter provides a full description of the trial, including the measures of health outcomes, resource use, and the collection of unit cost data across countries in the study. In Chapter 6 the main statistical methods identified in Chapter 4 are expanded and applied to the ISAT dataset providing a comparison across methods.

Chapter 7 combines information from the ISAT study and the theoretical foundations set in Chapter 3 to conduct a simulation study to explore the economic issues identified in this thesis when analysing cost-effectiveness analyses alongside multinational clinical trials.

Chapter 8 makes use of the research output from previous chapters to conduct the cost-effectiveness analysis alongside the ISAT study. Finally Chapter 9 provides a summary and

some conclusions from the output arising from this thesis. The chapter also includes some suggested directions for future research in the area.

Chapter 2: Economic evaluation alongside multinational clinical trials

2.1 Introduction

Chapter 1 provided the background to this thesis and the rationale for conducting cost-effectiveness analyses alongside multinational clinical trials. It was shown that although there are strong advantages to conduct such studies to inform decision making, there are also challenges to the presentation and reporting of these studies across countries. Subsequent chapters deal with some of these challenges, but first this chapter provides a preliminary background to the economic evaluation paradigm.

The chapter defines economic evaluation in health care decision making, and in particular in the context of multinational clinical trials. The aim is to provide the reader with the preliminary background in economic evaluation to help the reading of the thesis.

2.2 Definition of economic evaluation

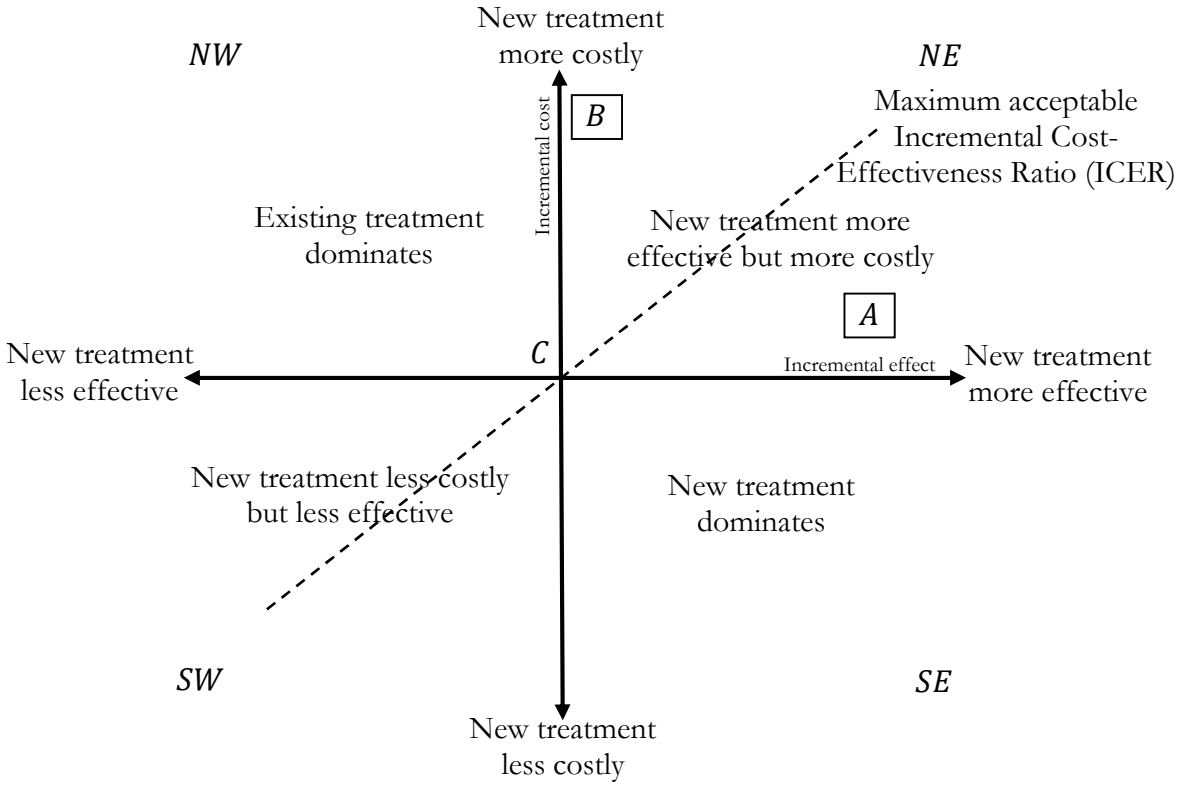
Economic evaluation seeks efficiency in health care decision making. The objective is to maximise health gains for the population given the scarce resources available. To achieve this aim, we compare the effectiveness and costs of an intervention with the best available alternative. We are interested thus in assessing whether the additional effect of a new treatment outweighs its additional costs. This is normally expressed using the incremental cost effectiveness ratio (ICER) which is simply the difference in costs divided by the difference in effects of the alternatives evaluated. The ICER can be formally expressed as follows:

$$ICER = \frac{\bar{C}_T - \bar{C}_C}{\bar{E}_T - \bar{E}_C} = \frac{\Delta C}{\Delta E} \qquad [2.1]$$

where $\bar{C}_T$ and $\bar{C}_C$ are the costs of the new and comparator treatments respectively, $\bar{E}_T$ and $\bar{E}_C$ the effectiveness in the new and comparator treatments respectively, and ΔC the difference in costs and ΔE the difference in effects between the alternatives.

The interpretation of the ICER can be represented graphically in the cost-effectiveness plane (CEP) in Figure 2.1 (Black 1990). The cost-effectiveness plane (CEP) is an incremental graph, i.e. expresses differences between the alternatives which may be positive or negative. The x-axis defines differences in effectiveness while the y-axis defines differences in costs between the alternatives. The origin of this plane represents the comparator alternative (C) which can take the form of conventional treatment, no treatment or placebo. Moving to the right of the origin of the CEP generates situations where the new treatment is more effective than the comparator. The opposite happens when moving to the left of the origin. Moving above the origin generates situations where the new alternative is more expensive than the comparator. Finally, moving below the origin indicates situations where the new alternative is cheaper than the comparator. Any point plotted in this plane represents an ICER, and any point along a line from the origin has consequently the same ICER.

Figure 2.1: The cost-effectiveness plane (Black 1990)



The cost-effectiveness plane has four quadrants (represented by compass points) and each quadrant identifies a possible combination of the costs and effects of the alternatives evaluated. In the southeast quadrant, the new treatment "dominates" the comparator since it is more effective and less costly. In the northwest quadrant of the plane the new treatment is dominated by the comparator as it is less effective and more costly. Dominant strategies are always preferred because they produce more health at a lower costs compared to dominated strategies. In the southeast and northwest quadrants the decision making process is straightforward.

More common and interesting are situations obtained in the northeast quadrant (where the new treatment is more effective but in turn more expensive than the comparator), and the southwest quadrant (where the new alternative is less effective but cheaper than the comparator). In both cases the decision to adopt will depend on whether the new treatment

represents good value for money, i.e. whether the additional effects of the new treatment are worth its additional costs. Point A in the northeast quadrant represents a situation where the new alternative is highly effective and not very expensive. A public health programme for smoking cessation is an example of this situation. There is evidence suggesting that these types of programmes are very effective and not very expensive in the long run (Shearer and Shanahan 2006). Hence, decision makers are likely to adopt interventions located near or at point A. However, point B in the same quadrant represents a situation where the new treatment is marginally more effective but significantly more expensive than the comparator. An example here would be the drug beta-interferon for multiple sclerosis patients (Forbes et al. 1999). In this case the decision maker may choose to spend the resources available more efficiently on alternatives that produce more health gains than the new treatment. As a result, points A and B in the same quadrant show clear differences in the decision making process and an explicit value is needed to determine which situation is the cost-effective alternative. This value should express the maximum amount of money we are willing to pay for health gains (the maximum acceptable ICER). This explicit value or "threshold" is represented by the dotted line that divided the CEP in Figure 2.1 into two halves. Those interventions that lie below this threshold are said to be cost-effective whereas the interventions located above this threshold are not.

In the CEP then, there are two types of uncertainty when performing economic evaluations of health interventions. The first uncertainty is the location of the new treatment under evaluation in the cost-effectiveness plane (Briggs and Gray 1999) whereas the second uncertainty is what an acceptable cost-effectiveness "threshold" is (Rawlins and Culyer 2004). The next section focuses on the x-axis of the cost-effectiveness plane and measuring outcomes in economic evaluation.

2.3 Measuring outcomes in economic evaluation

Depending on how outcomes are measured, economic evaluation is classified into five main categories, with a summary presented in Table 2.1.

Table 2.1: Types of economic evaluation

Type of economic evaluation	Outcomes
Cost-consequence	A range of outcome measures
Cost-benefit	Monetary terms
Cost-minimisation	Equal effectiveness and safety of interventions
Cost-effectiveness	Natural units (e.g. life-years, cases detected)
Cost-utility	Quality-adjusted life years (longevity and quality of life)

A cost-consequence analysis presents the results from an economic evaluation in a separate and disaggregated manner. Therefore, costs and outcomes are not combined into an incremental cost-effectiveness ratio and consequently, a cost-consequence analysis can be conducted with any outcome measure available. The advantage of this approach lies in its simplicity, costs and outcomes are presented to the decision maker in a disaggregated manner and for different scenarios. This type of evaluation assumes that the decision maker is able to synthesise and process the information from the study in some way. It hence assumes that decision makers know how properly use this information in their setting. As this is a strong assumption, the number of cost-consequences analyses compared to other type of economic evaluation is relatively small.

When health outcomes are measured in monetary terms both costs and effects are expressed in some monetary value the study is called a cost-benefit analysis (Pauly 1995). In principle, this type of study simplifies decision making, since we just need to evaluate whether the monetary benefits are greater than the costs to adopt the new alternative. Nevertheless, cost-

benefit analyses share similar challenges to other types of economic evaluations, especially in the presence of a budget constraint and when dealing with uncertainty in cost-benefit results. In addition, measuring health outcomes in monetary values has introduced significant unresolved challenges to the cost-benefit methodology (McIntosh et al. 2010). Therefore, the number of cost-benefit analyses available in the literature is limited compared to any other form of economic evaluation (Elixhauser et al. 1998).

In a cost-minimisation analysis the effectiveness and safety of the alternatives evaluated are similar and thus the lowest cost alternative is the treatment of choice. However, clinical trials are seldom designed to achieve equivalence between interventions. Therefore, no evidence of a significant difference in effectiveness does not necessarily mean that a difference does not exist. Hence, it is the joint distribution of cost and effects combined that should be estimated when reporting cost-effectiveness results (Briggs and O'Brien 2001). Therefore, cost-minimisation analyses should be used with caution and the analysis should also present the incremental cost-effectiveness ratio and the uncertainty around this estimate[3].

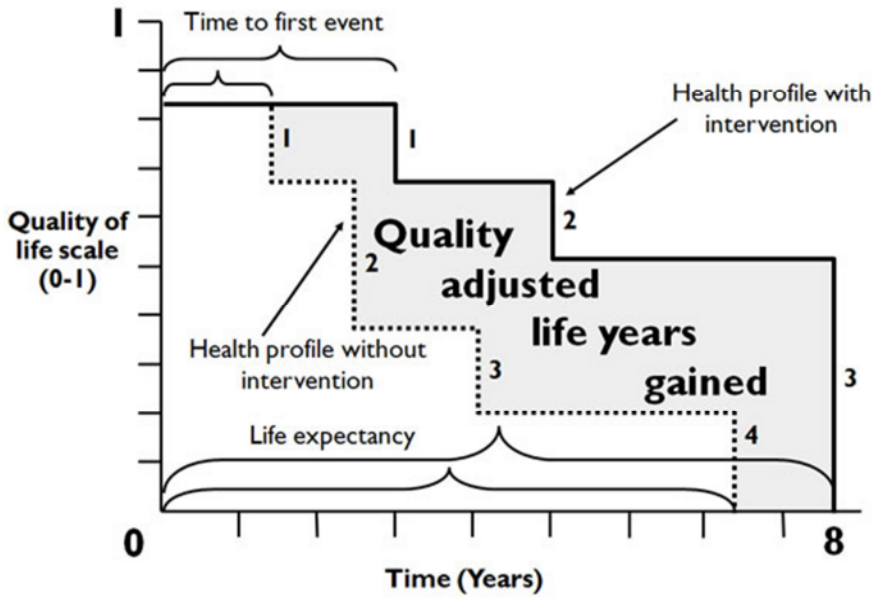
In cost-effectiveness analyses, health outcomes are measured in natural units such as the number of cases detected or number of life years gained or saved. The incremental cost-effectiveness ratio in this type of study will depend entirely on the units of outcome. For example, if the effectiveness of a screening program has been measured using number of cases detected, the ICER is expressed as cost per case detected. Although this indicator allows us to compare different types of screening programmes, comparison with other clinical areas is limited. It is therefore, very difficult or almost impossible to compare the results from economic evaluations where the results are expressed in different units.

[3] Handling uncertainty in economic evaluation of health care interventions is described in Section 2.6 later in the chapter.

To improve the comparability of cost-effectiveness results across a wide range of clinical areas, a common health outcome measure can be used across studies. This measure must capture important dimensions of the health status of patients such as quality and quantity of life. Multi-disciplinary groups of researchers have suggested the quality-adjusted life year (QALY) as a potential outcome measure to include in all economic evaluations (Drummond et al. 2005). QALYs combine quality of life and life expectancy into a single measure of effectiveness.

The idea behind calculating QALYs is shown using a hypothetical example in Figure 2.2 (Gray et al. 2010). The graph represents life expectancy (in years) on the x-axis and quality of life in a scale from 0 to 1 on the y-axis. The quality of life scale normally represents a death state indicated by 0 and a full health state indicated by 1. The figure indicates a health profile without an intervention (dotted line) and a health profile with an intervention (black line) for a hypothetical patient. From Figure 2.2 we can see that effectiveness can be measured in a number of ways. The number of events suffered by the patient if no intervention is implemented is four whereas only three if the intervention is adopted. The graph shows that the time to first event and life expectancy are shorter in the health profile without intervention. The area beneath each health profile is the number of QALYs associated with each profile. Note that all the dimensions just described (number of events, time to event, life expectancy and quality of life) are incorporated into this single measure. When comparing the effectiveness between two interventions using QALYs, the parameter of interest is the area between the two health profiles or the number of QALYs gained. Economic evaluations that use QALYs as the main outcome measure are called cost-utility analyses.

Figure 2.2: Quality-adjusted life years gained from an intervention (adapted from (Gray et al. 2010))



To calculate QALYs, therefore, survival information and health-related quality of life are needed. Survival data is straightforward to collect in clinical trials, and questionnaires are often used to collect health-related quality of life data. In the UK and continental Europe, one of the most widely used questionnaires in economic evaluation to estimate quality of life is the EQ-5D (EuroQol Group 1990). This instrument provides information on five health-related quality of life dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three possible answers: no problems, some problems or severe problems, yielding a combination of 243 possible health state descriptions. A set of valuations for each health state is available, which was estimated using an algorithm derived from a sample of 2997 individuals in England, Scotland, and Wales. The EQ-5D measure was developed specifically for use in economic evaluation, and the minimum and maximum tariff scores are -0.594 and 1, respectively (Dolan 1997).

The National Institute for Health and Clinical Excellence (NICE) recommends the use of QALYs as the main outcome measure in the submissions of economic evaluations (National Institute for Health and Clinical Excellence 2008). NICE is the organisation in charge of

creating guidance and guidance on health care treatments available in the UK National Health Service (NHS).

In economic evaluation, information on the effectiveness of the alternatives evaluated must be accompanied by robust and reliable estimates. Effect data may be obtained from several sources and then combined using methods of evidence synthesis (Egger, M. et al. 2001) or from a single large study. In the latter case, this information should come from well-designed randomised clinical trials that control for any potential biases present in most epidemiological studies (Friedman et al. 1998). When effectiveness data are available from a clinical trial, then analysts have access to patient-level data and statistical methods can be implemented to present the effectiveness results from the study.

The next section concentrates on the y-axis of the cost-effectiveness plane and the measurement of costs in economic evaluation.

2.4 The measurement of costs in economic evaluation

Costs in the economic evaluation of health care technologies are simply the product of the volumes of resources used and their associated unit costs. An appropriate analysis of costs should include all the categories of resource use consumed by the patients during the lifespan of the disease of interest. Examples of resource use categories include the number of general practitioner visits or the number of days in an intensive therapy unit.

In a randomised trial framework, resource use information is often collected as part of the study, in which case patient-level data are available. The collection of resource use data is normally conducted by means of patient-questionnaires or hospital/medical records. Ideally

questionnaires and the categories of resource use to be collected should be discussed during the development of the original study protocol. With patient-level data, standard statistical analysis can be implemented to analyse and present the results of the economic evaluation (Willan and Briggs 2006). Therefore, inference to detect significant differences in resource use between groups in the trial, similar to the effectiveness analysis, can be used.

The second component of cost is the unit cost associated with the category of resource use. Resources used should be valued at their opportunity cost (Brouwer et al. 2001). The opportunity cost is the value of the benefits forgone of the best alternative use. The concept of opportunity cost is best understood using an example. Imagine, a decision maker who needs to select between two health care policies: a new vaccination programme for infants to reduce mortality, and the construction of a new hospital. The opportunity cost is the value of all the benefits forgone because one of the policies is implemented. Suppose, the decision maker chooses to build the hospital, the opportunity cost is the value of all the life-years lost because the vaccination programme will not be executed. The rationale being that the benefits of the new hospital will outweigh the benefits of the vaccination programme.

In a competitive market, opportunity costs are expressed by market prices. However, health care is far from being a competitive market and hence caution needs to be exercised by analysts when selecting appropriate estimates of unit costs (Donaldson and Gerard 2005). In the UK, there are several national publications updated every year with information on prices and tariffs that are considered as accurate representations of health care resource use opportunity costs (Curtis 2009; Department of Health 2006).

The type of resource use collected and the presentation of the cost results define the perspective of the economic evaluation. A societal perspective includes three main categories

of costs: direct health care costs, informal care costs and employment-related costs (Brouwer et al. 2001). Direct health care costs refer to those costs directly relevant to the health care system. Informal care is the unpaid time provided by relatives, friends, or volunteers caring for patients and employment-related costs include costs associated with reduced productivity, sickness absence, early retirement or premature mortality. All categories of costs should be presented separately when reporting the results from the study to facilitate interpretation by health care decision makers. Besides, both informal care and employment-related costs are subject to a number of assumptions, and therefore combining estimates of costs from all categories may not be relevant to the different agents involved in the decision making process (Sculpher 2001).

In theory, all costs and consequences of the alternatives evaluated from all perspectives should be included in the analysis. In practice, the perspective of the study depends on who is making use of the information provided. For example, a particular agency such as NICE will typically be interested only in the health care costs of the alternatives evaluated whereas a private health insurer may also be interested in employment-related costs.

In a multinational clinical trial, unit costs are needed from different countries and this introduces some interesting characteristics to the analysis. Countries have different price levels and often have different currencies. Researchers need to consider these aspects when collecting and analysing unit cost data across countries. Chapters 4 and 5 of this thesis deal with this issue.

2.5 The importance of time horizon and the discount rate

An economic analysis should use a long enough time horizon to capture all costs and benefits related to the interventions evaluated. In most cases this time period lasts several

years. For instance, the outcome and costs of cancer screening programmes and other public health interventions are not observed immediately and analysts will need to include future costs and outcomes in the analysis. Future costs and outcomes in an economic evaluation are expressed in present values, that is, the point when a particular decision needs to be made. The process of expressing costs and outcomes from the future in present values typically involves discounting, based on the premise that individuals have a positive time preference (Cairns 2001). This premise suggests than in general individuals prefer to postpone undesirable events such as costs and ill health until the future, and also prefer to have desirable events such as profits and health in the present. As a result, costs and outcomes occurring during different time periods must be discounted and converted into present values. In the UK, the discount rate is provided by the Treasury and currently costs and outcomes incurred in the future are discounted at an annual rate of 3.5% (HM Treasury 2008). In the US, the "U.S. panel for cost-effectiveness analysis" recommends an annual rate of 3% for costs and outcomes (Gold et al. 1996).

2.6 Handling uncertainty in economic evaluation of health care interventions

The results from economic evaluation studies are subject to uncertainty. Uncertainty is present in all phases of an economic evaluation and as a result, steps needs to be taken to reduce and quantify the amount of uncertainty in a particular economic evaluation. Two types of design have led the economic evaluation literature for the last twenty years: cost-effectiveness analyses alongside clinical trials and economic evaluation using a decision analytic framework (Briggs and Gray 1999).

In cost-effectiveness analyses alongside a clinical trial, patient-level data on resource use and outcomes are collected with the aim of obtaining estimates such as the number of QALYs

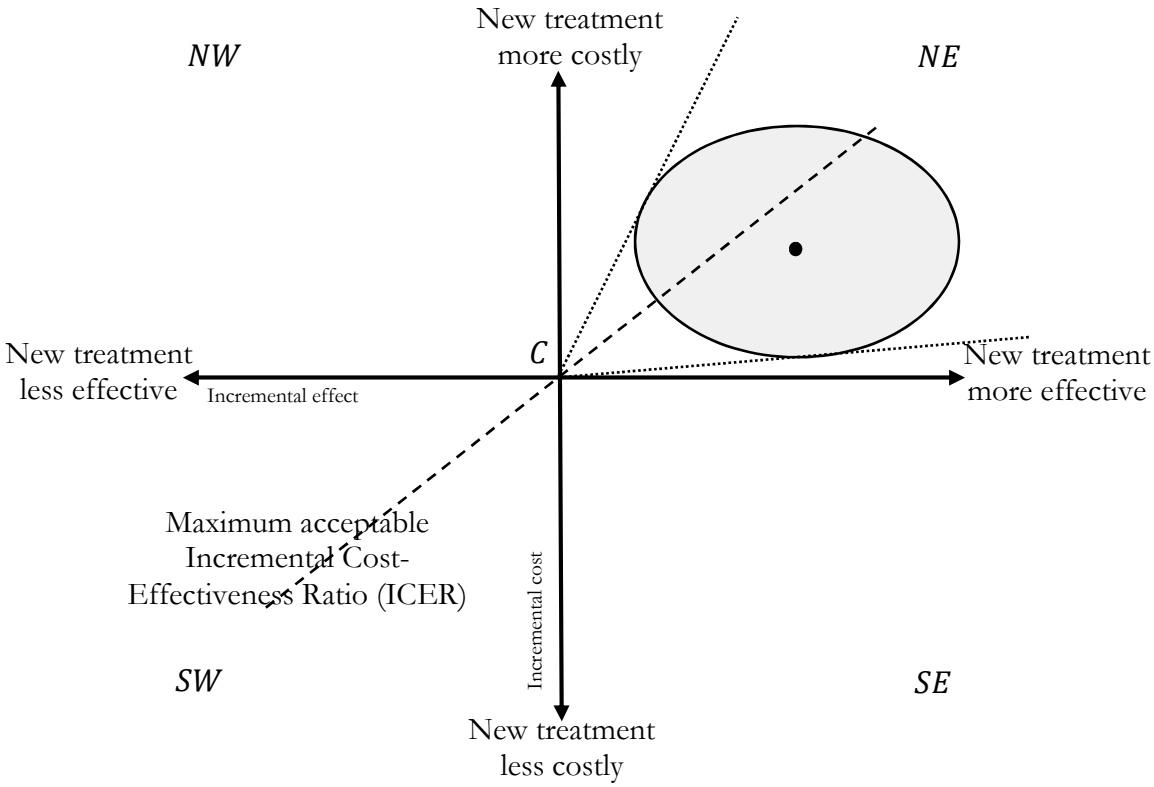
gained or the cost difference between two therapies. In this case, uncertainty occurs from the sampling process used to recruit a given number of participants to the study. This level of uncertainty is determined by the size of the sample, with smaller samples subject to a higher degree of uncertainty. Thus this type of uncertainty is quantified through the calculation of confidence intervals around the point estimates of health outcomes, resource use and costs during the statistical analysis.

Although economic evaluation alongside clinical trials is now common, a large number of economic evaluations synthesise evidence from different sources in a decision-analytic type framework (Briggs et al. 2006a). These evaluations take the form of economic models that require information on a range of parameters to estimate health outcomes, resource use and costs. In this case, the parameters and the structure of the model are subject to uncertainty and methods such as probabilistic sensitivity analysis (PSA) are implemented to capture and present the uncertainty around the results from the model. PSA typically involves repeatedly and simultaneously sampling parameter values using specified distributions to obtain empirical distributions that can be used to calculate the uncertainty associated with the parameters (Briggs et al. 2006a).

Statistical analyses in clinical trials and probabilistic sensitivity analyses in decision models provide similar information to present the uncertainty around cost-effectiveness results. Figure 2.1 introduced the cost-effectiveness plane as a graphical tool to aid the presentation and interpretation of cost-effectiveness results. Any point in this plane represents an incremental cost-effectiveness ratio and the uncertainty around this estimate can be represented using confidence intervals. However, the calculation of confidence intervals of a ratio poses statistical problems when the denominator is close or equal to zero. There are several parametric and non-parametric methods available to calculate confidence intervals

around a ratio (Briggs 2001) and the graphical intuition of these methods is presented in Figure 2.3.

Figure 2.3: Uncertainty on the cost-effectiveness plane and the calculation of confidence interval for the incremental cost-effectiveness ratio

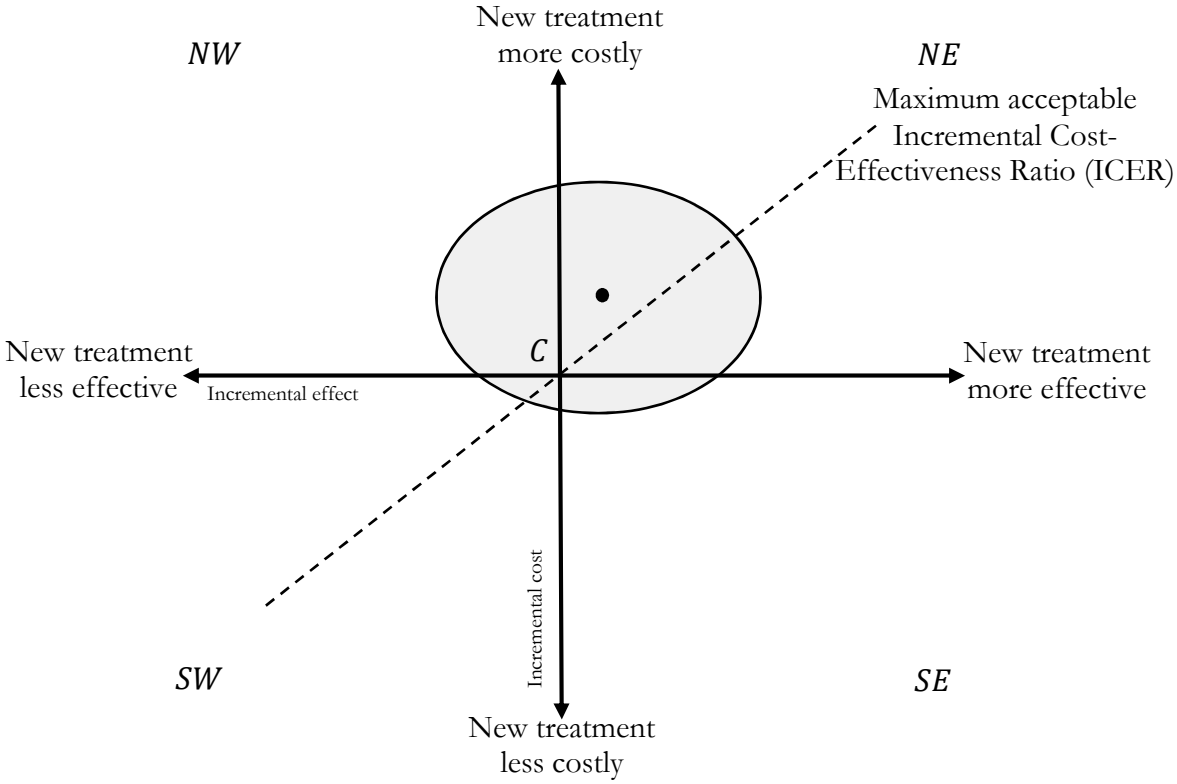


The methods assume that there is a joint density of costs and effects around the ICER that is represented by the shaded area inside the ellipse. This area is unknown *a priori* but confidence intervals around the ICER can be drawn parametrically using Fieller's theorem (Willan and O'Brien 1996) or non-parametrically using bootstrapping (Briggs et al. 1997), with evidence suggesting that both types of methods provide similar results (Briggs 2001). Figure 2.3 shows an example of confidence intervals around the ICER (black dot) using the two dotted lines that are tangent to the elliptical shape[4]. In Figure 2.3 the joint density of

[4] There is a parametric method that assumes that the joint density of costs and effects has an elliptical shape, with the confidence intervals constructed using lines tangent to this shape (van Hout et al. 1994). However, large Monte Carlo simulations have shown that the Fieller's and the bootstrap methods outperformed the ellipse approach (Briggs et al. 1999).

costs and effect is located on the northeast quadrant of the CEP and there is no overlap of this density to other quadrants of the plane. However, when this joint density is located in all four quadrants as illustrated in Figure 2.4, there are statistical problems in calculating confidence intervals around the ICER.

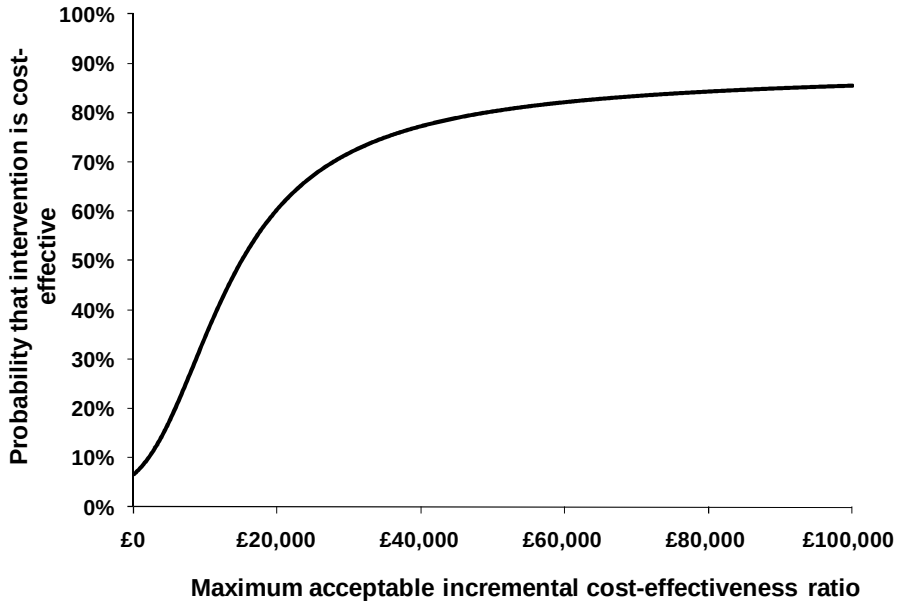
Figure 2.4: Uncertainty on the cost-effectiveness plane and the calculation of confidence interval for the incremental cost-effectiveness ratio



When confidence intervals around the ICER cannot be meaningfully computed, health economists have proposed the use of cost-effectiveness acceptability curves (CEACs) to present the uncertainty around the results (Lothgren and Zethraeus 2000). To construct CEACs, we need to make use of the concept of a threshold or maximum amount of money we are willing to pay for health gain, as described in Section 2.2. The idea is to calculate the proportion (probability) of the shaded area presented in Figure 2.4 that lies below a particular threshold starting from a threshold of zero to a threshold of infinite. At zero the

threshold equals the horizontal axis and the decision maker is indifferent to health gains or losses and is influenced only by whether the intervention is cost-saving. At infinity the threshold equals the vertical axis and the decision maker is indifferent to costs and persuaded only by the effectiveness of the new treatment. Plotting the proportion of the shaded area in Figure 2.4 to the different levels of threshold we obtain the CEAC presented in Figure 2.5. This figure indicates the probability that the intervention is cost-effective for different values of willingness to pay for health gain. In this case the graph suggests that the probability of the intervention being cost-effective lies asymptotically to 80% for values of willingness to pay above £40,000.

Figure 2.5: Cost-effectiveness acceptability curve



2.7 The net benefit framework

In recent years, health economists have proposed a linear transformation of the ICER into an incremental net benefit approach to allow the use of econometric techniques to model cost-effectiveness (Hoch et al. 2002). To calculate net benefits, we need to make use of the decision rule or threshold and a re-arrangement of expression 2.1 (Section 2.2). Net benefits

can be expressed in a monetary scale (net monetary benefit: NMB) or on the effect scale (net health benefit: NHB) as indicated in expression 2.2:

$$NMB = R_c \Delta E - \Delta C$$

$$NHB = \Delta E - \frac{\Delta C}{R_c} \qquad [2.2]$$

where R_c is the decision rule or threshold, ΔE the effect difference and ΔC the cost difference. One of the intrinsic advantages of net benefits compared to ICERs is its linear nature and as a result its interpretation. Values of the incremental net benefit above zero indicate the new intervention should be adopted because indicates that the new intervention has an associated higher net benefit than the comparator. When patient-level data on costs and outcomes are available, a net benefit can be calculated for each patient in the trial. Therefore, net benefits can be modelled using regression methods and Chapter 6 in this thesis applies a regression framework when comparing different approaches to report cost-effectiveness analyses alongside multinational clinical trials.

Cost-effectiveness acceptability curves can be derived parametrically from the net benefit statistic. The histogram of the net benefit statistic is approximately normally distributed, and CEACs are calculated from the p-value of the net benefit being positive (Briggs 2001).

2.8 The role of multinational clinical trials in economic evaluation

In multinational clinical trials information on costs and outcomes is collected from different settings. Chapter 1 set out that conducting a clinical trial in a multinational basis has the potential of improving the generalisability of the study. In addition, in multinational clinical trials the recruitment rate and the sample size available for analysis increases, reducing the time needed to collect data, and be able to analyse and present the results of the trial.

Nevertheless, having data from different countries introduces interesting challenges in the analysis of these studies. For example, to calculate costs, unit cost data need to be collected across countries and this may be a time consuming and expensive procedure. In addition, the statistical methods to analyse and present the cost-effectiveness results from these studies need to recognise the multinational nature of the study in the analysis. Issues with unit costs and statistical methods to analyse this type of studies will be explored in depth in subsequent chapters of this thesis.

2.9 Summary and conclusions

This chapter has provided an introduction to economic evaluation of health care interventions in multinational clinical trials. The different components of an economic evaluation were introduced with the aim of providing the reader with the appropriate background to follow the remainder of the thesis. The methods to report and present cost-effectiveness results described in this chapter, in particular the use of the cost-effectiveness plane, the net-benefit approach and cost-effectiveness acceptability curves are all examined and implemented in subsequent chapters.

The next chapter introduces the economic theory used in this thesis to explain economic evaluation alongside multinational clinical trials.

Chapter 3: Theoretical issues relating to the production function of a health care technology in multinational clinical trials

3.1 Introduction

The rationale for conducting clinical trials at a multinational level and the basis for economic evaluation alongside such studies were introduced in Chapter 1 and 2 respectively. It was argued that this type of study design has substantial advantages for researchers but also suffers from drawbacks that need further research before they are fully understood. Chapter 1 suggested that producers of health combine resources to achieve a certain level of output with the resources available. It is certainly true that this combination is likely to be different across countries participating in a common trial. Economic theory has proposed a set of tools that explains the behaviour of a producer, and this chapter introduces the concept of the production function of a health care technology and its relationship to multinational economic evaluations.

The idea behind this chapter is to provide a formal definition of a production function and detailed information of the main components and characteristics of this function. The concepts explained in this chapter provide theoretical foundations for the remainder of the thesis and will prove particularly useful in Chapter 7 and the simulation study conducted as part of this work.

The explanations that follow include an intuitive view of these economic tools and also a mathematical clarification. The algebra in this chapter is used in the simulation study later, and when appropriate full mathematical derivation for a particular concept is provided. The formulae and mathematics followed here can be found in standard microeconomics text at an undergraduate and graduate level (Folland et al. 2001; Varian 1992; Varian 1999).

The concept of the production function can be applied to almost all economic areas, but the examples used in this chapter relate specifically to health care. In the next section, the notion of how to represent a health care technology using production function tools is introduced. One of the key parameters in the simulation study in Chapter 7, the elasticity of substitution between input factors, is also introduced in that section. The chapter continues with a description of the cost minimisation problem faced by health care decision makers when allocating resources followed by an introduction to the measurement of efficiency in health care. The chapter finishes with a hypothetical multinational clinical trial example to explain the concepts covered in this third chapter.

3.2 Technology

3.2.1 Inputs and outputs

The theoretical perspective in this chapter assumes that there is a producer in each country interested in achieving a level of output given the resources that he or she has available. A producer could be a general practitioner in a primary care setting, a chef in restaurant or the line manager at a laundry department. We will use the term producers or producers of health care in the accompanying examples in this chapter.

The aim of producers of health care is to allocate the available resources efficiently to produce health. Resources are scarce and producers face several technical constraints to achieve the best possible level of health for the population. Technical constraints relate not only to medical equipment that may be available at any point in time but also includes labour and land. Medical equipment and other technology are often called capital goods in the economic literature.

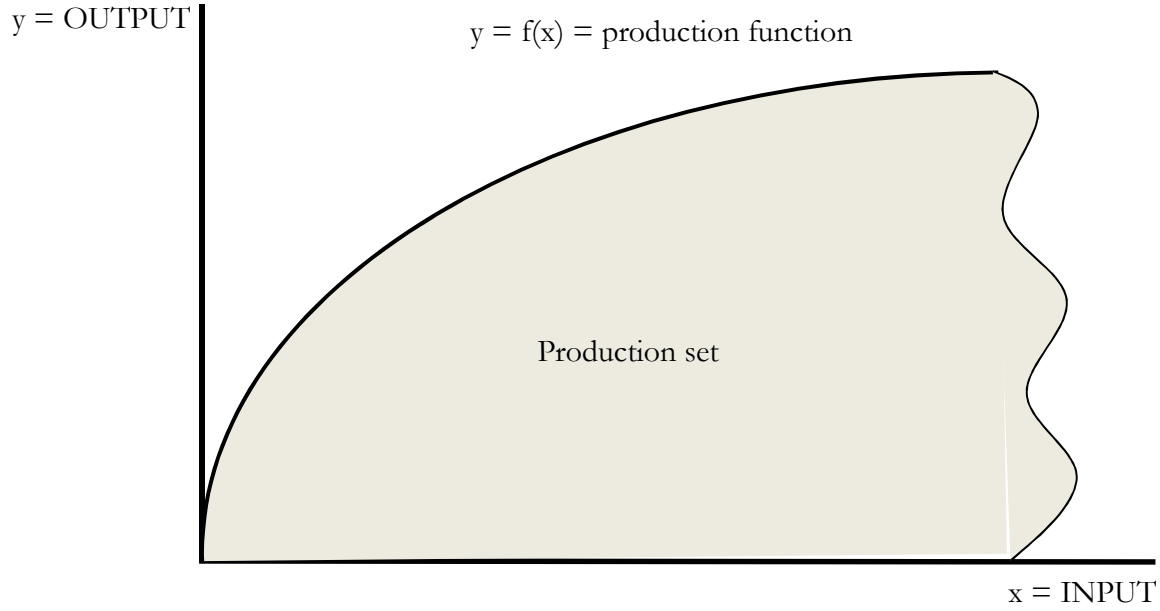
Labour, land and capital goods are examples of inputs or factors of production. Typically producers are able to combine these inputs to achieve a certain level of output. We can think of output in flow terms which in the case of health care can be measured. Examples of output might include "the number of successfully treated patients with stroke", "the number of deaths from a specific disease" or the "reduction in infant mortality" during a particular time period. Notice the time dimension when defining the measure of output. Time is important when describing the production process because decision makers can adapt the combination of inputs in response to technical innovations and other changes occurring during a particular time period[5].

3.2.2 Technological constraints

Health care decision makers face technological constraints that define the feasible combination of inputs required to produce a certain level of output. Hence producers of health need to study all the possible combinations of inputs and outputs that are technologically feasible. This set of combinations is referred to in economic theory as a production set. Figure 3.1 shows an example of a production set for the case of one input x and one output y. This production set shows pairs of points (x, y) that are technologically feasible and indicates that it is technologically possible to produce y amount of output if we have available x units of input. The function that describes the frontier of this production set is called the production function and indicates the maximum quantity of output that can be produced given the amount of input.

[5] In addition, time is the key feature when evaluating returns and economies of scale.

Figure 3.1: A production set and a production function

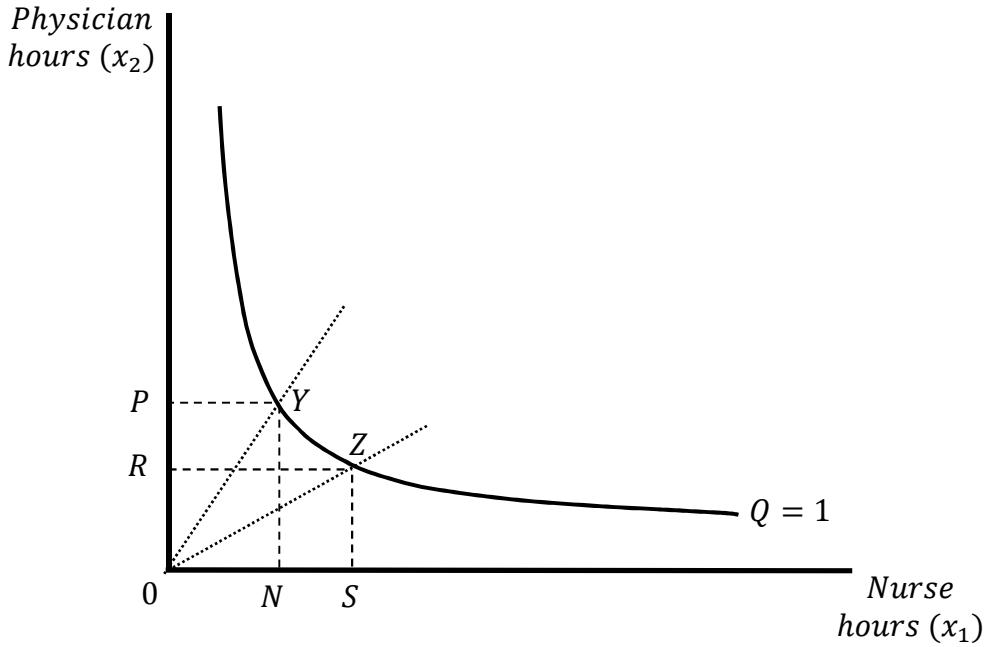


The concept of the production function can be extended to the case of several inputs. If we have two inputs x_1 and x_2, the production function $y = f(x_1, x_2)$ estimates the maximum amount of y that can be obtained if we use x_1 units of the factor 1 and x_2 units of the factor 2. The relationship between inputs and the associated level of output in a two-input world can be represented by an isoquant. An isoquant indicates all the possible combinations of inputs 1 and 2 that can produce a certain amount of output.

Figure 3.2 shows an example of an isoquant that represents the possible combination of physician hours (x_2) and nurse hours (x_1) to successfully treat one case or patient. Any point along the isoquant Q indicates a possible combination of inputs to treat one case. For example, one case can be treated by the ratio 0P/0N which is the slope of the line 0Y or by the ratio 0R/0S which is the slope of the line 0Z. At point Y, the amount of physician hours needed is larger than nurse time to treat a case, whereas at point Z the opposite occurs. Therefore, the isoquant indicates that there is some level of substitution between the inputs. However, nurses cannot always replace a physician for a certain tasks, such as prescribing

medication or performing an operation, and as we substitute physician for nursing times we are likely to reach a point where no substitution is possible. The degree of substitution that is possible in practice dictates the shape and the slope of the isoquant.

Figure 3.2: An isoquant indicating the degree of substitution between physician and nurses



The isoquant depicted in Figure 3.2 refers to the case of a Cobb-Douglas production function. The Cobb-Douglas case has the form $f(x_1, x_2) = Ax_1^a x_2^b$ and represents an isoquant that allows substitution between both inputs. This production function has interesting mathematical properties that facilitate the estimation of well-behaved isoquants. A well-behaved isoquant fulfils a set of properties that includes monotonicity, convexity and regularity. These sets of properties ensure that the isoquant makes sense from an economic point of view. A formal description of each property is provided in Appendix 1.

In the Cobb-Douglas production function, the parameter A measures the scale of production, i.e. the amount of output if we use one unit of each input. The parameters a and b measure the response of the output to changes in the amount of inputs. The Cobb-

Douglas function has been widely used by economists to represent production and utility functions since they were proposed (Cobb and Douglas 1928). Although it may seem a simple function at first, it has been shown that it frequently represents real life accurately (Aiyar and Dalgaard 2009).

3.2.3 The marginal product and the technical rate of substitution

If a particular production function is observed, producers of health can explore how the output changes when one of the inputs increases or decreases. Similarly, producers of health are interested in knowing the amount by which it would be necessary to increase one input and reduce others to keep the same level of output. The marginal product and the technical rate of substitution answer these questions.

Let us assume that we are operating at some point (x_1, x_2) for a particular production function $y = f(x_1, x_2)$, and would like to use more of factor 1 while keeping factor 2 constant. The marginal product is the additional amount of output obtained by increasing one unit of factor 1 and keeping factor 2 constant. Using algebra, the marginal product can be expressed as:

$$\frac{\Delta y}{\Delta x_1} = \frac{f(x_1 + \Delta x_1, x_2)}{\Delta x_1} = \frac{\partial f}{\partial x_1} = MP_1(x_1, x_2) \qquad [3.1]$$

where Δ indicates a change in y and x_1 and ∂f indicates the first derivative.

It is expected that the marginal product will increase each time we use more of factor 1, but at a decreasing rate. This is known as the law of diminishing marginal product and only applies when one of the factors, in this case factor 2 is kept fixed.

Assume now that we are operating at some point (x_1, x_2) for a particular production function $y = f(x_1, x_2)$, and we would like to use extra units of factor 1 but remove some of factor 2 to keep the output constant. This is just the slope of the isoquant in Figure 3.2 and is known as the technical rate of substitution (TRS)[6]. Using the example of physician and nurses, the producer of health may be interested in reducing/increasing the time for a physician in a particular case at the expense of increase/reduce nursing time that he or she has available. This often occurs in health care when dealing with management, budget, or logistic constraints.

Using algebra, if small changes in the inputs are defined as $dx = (dx_1, dx_2)$, the associated change in output can be expressed as:

$$dy = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2$$

If $dy = 0$, so changes dx_1 and dx_2 in factor 1 and 2 keep the output constant (adjustment along the isoquant). We can solve as:

$$TRS(x_1, x_2) = \frac{dx_2}{dx_1} = -\frac{\partial f / \partial x_1}{\partial f / \partial x_2} = -\frac{MP_1(x_1, x_2)}{MP_2(x_1, x_2)} \quad [3.2]$$

Figure 3.2 showed that the slope of the isoquant decreases as we move from x_2 to x_1. This suggests that as the amount of x_2 reduces, increasing amount of x_1 need to be substituted to keep the output constant. Using our example of health labour, as the amount of physician hours are reduced, increasing number of nursing hours are needed to treat a patient successfully. This is known as diminishing technical rate of substitution and explains the negative slope of the isoquant.

[6] The technical rate of substitution is also called the marginal rate of technical substitution in some textbooks (Folland et al. 2001).

An example of how to derive the TRS from a Cobb-Douglas function is given as follows. The Cobb-Douglas function is defined as $f(x_1, x_2) = Ax_1^a x_2^b$ and calculating the first derivative as explained in equation [3.2] we obtain:

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = Aax_1^{a-1}x_2^b$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = Abx_1^a x_2^{b-1}$$

Then,

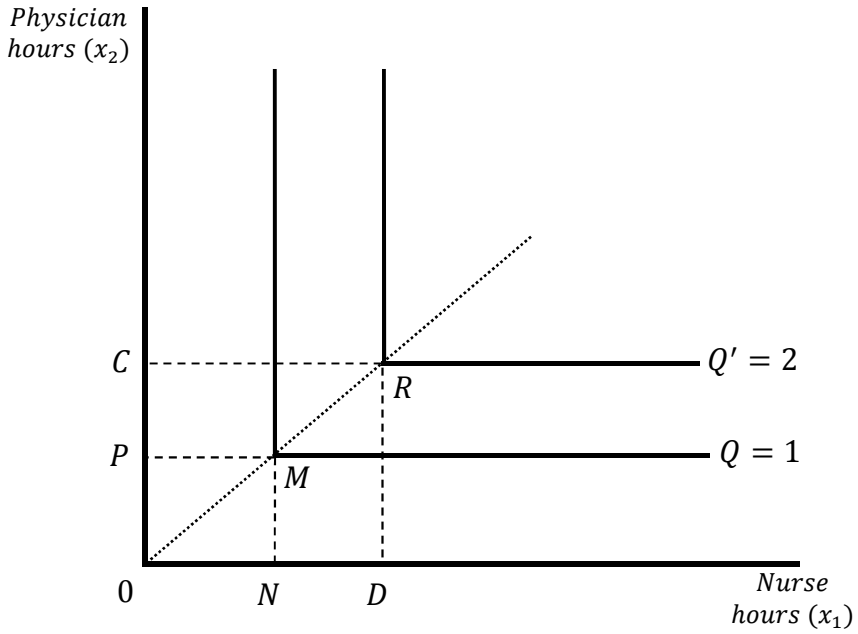
$$TRS = -\frac{\partial f/\partial x_1}{\partial f/\partial x_2} = -\frac{a}{b}\frac{x_2}{x_1} \qquad [3.3]$$

3.2.4 Additional examples of isoquants

Figure 3.2 presented the case of a Cobb-Douglas production function but isoquants can take various forms and in this section two additional examples are introduced. Figure 3.3 shows a fixed proportions or Leontief production function defining the relationship between doctors and nurses. This production function takes the following form $f(x_1, x_2) = min\{x_1, x_2\}$ and indicates the case where no substitution is possible between the inputs. The isoquant shows that there is only one possible combination of physician and nurse time to achieve one successful treated case ($Q = 1$) with 0P physician hours and 0N nurse hours. This type of isoquant is not uncommon in health care when some medical tasks need to be conducted by well trained physicians and hence a fixed amount of hours from the physician is required such as cardiac surgery for instance. Adding extra nurses for the care of patient will not improve the output unless the physician hours are also increased. For instance, if a producer wants to achieve two successfully treated cases ($Q' = 2$), he or she needs to allocate physician and nurse hours according to the ratio (0C/0D). The possible combinations of doctors and nurses are summarised by the production line that passes through the corner

points M and R. The slope of this line is the ratio of doctors to nurses needed to achieve a particular level of output.

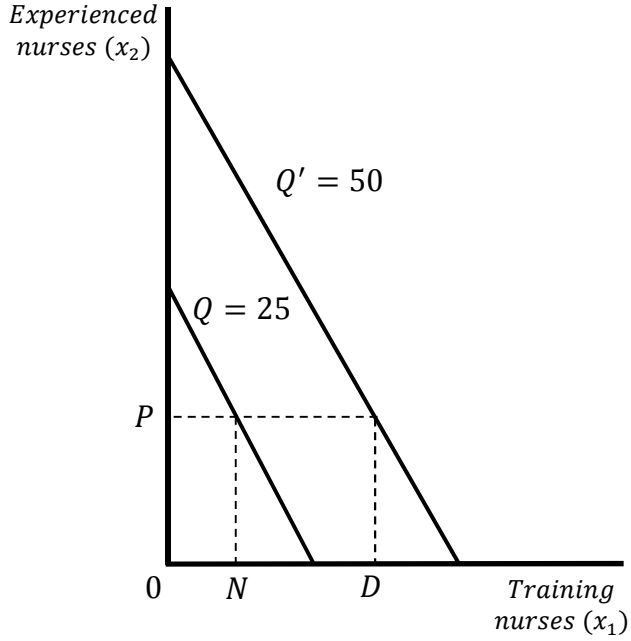
Figure 3.3: Fixed proportions or Leontief technology isoquants between physician and nurses



The next example in Figure 3.4, presents the case of isoquants that are perfect substitutes. In this example the total amount of output produced depends only on the total amount of inputs available. For instance, the distribution of medicines in a specific hospital inpatient department is conducted by two types of nurses, experienced and training nurses. Assuming there are no serious side effects on the drugs given, it does not matter who distribute the medicines as long as we have a total number of nurses to cover the whole department. In figure 3.4 the total number of nurses needed to distribute medicines to 25 inpatients (Q) is given by adding 0P experienced nurses and 0N training nurses. Let us assume that the number of experienced nurses in this inpatient department is fixed but the number of trainees can change depending on the need of the department. If medicines had to be distributed to 50 inpatients (Q'), then the total number of nurses needed would be 0P+0D.

Therefore, the production function of a perfect substitute isoquant takes the following expression $f(x_1, x_2) = ax_1 + bx_2$. Note that this type of isoquant has constant slope and therefore constant technical rate of substitution $-a/b$.

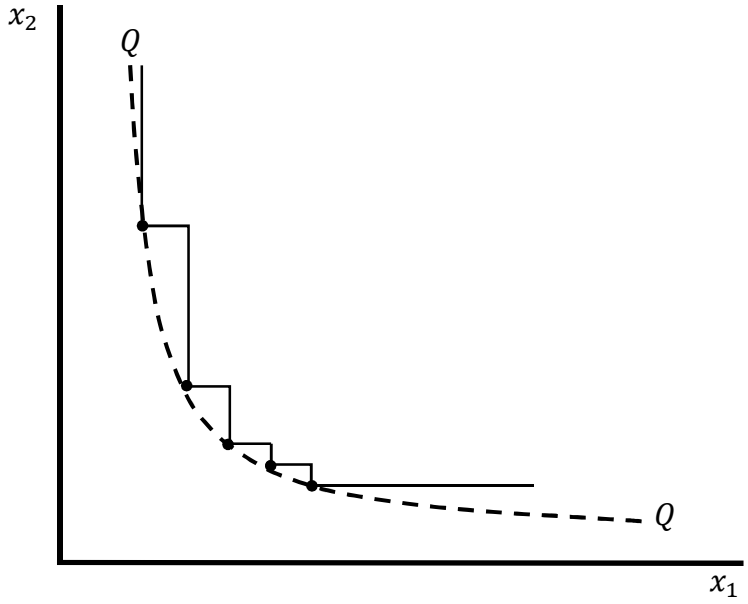
Figure 3.4: Perfect substitutes isoquants between experienced and training nurses



3.2.5 Parametric representation of a technology

All possible combinations of inputs needed to produce a given level of output are rarely observed. Instead the observable combinations can be parameterised under a known functional form. The idea behind this parametric approximation is illustrated in Figure 3.5, which shows how different combination of inputs can be represented by the dotted line production function.

Figure 3.5: Parametric approximation of an isoquant



This parametric approximation is a powerful tool as it allows analysts to estimate production functions in a real-life setting. Although production functions can take various forms, economists have found some functional forms that fit reasonably well real production plans. The production functions introduced earlier in the chapter are some examples. This parametric approach to combinations of inputs will be of great assistance in the simulation study conducted in Chapter 7.

3.2.6 The elasticity of substitution

So far, we have defined and explained that an isoquant can represent combinations of inputs to achieve a level of health output. In theory then, the producer of health can move along the isoquant and select the amount of input-factor resources. The producer knows that in order to keep the same level of output, to increase the amount of one of the inputs, entails reducing the other input in some magnitude. This information is provided by the slope of the isoquant and the TRS as described in the previous section. However, the TRS does not

provide any information on how easy or difficult is for a particular producer of health to substitute between factors. This information is provided by the elasticity of substitution and a definition is provided next.

The elasticity of substitution measures the curvature of the isoquant. i.e. the percentage change in the factor ratio divided by the percentage change in the technical rate of substitution (TRS) with the output held fixed. It indicates how responsive the factor ratio is to changes in the TRS. Expressing $\Delta(x_2/x_1)$ as the change in the factor ratio and ΔTRS the change in the TRS, the elasticity of substitution is:

$$\sigma = \frac{\frac{\Delta(x_2/x_1)}{x_2/x_1}}{\frac{\Delta TRS}{TRS}} \qquad [3.4]$$

The elasticity of substitution provides information about how the ratio of factor inputs changes as the slope of the isoquant varies. For instance, if a small change in the slope results in a large change in the factor ratio, the elasticity of substitution is large.

Assuming the percentage changes are small, and applying limits to equation [3.4] as Δ approximates to zero yields the following form for the elasticity of substitution:

$$\sigma = \frac{TRS}{(x_2/x_1)} \frac{d(x_2/x_1)}{dTRS} \qquad [3.5]$$

For convenience, it is useful to express σ using the logarithmic derivative. If we define $y = g(x)$, the elasticity of y with respect to x is the percentage change in y when a small percentage change in x occurs. This can be expressed as:

$$\varepsilon = \frac{\frac{dy}{y}}{\frac{dx}{x}} = \frac{dy}{dx}\frac{x}{y}$$

And using the chain rule for partial derivatives (Chiang A.C. and Wainwright K. 2005), it can be shown that

$$\frac{dln\ y}{dln\ x}\frac{dln\ x}{dx} = \frac{dln\ y}{dx}$$

Which after calculating the partial derivatives $dln\ x/dx$ and $dln\ y/dx$ produces

$$\frac{dln\ y}{dln\ x}\frac{1}{x} = \frac{1}{y}\frac{dy}{dx}$$

Therefore ε can be expressed as:

$$\varepsilon = \frac{dln\ y}{dln\ x} = \frac{x}{y}\frac{dy}{dx}$$

Finally we can substitute this expression into equation [2.5] where $y = g(x) = (x_2/x_1)$ and $x = TRS$ and obtain[7]:

$$\sigma = \frac{dln(x_2/x_1)}{dln|TRS|} \qquad [3.6]$$

Using the Cobb-Douglas production function, we can calculate the elasticity of substitution using the TRS calculated before. The technical rate of substitution of a Cobb-Douglas in equation [2.3] was estimated to be:

$$TRS = -\frac{a}{b}\frac{x_2}{x_1}$$

Rearranging the expression we get

$$\frac{x_2}{x_1} = -\frac{b}{a}TRS$$

and applying logarithms in each side gives

[7] The absolute values for the technical rate of substitution in the denominator transform the TRS into a positive number and therefore the logarithm makes sense. In the previous section we showed that the slope of the isoquant is negative.

$$ln\frac{x_2}{x_1} = ln\frac{b}{a} + ln|TRS|$$

Finally, it can be easily calculated that

$$\sigma = \frac{dln(x_2/x_1)}{dln|TRS|} = 1$$

3.2.7 The constant elasticity of substitution (CES) production function

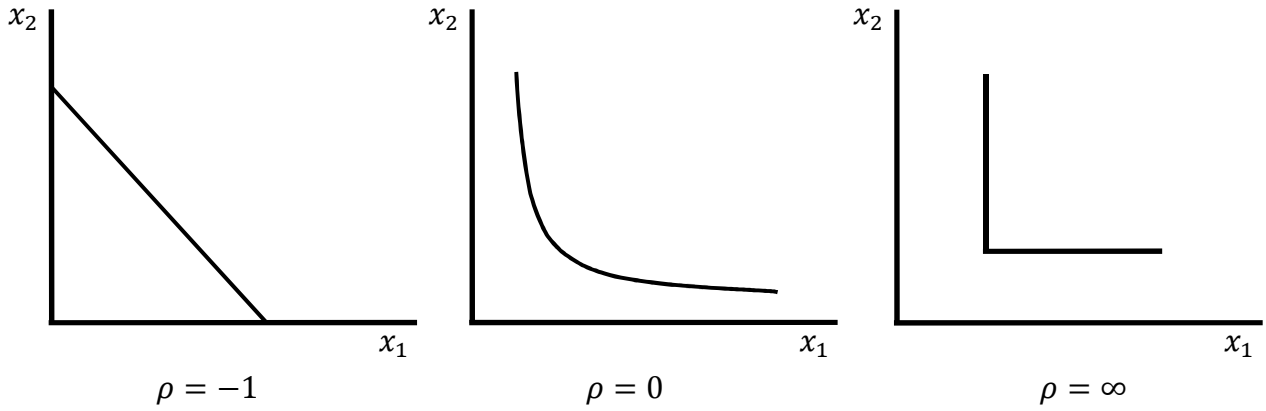
The CES production function is a useful functional form suggested in its final form by Arrow and colleagues (Arrow et al. 1961). The CES production function will be the production function assumed to model the behaviour of the isoquants in each country in Chapter 7. A description of the mathematical properties and the intuition why this production function is so useful is given next. The mathematical representation of this production function takes the following form:

$$y = \gamma\left[\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right]^{-1/\rho}$$

where γ is the efficiency parameter, δ is the distribution parameter, ρ is a substitution parameter and x_1 and x_2 factors of production. The efficiency parameter γ $(0 < \gamma < \infty)$ indicates the state of the technology, i.e. a change in the parameter γ changes the output for any given sets of inputs in the same proportion. The distribution parameter δ $(0 < \delta < 1)$ indicates the relative factor shares to achieve a level of output. The final substitution parameter ρ $(-1 < \rho \neq 0)$ helps determine the value of the elasticity of substitution.

The CES production function includes the three examples introduced at the beginning of the chapter depending on the value of ρ and this is depicted in Figure 3.6.

Figure 3.6: The CES production function depending on rho



The assumed constant elasticity of substitution is helpful to explore different levels of substitution effects for a particular elasticity of substitution. To demonstrate the constant elasticity property we define $y = g\big(f(x_1, x_2)\big)$ such as $\frac{\partial y}{\partial x} = g'\big(f(x_1, x_2)\big) * f'(x_1, x_2)$ with $f(x_1, x_2) = (\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho})$. The TRS in the CES production function is given then by

$$
\begin{aligned}
TRS &= -\frac{\partial f/\partial x_1}{\partial f/\partial x_2} \\
&= -\frac{\frac{1}{\rho}\gamma\left[\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right]^{\frac{-1}{\rho}-1} * (-\rho)\delta x_1^{-\rho-1}}{\frac{1}{\rho}\gamma\left[\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right]^{\frac{-1}{\rho}-1} * (-\rho)(1-\delta)x_2^{-\rho-1}} \\
&= -\frac{\delta}{(1-\delta)}\left(\frac{x_1}{x_2}\right)^{-\rho-1} = -\frac{\delta}{(1-\delta)}\left(\frac{x_2}{x_1}\right)^{1+\rho}
\end{aligned}
\qquad [3.7]
$$

Working with expression [3.7] and applying absolute terms in both sides we obtain:

$$|TRS| = \frac{\delta}{(1-\delta)}\left(\frac{x_2}{x_1}\right)^{1+\rho}$$

This is the same as:

$$\frac{x_2}{x_1} = \left[\frac{(1-\delta)}{\delta}|TRS|\right]^{\frac{1}{1+\rho}}$$

and taking logs,

$$ln\frac{x_2}{x_1} = \frac{1}{1+\rho} ln\frac{(1-\delta)}{\delta} + \frac{1}{1+\rho} ln|TRS|$$

Using the definition introduced in equation [3.6] and applying the logarithmic derivative

$$\sigma = \frac{dln(x_2/x_1)}{dln|TRS|} = \frac{1}{1+\rho} \qquad [3.8]$$

Equation [3.8] indicates that once the parameter ρ has been defined, the elasticity of substitution of a CES production function is constant. This is a useful property because it is possible to define a particular level of elasticity of substitution and hence a level of responsiveness of the substitution between factors. This property will be used in Chapter 7 and the simulation study to determine how different levels of substitution effect influence estimates of incremental costs across countries in a multinational trial.

3.3 Costs

3.3.1 Cost minimisation

Once a production set and the associated production function are identified, producers need to decide how to minimise the costs of producing a particular level of health. This cost-minimisation problem should not be interpreted as "decision makers should just minimise costs". The correct interpretation is "decision makers should minimise costs to produce a certain level of health output". For example, let's assume that a decision maker decides to minimise costs to achieve a level of output y and he or she chooses the cost c_1. If there is a cheaper feasible way of producing y with for instance a cost $c_2 < c_1$, then the decision maker is not allocating resources efficiently. The extra costs $(c_1 - c_2)$ could be invested in an alternative activity and for another group of patients and potentially generate an extra level of health Δy increasing the overall level of health of the population.

The cost-minimisation problem can be represented in the two-input world followed in this chapter. We have two factor of productions or inputs x_1 and x_2 and the associated prices to each of these input factors are ω_1 and ω_2 respectively. The producer of health care wants to find out the cheapest way of producing the level of health output y. If the production function $f(x_1, x_2)$ defines the maximum amount of output attainable given the level of inputs, the problem can be written as:

$$\begin{aligned} \min_{x_1,x_2} \quad & \omega_1 x_1 + \omega_2 x_2 \\ subject\ to: \quad & f(x_1, x_2) = y \end{aligned}$$

In this minimisation problem, $c(\omega_1, \omega_2, y) = \omega_1 x_1 + \omega_2 x_2$ is known as the cost function and indicates the minimum cost needed to produce y when the input prices are ω_1 and ω_2 for factor 1 and factor 2 respectively. It is useful to plot both costs and technological constraints or isoquants in the same graph. To make the graph easier to understand, we may consider plotting all the combinations of inputs that yield the same level of costs C. This can be written as:

$$C = \omega_1 x_1 + \omega_2 x_2$$

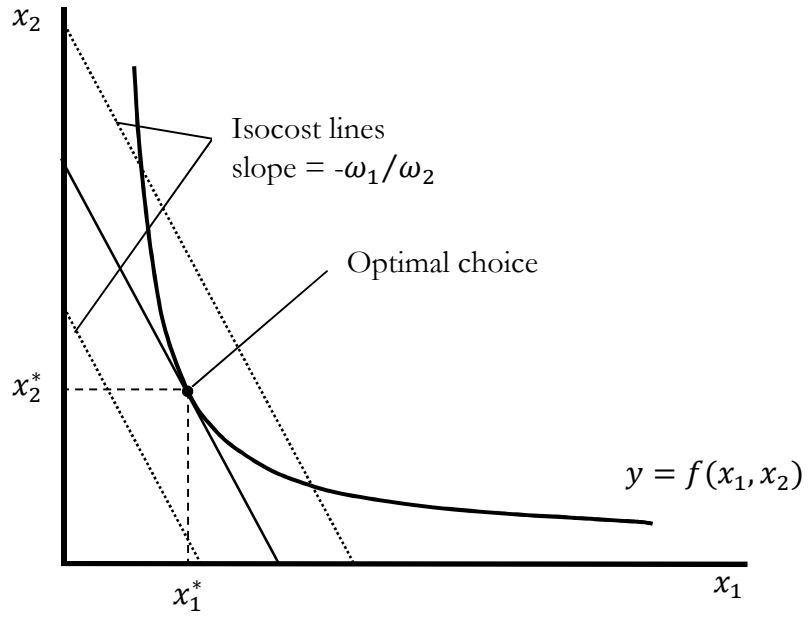
And rearranging,

$$x_2 = \frac{C}{\omega_2} - \frac{\omega_1}{\omega_2} x_1 \qquad [3.9]$$

Equation [2.9] has a slope of $-\omega_1/\omega_2$ and a vertical intercept C/ω_2. Varying the cost C we can obtain a family of isocost lines. Along the isocost line, we have the same cost C and hence higher lines mean higher costs. Figure 3.7 includes both the isoquants and the isocost

lines. The graph shows the cost-minimisation problem for the decision maker visually, i.e. find the point in the isoquant that has the lowest isocost line and therefore the lowest costs.

Figure 3.7: The cost-minimisation problem for the health care decision maker



The optimal solution for the cost-minimisation problem can be provided using the Lagrange multiplier optimisation method (Chiang A.C. and Wainwright K. 2005). In this case the Lagrangian is defined as:

$$L = \omega_1 x_1 + \omega_2 x_2 - \lambda(f(x_1, x_2) - y)$$

Differentiating with respect to x_1, x_2, and λ provides the first-order conditions:

$$\omega_1 - \lambda \frac{\partial f(x_1, x_2)}{\partial x_1} = 0$$

$$\omega_2 - \lambda \frac{\partial f(x_1, x_2)}{\partial x_2} = 0$$

$$f(x_1, x_2) - y = 0$$

The last condition indicates the constraint. Dividing the first condition by the second condition and rearranging we obtain:

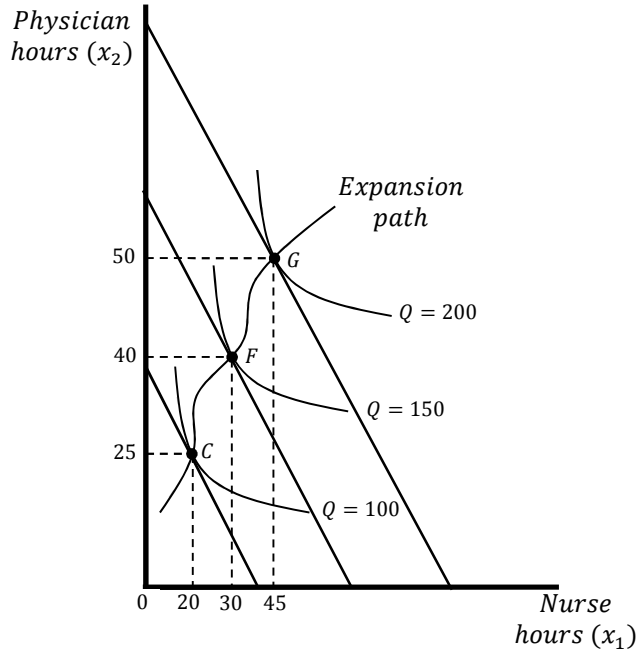
$$\frac{\omega_1}{\omega_2} = \frac{\partial f(x_1, x_2)/\partial x_1}{\partial f(x_1, x_2)/\partial x_2} \quad [3.10]$$

Equation [3.10] suggests that the technical rate of substitution should be the same as the ratio of the input prices. This is depicted in Figure 3.8 by a black dot indicating the optimal choice solution. This is the solution that is used in Chapter 7 to determine how input factors are combined to produce a successfully treated patient across countries in the simulation study.

3.3.2 Deriving the cost function and expansion paths for the producer

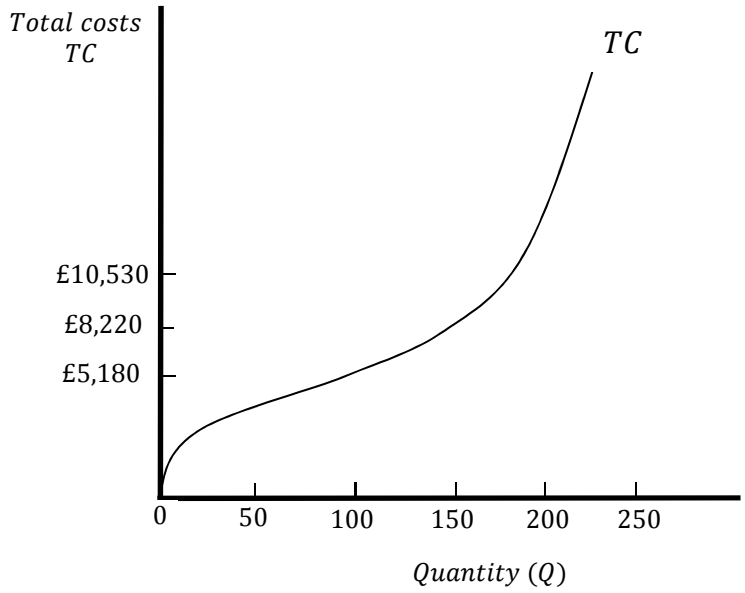
The cost function describes the relationship between output and costs. In this section we present the cost-minimisation problem which shows how producers choose the optimal level of output given the relative prices of the inputs. Figure 3.8 presents three isoquants and isocost lines with optimal solutions from a cost-minimisation problem for three different levels of output. We continue with the example of combining clinicians and nurses to treat a group of patients successfully. The lower isoquant indicates possible combinations of health labour to produce 100 successful cases. The higher isoquants are able to produce more with 150 and 200 successful patients respectively. Each of the optimal solutions C, F and G indicate the minimum cost of producing each output level. The cost function is given by $c(\omega_1, \omega_2, y) = \omega_1 x_1 + \omega_2 x_2$ and we assume that the price of a doctor's hour is ω_2=£180 and for the nurse's hour is ω_1=£34. Then producing 100 cases requires 25 hours of physician and 20 hours of nurses costing £5,180 in point C. Similarly in point F for 150 cases the total cost would be £8,220 and, in point G for 200 cases the total cost would be £10,530.

Figure 3.8: The cost-minimisation problem for three levels of output between doctors and nurses



The line that joins the optimal points C, F and G is called the expansion path for this producer. The expansion path indicates the minimum cost for each level of output required. Therefore the expansion path holds information on the total costs and the average cost (cost/quantity) of producing each level of output. Plotting the information on minimum costs and the associated level of output we can derive the total cost function for this producer which has a S-shape and it seems to represent well the behaviour of most firms (Figure 3.9).

Figure 3.9: The derived total cost function



This cost function is also called a frontier and represents the minimum possible cost of producing a level of output. However, in many situations the producer operates at levels above the frontier and so incurs some inefficiencies. The next section provides an introduction to the measurement of efficiency in health care.

3.4 Measures of efficiency in health care

Once the cost-minimisation problem is solved, the producers of health care act efficiently because they are providing the chosen level of health with a combination of inputs that minimise costs. However, there are situations where producers deviate from the optimal, introducing inefficiencies in the allocation of resources.

Figure 3.10 shows the cost minimisation problem again as described above. The isoquant QQ indicates the possible combination of inputs x_1 and x_2 to produce a level of health y. The line BB indicates the isocost line and indicates the possible combination of inputs that yield the same cost C with input prices ω_1 and ω_2.The optimal and efficient solution is achieved at point P_1 where the ratio of input prices $-\omega_1/\omega_2$ is equal to the technical rate of

substitution. At point P_2 although the producers are achieving y level of health, they are not minimising costs. A reduction in costs $P_2P_2^*$ is possible that involves a trade-off between inputs and places the producer at point P_1. Situation P_2 is called allocative inefficiency, where allocative efficiency measures efficiency with respect to prices (Jacobs et al. 2009). The price efficiency of P_2 is measured by the ratio OP_2^*/OP_2.

Figure 3.10: Allocative efficiency in the two-input world

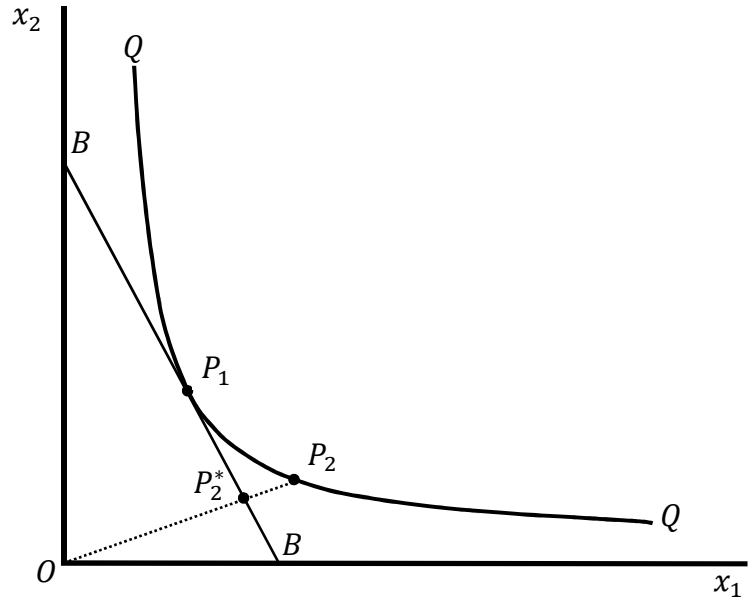
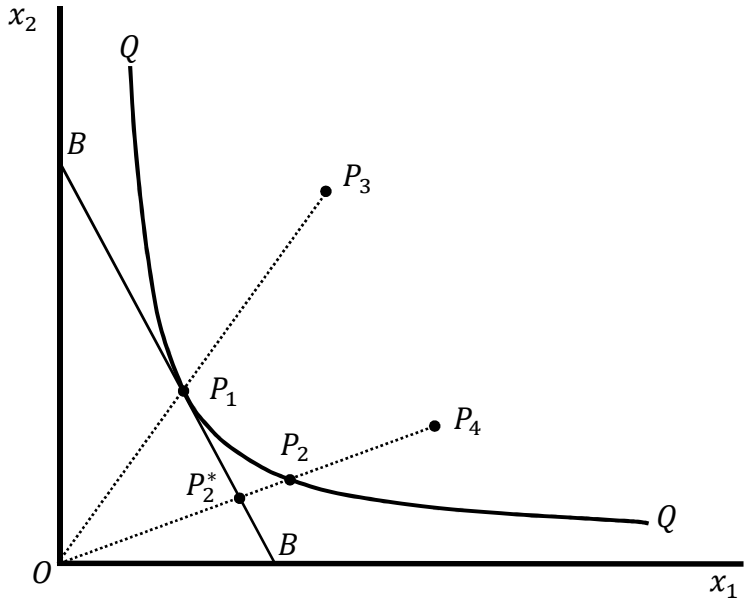


Figure 3.10 presented a situation where the producer was placed on the isoquant, indicating a technically feasible combination of inputs to produce the level of health y. However the decision maker may be operating at a point inside the isoquant yielding another type of inefficiency termed technical inefficiency (Jacobs et al. 2009). Figure 3.11 expands our presentation and includes two new points P_3 and P_4. At point P_3 the decision maker still produces a level of health y and at the optimal relative price $-\omega_1/\omega_2$, but is not choosing the efficient technical combination of inputs. Some degree of technical inefficiency is indicated by the ratio OP_1/OP_3. At point P_4, the decision maker is not choosing the efficient

technical combination of inputs to produce the level of health y and is not purchasing at the optimal relative price. Therefore point P_4 presents both allocative and technical inefficiency.

Figure 3.11: Allocative and technical efficiency in the two input world



3.5 Theory of production in the economic evaluation of multinational clinical trials

The following section applies the theoretical foundations described above to the case of a multinational clinical trial. If producers of health care behave according to the theory of production, then it is possible to explain differences in the total costs per patient of a health technology across countries, using the concepts introduced in this chapter. The aim of this section is to show then that the total cost per patient in a particular country is likely to be different due to differences in relative prices, substitution effects or a combination of both. This section will provide the reader with the theoretical foundations for the simulation study conducted as part of this thesis in Chapter 7.

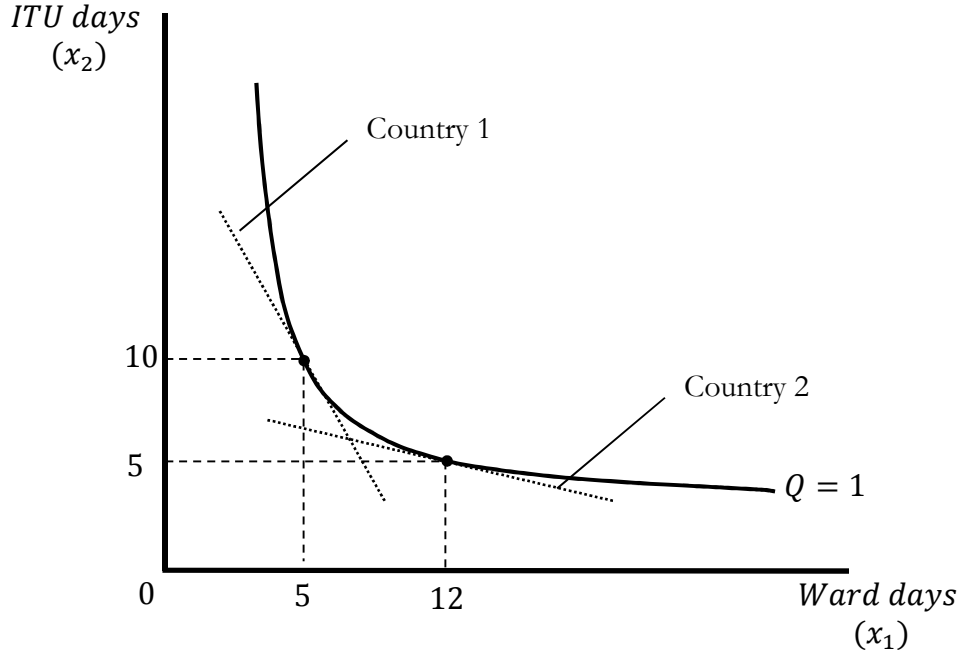
Using a hypothetical numerical example, three scenarios are introduced: the first scenario assumes the same relative substitution of inputs in each country but different relative prices; the second scenario assumes equal relative prices but different substitution of inputs in each country; and the final scenario assumes different substitution of inputs and relative prices in each country.

The description that follows assumes that there is a producer in each country interested in combining two inputs of a new health care technology to achieve one successfully treated patient. The two inputs that affect the production function of this technology are the number of days in a hospital ward (x_1) and the number of days in an intensive therapy unit (ITU) (x_2). A multinational clinical trial has been conducted to determine the cost-effectiveness of this new technology compared with current practice. To simplify the current exposition, only two countries are assumed to participate in the multinational clinical trial. The trial collected information on the number of days in ward and ITU and also the associated unit cost data. Both countries have a production function $y = f(x_1, x_2)$ and a cost function $c(\omega_1, \omega_2, y) = \omega_1 x_1 + \omega_2 x_2$ where ω_1 is the price of one day in a hospital ward and ω_2 is the price of one day in ITU.

The first scenario presented in Figure 3.12 illustrates the case where both countries behave according to the same isoquant and therefore have similar substitution effects between the inputs. However, in country 1, ω_1=£500 and ω_2=£1,000 whereas in country 2, 1, ω_1=£500 and ω_2=£1,500. In this example the price of one day in hospital ward is the same between countries but the price of one day in ITU is more expensive in country 2. Therefore, the ratio of input prices or relative price is higher in country 1 and this is indicated by the isocost line with a higher slope. The optimal solutions in each country are derived through a cost-

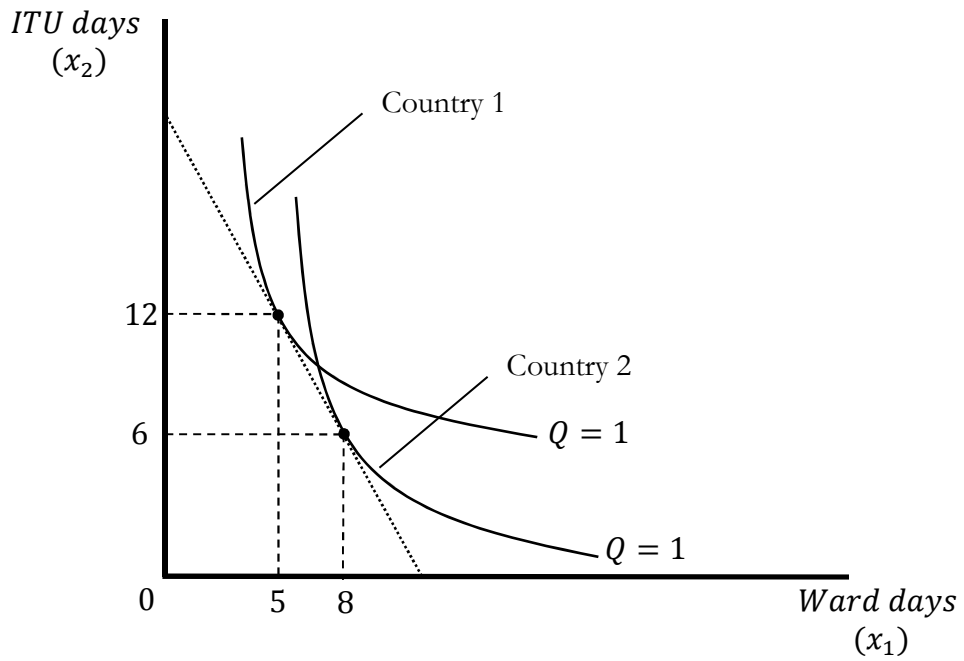
minimisation problem and are represented by the black dots in the figure. In country 1 the optimal solution suggests that a successfully treated patient requires 5 days in a ward and 10 days in ITU whereas in country 2 the optimal solution is 12 days in ward and 5 days at an ITU facility. The total cost per patient in country 1 is estimated to be £12,500 and in country 2 this estimate is £13,500.

Figure 3.12: Scenario with the same isoquant in each country but different relative prices between countries



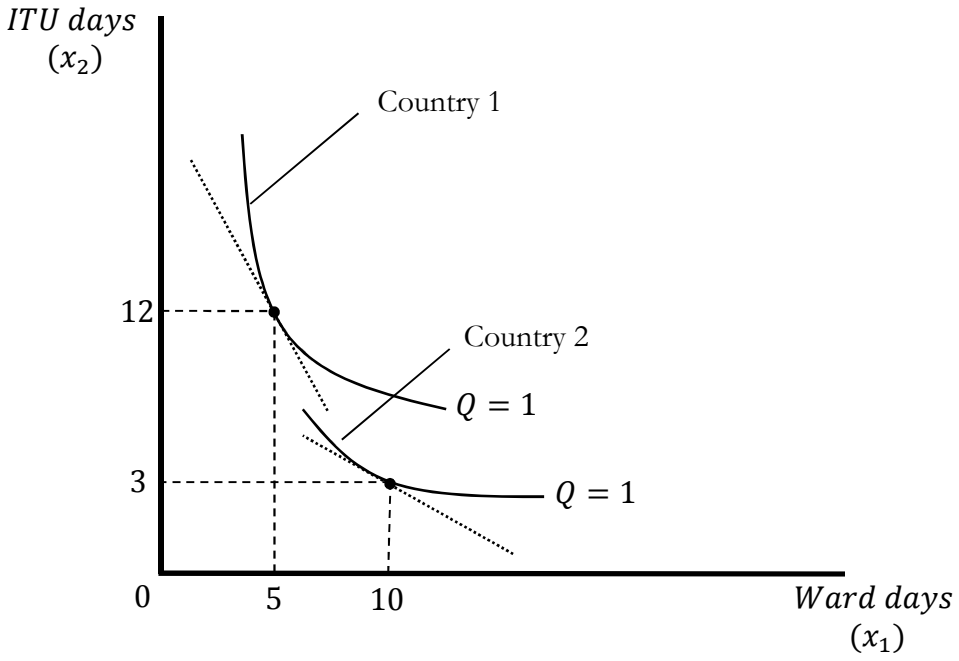
The second scenario in Figure 3.13 presents a situation with similar relative prices but a different substitution effect in each country. The different substitution effect is represented by a different isoquant in each country. This situation is possible when, for example, the producer in each country faces a different elasticity of substitution. The input prices are ω_1=£500 and ω_2=£1,000 and the associated relative price is 1/2. The optimal solution is indicated by the black dots and suggests that in country 1 the number of ward and ITU days needed are 6 and 12, respectively whereas in country 2 are 8 and 6 respectively. This yields a total cost per patient of £14,500 and £10,500 in country 1 and 2 respectively.

Figure 3.13: Scenario with the different isoquants and similar relative prices between countries



The last scenario represented in Figure 3.14 indicates the case where both the relative prices and substitution effects are different between the countries. The input prices are assumed to be similar to scenario 1. The scenario has been exaggerated to illustrate the case of having an obvious substitution effect differences between countries. The optimal solutions indicated by the black dots are 5 days in ward and 12 days in ITU in country 1 and 10 days in ward and 3 days in ITU in country 2. This has an associated total cost per patient in country 1 of £14,500 and in country 2 of £9,500.

Figure 3.14: Scenario with different isoquant and relative prices between countries



3.6 Conclusion

This chapter has provided some definitions of the theoretical foundations of a production function of a health care technology and its relationship to economic evaluation alongside a multinational clinical trial. It was shown that if countries behave according to the theory described here, the same output may be produced with different input combinations, different factor prices, and different costs per patient. These differences may be driven by dissimilar relative prices and substitution effects across the countries participating in a multinational clinical trial. This has important implications for analysts dealing with international data and conducting economic evaluations across several countries. In particular a failure to recognise differences in the underlying production function and cost function of the health technologies under evaluation may lead to inappropriate methods of analysis and conclusions. The remaining chapters of the thesis will explore these implications for current methods of analysing and presenting cost-effectiveness analyses alongside multinational clinical trials.

The next chapter presents the current situation concerning cost-effectiveness alongside multinational clinical trials and identifies the multinational characteristics of these studies. This is an aspect that has received little attention in the literature but is of great interest to health economists.

Chapter 4: A structured review on cost-effectiveness analyses alongside multinational clinical trials

4.1 Introduction

This chapter identifies published cost-effectiveness studies alongside multinational randomised controlled trials. A structured review is carried out and describes in detail the multinational nature of these studies, the reporting of cost-effectiveness results and the costing and outcomes methods applied during their analyses.

Published multinational economic evaluations have been explored in different literature searches over the last couple of years. Barbieri and colleagues for instance identified 46 multinational economic evaluations of pharmaceuticals in Europe and explored the variability of cost-effectiveness estimates across countries from 1998 to 2001 (Barbieri et al. 2005). The flexibility of the analysts to choose a method that permits variability in resource use across countries and the real value of cost-effectiveness thresholds across jurisdictions were identified as the main factors influencing variability in cost-effectiveness results. In another review, Halliday and Darba explored how cost data were reported in 54 multinational economic evaluations for a period of ten years until 2002 and found wide variability in the quality of the reporting of these studies (Halliday and Darba 2003). Barbieri et al. and Halliday and Darba included both trial-based studies and economic models in their reviews.

Torti and colleagues identified 23 studies in a literature search from 1995 to 2004 of multinational economic evaluation alongside clinical trials in the area of cardiology and extracted the type of costs included in the analysis, the costing strategies adopted and the extrapolation of costs and outcomes beyond the clinical trial period in these studies (Torti,

Jr. et al. 2006). The authors concluded that more guidance is needed as the design and the analysis conducted differed widely across studies. The review of generalisability by Sculpher and colleagues, first introduced in Chapter 1 of this thesis, included a systematic review of economic evaluations undertaken alongside multicentre randomised controlled trials from 1994 to 2000 (Sculpher et al. 2004b). Multinational economic evaluations alongside clinical trials are by definition multicentre studies and hence 21 studies of this type were retrieved in their review. The review concentrated on identifying the type of cost analysis performed and reported that around 50% of the studies applied a single set of unit costs to resource use in all countries.

The outcomes from these previous literature searches suggest that articles presenting empirical evidence and methodological developments are currently available. As a result this review concentrates on identifying and extracting both types of studies. The first group of studies identifies different methodological approaches to analyse economic evaluations alongside multinational clinical trials. These methods will be described and tested in detailed using real data in subsequent chapters, hence only a brief description of these approaches is included in this review. The empirical evidence review describes the methods used to collect outcomes, resource use and unit costs, the approaches to the analysis and the reporting of the results.

The objectives of this chapter are therefore to conduct a structured review of cost-effectiveness analysis alongside multinational randomised controlled trials to update current literature searches, extract new data not yet retrieved by previous reviews and find all relevant applied and methodological evidence in analysing and reporting cost-effectiveness analysis alongside multinational clinical trials.

4.2 Methods

The methodology in this review follows the standard procedure for identifying relevant evidence in a particular topic. As one of the main criticisms of a general review is its lack of transparency, a structured review was conducted (Light and Pillemer 1984). In this type of review, the evidence is searched, synthesized, and presented in a transparent manner with its methods and results clearly explained and reported. The structure of this review includes the following headings: inclusion and exclusion criteria for selecting papers, search strategy and databases involved, fields and information extracted, presentation of results of the review and a discussion.

4.2.1 Inclusion/exclusion criteria

Articles were included if they reported cost-effectiveness analyses alongside a multinational randomised controlled trial with individual patient-level data (IPD) on resource use and health outcomes in more than one country. To identify other relevant articles, papers were also included if analytical strategies either for resource use, health outcomes or both that could assist in the performance of cost-effectiveness analysis alongside multinational clinical trials were explored. To facilitate the reporting of the review, the identified evidence was categorised in two groups: empirical or applied evidence, and methodological studies.

These inclusion criteria ensured retrieval of, for example the main cost-effectiveness results from a trial, and a second publication using the same trial data to explore, a methodological question such as different methods of analysing cost-effectiveness analysis alongside multinational clinical trials. In this case, the former article would be categorised as empirical whereas the latter one would be categorised as methodological.

Cost-effectiveness analyses alongside non-randomised multinational clinical trials, economic models, published abstracts of conference proceedings, research protocols, discrete choice experiments and willingness to pay studies were not included in the review.

4.2.2 Search Strategy

The core search strategy was performed in OvidSP and the NHS Economic Evaluation database (NHS-EED), two large web-based electronic databases. OvidSP is the result of the recent merge of the two well-known interfaces Ovid and SilverPlatter. The selected databases within OvidSP included in this review were MEDLINE and EMBASE. Some overlapping between these two databases is expected but OvidSP includes features to identify and remove possible duplicates. The NHS-EED is one of the databases from the Centre of Review and Dissemination at the University of York which performs rigorous structured abstracts of economic evaluation of health care technologies worldwide (NHS Centre for Reviews and Dissemination 2001). The NHS-EED searches not only general databases but also includes hand searches of specific journals, and working papers. Hence, NHS-EED helped to identify references not indexed in MEDLINE or EMBASE. In addition cross-reference searching of eligible papers was also carried out.

To identify future research output and be able to update the current review, the main search strategies will be repeated regularly. Electronic alerts of the table of contents from new issues of the main health economics and medical journals will be received on a regular basis[8]. The references from a recent international conference presentation of a comparison of methods of analysing and reporting multi-country studies were consulted to determine

[8] The list of journals is as follows: Health Economics, Journal of Health Economics, Medical Decision Making, Medical Care, Value in Health, Applied Health Economics and Health Policy, Journal of Health Services Research and Policy, European Journal of Public Health, Pharmacoeconomics, British Medical Journal, The Lancet, New England Journal of Medicine, JAMA, Applied Statistics, Statistical Methods in Medical Research and Biometrics.

suitable search terms for the search strategy[9]. This provided a good overview of the appropriate terminology to achieve successful results in the identification of articles. The use of terms that potentially could retrieve a large list of unsuccessful results, such as "country/ies" or "multicenter/centre" was avoided[10]. The search terms were programmed to scan the original title, the abstract and the subject headings in each electronic database. A description of the search strategies is detailed in Box 4.1.

Articles published from January 1996 to December 2007 were included in this review. Multinational trials have been conducted regularly since the mid-nineties and a similar pattern was expected in published economic evaluations. This assumption was supported by Sculpher and colleagues (Sculpher et al. 2004b) which reported 21 multinational economic evaluations alongside multinational trials during the period 1994 to 2000 with 81% of these studies published from 1996 onwards.

Box 4.1 : Search strategies used in OvidSP and NHS-HEED

1. multinational trial$1 or multinational clinical trial$1
2. multinational randomi#ed controlled trial$1 or multinational RCT
3. multinational stud$
4. multicountr$ or multi-countr$
5. substud$ or sub-stud$
6. cost-effectiveness analysis or cost$ or cost-analysis or economic evaluation or cost-utility analysis or cost-benefit analysis

$ and $1 (* in NHS-HEED) refer to the limited and unlimited truncation symbol. # (not allowed in NHS-HEED) refers to a wild card character.

[9] Rivero-Arias O, Wolstenholme J, Gray A. Analysing and reporting resource-use and outcome measures in multi-country studies: a comparison of methods. Presented as an oral presentation at the International Health Economics Association (iHEA) Conference in Barcelona, July 2005.

[10] Some authors use the term "multicenter/centre" to refer to multinational studies and omitting this term in the search strategy could have left potential evidence unidentified. However, it was expected that cross-references would help to identify these articles and minimise any potential bias.

4.2.3 Data extraction

The title and abstract of all papers identified were reviewed and if they fulfilled the inclusion criteria, a complete printed document was electronically downloaded or ordered. For ambiguous results, a full document was also retrieved and examined. Details of all articles were imported into the software Reference Manager (Institute for Scientific Information 2005).

A keyword proforma was developed for the extraction of the data. The main information extracted included:

- The overall sample size of the trial
- The number of participating countries and their corresponding contribution to the sample size
- The categories of resource use collected
- The sources for unit costs, the country of origin and adjustments for missing unit cost information
- The sources and methods to calculate costs
- The currency adjustment adopted
- The sources and methods used in the effectiveness analysis
- The reporting of the incremental cost-effectiveness (ICER) results
- The approaches evaluated in the methodological studies

The information from the proforma form was included in a spreadsheet and coded accordingly for statistical analysis. For details on the keyword template proforma and the list of keywords available after the literature search see Appendix 2. All analyses were carried out in Excel 2007 and STATA 10 (StataCorp 2007).

The World Bank income classification that divides countries according to 2007 Gross National Income (GNI) per capita was used in the results section: low income, $935 or less; middle income, $936 - $11,455; and high income, $11,456 or more. In order to identify trends across the study period of the review, estimates were also calculated and presented from the periods 1996 to 2000 and from 2001 to 2007.

The terminology proposed by Reed and colleagues to summarise the analytical approach used in multinational economic evaluations alongside trials is adopted here (Reed et al. 2005b). A summary of this terminology was first presented in Chapter 1. Studies are classified first depending on whether patients in a single country (or a subset of countries) or patients in all countries contribute to the sample size for the effectiveness and resource use data used in the analysis. A ***fully pooled*** analysis is based on effectiveness and resource use data from patients in all countries. A ***partially split*** analysis is based on estimates of effectiveness data from all countries but resource use are based on a single country (or a subset of countries). In a ***fully split*** analysis, estimates of effectiveness and resource use are based on the same group of patients in a single country (or a subset of countries). Studies are then further classified depending on the costing method performed. If unit cost estimates from each particular country are applied to the resources use in those countries it is described as a ***multicountry costing***. In a ***one-country costing***, a unit cost estimate from one country is applied to resource use in all countries, a subgroup of countries or a single country.

4.2.4 Presentation of results of the review

Descriptive statistics and standard statistical inference were adopted to report the results of this chapter. A summary and detailed results from both methodological and applied evidence is presented.

4.3 Results

4.3.1 Summary descriptive information of the overall review

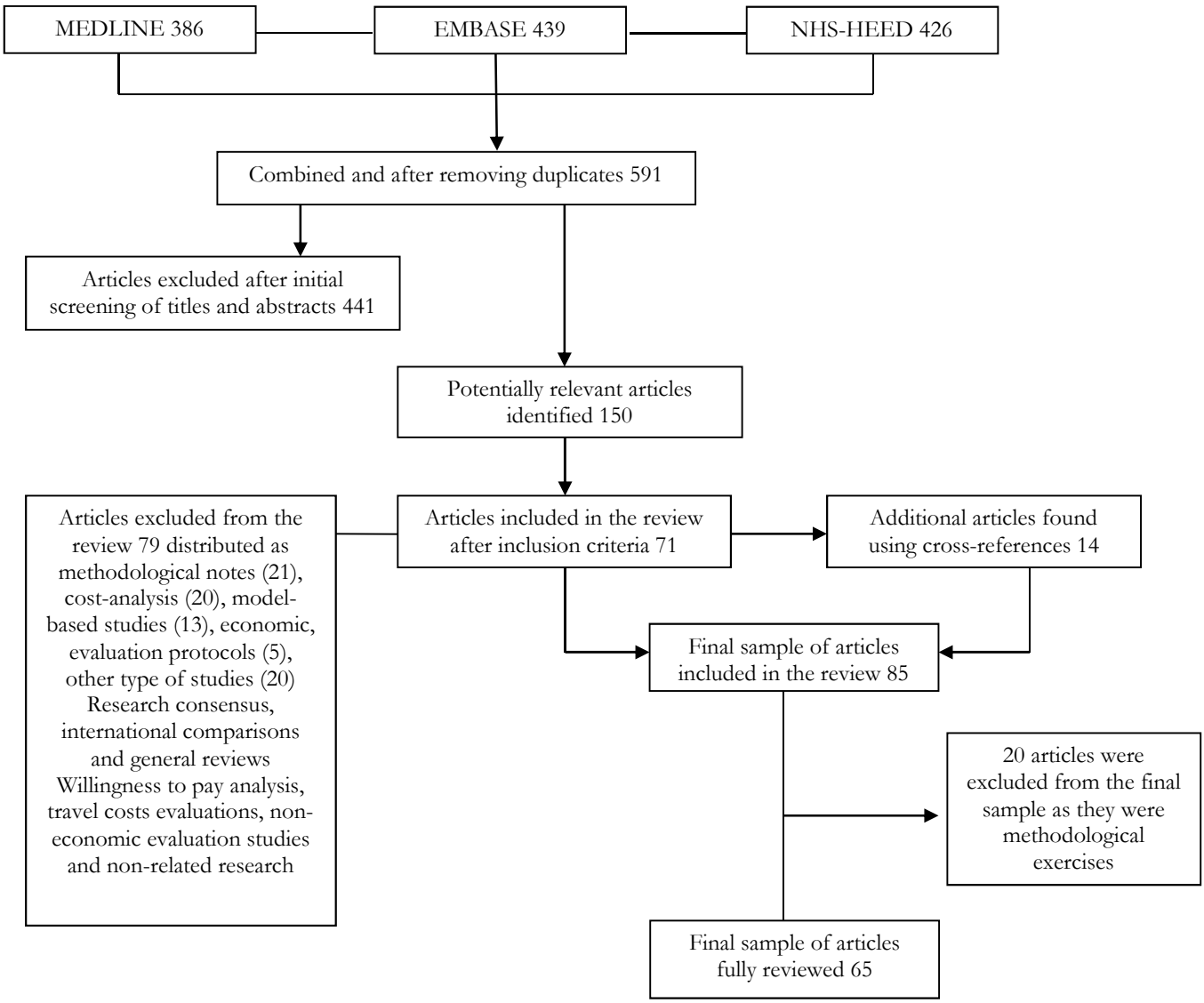
The total number of identified articles for the search strategies used in each of the electronic databases is reported in Table 4.1. A schematic diagram of the overall identification and review process is presented in Figure 3.1.

Table 4.1: Number of articles indentified during the main search strategy in MEDLINE, EMBASE and NHS-HEED

	Main Search Strategy†	MEDLINE	EMBASE	NHS-HEED
1	multinational trial$1 or multinational clinical trial$1 or multinational randomi#ed controlled trial$1 or multinational RCT	215	232	84
2	multicountr$ or multi-countr$ or multinational	2086	2095	330
3	substud$ or sub-stud$	1481	1487	31
4	cost-effectiveness analysis or cost$ or cost-analysis or economic evaluation or cost-utility analysis or cost-benefit analysis	175817	207368	25971
1 and 4		43	54	78
2 and 4		243	284	319
3 and 4		100	101	29
Total‡		**386**	**439**	**426**

† Search carried out from January 1996 to December 2007; ‡ these figures include possible duplicates which were identified and removed at a later stage using tools from OvidSP and Reference Manager.

Figure 4.1: Schematic diagram of overall identification and review process



A total of 591 articles were identified from the original search strategy. From those, 441 articles were excluded as they clearly did not fulfil the inclusion criteria. The majority of these articles were economic evaluation models, cost-analysis using data from randomised and non-randomised designs and non-related research. Full printed copies were obtained from 150 potentially relevant articles. After careful evaluation, 79 articles were excluded from the review, leaving 71 papers fulfilling the inclusion criteria. Detailed reasons for these exclusions are presented in Figure 4.1. Cross-reference checks from the identified studies yielded another 14 relevant articles. The final sample of this review therefore comprises 85

papers with 65 related to applied evidence and 20 to methodological research. A list of the references of this final sample is presented in Appendix 3.

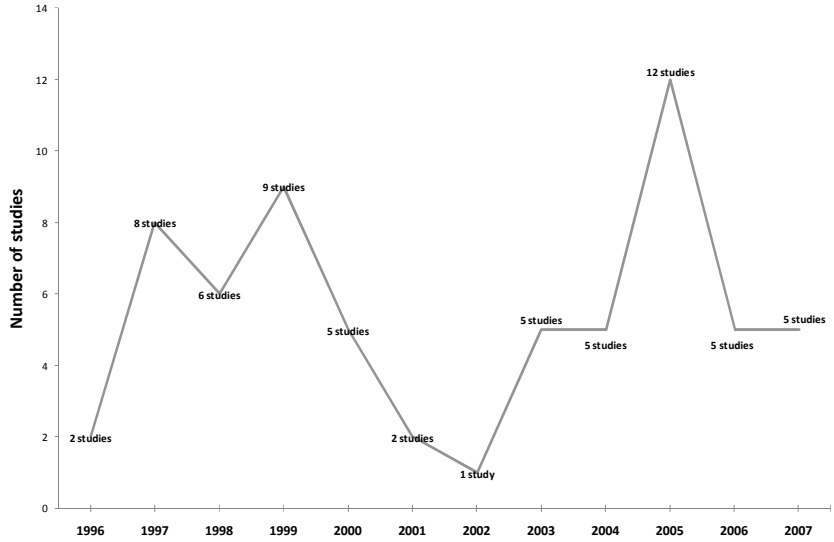
4.3.2 Summary of results from the methodological evidence review

Out of the 20 relevant articles, the majority (18) of the methodological studies identified dealt with issues related to unit costs and statistical methods to analyse and report multinational cost-effectiveness studies. Economic issues in multinational and multicentre cost-effectiveness analysis were investigated in two papers. The next section presents detailed results of the methodological evidence review. In addition, chapter 6 provides an in-depth explanation of the main statistical methods in practice.

4.3.3 Summary of results from the applied evidence review

The frequency of studies published annually since 1996 is presented in Figure 4.2. The chart shows no clear trend in the frequency of multinational economic evaluation publications over the years.

Figure 4.2: Frequency of multinational economic evaluations published from 1996 to 2007



4.3.3.1 The multinational nature of the studies

Detailed information on the number of participating countries and their contribution to the overall sample size was reported in 20/65 (31%) of studies. 29/65 (45%) of publications provided limited information about their international dimension, e.g. articles which mentioned the number of participating countries but provided no information on which countries were included or the contributing sample size from each country. Finally, 16/65 (25%) of studies did not report any information on the multinational nature of the study.

A detailed examination of countries contributing data and the number of studies to which data were contributed is presented in Table 4.2. The overall sample size from all the randomised controlled trials identified was estimated to be 172,401 patients. Country-specific sample size was reported for 74,852 (43%) of patients, therefore for 97,549 (57%) of participants the country contribution was unknown. Seventy-two percent (53,852) of the patients in studies reporting country-specific sample size came from high income countries, and 15% (11,070) middle income countries. The majority of middle income countries participating in these studies did so during 2004 and 2005. Centres from the United Kingdom participated in 23 of the studies, making the UK the country with the most frequent involvement in research of this type. However, the United States recruited the largest number of patients across all countries identified in the review.

Table 4.2: Description of countries contributing data and number of studies to which data were contributed using the World Bank classification category

Country	Number of patients reported	Percentage contribution from overall sample size	Number of studies
High income countries			
US	19926	11.56%	18
Canada	5042	2.92%	14
Sweden	5029	2.92%	16
UK	3266	1.89%	23
Italy	2639	1.53%	11
Norway	2518	1.46%	11
Netherlands	2284	1.32%	11
Denmark	2278	1.32%	14
Finland	2069	1.20%	11
Germany	1989	1.15%	17
France	1281	0.74%	21
Greece	966	0.56%	3
Spain	925	0.54%	18
Australia	744	0.43%	9
Hungary	584	0.34%	4
Belgium	466	0.27%	11
Czech Republic	379	0.22%	3
Iceland	314	0.18%	3
Austria	218	0.13%	6
Luxembourg	185	0.11%	1
Slovak Republic	184	0.11%	1
Switzerland	174	0.10%	5
New Zealand	144	0.08%	2
Israel	85	0.05%	3
Portugal	72	0.04%	3
Singapore	46	0.03%	1
Ireland	38	0.02%	2
United Arab Emirates	7	0.00%	1
Total high income	**53852**	**31.24%**	
Middle income countries			
Russia	3146	1.82%	3
South Africa	2980	1.73%	7
Argentina	2291	1.33%	5
India	718	0.42%	1
Colombia	474	0.27%	1
Brazil	401	0.23%	4
Poland	348	0.20%	2
Cuba	254	0.15%	1
Albania	108	0.06%	1
Egypt, Arab Rep.	108	0.06%	1
Mexico	67	0.04%	3
Turkey	46	0.03%	1
Venezuela, RB	33	0.02%	1
Malaysia	32	0.02%	1
Jordan	28	0.02%	1
Thailand	20	0.01%	1
Sri Lanka	8	0.00%	1
Chile	4	0.00%	2
Peru	2	0.00%	1
Uruguay	2	0.00%	1
Total middle income	**11070**	**6.42%**	
Low income countries			
Zimbabwe	611	0.35%	1

contd.

Country	Number of patients reported	Percentage contribution from overall sample size	Number of studies
Uganda	584	0.34%	1
Nigeria	466	0.27%	1
Pakistan	297	0.17%	1
Bangladesh	196	0.11%	1
Malawi	109	0.06%	1
Yemen Rep.	109	0.06%	1
Ghana	106	0.06%	1
Sierra Leone	34	0.02%	1
Total low income	**2512**	**1.46%**	
Other(s)*	7418	4.30%	
Total country sample size reported	74852	43.42%	
Unknown country contribution	97549	56.58%	
Total sample size from all RCTs	172401		

* Some articles reported sample size by regions but did not clarify the participating countries (e.g. Europe, South America, Western Europe, North America...)

Table 4.3 provides estimates on the multinational nature of the studies included in the review. A total of 31/65 (48%) studies were conducted entirely in high income countries and contributed 63,693 (37%) patients to the overall reported sample size across all studies. The mean (range) number of countries per study was estimated to be 7 (2-33) across all studies and 5.5 (2-15) across studies in high income countries, a non significant difference (p=0.1794). The mean (range) sample size per study was estimated to be 2655 (44-14,703) across all studies and 2055 (44-10,305) across studies in high income countries, a non significant difference (p=0.4039).

Table 4.3 also shows the mean (range) proportion of patients from the largest contributing country by study period. Non significant differences were found between groups and across time period (1996-2000: p =0.900, 2001-2007: p=0.5726). The largest contributing country recruited on average around 38% of all patients participating in the study.

Table 4.3: Descriptive statistics of the multinational nature of the studies included in this review

	n	All studies	n	Studies conducted entirely in high income countries
Total number of studies		65		31
Total number of participating countries		57		28
Total sample size across studies		172401		63693
Mean (range) number of countries per study	48	7 (2-33)	30	5.5 (2-15)
Mean (range)sample size per study	65	2655 (44 - 14703)	31	2055 (44 - 10305)
Mean (range) sample size per country	199	376 (1 - 3964)	57	448 (1 - 3522)
Mean (range) proportion of patients from the largest contributing country*				
Overall				
1996-2000	8	0.37 (0.20 – 0.60)	6	0.36 (0.20 – 0.60)
2001-2007	12	0.37 (0.11 – 0.75)	5	0.43 (0.11 – 0.75)

*These parameter could only be calculated using studies that reported detailed information on the countries participating and contributing sample size.

Table 4.4 shows the unit cost collection approach used in the studies. 39/63 (62%) studies collected unit costs data from one country and had an associated mean (range) of 5.6 (2 – 16) countries per study. Studies collecting unit cost information from two to four countries had a mean (range) of 5.4 (2 -12) countries per study. Thirteen studies (21%) collected unit cost data from more than four countries and 11.9 countries participated in these studies on average. An analysis of variance confirmed a significant difference in the mean number of countries per study among groups ($F=6.08$, $p<0.01$). No differences in the collection of unit costs were found across time.

Table 4.4: Unit costs information used in the costing analysis in the studies included in the review

Collection of unit cost information	Number (%) of studies*	Mean (range) number of countries in the study
Overall		
Unit cost from one country only	39 (62%)	5.6 (2 – 16)†
Unit cost from 2-4 countries	11 (17%)	5.4 (2 – 12)†
Unit cost from 5+ countries	13 (21%)	11.9 (5 – 33)†
1996-2000		
Unit cost from one country only	19 (66%)	5.1 (2 – 15)
Unit cost from 2-4 countries	5 (17%)	3.6 (2 – 6)
Unit cost from 5+ countries	5 (17%)	7.4 (5 – 15)
2001-2007		
Unit cost from one country only	20 (59%)	5.9 (2 – 16)††
Unit cost from 2-4 countries	6 (18%)	6.8 (4 – 12)††
Unit cost from 5+ countries	8 (24%)	14.8 (6 – 33)††

* Two countries did not provide detailed data in the collection of unit costs across countries. † $F=6.08$, $p<0.01$. †† $F=5.15$, $p=0.01$

Table 4.5 reports the type of unit costs data collected within each country. The majority of studies collected national average unit costs or used mixed sources to perform the cost analysis. It seems there has been a move from using mixed sources towards national average unit costs after 2000 but no statistical difference between periods was detected to validate this result ($\chi 2=8.59$, $p=0.072$). In multi-country costing approaches, unit costs collected in more than one country may be expressed in different currencies. Therefore, currencies are commonly converted into a common base before conducting the costing analysis. Out of the 12 studies using a multi-country costing method, seven applied Gross Domestic Product Purchasing Power Parity (GDP PPPs)[11], four applied exchange rates and one study did not report the conversion used.

Table 4.5: Unit cost collection approach used in the studies included in the review

	Single centre	Multiple centre unit costs	National average unit costs	Mixed Sources	Not clear	Number of studies
Overall	**3 (5%)**	**3 (5%)**	**35 (54%)**	**21 (32%)**	**3 (5%)**	**65**
1996-2000	2 (7%)	2 (7%)	11 (37%)	12 (40%)	3 (10%)	**30**
2001-2007	1 (3%)	1 (3%)	24 (69%)	9 (26%)	0 (0%)	**35**

[11] A brief explanation of exchange rates and purchasing power parity is given in the next section.

4.3.3.2 The collection of effectiveness data

A total of 58/65 (89%) studies used trial-wide health outcome data to conduct their effectiveness analysis, indicating the dominance of fully pooled and partially split methods in this type of study (Table 4.6). Although no differences were found in the analysis and reporting of effectiveness data across periods, the number of studies using health-related quality of life data seems to have increased over the years. Quality-adjusted life years (QALYs) were used as the main outcome measure in nine studies with the EQ-5D UK tariff, the EQ-5D thermometer and mapping algorithms being the main approaches used to calculate utility values.

Table 4.6: Health outcome data collected and type of effectiveness analysis conducted in the studies

Summary of effectiveness analysis	Overall	1996-2000	2001-2007
Handling of the effectiveness data			
Trial-wide effectiveness data used	58 (89%)	26 (87%)	32 (91%)
Sub-sample of cases used	7 (11%)	4 (13%)	3 (9%)
Health outcomes used in the analysis			
Disease-specific	28 (43%)	16 (53%)	12 (34%)
Life-years	23 (35%)	11 (37(%)	12 (34%)
HRQoL			
- EQ-5D	9 (14%)	1 (3%)	8 (23%)
- SF-36	2 (3%)	1 (3%)	1 (3%)
- Other	3 (5%)	1 (3%)	2 (6%)
Utility calculation in studies presenting quality-adjusted life year (QALYs)			
- EQ-5D UK tariff	4 (44%)	-	4 (57%)
- EQ-5D thermometer	2 (22%)	1 (50%)	1 (14%)
- Utility mapping	3 (33%)	1 (50%)	2 (29%)

HRQoL: Health-related quality of life; EQ-5D: EuroQol 5 dimension generic instrument; SF-36: Short-form 36 items generic instrument; QALYs: Quality-adjusted life years

4.3.3.3 The reporting of cost-effectiveness analysis

Fully pooled, one-country costing with 66% (43/65 manuscripts) and fully pooled, multi-country costing with 15% (10/65 publications) were the preferred methodological approaches adopted to analyse and report the results from these studies (Table 4.7). Fully and partially split were less common, being used in studies where the proportion of patients

from the largest contributing country was greater than 45% on average. Although no significance differences between periods were detected ($\chi 2=7.31$, $p=0.121$), a description of the methodological approaches by year of publication is presented in Figure 4.3.

Table 4.7: Ways in which cost-effectiveness data were analysed in the studies using Reed et al 2005 (Reed et al. 2005a) classification

Classification	Overall Freq (%)	1996-2000 Freq (%)	2001-2007 Freq (%)	n, mean (range) proportion of patients from the largest contributing country†
Fully pooled, one-country costing	43 (66%)	19 (63%)	24 (69%)	9 studies ; 0.32 (0.11 – 0.44)
Fully pooled, multicountry costing	10 (15%)	2 (7%)	8 (23%)	6 studies; 0.36 (0.20 – 0.54)
Fully split, one-country costing	6 (9%)	4 (13%)	2 (6%)	3 studies; 0.46 (0.24 – 0.75)
Partially split, one-country costing	5 (8%)	4 (13%)	1 (3%)	2 studies; 0.49 (0.40 – 0.60)
Partially split, multicountry costing*	1 (2%)	1 (3%)	0	-
Total	**65**	**30**	**35**	

* This classification is possible if health outcomes have been analysed using data from all countries but costs have been analyses using a subset of countries; †This parameter could only be calculated using studies that reported detailed information on the countries participating and contributing sample size.

Figure 4.3: Description of methodological approaches used in the studies by year of publication

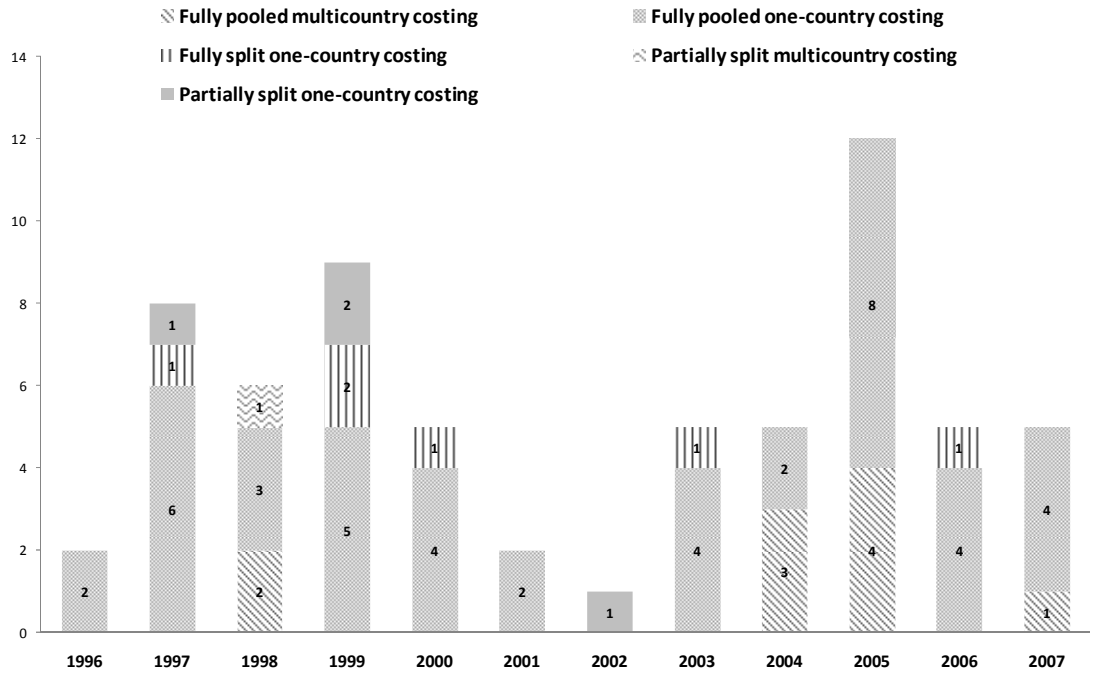


Table 4.8 shows that country-specific incremental cost-effectiveness ratios (ICERs) were reported in 10/65 (15%) studies and 25/65 (38%) studies reported one estimate of ICER for all countries. ICERs were not reported in 24/65 (37%) articles, but this was particularly a characteristic in studies published before the year 2000 (16/30 (53%) studies) with the reporting improving significantly after then with only 8/35 (23%) studies not combining costs and effects.

Table 4.8: Presentation of country-specific incremental cost-effectiveness ratio (ICER) results in the applied studies

Classification	Overall	1996-2000*	2001-2007*
One estimate of ICER was reported for all countries	25 (38%)	8 (27%)	17 (49%)
Not reported	24 (37%)	16 (53%)	8 (23%)
Country-specific ICERs were reported for each country†	10 (15%)	2 (7%)	8 (23%)
One estimate of ICER for a particular country	6 (9%)	4 (13%)	2 (6%)
Total	**65**	**30**	**35**

† The study by Simon et al. (Simon et al. 2006) divides group of patients by Gross National Income (GNI) and does not report ICERs by country but GNI regions.
*Statistical significant difference between periods $\chi 2 = 9.85$, $p = 0.020$

A similar pattern was observed in the studies when handling uncertainty for the ICERs. Table 4.9 suggests a significant improvement when reporting measures of uncertainty for ICERs after the year 2000.

Table 4.9: Handling uncertainty for the incremental cost-effectiveness ratio (ICER) estimates in the applied studies

Classification	Overall	1996-2000*	2001-2007*
No uncertainty presented	37 (57%)	23 (77%)	14 (40%)
Deterministic sensitivity analysis	2 (3%)	1 (3%)	1 (3%)
Uncertainty handled through statistical methods	26 (40%)	6 (20%)	20 (57%)
Total	**65**	**30**	**35**

* Statistical significant difference between periods $\chi 2 = 9.40$, $p = 0.009$

Detailed results of the cost-effectiveness analyses performed in multi-country studies and studies reporting country-specific ICERs are presented in Appendices 4 and 5 respectively.

4.3.3.4 Funding of studies

A total of 53/65 (82%) of the studies identified in this review were funded by the pharmaceutical industry; 5/65 (7%) were funded by non-industry organizations, and the remainder 7/65 (11%) did not specify the funding received. The studies funded by the pharmaceutical industry comprise 135,521 (79%) of the overall sample size from all the randomised controlled trials and were conducted exclusively in high income countries in 24/39 (62%) studies.

4.3.4 Detailed results of the methodological evidence review

4.3.4.1 Design, collection and analysis of unit costs in multinational clinical trials

Problems related to the collection and use of unit costs have been commonly confronted by analysts undertaking empirical multinational trials. As a result, this has also been the focus of a number of methodological studies. In particular, methodological studies have considered how to handle missing country unit costs, how many countries and centres to collect data from, and the use of national average or centre-specific unit costs.

4.3.4.1.1 Imputing missing country unit costs

Missing unit cost data for particular categories of resource use are likely to cause difficulties when performing any cost analysis. In multinational trials, the likelihood of a missing response is higher since the number of unit cost needed to calculate country-specific costs in the trial increases systematically. Schulman and colleagues (Schulman et al. 1998) proposed a framework to assist imputing missing unit costs for health care services as part of a multinational clinical trial on subarachnoid haemorrhage patients. Procedure and inpatient costs were the categories of resource use included in this study. Two methods depending on whether unit cost information is available for at least one country participating in the clinical

trial (market-basket approach) or whether unit cost information is not available in any country (physician-work and practice-expense resource-based relative value units) were introduced. Health care service utilisation and their associated unit costs were collected using case record forms by each trial coordinating centre across countries. Fifteen countries participated in the trial but unit cost data were collected in the seven countries with the largest enrolment and users of the technology under evaluation.

In the market-based method, an index table was created and reflected the relative cost of a basket of resources for pairs of countries (i.e. relative costs between country A and B or relative costs between country C and D). The basket was constructed using six health care services with complete unit cost data for the seven countries. A total cost for each health care service was calculated by multiplying the unit cost of each service by the mean resource use utilisation observed in the trial. The total costs of the six services were summed to create a market basket total cost. All costs were expressed in US dollars using power purchasing parities. The relative cost of one country compared to another was then calculated using a ratio between the market-basket costs for each combination of countries. In total 49 pairwise comparisons were performed and a complete set of pairwise comparisons for the index table was built.

Missing unit costs for a particular health care service and country were imputed by multiplying the unit cost information from available countries by the appropriate value from the index table indicating the relative cost of the country with no information compared to the country with information. The product of these calculations were summed and divided by the number of available countries for that particular health care service to impute the missing unit cost.

The second method introduced a set of relative value units originally developed in the US to measure resource intensity in medical services (Hsiao et al. 1988). Physician-work relative value units were available to indicate the physician time and skills involved in the procedure and practice-expenses relative value units to indicate the physician practice expenses for the procedure. Regression methods were then applied to translate the set of relative value units into country-specific unit costs when unit cost data were unavailable for all countries.

Costing hospital admissions by specific diagnosis has been performed in several economic analyses alongside clinical trials. Ideally a unit cost by diagnosis for every hospital admission incurred in the trial should be collected to perform a reliable cost analysis; however this is not normally feasible because additional time and funding are required. Hence, analysts must decide the number of hospital diagnosis unit costs that will be collected during the trial. In a multinational clinical design, this number needs to be carefully selected because the effort required to collect this information is more demanding.

Glick and colleagues (Glick et al. 2003) proposed a framework to impute the unit costs for hospital diagnosis in multinational studies where cost data were not collected for a particular diagnosis either because it was not planned to or because it was not possible to collect. The authors explored the use of US diagnostic-related group (DRG) relative values (1999) as predictors of missing hospital diagnosis unit costs in four European countries. The relative values included relative weights, arithmetic mean length of stay (LOS) and geometric mean LOS. Regression models were estimated using different combinations of relative values as covariates and dummies for the four countries participating in a trial as explanatory variables. Arithmetic LOS predicted hospital diagnosis unit cost more accurately than the other two relative values. The predictive model using arithmetic LOS and dummies for the countries was then modified with different thresholds of missing hospital diagnosis information and

number of countries to explore how many hospital diagnosis categories and countries are needed to obtain reliable unit costs. An imputation error statistic was constructed to determine the accuracy from each imputation model. It was shown that collecting fewer unit costs in more countries, rather than more unit costs in fewer countries, minimised the imputation error.

4.3.4.1.2 Designing unit cost forms

An overview of causes and factors influencing hospital cost variation across centres at a national and international level was given by Goeree and colleagues (Goeree et al. 1999). In this study, a conceptual framework for selecting hospitals for collecting unit cost information in multinational economic evaluations was presented. The authors suggested that researchers involved in the design of this type of study always face three questions: 1) what number of hospitals should be selected to collect the unit cost estimates; 2) what sampling method should be used to select the hospitals; 3) and what desired level of geographical representativeness is required to inform decision-makers. Answers to these questions were explored using data from 72 hospitals across six countries from a multinational clinical trial of prelabour rupture of the membranes at term. The majority of patients were recruited from Canada, UK and Australia and due to budget constraints, detailed unit cost data was collected from 12 hospitals in those three countries.

The total cost per patient for each of the four management strategies in the trial was calculated using three sets of unit costs. The first set used unit cost data from one hospital in one country selected on convenience and accessibility, the second set used a sample of hospital unit costs in each country based on accessibility and the third set used a sample of hospital unit costs in each country based on the size and the teaching status of hospitals.

When a sample of hospital unit costs was selected, the midpoint between the highest and lowest unit cost estimate was used in the cost analysis. Following the classification of methodologies in multinational economic evaluation described above, the authors performed several one-country costings, i.e. resource use were pooled across countries and multiplied by a set of unit costs from one country. The authors explored within country variation performing a similar analysis using data from Canada alone.

Both analyses showed that, depending on the number and type of hospital used in the set of unit costs, median cost estimates differed between 30 to 55% across the four management strategies evaluated. The authors concluded that researchers should collect unit costs sets from as many hospitals within and across countries as possible to fully capture price differences, which can be incorporated later when conducting the cost analysis.

4.3.4.1.3 Multinational costing methods

A comparison of hospitalisation costing methods in multinational trials was carried out by Reed and colleagues (Reed et al. 2003). In this study three costing approaches were investigated in a multinational trial exploring different treatments for congestive heart failure. A unit cost survey was conducted by local analysts in each country to collect cost data for hospital diagnosis categorised using US DRG codes (1999). The first costing method assigned a DRG to every patient's admission and attached one of the collected DRG unit costs. Length of stay (LOS) for patient's admission was also routinely collected during the trial. The second and third method calculated a cost per patient by multiplying a daily cost for a diagnosis in a particular country by the LOS observed for that patient's admission. The daily cost was computed by dividing the DRG unit cost in a particular country by either a mean LOS or a median LOS for that DRG in a particular country. The daily cost methods

were further adjusted to control for the intensity of use during a hospitalisation across countries. The results showed that adjustment methods that considered variability in LOS across patients, although providing better variance estimations, need larger sample sizes in trials to maintain statistical power.

4.3.4.1.4 Average versus centre-specific unit costs

The implications of using average or centre-specific unit costs in a general cost-effectiveness analysis framework have not been widely explored. It is expected though that these two approaches are likely to produce different results if unit costs vary greatly across centres within a country. In a multinational setting, these differences may be even more marked, as centres are distributed across countries and unit costs may vary much more across countries than within them. Economic theory suggests that there is an underlying production function for health care in each centre that relates volumes of resource use and the costs of these resources to achieve a certain level of output. If this function truly exists and centres are cost-minimising, then each centre will respond to price differences by substituting more expensive inputs for less expensive ones to maintain this level of output. Therefore, if individual centres are responsive to relative price differences and behave according to economic theory, using average unit costs may fail to incorporate this substitution effect and will as a consequence overestimate the total costs for a particular treatment.

This assumption was explored by Raikou and colleagues in a simulation exercise (Raikou et al. 2000). The authors simulated a production function with two inputs and two prices for a number of centres and created two scenarios: 1) centres were responsive to price changes and respond to incentives as predicted by economic theory, and 2) centres were unresponsive to price changes. The responsiveness of the centres in scenario 1, i.e. how

centres responded to input unit cost changes, was measured through the elasticity of substitution. The influence of uniform, normal and logistic distributions for input unit costs was explored in each simulation. The elasticity of substitution was also varied within a range from 0.1 to 10. Finally, once a unit cost change occurred in one of the inputs, centres were allowed to achieve the new allocation of input volumes without uncertainty (deterministic) or with it (stochastic). The overall costs were calculated using average or centre-specific unit costs in each simulation. An empirical distribution of the difference between methods was created and a p-value associated to each difference calculated.

The simulation results showed that in scenario 2 there were no differences between the costing methods, as no obvious economic relationship between input unit costs and input volumes exists. Scenario 1 with deterministic relationship between input unit costs and volumes showed statistical differences between methods for relatively low values of elasticity of substitution and distribution for prices. However, with a stochastic relationship between input unit costs and volumes the results were similar to scenario 2 as the uncertainty around the new allocation of input volumes increases. This final result suggests that when the allocation of input volumes after price changes is subject to some degree of uncertainty, it replicates a situation where no obvious economic relationship between input unit costs and input volumes exist.

4.3.4.1.5 Currency conversion methods and the use of exchange rates

In a multinational cost-effectiveness analysis, unit costs collected from different countries may be expressed in different currencies. Before attempting any cost analysis, unit cost data need to be converted to a common currency using, for example, exchange rates. However, countries produce volumes of goods and services at price levels that vary greatly across

countries. Hence, to perform meaningful comparisons across locations, unit cost data not only needs to be expressed in the same currency but also at a uniform price level. This transformation would ensure that if unit costs are different in a country compared to another, this difference is due only to differences in the volumes of goods and services produced in the countries.

Unfortunately, converting into a common currency using exchange rates reflects both differences in the volumes produced and price level across countries[12] (Eurostat-OECD 2006). Purchasing power parities (PPPs) are an alternative that corrects the shortcomings of using exchange rates. PPPs are the result of the ratio of the prices in national currencies of the same good or service in different countries. For example, if the price of an ice-cream in UK is 2.96 pounds and in the United States is 1.80 dollars, the PPP for ice-cream between UK and the United States is 2.96 pounds to 1.80 dollars or 1.64 pounds to the dollar. In other words, for every dollar spent on ice-cream in the United States we need to spend 1.64 dollars in the UK to achieve the same quantity and quality of ice-cream. The comparison of volume of ice-cream in both countries can be performed by 1) converting the total expenditure of ice-cream in the UK into dollars (i.e. dividing it by 1.64) or 2) converting the total expenditure of ice-cream in the United States into pounds (i.e. multiplying it by 1.64). PPPs are calculated not only for individual good and services but also for product groups and different level of aggregation (Eurostat-OECD 2006). In the area of health for instance,

[12] Purchasing power parity theory suggests that national price levels once converted to a common currency using exchange rates should be equal (Rogoff 1996). Market forces will eliminate any arbitrage opportunity ensuring that the price of a particular good will be the same across countries where the good is traded at least in the medium and long term. This process is known as the "law of one price". This theory requires a set of assumptions: all goods are internationally tradable; tariffs, VAT, transport costs and related costs do not create price differences across countries; and demand and supply for currency is achieved completely by international trade in goods. However, current international trading makes these assumptions difficult or almost impossible to hold. Supply and demand for currency are influenced by currency speculation, interest rates, governmental intervention and capital investment among countries in addition to just the international trade in goods. Besides, many good and services, for example in health care and education are not traded internationally. As a consequence exchange rates do not reproduce the relative purchasing power of currencies in their national markets.

medical care PPPs have been developed to convert health care expenditure in different countries (Amerini 2000).

The impact of using these conversion methods in a multinational context was explored by Wordsworth and Ludbrook (Wordsworth and Ludbrook 2005). Exchange rates, Gross Domestic Product (GDP) PPPs, medical care PPPs and a technology specific PPPs were compared using a cost-effectiveness dataset in the area of dialysis. Technology specific PPPs were constructed using a common dialysis resource-related basket for each country. Data to compose the basket were collected using a questionnaire and site visits in the participating centres across countries. The results illustrated that the mean cost per patient across conversion methods and centres varied. However, no measures of variability were presented to understand the magnitude of these variations. The authors argued that because a more homogeneous basket of items is used to create a technology specific PPPs, it has the potential of better removing price differences across countries for that technology. Nevertheless, more empirical research needs to be conducted to understand the implications of using each of the conversion methods before recommending a standard in multinational trials.

4.3.4.2 Statistical methods in cost-effectiveness analysis alongside multinational clinical trials

Three types of analytical strategies have been suggested to investigate cost-effectiveness analysis alongside multinational clinical trials. The first method uses standard tests for heterogeneity to assess whether a pooled estimate is appropriate to calculate cost-effectiveness estimates in multinational trials. The second approach uses regression models to calculate cost-effectiveness estimates using trial-wide data, controlling for the multinational nature of the dataset by including the variability among countries in the

estimation process. The third suggestion adjusts the baseline cost-effectiveness results conducted in a particular location to inform cost-effectiveness in another location.

4.3.4.2.1 Test for heterogeneity and statistical test of treatment effect

Cook and colleagues (Cook et al. 2003) assessed the feasibility of combining data from different countries using a pooled estimate to present cost-effectiveness results. The 4S Simvastatin trial provided the data for this paper (Scandinavian Simvastatin Survival Study Group 1994) with patients from Denmark (n=713), Finland (n=868), Iceland (n=157), Norway (n=157) and Sweden (n=1681). A one-country costing approach was carried out in this methodological paper. The authors evaluated mortality rates, hospitalisation rates, incremental cost-effectiveness ratios and incremental net benefits using a quantitative and qualitative statistical test of treatment effects among countries. It was shown that in the absence of heterogeneity of treatment effect across countries, a pooled estimate may be appropriate to present the results from a multinational study. However, as highlighted by the authors, the lack of power to detect differences for countries with small sample sizes is a serious limitation of this method.

4.3.4.2.2 The use of Ordinary Least Squares (OLS)

The use of ordinary least squares (OLS) regression methods to control the multinational nature of the data was first suggested by Willke and colleagues (Willke et al. 1998a). The authors used data from a multinational randomised controlled trial of patients with subarachnoid haemorrhage (n=833) (Glick et al. 1998) and performed a multi-country costing approach. Data on resource use and outcomes from five anonymous countries were included in this analysis but the sample size in each country was not reported. Country-specific incremental cost-effectiveness ratios (ICERs) based on the incremental costs and

effects coefficients from a structural regression model were estimated. The model estimated the cost of an episode of care for an individual patient on a set of covariates, a treatment group indicator variable and the health outcome of the episode. The treatment effect on the health outcome of the episode was modelled using a separate equation with the same set of covariates and the treatment group indicator. The presence of endogeneity bias as a result of including the health outcome in the cost equation was fully explored before estimating the equations. Dummies for each country and interaction terms were further included in both the cost and outcome equations to estimate country-by-outcome and country-by-treatment interactions terms. The results showed the larger variability of country-specific ICERs when resource use, unit costs and treatment effect were allowed to vary across countries using the interaction terms.

The use of OLS to examine differences in practice patterns across countries in a multinational randomised controlled trial was explored by Koopmanschap and colleagues (Koopmanschap et al. 2001). This paper suggested adjusting country-specific resource use and outcome data to undertake a cost-effectiveness analysis alongside multinational trials. The adjustment replicates a situation that would result if all patients had been treated in only one country from the trial. To demonstrate this, the authors used a multinational clinical trial of breast cancer involving participants from the Netherlands (n=41), Belgium (n=48) and France (n=60).

Quality of life and resource use information of hospital, non-hospital and medication as well as country-specific unit costs were collected alongside this trial. Six categories of hospital resource use were included as dependent variables in six OLS regression models against a set of baseline characteristics, dummies for countries and interaction terms. These equations

identified for each of the hospital categories, medical practice differences across the countries detected by the coefficient from the country dummies in the models.
The results indicated significant mean differences across countries on the number of consultations, number of day care and number of laboratory procedures. The mean resource use of the categories, where no differences across countries were observed, was multiplied directly by country-specific unit costs. The three resource use categories that showed differences in medical practice across countries were averaged first and then corrected using the appropriate dummy coefficient to estimate the level of resource use that would have been seen if the trial had been run in that particular country. The corrected mean resource use was then multiplied by the country-specific unit cost. Finally, the cost per patient was calculated summing the cost of all resource use categories for each patient.

A similar approach was carried out for the quality of life data but no differences across countries were shown in the regression models and therefore a pooled approach was adopted. Although this approach has the potential of correcting and producing country-specific cost-effectiveness estimates, because average figures are used to calculate the cost per patient, no measures of uncertainty can be estimated and this is a serious drawback of this approach.

4.3.4.2.3 The use of multilevel models (MLM)

Grieve and colleagues (Grieve et al. 2005) compared OLS and MLM approaches to explore cross-country variation in length of stay and total cost using data from a stroke study among 13 centres in 11 European centres (Grieve et al. 2001). A very complete multi-country costing exercise was conducted in each of the participating centres. The effects of including centre-level covariates in the regressions to explain differences in costs and effects across

centres were also explored. A random intercept model was selected using centres as a random effect. Multilevel models were performed assuming a normal distribution for residuals or a gamma distribution for resource use and costs. The authors concluded that OLS resulted in incorrect inferences as it does not recognize the hierarchical structure of the dataset in the estimation process. Because the model does not recognize that individuals are nested within centres, it ignores the variability among centres when estimating and hence underestimates the standard error of coefficients. Including country dummies to estimate the magnitude of cost and resource use variation is possible but OLS is unable to quantify the variation across centres and also unable to provide an explanation for these variations.

This study also showed that the MLM with a gamma distribution of the dependent variable fitted more accurately the data than the MLM with the normality assumption on the residuals. The authors published an expanded discussion on this last result in a separate paper (Thompson et al. 2006) that showed again the superiority of using gamma instead of normality assumptions when using MLM to report cost-effectiveness results alongside multinational clinical trials.

Chapter 1 explained the exchangeability assumption and the use of shrinkage estimates in MLMs. It was introduced that the exchangeability assumption is a judgment from analysts based on their experience in the area. When this assumption was valid, country-specific cost-effectiveness estimates could be presented using shrinkage estimates.

Grieve and colleagues (Grieve et al. 2007) and Manca and colleagues (Manca et al. 2007) explored the exchangeability assumption and investigated the use of a bivariate hierarchical model to estimate incremental net benefit including centre-level covariates to explain systematic differences across countries. Using this approach, the average cost-effectiveness

estimates are exchangeable but only after controlling using the centre-level covariates. The results from these studies showed that MLM needs to be used with caution when systematic differences across countries exist.

The use of shrinkage estimates to report country-specific costs using multilevel modelling in multinational clinical trials was explored in a study by Willan and colleagues (Willan et al. 2005). Using data from a randomised trial in patients with acute myocardial infarction (Assessment of the Safety and Efficacy of Tenecteplase Group 2001) and a one-country costing approach, the authors explored the use of shrinkage estimation to report country-specific results for treatment differences in costs. Shrinkage estimation can take various forms and an intensive discussion on the use of shrinkage estimation to report country-specific costs and effects in multinational clinical trials was provided by Pinto and colleagues (Pinto et al. 2005). This study compared different approximations to calculate shrinkage estimates in multilevel models using the same dataset as Willan and colleagues (Assessment of the Safety and Efficacy of Tenecteplase Group 2001) and simulation.

Manca and colleagues (Manca et al. 2005b) modelled patient-level monetary net benefits and compared the results from an OLS and MLM models in a multicentre study carried out in 25 centres in England. Although this study was not based on a multinational trial, the methods can be applied to the multinational case as the authors indicated in the article. A random intercept and slope models were estimated assuming centre and treatment effects as random effects in the MLM model. This study used a simplified dataset from a randomised controlled trial comparing laparoscopic assisted hysterectomy with standard hysterectomy (Sculpher et al. 2004a). A multi-centre costing using a mix of centre-specific and national average unit costs was undertaken. It was concluded that MLM can provide more appropriate average cost-effectiveness estimates and associated standard errors. It also

highlighted the potential of MLM to estimate location-specific cost-effectiveness estimates using shrinkage estimates.

An in depth analysis of multilevel models using appropriate specification to deal with skewed cost data, interactions between baseline covariates and treatment and the use of binary outcomes as effectiveness measures was reported in a recent paper (Willan and Kowgier 2008)[13]. In addition the authors illustrated the use of expected value of information theory to inform whether the evidence from the clinical trial is sufficient for making a decision in a particular country.

4.3.4.2.4 Using trial data from a particular country to inform cost-effectiveness estimates in another country

An example of using trial data from a study performed in one country to estimate country-specific cost-effectiveness estimates in another country was illustrated in a paper by Menzin and colleagues (Menzin et al. 1996). The authors used data from a US clinical trial in the area of cystic fibrosis to inform decision makers in France, Germany, Italy and the UK. Main resource use categories were identified and country-specific unit cost weights were obtained. A comparison of unit costs provided information to identify resources that varied across countries. The number of hospitalisations and length of stay were evaluated using hospital records in each country and were classified as the main parameters that reflect practice patterns. The main effectiveness outcome, reduction of respiratory tract infections (RTI), was assumed to be generalisable to other jurisdictions. The effect of the therapy under evaluation on the costs of RTI-related care in the US clinical trial was estimated by multiplying the difference in the mean number of RTI patients between treatment groups by an estimate of the average cost of treating an RTI. The same analyses were conducted for

[13] This review covers evidence from 1996 to 2007. However, this paper was published online in November 2007 and printed in the seventh issue of Health Economics in 2008, and was therefore selected for the review.

Germany and the UK as no differences in practice patterns were achieved in these two countries when compared to the US. In the case of France and Italy, the average cost of treating an RTI was calculated using a weighted average of local inpatient and outpatient treatment cost, where the weight corresponded to the likelihood of being treated in each of these settings. The weights were collected in these two countries using observational data and expert opinion.

4.3.4.2.5 Algorithm to select appropriate methods to analyse and translate results from multinational clinical trials

A recent paper (Manca and Willan 2006) presented a pathway or algorithm to assist researchers in selecting the appropriate analytical strategy to analyse and report economic evaluations alongside multinational clinical trials. The algorithm introduced different scenarios depending on whether the country of interest participated in the trial and whether individual patient-level data (IPD) are available. Hence, the authors identified four main scenarios: 1) IPD are available but the country of interest did not participate in the trial; 2) IPD are available and the country of interest participated in the trial; 3) IPD are unavailable and the country of interest did not participate in the trial; and 4) IPD are unavailable but the country of interest participated in the trial. For scenarios 1, 3 and 4, it was suggested that modelling to extrapolate results in combination with methods from evidence-based synthesis to estimate important parameters could be used. In scenario 2, regression methods and in particular the use of multilevel models was proposed as the alternative to deal with the hierarchical nature of multinational datasets. A brief description of potential factors that affect generalisability of the results of cost-effectiveness analysis was also presented.

4.3.4.2.6 Evidence of within and between country variations of health care costs in multination clinical trials

Only one paper identified in this review explored within-country health care costs variation across centres using data from a multinational antenatal randomised controlled trial (Hutton et al. 2004). The WHO antenatal care study investigated the effectiveness of a new antenatal programme in reducing the number of antenatal visits in low-risk pregnancies to four visits (Villar et al. 2001). Centres from Argentina, Saudi Arabia, Cuba and Thailand participated in the trial. Unit cost data were collected in all countries but an exhaustive unit cost collection was conducted in Cuba and Thailand. Using resource use data from the trial and unit costs from these two countries, Hutton and colleagues explored how maternity health care costs varied across centres within each country.

This paper argued that it is not possible to understand overall cost differences across arms of a clinical trial (aggregate level) without an understanding of how costs vary across the centres participating in the trial (disaggregated level). Using descriptive statistics techniques, variations in the average number of outpatient visits, variations in the average number of inpatient visits, variations in the unit costs for outpatient visits and variations in the total cost per pregnancy across centres in Cuba and Thailand were investigated. Statistical inference was not conducted and only point estimates of averages were reported to determine differences across centres within and across countries.

The results showed a similar average number of outpatient visits across centres within-countries in each arm of the trial. However, the average number of inpatient visits varied across centres within each arm of the trial in the case of Cuba. This variation was explained to be related to random operational differences and patient case mix. Unit costs differences across centres in both countries were reported and related to the facility efficiency and the

level of throughput in each centre. The differences in unit costs across centres were the main determinant explaining the variability in the total cost per pregnancy found across sites in both countries. The authors identified important implications of these issues for the randomisation process when conducting cost-effectiveness analysis alongside multinational trials. Randomisation of patients using clustering methods by centres and including a representative number of centres in each country would help obtaining more reliable results and a better understanding of the variation across centres within each country.

The article by Hutton and colleagues reported and described differences in resource use and unit cost estimates across centres in Cuba and Thailand. However, between countries differences were not explained in that study.

No other methodological studies investigating between country differences in health care costs in multinational clinical trials were identified in this review.

4.3.5 Detailed results of the applied evidence review

For a complete description of the information extracted from each article included in the applied review see Appendix 2.

4.3.5.1 Studies with no acknowledgement of multinational nature

The multinational nature of the study was not acknowledged anywhere in the manuscript in 12 (19%) articles (Armstrong and Ruckley 1997; Aspelin et al. 2005; Bennett-Guerrero et al. 1997; Dasbach et al. 1999; Davey et al. 1999; Griffin et al. 2001; Herman et al. 2003; Hillner et al. 2000; Hull et al. 1997; Nord et al. 1997; Peeters et al. 2005; Weaver et al. 2000).

4.3.5.2 Studies using adjustments to control for the multinational nature

Regression analysis was used to control for the multinational nature in 5 (8%) of the studies indentified in the applied evidence review. Briggs and colleagues (Briggs et al. 2006b) used a standard regression model for costs and effects and included a dummy variable to allow adjustments for the UK using trial-wide data in a clinical trial of asthmatic patients. A generalised linear model with a log link function and dummies for each participating country in the trial was adopted in two studies (Fernandez et al. 2005; Radeva, I et al. 2005). Drummond and colleagues (Drummond et al. 2003) tested country-by-treatment interactions using analysis of variance to determine the homogeneity of resource use across countries before using a pooled estimate. In this paper, no country-by-treatment interaction was observed.

A multilevel model with a random effect for country was adopted by Willan and colleagues (Willan et al. 2006) to estimate differences in quality-adjusted survival time and total costs between treatment arms in a multinational clinical trial of patients with dementia.
Tests for heterogeneity to detect country-by-treatment interactions in the larger components of a cost analysis of a multinational clinical trial in patients with advanced colorectal cancer were carried out in one study (Groener et al. 1999). The authors reported no heterogeneity in these larger components and hence concluded that a fully pooled, one-country costing could be performed for all countries in the trial if local unit costs were available.

Adjustments in the cost analysis of a multinational clinical trial to reflect country-practice patterns were identified in one of the studies (Levy et al. 1998). The aim of the paper was to conduct an economic evaluation for a French setting using a multinational clinical trial of heart failure therapy. The cost analysis performed included only two categories: the cost of the treatment and hospital costs. Levy and colleagues acknowledged that length of stay

duration collected in the trial was significantly different across countries and that using a trial-wide data on this variable would be misleading. For that reason, the authors decided not to use length of stay duration in the hospitalisation costs calculations. Instead hospital costs were calculated by multiplying the number of admissions obtained from the trial data by a French unit cost per admission.

4.3.5.3 Studies using sensitivity analysis to validate the results from the analysis

Sensitivity analysis was used in 7 (11%) of the applied studies to explore the impact on the results of using data from different countries. The type of sensitivity analysis varied across studies but similarities were found depending on the data available. Special attention was given to the analysis of unit costs from different countries. Studies with reliable unit cost data from different countries compared the results of using different approaches to the costing analysis, for example fully pooled, multi-country costing in comparison with fully pooled, one-country costing (Henderson and Brown 1999; Reed et al. 2005a). In these studies, no differences were found between the comparisons performed but it is likely that this is not always going to be the case. If the unit cost variability across countries participating in a clinical trial is large, then it is expected that the results of multiplying resource use by a multi-country or a one-country costing will be different.

In another study, Gafni and colleagues (Gafni et al. 1997) varied the unit costs used in each country using low and high values available and compared the results with the baseline case. Cost differences across countries were observed, but using different sets of unit costs in the incremental analysis did not change the baseline results in each country.

Sensitivity analysis was also performed to explore how fully pooled or fully split analyses affect cost-effectiveness results (Brown and Armstrong 2003; Detournay et al. 2000; Henderson and Brown 1999; Lofdahl et al. 2005; Simon et al. 2006). In the case of a fully split analysis, estimates of effectiveness and resource use were based on a single country (Detournay et al. 2000; Lofdahl et al. 2005) or on a specific characteristic. For the latter, Brown and colleagues (Brown and Armstrong 2003) and Henderson and Brown (Henderson and Brown 1999) divided their sample in groups of countries with similar catheterization rates in a clinical trial of unstable angina as small number of patients were recruited in some countries. In another study Simon and colleagues grouped countries by gross national income to perform a cost-effectiveness analysis of women with pre-eclampsia (Simon et al. 2006). The authors stated that eclampsia was a rare event which yielded relatively small sample size in some countries and prevented the calculation of country-specific cost-effectiveness ratios.

4.3.5.4 Studies using imputation methods to estimate missing country unit cost

A market-basket approach (Nathwani et al. 2004; Radeva, I et al. 2005; Reed et al. 2004a; Reed et al. 2004b; Reed et al. 2005a) and averaging across available country unit costs (Fernandez et al. 2005; Glick et al. 1998) were the two methods used to deal with missing unit costs for a particular country.

4.3.5.5 Studies reporting evidence of resource use and effectiveness differences across countries

The majority of papers 50 (81%) from the applied review recognized that cost and effect differences across countries exist and are a potential challenge when performing the cost-effectiveness analysis of these studies. Some studies found significant differences in

outcomes and resource use across settings. An interesting early discussion on this issue was provided by Jonsson and colleagues (Jonsson et al. 1999), which suggested that if strong evidence is available on similarities in resource use consumption for a particular intervention across countries, a fully pooled, one country costing is justified. It was also emphasised that not only resource use consumption differences across locations should capture the attention of analysts: for instance, if the rate of coronary artery bypass grafts (CABGs) in coronary heart disease is different across countries this should be considered when interpreting the results of the study.

Mark and colleagues (Mark et al. 2000) carried out an economic analysis of US participants alone from a large randomised controlled trial of patients with acute coronary syndrome. Significant differences in baseline characteristics between US and non-US patients were reported. The main clinical results also showed treatment differences across regions with the larger benefit in the US, followed by modest differences in Canada and Western Europe and smaller treatment differences in Latin America and Eastern Europe. The authors could not state any reasons for these differences.

On the resource use side, Nathwani and colleagues (Nathwani et al. 2004) carried out an independent analysis between the European and the South American samples in a randomised controlled trial of Gram-positive bacterial infections. A preliminary univariate analysis of the data was carried out and found large differences in length of stay and hence overall costs across regions. In addition, differences in baseline characteristics were reported between European and South American patients. The authors used this argument to validate the choice of analysis.

4.3.5.6 Studies using flawed or confusing arguments to support the choice of analysis

Although some authors recognized the problem of cross-country differences, some provided limited evidence on the arguments to support the choice of analysis. For example Bachert and colleagues and Canonica and colleagues (Bachert et al. 2007; Canonica et al. 2007) pooled data from countries with similar health care systems. However, no statistics were presented to support the similarities of systems. In another study, Wimo and colleagues (Wimo et al. 2003) performed a fully pooled, one-country costing and suggested that the results of the study could be generalised to the other countries included in the original dataset as a similar cost per patient was found across countries. But this finding was made more likely by the fact that only Swedish unit costs were used in the study. A final interesting example was the work by Boersma and colleagues (Boersma et al. 2006), where it was advocated that multiplying trial-wide resource use by unit costs from one country can be considered a "country-adaptation". However, no details of the meaning of "country-adaptation" were provided by the authors to support this statement.

4.4 Discussion

This chapter identifies all relevant applied and methodological evidence in cost-effectiveness analysis alongside multinational clinical trials from 1996 to 2007 using a structured review. Sixty-five applied economic evaluations and 20 methodological papers were identified.

4.4.1 Applied review

Around 57% of the overall sample failed to describe the contribution of the countries participating in the trial, indicating poor reporting of the multinational dimension. In addition, around 20% of the articles did not recognise the multinational nature of the study in any section of the manuscript.

In most studies, one country contributed a large proportion of the overall sample size. This was clearly influential when deciding on the method to analyse cost-effectiveness data. Studies where the largest country contributed less than 36% of the total sample size preferred a fully pooled costs and effects approach to conduct their analyses. However, studies where one country contributed 45% or more of the total generally split the data instead of pooling across all countries.

Thirty-one of the 65 studies identified were conducted exclusively in high income countries and the remainder also recruited heavily from high income countries. A weak trend of countries from middle income countries participating in these studies was identified during the years 2004 and 2005. Nevertheless this trend disappears after that period. All the low income countries identified in this review participated in just one of the studies (Simon et al. 2006): the study by Simon and colleagues comparing the use of magnesium sulphate for the treatment of pre-eclampsia. This low participation rate from low income countries perhaps explains why disability-adjusted life years (DALYs) were not used as the primary outcome measure in any of the studies.

The results from this review suggest that future guidelines on transferability of economic evaluations across settings should emphasise the importance of reporting the number and identity of countries and their contribution to the overall sample size in cost-effectiveness analyses alongside multinational clinical trials. Both analysts and decision makers will benefit from this improvement. From the analyst side, reporting participating countries and contributing sample size will improve the overall quality of the study. Decision makers will also greatly benefit from this improvement as they will be better informed on the relevance of the study to their particular setting.

4.4.2 Methodological review

Methodological developments in cost-effectiveness analysis alongside multinational clinical trials have occurred quite rapidly over the last ten years. Health economists have mainly concentrated on issues related to unit cost data and statistical methods to deal with multinational datasets.

Missing unit cost information is a practical problem in each multinational economic evaluation and therefore has found a direct application in the field. The market-basket approach, for example, was used in five of the applied studies to complete missing unit cost information. Regression analysis has been advocated as the tool to control and adjust for the multinational nature in these studies. Although tests for homogeneity were also suggested, the lack of power to detect differences associated with this method has limited its applicability in the empirical evidence. Standard regression analysis has the advantage of computing the cost-effectiveness parameters controlling for country-specific characteristics using a set of baseline covariates. This method estimates country-specific parameters but overlooks the fact that patients are grouped within countries and hence ignores the variability among countries. Ignoring this variability may result in an underestimation of the standard error associated with the coefficients of the regression.

Multilevel modelling is a method that recognizes the hierarchical nature of multinational datasets and provides variance estimation for all levels in the hierarchy. A comparison of single-level (standard regression model) and multilevel methods was described in two methodological articles (Grieve et al. 2005; Manca et al. 2005b). Nonetheless, few analyses have been carried out using a multilevel approach and more applied evidence is needed before these models are fully recommended. Moreover, this type of model assumes the parameters estimated, such as incremental costs or effects, are exchangeable across countries,

i.e. these parameters are similar but not identical across countries. However if systematic differences across countries exist this assumption may not be valid.

There may be reasons to question the exchangeability assumption above. The study from Hutton and colleagues (Hutton et al. 2004), for example, suggests that the level of cost variability within arms in a multinational randomised controlled trial were not related to the intervention but due to other factors such as patient's case mix, the facility's efficiency and the level of throughput in each centre. These factors are likely to be country and health care specific and related to a specific production function in each country. The empirical results section also showed some examples of differences in clinical and baseline characteristics between patients across countries (Mark et al. 2000) and resource use differences across countries (Nathwani et al. 2004) in a multinational trial setting.

4.4.3 Limitations of the review

Individual patient-level data from multinational clinical trials are likely to inform parameters in model-based studies, which aim to inform cost-effectiveness in a particular country. This type of study was not included in the current review and currently little is known about the quality of these studies. Future reviews should extract relevant information that helps fill the current gap in the literature on this issue.

To find the appropriate search strategy, general terms that could potentially retrieve a large number of unsuccessful results were avoided. Search terms were selected using mainly keywords from applied and methodological studies from preliminary reviews. Therefore, terms like multicentre and international were not included in the search strategy. It is possible that some literature were not found in the selected strategy but cross-references helped to identify relevant literature and minimise any potential bias.

This review attempted to identify whether the reporting of the multinational nature of these studies had improved over the years. Although some trends were observed between time periods, few significant differences were detected, probably because of the small sample size of 65 studies in the review. However, the significant improvement in the reporting of ICERs and handling of uncertainty after 2000 suggests that guidelines on the conduct of economic evaluation of health care technologies have had some effect.

Only one assessor evaluated the evidence included in the final sample of this review and although the search was performed in a transparent and structured manner, the final sample included in the review cannot be considered exhaustive. However, the large number of references extracted compared to other reviews suggests that if any relevant evidence was left unidentified, it is unlikely to cause bias in the results.

4.4.4 Summary

Cost-effectiveness analyses alongside multinational clinical trials are available but the current reporting of the multinational nature of these studies is very poor. Robust and well-founded statistical methods are available to analyse these studies. Missing unit cost data is very common in this type of analysis; however the implications of using unit costs from the available countries to calculate costs and cost-effectiveness of the treatments under evaluation are not well understood.

The methods identified here will be expanded in Chapter 6 using a large randomised controlled trial of patients with subarachnoid haemorrhage. The next chapter provides detail information about this case study.

Chapter 5: The Medical Research Council (MRC) International Subarachnoid Aneurysm Trial (ISAT) case study

5.1 Introduction

This chapter introduces the case study used to explore the methods identified in Chapter 4, and provides relevant data for the simulation exercise conducted in Chapter 7. It also introduces the disease area, which is the focus of the case study.

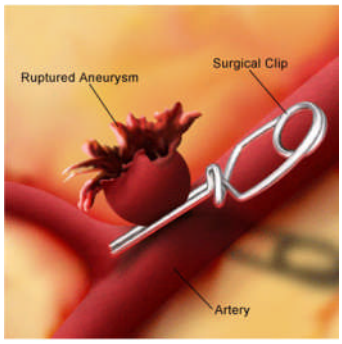
Aneurysmal subarachnoid haemorrhage (aSAH) is a serious life-threatening condition caused by a rupture of an aneurysm in a blood vessel over the surface of the brain damaging the brain tissue. An aneurysm is a balloon-like lump or weakening of an arterial wall. The reduction of blood supply can cause permanent brain damage leading to disruption or loss of brain functionality and probable death.

Until 1992, when the Guglielmi detachable platinum coil (GDC) was introduced to Europe, there was no alternative to craniotomy and neurosurgical clipping to secure the aneurysm and prevent re-bleeding. Craniotomy requires general anaesthesia and involves manual opening of the scalp where the aneurysm is located. The aneurysm is then sealed and shut using a surgical clip and then the scalp is stitched to its original place (Figure 5.1a). In an endovascular intervention, a catheter is inserted into an artery in the leg and is guided through the blood vessels network to the aneurysm. Platinum coils are used to block the flow of blood into the aneurysm (Figure 5.1b).

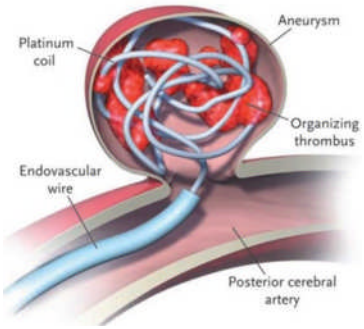
In 2002 the International Subarachnoid Aneurysm Trial (ISAT), a large Medical Research Council (MRC) funded prospective multinational randomised controlled trial, reported that endovascular coiling yields better clinical outcomes than neurosurgical clipping at one year,

and these results have significantly changed clinical practice in both UK and overseas centres (Molyneux et al. 2005).

Figure 5.1: Neurosurgical clipping (a) and endovascular coiling (b) diagrams



(a)



(b)

aSAH is a significant cause of death and severe disability in relatively young patients (mean age 52), with an annual incidence of 6 to 12 per 100,000 in Western countries (Linn et al. 1996). The natural history of the disease suggests that 30% of patients will die within 24 hours of the bleed and a further 25% to 30% of patients will re-bleed and die over the next four weeks if no treatment is given (Wermer et al. 2005). 50% of survivors will be dead by six months without treatment (Alvord et al. 1972). A total of 74,807 quality-adjusted life years (QALYs) were estimated to be lost due to aSAH in 2005, and the estimated annual cost to the UK NHS £168.2 million in 2005 (Rivero-Arias et al. 2010). The consequences for patients and families of an event of this type can be devastating, particularly if patients are left with severe disabilities. It has been estimated that around £300 million in productivity losses were lost due to this condition in 2005 (Rivero-Arias et al. 2010).

5.2 An overview of the ISAT clinical investigation study

5.2.1 Aim of the study

ISAT was developed to reduce the uncertainty on whether endovascular coiling was effective and cost-effective compared with neurosurgical clipping to treat patients with aSAH. Full details of the clinical characteristics of ISAT are reported elsewhere (Molyneux et al. 2002; Molyneux et al. 2005) and therefore only a summary is included here. The primary objective of ISAT was to investigate whether endovascular compared to neurosurgical treatment reduced the proportion of patients dependent or dead at one year by 25%. The trial aimed to recruit 2,500 patients to achieve a 90% power at 0.01% significance level of detecting this proportion at one year. Secondary objectives in the study also included a quality of life assessment of patients and a long-term cost-effectiveness analysis.

ISAT has a multinational design and eleven countries collected resource use and health outcomes data from 46 centres. A summary of the centres included in the study and the distribution of patients across countries and randomisation group are presented in Table 5.1.

Table 5.1: Centres and distribution of patients across countries participating in ISAT

Countries	Endovascular (n=1,073)	Neurosurgery (n = 1,070)	Total
UK	**809**	**835**	**1,644**
- Oxford	122	127	249
- Edinburgh	32	31	63
- Nottingham	134	138	272
- Cardiff	27	30	57
- Atkinson Morley	17	20	37
- Newcastle	105	112	217
- Bristol	28	31	59
- Hurstwood Park	34	35	69
- Manchester RI	47	46	93
- Salford	71	58	129
- Kings (London)	61	70	131
- Birmingham	31	29	60
- Plymouth	13	12	25
- Glasgow	36	34	70
- Queen Sq.	1	2	3
- Liverpool	26	28	54
- Cambridge	1	3	4
- Sheffield	9	10	19
- Hull	3	4	7
- Belfast	9	13	22
- Royal London	1	2	3
- Oldchurch	1	0	1
Sweden	**64**	**59**	**123**
- Gothenburg	53	47	100
- Uppsala	11	12	23
Germany	**62**	**56**	**118**
- Freiburg	28	24	52
- Mainz	12	12	24
- Dresden	8	7	15
- Wurzburg	5	6	11
- Augsburg	6	6	12
- Hamburg	3	1	4
Canada	**48**	**41**	**89**
- Montreal	18	9	27
- Toronto	30	32	62
Finland	**34**	**29**	**63**
- Turku	34	29	63
France	**25**	**18**	**43**
- Paris	18	13	31
- Amiens	1	1	2
- Lille	1	0	1
- Montpellier	1	1	2
- Rennes	3	2	5
- Saint-Anne	1	1	2
Australia	**14**	**13**	**27**
- Perth	14	13	27
Denmark	**13**	**12**	**25**
- Odense	13	12	25
Switzerland	**4**	**4**	**8**
- Lausanne	4	4	8
USA	**0**	**2**	**2**
- Baltimore	0	2	2
Czech Republic	**0**	**1**	**1**
- Prague	0	1	1

Baseline data collected at randomisation for the 2,143 included in the study are presented in Table 5.2. There were no significant imbalances between the two groups. The median age in both groups was 52 with most patients suffering aSAH being females. The WFNS grade is a clinical scale that grades consciousness in patients with aSAH where 1 indicates good level of consciousness and 5 indicates severely disabled. Some patients could not be assessed and a level 6 is used then. In ISAT the majority of patients in both groups were in good level of consciousness.

Table 5.2: Baseline characteristics whole sample

	Endovascular (n=1,073)	**Neurosurgery (n = 1,070)**
Male sex	400 (37%)	398 (37%)
Age years*	52 (18-87)	52 (19-84)
WFNS grade		
1	674 (63%)	661 (62%)
2	269 (25%)	280 (26%)
3	66 (6%)	68 (6%)
4	38 (4%)	36 (3%)
5	11 (1%)	9 (1%)
6 (not assessable)	15 (1%)	16 (1%)
Number of aneurysm detected		
1	836 (78%)	850 (79%)
2	173 (16%)	170 (16%)
3	44 (4%)	35 (3%)
≥4	20 (2%)	15 (1%)
* Median (range)		

Table 5.3 shows selected baseline characteristics for patients across the five countries who contributed the majority of patients in ISAT, i.e. UK, Sweden, Germany, Canada and Finland. No significant differences between groups were detected using each individual's country data. This result seems to suggest that baseline characteristics among patients are similar across countries in ISAT. However, caution needs to be exercised when making subgroup comparisons from each country. This is extensively explained in the next Chapter.

Table 5.3: Selected baseline characteristics by country

	Endovascular	Neurosurgery
UK (n=1,644)		
Male sex	292 (36%)	305 (37%)
Age years*	52 (18-87)	52 (20-84)
WFNS grade		
1	546 (68%)	546 (65%)
2	185 (23%)	212 (25%)
3	50 (6%)	55 (7%)
4	18 (2%)	16 (2%)
5	5 (1%)	0 (0%)
6 (not assessable)	5 (1%)	6 (1%)
Sweden (n=123)		
Male sex	21 (33%)	26 (44%)
Age years*	53 (21-72)	55 (29-75)
WFNS grade		
1	36 (56%)	38 (64%)
2	19 (30%)	14 (24%)
3	1 (2%)	2 (3%)
4	6 (9%)	4 (7%)
5	1 (2%)	0 (0%)
6 (not assessable)	1 (2%)	1 (2%)
Germany (n=118)		
Male sex	17 (35%)	11 (27%)
Age years*	53 (30-77)	52 (19-76)
WFNS grade		
1	23 (37%)	18 (32%)
2	16 (26%)	18 (32%)
3	5 (8%)	4 (7%)
4	5 (8%)	2 (4%)
5	5 (8%)	6 (11%)
6 (not assessable)	8 (13%)	8 (14%)
Canada (n=89)		
Male sex	17 (35%)	11 (27%)
Age years*	55 (21-84)	53 (19-76)
WFNS grade		
1	24 (50%)	22 (54%)
2	15 (31%)	8 (20%)
3	5 (10%)	4 (10%)
4	4 (8%)	5 (12%)
5	0 (0%)	2 (5%)
6 (not assessable)	0 (0%)	0 (5%)
Finland (n=63)		
Male sex	22 (65%)	18 (62%)
Age years*	50 (29-66)	49 (30-67)
WFNS grade		
1	15 (44%)	14 (48%)
2	16 (47%)	13 (44%)
3	3 (9%)	2 (7%)
4	0 (0%)	0 (0%)
5	0 (0%)	0 (0%)
6 (not assessable)	0 (0%)	0 (0%)

* Median (range)

5.2.2 Primary and secondary outcomes

The primary outcome used in the ISAT study was the modified Rankin Scale (mRS). The mRS is one of the most commonly used health outcome tools for clinical evaluation of patients in the area of stroke (Wolfe et al. 1991). The mRS evaluates a patient's level of handicap by looking at activity limitations and lifestyle changes. The score is derived from an interview with a patient performed either by a nurse or a clinician. The scale spans 6 grades from 0 to 5, with 0 representing no symptoms at all, and 5 representing severe disability, with 6 being used for death.

The EQ-5D instrument was included as a secondary outcome to measure quality of life. Detailed resource use data attributable to endovascular and neurosurgery interventions were collected using an extensive case record form. In addition, questionnaires on employment status and hospital readmissions were also included. Section 5.3 describes in detail the EQ-5D instrument and the resource use data available.

Outcome measures and resource use data collection were carried out at two months and one year after randomisation and annually subsequently. It was not possible to collect baseline outcome data because of the severity of the patient after the collapse.

5.2.3 Statistical analysis

Tests for interactions using relative risks, absolute risk reduction and relative risk reduction with associated 95% parametric confidence intervals (CIs) were used to report the results of the study up to one year. The Kaplan-Meier method was used to analyse time to death for each group, and differences in the survival distribution between groups were explored using the log-rank test. The main effectiveness analysis pooled data from all countries.

5.2.4 Clinical results of the study

1073 and 1070 patients were randomised to endovascular coiling or neurosurgical clipping respectively (Table 5.1). Missing data in the ISAT study was not a serious problem, with the amount of missing information to conduct the primary effectiveness analysis less than 1%.

Table 5.4 shows the number of patients for each mRS level at two months and at one year. The proportion of patients dependent or dead is represented by participants in levels 3 to 5 of the mRS. 278 (26.1%) of 1065 patients allocated to endovascular treatment were known to be dead or dependent at two month compared with 392 (36.9%) of 1063 allocated to neurosurgery, a relative risk = 0.71 with 95% CIs 0.62-0.80. 250 (23.5%) of 1063 patients allocated to endovascular treatment were known to be dead or dependent at one year compared with 326 (30.9%) of 1055 patients allocated to neurosurgery, an absolute risk reduction of 7.4% (95% CIs 3.6-11.2) and a relative risk reduction of 23.9% (95% CIs 12.4-33.9).

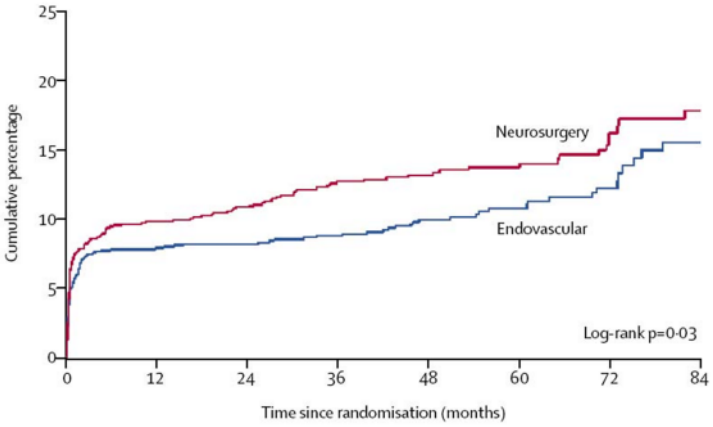
Table 5.4: Clinical outcome at 2 months and 1 year (Molyneux et al. 2005)

	2 month outcome		**1 year outcome**	
	Endovascular (n=1065)	**Neurosurgery (n =1063)**	**Endovascular (n=1063)**	**Neurosurgery (n=1055)**
Modified Rankin Scale				
0 No symptoms	203 (19.1%)	144 (13.6%)	260 (24.5%)	187 (17.7%)
1 Minor symptoms	310 (29.1%)	273 (25.7%)	301 (28.3%)	292 (27.7%)
2 Some restriction in lifestyle	274 (25.7%)	254 (23.8%)	252 (23.7%)	250 (23.7%)
(0-2 inclusive)	787 (73.9%)	671 (63.1%)	813 (76.5%)	729 (69.1%)
3 Significant restriction in lifestyle	107 (10.1%)	189 (17.8%)	107 (10.1%)	141 (13.4%)
4 Partly dependent	34 (3.2%)	46 (4.3%)	30 (2.8%)	42 (4.0%)
5 Fully dependent	62 (5.8%)	73 (6.9%)	28 (2.6%)	38 (3.6%)
6 Dead	75 (7.0%)	84 (7.9%)	85 (8.0%)	105 (9.9%)
(3-6 inclusive)	278 (26.1%)	392 (36.9%)	250 (23.5%)	326 (30.9%)

Figure 5.2 presents the cumulative mortality for each treatment up to seven years with the annual number at risk. It shows that the number of deaths in the neurosurgical intervention was always higher than in the endovascular group (log-rank test = 0.03). Given the nature of

the results, the ISAT Steering Committee stopped recruitment before the original target size was completed.

Figure 5.2: Kaplan-Meier cumulative mortality to 7 years (Molyneux et al. 2005)



Annual number at risk (deaths):								
Endovascular	1073 (85)	974 (3)	887 (5)	717 (8)	541 (4)	373 (5)	215 (6)	103
Neurosurgery	1070 (105)	944 (10)	842 (16)	663 (3)	503 (3)	340 (7)	192 (3)	98

5.2.5 Conclusion of the ISAT study

The results of the study suggested that endovascular treatment yielded better health outcomes at one year follow-up when compared to neurosurgical intervention in patients with aSAH and a suitable aneurysm anatomy for both treatments. The results from the survival analysis showed that this effect is likely to remain in the long-run.

5.2.6 Implications of the ISAT study

The ISAT study generated great controversy and discussion in the worldwide neurosurgical community when the first publication was released (Britz et al. 2003; Brown et al. 2003; Leung et al. 2003; Mohr 2003; Molyneux and Kerr 2003; Sellar and Whittle 2003). However the results from a second publication from the ISAT Collaborative Group confirmed the conclusions reached in the original publication (Molyneux et al. 2005). Moreover, a recent

report has shown that the management of aSAH patients has substantially changed in the UK and other countries since ISAT was published (Clinical Effectiveness Unit 2006). There are however, some important questions such as the risk of re-ruperture and re-bleeding in endovascular patients that need to be addressed in future research to support the endovascular technique. It is expected that the ongoing long-term data collection will answer some of these questions. Another subject that remains unanswered is the cost-effectiveness of endovascular versus neurosurgical intervention. ISAT collected both resource use and health outcome to perform the economic evaluation of these techniques. However, the multinational nature of the study introduces challenges in estimating cost-effectiveness, compared to an economic evaluation undertaken in one hospital for instance, or across several centres in one country.

5.3 Data available to conduct the economic analysis of the ISAT study

5.3.1 Resource use data collected

Before the trial began, all categories of hospital resource use likely to be important were identified by the health economics team from previous research in the area. Trial case record forms (CRFs) were then designed to record the volume of resources used by each patient in each category, including type and number of hospitalisations and procedures, procedure duration in minutes, number of coils used in endovascular procedures, and lengths of stay in intensive care and in general wards. Hospital readmissions for further treatment or following complications or adverse events were documented, as was any follow up angiography or imaging. Information was also recorded on patient discharge to rehabilitation. Employment related information was collected and is reported separately from health service costs. Some information was recorded on the domestic circumstances of patients, but not on informal care received.

5.3.2 Unit cost data collection

A detailed unit cost data collection was conducted in the UK to perform an exhaustive cost analysis in the UK (Wolstenholme et al. 2008). Additional country-specific unit costs for the main cost drivers were also collected from the main participating countries to perform the cost-effectiveness analysis.

To identify the main cost drivers for each procedure, around 250 participants from the Oxford centre were reviewed in a detailed micro costing exercise. Staff from the operating theatre and the John Radcliffe Hospital Finance Department collected detailed information on the resource use and unit costs associated with both interventions. Staffing types and numbers, unit cost of coils, and cost per in-patient day were identified as the main cost drivers and were collected in the country-specific unit cost questionnaires.

5.3.3 Unit cost data collected in the United Kingdom only

Unit costs for staff time in theatre were obtained from national salary tables, and equipment costs were provided by the manufacturers. The cost per inpatient day in a district general hospital (DGH) and the cost per day in a rehabilitation clinic were taken from standard publications (Curtis 2008; Department of Health 2006).

The unit costs of equipment and consumables used by both procedures were provided by manufacturers and by the Oxford Radcliffe Hospitals NHS Trust finance department. Information on the purchase price and the useful life of each piece of equipment (in years) used by both procedures was also provided to calculate the equivalent annual cost (Drummond et al. 1997). The equivalent annual cost was then discounted over the useful life of the equipment at the standard annual discount rate of 5%.

For the endovascular procedure, consumables used were assumed to be the same for all patients, but patient specific data were available on the use of thrombolytic agents and the number of coils. For the neurosurgical procedure, consumables and the cost of anaesthetics were assumed to be the same for all patients and were included as a lump sum, but were varied according to whether temporary clips were used or not. The average cost for anaesthetics was calculated separately for endovascular and neurosurgical patients; based on time in theatre, and included with the cost of consumables.

The unit cost of an angiogram was provided by the hospital finance department; the cost of a hospital day case was included for each angiogram. The unit costs of a computed tomography (CT) and magnetic resonance imaging (MRI) were taken from national sources.

5.3.4 Country-specific unit cost data collection

A questionnaire was designed asking all ISAT centres to provide local information on the normal numbers and types of staff present for both procedures, mean cost of coils, and cost per inpatient day in an intensive therapy unit (ITU), neurology ward and neurosurgery ward for that centre. Overseas centres were also asked to provide information on the average annual salary including employer contribution for staff present during the operations. The unit cost survey template used to collect these data is presented in Appendix 6.

Eight of the twenty-two (36%) centres completed and returned the questionnaire in the UK; two centres subsequently merged, resulting in seven independent responses, these centres are representative of all the UK based ISAT centres in terms of geographical dispersion and hospital size. An average of the cost per coil and the cost per day in an ITU provided by these centres was calculated to estimate the average unit costs for these two resource categories in the UK.

Overseas questionnaires were circulated to centres in Sweden, Canada, Finland and Germany, the four countries which recruited at least 20 patients per arm. Three (one centre for each country) of the twenty overseas centres (15%) completed and returned the unit cost questionnaire. None of the centres from Canada completed the questionnaire. Although frequent reminders were sent to the centres during a time-span of six months, the lack of funding and interest from overseas centres to participate in further data collection prevented the collection of additional unit cost data. As a result, Canadian patients were valued using UK unit cost data in the following chapters.

All unit costs were first inflated to 2008 prices using the health inflation index from the Department of Health in the UK (Department of Health 2008) and from Eurostat for Sweden, Germany and Finland (EUROSTAT 2009). Then unit costs were expressed in 2008 US dollars using GDP purchasing power parities (PPPs) provided by the OECD website to eliminate any price difference across countries and ensure comparability (Eurostat-OECD 2006).

5.3.5 Unavailable unit costs (market-basket method)

Not all the items requested in the survey sent to overseas centres were returned completed. Hence, missing information was imputed using the market-basket approach identified and described in the structured review in Chapter 4 (Schulman et al. 1998). In addition, the questionnaire did not include questions on consumables, imaging or investigations as no extra funding was available for such an exercise. However, the pilot study did not identify these categories as main cost drivers and it was assumed after consultation with the lead investigators from ISAT that this would be the case also in the remainder of countries. These unit cost estimates were collected in the UK and imputed using the market-basket method for Sweden, Finland and Germany.

Four categories of resource use with complete unit cost data in all countries were identified to create the basket in each country. These categories were consultant/radiologist operating time, nursing operating time, number of days in a neurosurgical ward and number of coils used in the endovascular procedure. The total cost of this basket was computed in each country using a weighted average of the unit cost in the country and the mean utilisation observed in the trial for all participants in the study. The mean time in theatre for consultant/radiologist and nurse was estimated to be 175 minutes, the mean number of days at a neurosurgical ward was 14.82 days and the mean number of coils was estimated to be 3.76 coils. Using this information and the unit cost associated to each resource use, the total cost of the basket for Sweden for example was calculated as follows: 175*3.02+175*0.43+14.82*881+3.76*542 = $15,707. Similarly for the remainder of countries the total cost of the basket were $14,830 for the UK, $13,777 for Finland and $15,720 for Germany.

The relative cost of each basket between pairs of countries was calculated to create an index table to assist imputing missing unit cost for a particular resource category. For instance, the relative index between Sweden and the UK was 1.06 (15,707/14,830) and between UK and Sweden was estimated to be 0.94 (14,830/15,707). These would be the indexes to use to impute any missing values. Because this combination had to be performed for each pair of countries, a total of 16 pairs were calculated and are shown in table 5.5.

Table 5.5: Relative medical cost indexes of the study countries

Index	**UK**	**Sweden**	**Finland**	**Germany**
UK	1.00	1.06	0.93	1.06
Sweden	0.94	1.00	0.88	1.00
Finland	1.08	1.14	1.00	1.14
Germany	0.94	1.00	0.88	1.00

Missing unit costs for a particular country were imputed by applying a weighted average among available country-specific unit cost using the indexes as weights. For example, the cost per day at an intensive therapy unit (ITU) was available for the UK ($2,753), Finland ($2,409) and Germany ($828) and missing in Sweden. To impute the cost per day at ITU in Sweden, the UK cost was multiplied by the British-Swedish index (1.06), the Finnish cost was multiplied by the Finnish-Swedish index (1.14) and the German cost was multiplied by the German-Swedish index (1.00). Then an average was calculated: (2,753*1.06+2,409*1.14+828*1.00)/3= $2,163. Table 5.6 presents the country-specific unit cost data available and the imputed unit costs using the market-basket method.

Table 5.6 presents the country-specific unit cost information available after adjusting for inflation and converted to $US PPP in 2008 prices. It also shows the imputed unit cost data using the market-basket method for missing information.

Table 5.6: Country-specific unit costs data collected, $US PPP 2008 (bold figures are values imputed using the market-basket method)

$US PPP 2008	UK	Sweden	Finland	Germany
Staff (per minute)				
Consultant Radiologist/Surgeon	10.09	3.02	1.83	2.88
Assistant Radiologist/Surgeon	1.39	**0.92**	0.40	0.82
Consultant Anaesthetist	10.09	2.47	1.83	**5.08**
Assistant Anaesthetist	1.39	**1.15**	**1.01**	0.82
Other Anaesthetic staff	0.75	**0.58**	0.30	0.60
Radiographer	0.74	0.55	0.30	**0.56**
Nurse	0.63	0.43	0.30	0.60
Ward Facility (per day)				
Intensive Therapy Unit (ITU)	2755	**2163**	2409	828
Neurology Ward	659	**664**	**583**	632
Neurosurgery Ward	678	881	746	828
Rehabilitation Unit	430	**465**	416	**465**
District General Hospital	410	**324**	187	**324**
Equipment (per minute)				
Endovascular	2.40	**2.54**	**2.23**	**2.54**
Neurosurgery	0.25	**0.26**	**0.23**	**0.26**
Consumables				
Coils (per coil)	778	542	624	755
Endovascular	1755	**1857**	**1629**	**1858**
Neurosurgery	1150	**1218**	**1068**	**1219**
Thrombolitic agent	1416	**1498**	**1314**	**1499**
Clip	260	**275**	**242**	**276**
Imaging and investigation				
Angiogram	1270	**1346**	**1180**	**1347**
CT Scan	240	**254**	**223**	**254**
MRI Scan	453	**480**	**421**	**481**

5.3.6 Intervention costs

For the endovascular procedure, theatre costs were calculated as the product of procedure duration and cost per minute of staff, equipment costs, the product of number of coils used per patient and cost per coil, the cost of thrombolytic agent if used, and a cost for all other consumables. The cost per patient for neurosurgery was calculated as the product of theatre

time and cost per minute of staff and equipment, plus the cost of consumables. If patients underwent a second procedure, this was costed using the same methods.

To calculate the cost per minute of staff time, information on the number of weeks worked per year, number of hours worked per week and the ratio of direct to indirect time in theatre for consultants was needed. This information was only available for the UK and similar estimates were assumed for the rest of the countries.

All costs incurred over a 12-month period were discounted at an annual rate of 3.5% following current recommendations (National Institute for Health and Clinical Excellence 2008).

5.3.7 Quality-adjusted life years (QALYs)

Quality-adjusted life years (QALYs) were used as the main outcome measure in the economic evaluation of ISAT. QALYs combine quality of life and mortality data into one single measure and it is the preferred health outcome by reimbursement agencies such as the National Institute for Health and Clinical Excellence in the United Kingdom (National Institute for Health and Clinical Excellence 2008). Expressing the results of a cost-effectiveness analysis[14] as cost per QALY gained ensures comparability across disease areas and medical specialties. QALYs were calculated as the area under the mortality adjusted health utility profile with straight line interpolation between utilities at each follow-up point (Billingham et al. 1999).

[14] When the results of a cost-effectiveness analysis are expressed in cost per QALY, the analysis is also referred as cost-utility analysis.

Quality of life was collected using the generic EuroQol EQ-5D instrument (EuroQol Group 1990). The EQ-5D covers five domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with three levels in each domain. Therefore, the EQ-5D produces a description of the health state of the patient using one of the 243 possible health states (3^5). Each of these health states can be converted into a utility using a tariff or value set estimated through time trade-off or visual analog scale methods (Dolan 2000). As population characteristics differ across countries, these tariffs have been already estimated for the UK, Finland and Germany as part of the EuroQol Group agenda (Szende et al. 2007). A Swedish value set is under development but current recommendation in this country suggests that the UK tariff should be used at the moment (Burstrom et al. 2006). There was no Canadian value set at the moment this thesis was conducted. However, a US value set has been recently estimated and was used in the Canadian subgroup (Shaw et al. 2005).

A country-specific value set was applied to country-specific EQ-5D data to describe differences in quality of life in each country between the endovascular and neurosurgical groups. However, across countries comparisons when using different value sets can be misleading as it is difficult to determine whether differences between the groups are the result of the intervention or due to country-differences in the valuation of the social tariff (Huang et al. 2007). Therefore when calculating QALYs the UK value set was applied to all countries as it is considered the most robust tariff to date (Szende et al. 2007) .
QALYs obtained over 12-month period were discounted at an annual rate of 3.5% following current recommendations (National Institute for Health and Clinical Excellence 2008).
Table 5.7 presents the coefficients of the EQ-5D value sets for the UK, Germany, Finland and the US.

Table 5.7: Coefficients for the estimation of the EQ-5D index values

Attribute	Level	UK	Germany	Finland	US
	Constant	-0.081	-0.001	-0.158	
	No problems (1)	0	0	0	0
Mobility	Some problems (2)	-0.069	-0.099	-0.058	-0.146
	Confined to bed (3)	-0.314	-0.327	-0.230	-0.558
	No problems (1)	0	0	0	0
Self care	Some problems (2)	-0.104	-0.087	-0.098	-0.175
	Unable to (3)	-0.214	-0.174	-0.143	-0.471
	No problems (1)	0	-	0	0
Usual activities	Some problems (2)	-0.036	-	-0.047	-0.140
	Unable to (3)	-0.094	-	-0.131	-0.374
	None (1)	0	0	0	0
Pain / discomfort	Moderate (2)	-0.123	-0.112	-0.111	-0.173
	Extreme (3)	-0.386	-0.315	-0.153	-0.537
	None (1)	0	0	0	0
Anxiety / depression	Moderate (2)	-0.071	-	-0.160	-0.156
	Extreme (3)	-0.236	-0.065	-0.196	-0.450
N3		-0.269	-0.323	-	-
D1					+0.140
I2-square					-0.011
I3					+0.122
I3-square					+0.015

N3: extra disutility associated with deviations from full health; D1: ordinal variables that represented the number of movements away from full health beyond the first (ranging from 0 to 4); I3: ordinal variable that represented the number of dimensions at level3 beyond the first. The square of the I3 term to allow for non-linearity in its association with the dependent variable; I2-squared: the square of I2, an ordinal variable that represented the number of dimensions at level 2 beyond the first.

No baseline EQ-5D data was collected due to the severity of the initial health state of the patient and therefore information was available from two months onwards. However, EQ-5D data were not collected for a proportion of patients at two months follow-up and lost to follow-up yielded missing data at each follow-up point. The EQ-5D was not originally included in the original trial protocol and its inclusion was decided when follow-up data had

started. Hence imputation techniques were used to report QALYs and Section 5.3.9 provides full description of the multiple imputation technique used.

5.3.8 Follow-up

Patients were followed up at two months post-intervention and on an annual basis thereafter. Currently data collection has been completed up to an eight year follow-up threshold.

5.3.9 Descriptive statistics and data analysis

Descriptive statistics and parametric t-tests to compare means between the endovascular and neurosurgical group were used. Contingency tables were compared using the Pearson's χ^2 or the Yate's corrected χ^2 test for low frequencies. Differences in proportions were detected using the recommended method by Altman and colleagues to deal with this type of data (Altman et al. 2000). Resource use and health outcomes are presented first for all participants in the study and then for the United Kingdom, Sweden, Germany, Canada and Finland. Resource use data are presented for the first episode of care and any subsequent procedure and follow-up. Resource use results are reported by specific resource category for all patients from whom information is available in that category.

Multiple imputation was used to impute missing EQ-5D data at any follow-up point (Rubin 1996). Missing data were imputed five times creating five separate datasets. Utilities at each follow-up point were model using gender, age and treatment arm as covariates and a predictive matching approach when estimating the imputations. QALYs were calculated in each dataset first and treatment effect between endovascular and neurosurgery was combined using Rubin's Rules (Rubin 1996). However, to obtain a QALY estimate for each

patient to be used in subsequent chapters, an average per patient across the five datasets was also computed and the treatment effect on this variable is also presented for comparison.

5.4 Results

5.4.1 Resource use

Table 5.8 shows the distribution of theatre staff present during the endovascular and neurosurgical procedures in the UK. No significant differences were detected across the seven centres in any of the procedures; endovascular χ^2=14.51, p=0.99 and neurosurgery χ^2=28.48, p=0.99. Therefore, an average across all centres was calculated to determine the theatre staffing present in the UK.

Table 5.8: Distribution of theatre staffing present during the procedures across centres in the UK

	Centre A	Centre B	Centre C	Centre D	Centre E	Centre F	Centre G	Total
Endovascular								
Consultant Radiologist/Surgeon	1	1	1	1	1	1	1	**7**
Assistant Radiologist/Surgeon	1	0	1	1	1	0	1	**5**
Consultant Anaesthetist	1	1	1	1	1	1	1	**7**
Assistant Anaesthetist	0	0	0	0	1	0	1	**2**
Anaesthetic ODP	1	1	1	1	0	1	1	**6**
Radiographer	2	1	2	2	2	1	2	**12**
Nurse grade G	0	0	1	0	0	0	0	**1**
Nurse grade F	1	1	1	1	2	1	0	**7**
Nurse grade E	1	1	0	0	0	1	2	**5**
Nurse grade D	0	1	0	1	0	1	1	**4**
Total	**8**	**7**	**8**	**8**	**8**	**7**	**10**	**56**
Yates χ^2								**14.51**
p-value								**0.99**
Neurosurgery								
Consultant Radiologist/Surgeon	1	1	1	1	1	1	1	**7**
Assistant Radiologist/Surgeon	1	1	1	1	1	1	1	**7**
Consultant Anaesthetist	1	1	1	1	1	1	1	**7**
Assistant Anaesthetist	0	1	0	1	1	1	0	**4**
Anaesthetic ODP	1	1	1	1	0	1	1	**6**
Radiographer	0	0	0	0	1	0	0	**1**
Nurse grade G	0	0	1	0	0	0	0	**1**
Nurse grade F	1	1	1	1	5	1	0	**10**
Nurse grade E	1	1	0	0	0	1	3	**6**
Nurse grade D	0	1	1	2	0	2	0	**6**
Total	**6**	**8**	**7**	**8**	**10**	**9**	**7**	**55**
Yates χ^2								**28.48**
p-value								**0.99**

ODP: Operating Department Practitioner

Table 5.9 shows the distribution of staffing present during the interventions for the four countries with available data. Some differences were shown in the distribution of staff across countries but no statistical differences were detected. It is interesting to note for example, that the German centre did not employ radiographers during the endovascular procedure although the remaining countries employed two on average. The substitution of anaesthetist staff for nurses in the neurosurgical intervention in Sweden is another example.

Table 5.9: Distribution of theatre staffing present during the procedures in the UK, Sweden, Finland and Germany

	UK	Sweden	Finland	Germany	Total
Endovascular					
Consultant Radiologist/Surgeon	1	1	1	1	4
Assistant Radiologist/Surgeon	1	0	1	1	3
Consultant Anaesthetist	1	1	1	0	3
Assistant Anaesthetist	0	0	0	1	1
Other Anaesthetic staff	1	0	1	1	3
Radiographer	2	2	2	0	6
Nurses	3	2	0	2	7
Total	9	6	6	6	27
Yates χ^2					**4.20**
p-value					**0.99**
Neurosurgery					
Consultant Radiologist/Surgeon	1	1	1	1	4
Assistant Radiologist/Surgeon	1	0	1	1	3
Consultant Anaesthetist	1	1	1	0	3
Assistant Anaesthetist	1	0	0	1	2
Other Anaesthetic staff	1	0	1	1	3
Radiographer	0	0	0	0	0
Nurses	3	4	2	2	11
Total	8	6	6	6	26
Yates χ^2					**2.10**
p-value					**0.99**

Tables 5.10-5.15 presents the mean resource use (standard deviation SD) associated with the first episode of care averaged over the number of patients using the resource category for all participants in the trial and for each country separately.

Significant differences in the total time in theatre between interventions were detected indicating longer times for the neurosurgical group within all countries. However, the results

from imaging and investigation post randomisation and length of stay suggested that the management of patients between groups was different in some resource categories for some countries.

The mean (SD) number of CT scans in the United Kingdom was 1.78 (1.56) in the endovascular group and 2.02 (1.28) in the neurosurgical group, a significant difference of 0.24 ($p<0.001$). Swedish and Finnish neurosurgical patients received on average more angiograms than endovascular participants. The opposite result was observed in Germany, where the mean (SD) number of angiograms was estimated to be 1.72 (1.01) in the endovascular group and 1.38 (0.56) in neurosurgical patients, a significant difference of an additional 0.34 in the endovascular group.

Endovascular patients spent on average significantly fewer days (2.43) on a ward than neurosurgical participants in the UK. Patients from Canada spent a mean (SD) number of days in ITU of 6.72 (5.58) and 11.08 (12.28) in the endovascular and neurosurgical group respectively, a significant difference of 4.36 days ($p<0.05$) more in the neurosurgical arm. It is interesting to note that when pooling the data from all countries, similar results to those using UK data alone are obtained. This indicates the influence of the UK sample size in determining the direction and significant of the difference when using the data from all patients.

Table 5.10: Resource use associated with the first episode of care for participants from all countries in the study (n=2,143)

	Endovascular (n=1,073)				Neurosurgery (n=1,070)			
Resource Category	**Number of patients using resource category**	**Mean resource use per patient (SD)**	**Min**	**Max**	**Number of patients using resource category**	**Mean resource use per patient (SD)**	**Min**	**Max**
Endovascular intervention (1st episode of care)								
-Duration (in minutes)	1045	138.98 (61.36)	45	585	38	135.18 (50.11)	60	300
-No of coils	1054	3.73 (2.54)	0	26	39	4.56 (2.59)	1	10
Neurosurgery intervention (1st episode of care)								
-Duration (in minutes)	10	207 (77.75)	120	360	995	215.40 (80.85)	60	960
Total duration (in minutes)*	1055	139.62 (61.84)			1033	212.44 (81.33)		
Subsequent intervention procedures before original discharge								
Endovascular procedures								
-Duration (in minutes)	21	174.10 (97.84)	60	440	30	139.84 (58.69)	60	300
-No of coils	21	2.57 (1.50)	0	5	30	3.43 (2.65)	0	10
Neurosurgical procedures								
-Duration (in minutes)	76	226.97 (82.00)	60	480	5	181.4 (90.80)	90	300
Imaging and investigations post-randomisation								
Number of angiograms	1040	1.26 (0.57)	1	7	1034	1.25 (0.47)	1	3
Number of CT scans*	976	1.98 (1.47)	1	15	999	2.20 (1.49)	1	9
Number of MRI	63	1.21 (0.51)	1	3	34	1.21 (0.41)	1	2
Length of stay first episode of care								
Number of ITU days	538	7.14 (8.09)	1	67	573	7.89 (8.20)	1	63
Number of ward days*	1045	14.08 (13.49)	1	242	1048	16.20 (15.01)	-5	316
Number of DGH days	434	25.76 (41.63)	1	345	503	32.10 (59.67)	1	731
Number of rehabilitation clinic days	89	69.60 (90.69)	2	650	123	93.67 (113.87)	1	666

* Statistical significant difference at $p<0.001$

Table 5.11: Resource use associated with the first episode of care for UK participants (n=1,644)

Resource Category	Endovascular (n=809)				Neurosurgery (n=835)			
	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max
Endovascular intervention (1st episode of care)								
-Duration (in minutes)	785	137.41 (55.69)	45	420	31	140.07 (50.51)	60	300
-No of coils	790	3.76 (2.49)	0	26	32	5.06 (2.54)	2	10
Neurosurgery intervention (1st episode of care)								
-Duration (in minutes)	10	207 (77.75)	120	360	771	213.81 (76.94)	60	595
Total duration (in minutes)*	795	138.28 (56.49)			802	210.96 (77.38)		
Subsequent intervention procedures before original discharge								
Endovascular procedures								
-Duration (in minutes)	12	154.67 (37.99)	105	240	19	152.11 (52.48)	60	255
-No of coils	12	2.58 (1.17)	0	4	19	3.53 (2.43)	0	8
Neurosurgical procedures								
-Duration (in minutes)	54	229.67 (73.54)	60	390	3	190 (105.36)	90	300
Imaging and investigations post-randomisation								
Number of angiograms	789	1.16 (0.41)	1	3	806	1.14 (0.37)	1	3
Number of CT scans*	735	1.78 (1.56)	1	8	773	2.02 (1.28)	1	9
Number of MRI	43	1.16 (0.48)	1	3	25	1.24 (0.44)	1	2
Length of stay first episode of care								
Number of ITU days	301	5.25 (5.91)	1	40	349	5.77 (5.94)	1	37
Number of ward days*	808	14.82 (11.92)	1	152	834	17.25 (15.31)	1	316
Number of DGH days	332	25.28 (42.93)	1	365	409	29.71 (51.97)	1	389
Number of rehabilitation clinic days	52	73.39 (101.73)	2	650	71	115.28 (132.32)	2	666

* Statistical significant difference at $p<0.001$

Table 5.12: Resource use associated with the first episode of care for Swedish participants (n=123)

Resource Category	Endovascular (n=64)				Neurosurgery (n=59)			
	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max
Endovascular intervention (1st episode of care)								
-Duration (in minutes)	64	135.47 (87.31)	60	420	2	1257.5 (60.10)	115	200
-No of coils	64	2.75 (1.77)	0	10	2	3.5 (0.70)	3	4
Neurosurgery intervention (1st episode of care)								
-Duration (in minutes)	0				57	219.56 (78.33)	90	435
Total duration (in minutes)*	64	135.47 (87.31)			59	217.46 (78.20)		
Subsequent intervention procedures before original discharge								
Endovascular procedures								
-Duration (in minutes)	1	80	80	80	4	112.5 (15)	90	120
-No of coils	1	5	5	5	4	4.5 (2.65)	1	7
Neurosurgical procedures								
-Duration (in minutes)	0				7	234 (61.46)	135	300
Imaging and investigations post-randomisation								
Number of angiograms**	63	1.30 (0.56)	1	3	53	1.55 (0.50)	1	2
Number of CT scans	64	2.23 (1.51)	1	7	58	2.29 (1.44)	1	6
Number of MRI	2	1	1	1	2	1	1	1
Length of stay first episode of care								
Number of ITU days	63	8.68 (7.44)	1	33	57	10.35 (7.26)	1	36
Number of ward days	58	9.62 (6.33)	1	41	51	9.71 (5.70)	1	24
Number of DGH days	45	26.12 (46.17)	1	213	42	29.96 (42.94)	1	239
Number of rehabilitation clinic days	5	32 (10.84)	22	49	14	73.18 (67.22)	8	246

* Statistical significant difference at $p<0.001$
** Statistical significant difference at $p<0.05$

Table 5.13: Resource use associated with the first episode of care for German participants (n=118)

Resource Category	Endovascular (n=62)				Neurosurgery (n=56)			
	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max
Endovascular intervention (1st episode of care)								
-Duration (in minutes)	59	154.66 (94.50)	60	585	3	76.67 (15.27)	60	90
-No of coils	62	3.05 (2.29)	0	12	3	2.33 (1.53)	1	4
Neurosurgery intervention (1st episode of care)								
-Duration (in minutes)	0				52	231.94 (130.46)	60	960
Total duration (in minutes)**	59	154.66 (94.50)			55	223.47 (131.72)		
Subsequent intervention procedures before original discharge								
Endovascular procedures								
-Duration (in minutes)	4	235 (163.83)	80	440	1	300	300	300
-No of coils	4	2 (2.45)	0	5	1	2	2	2
Neurosurgical procedures								
-Duration (in minutes)	6	216.67 (132.43)	80	455	2	168.5 (101.12)	97	240
Imaging and investigations post-randomisation								
Number of angiograms**	57	1.72 (1.01)	1	7	52	1.38 (0.56)	1	3
Number of CT scans	52	3.44 (2.81)	1	15	51	3.81 (2.33)	1	9
Number of MRI	3	1	1	1	2	1	1	1
Length of stay first episode of care								
Number of ITU days	61	14.03 (12.30)	1	62	56	17.63 (10.59)	2	43
Number of ward days	46	9.78 (8.93)	1	33	50	11.04 (11.47)	1	61
Number of DGH days	17	41.18 (25.34)	11	98	10	76.35 (73.32)	8	207
Number of rehabilitation clinic days	17	94.85 (98.24)	15	411	25	59.88 (64.07)	1	284

** Statistical significant difference at $p<0.05$

Table 5.14: Resource use associated with the first episode of care for Canadian participants (n=89)

Resource Category	Endovascular (n=48)				Neurosurgery (n=41)			
	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max
Endovascular intervention (1st episode of care)								
-Duration (in minutes)	48	151.04 (59.35)	60	305	1	130	130	130
-No of coils	48	4.44 (3.43)	0	14	1	1	1	1
Neurosurgery intervention (1st episode of care)								
-Duration (in minutes)	0				38	268.29 (70.63)	150	480
Total duration (in minutes)*	48	151.04 (59.35)			39	264.74 (73.13)		
Subsequent intervention procedures before original discharge								
Endovascular procedures								
-Duration (in minutes)	0				0			
-No of coils	0				0			
Neurosurgical procedures								
-Duration (in minutes)	4	280 (143.53)	150	480	0			
Imaging and investigations post-randomisation								
Number of angiograms	47	2.11 (0.89)	1	5	41	1.95 (0.55)	1	3
Number of CT scans	44	2.52 (1.61)	1	7	40	3.05 (2.24)	1	9
Number of MRI	5	1.2 (0.44)	1	2	3	1	1	1
Length of stay first episode of care								
Number of ITU days**	47	6.72 (5.58)	1	20	39	11.08 (12.28)	1	63
Number of ward days	48	11.81 (8.84)	1	36	38	15.84 (21.51)	1	134
Number of DGH days	6	15.83 (12.91)	2	36	7	125 (258.88)	2	731
Number of rehabilitation clinic days	3	36 (7)	31	44	4	31.5 (19.89)	15	59

* Statistical significant difference at $p<0.001$
** Statistical significant difference at $p<0.05$

Table 5.15: Resource use associated with the first episode of care for Finnish participants (n=63)

Resource Category	Endovascular (n=34)				Neurosurgery (n=29)			
	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max	Number of patients using resource category	Mean resource use per patient (SD)	Min	Max
Endovascular intervention (1st episode of care)								
-Duration (in minutes)	34	128.68 (57.67)	60	300	1	120	120	120
-No of coils	34	3.91 (2.77)	0	12	1	1	1	1
Neurosurgery intervention (1st episode of care)								
-Duration (in minutes)	0				28	174.46 (49.54)	80	280
Total duration (in minutes)**	34	128.68 (57.67)			29	172.59 (49.69)		
Subsequent intervention procedures before original discharge								
Endovascular procedures								
-Duration (in minutes)	2	75 (21.21)	60	90	4	93.75 (39.45)	60	135
-No of coils	2	1.5 (0.71)	1	2	4	1	1	1
Neurosurgical procedures								
-Duration (in minutes)	2	185 (77.78)	130	240	0			
Imaging and investigations post-randomisation								
Number of angiograms*	34	1.15 (0.36)	1	2	29	2.03 (0.42)	1	3
Number of CT scans	33	1.67 (1.02)	1	6	28	1.71 (1.01)	1	4
Number of MRI	1	1	1	1	0			
Length of stay first episode of care								
Number of ITU days	11	4.64 (5.02)	1	13	28	2.25 (2.63)	1	12
Number of ward days	34	10.77 (5.50)	2	32	29	11 (4)	5	21
Number of DGH days	21	31.86 (72.98)	1	345	22	19.18 (18.96)	1	64
Number of rehabilitation clinic days	3	33 (37.99)	4	76	0			

* Statistical significant difference at $p<0.001$
** Statistical significant difference at $p<0.05$

Table 5.16 presents the risk of readmission for any operative procedure after original discharge at each follow-up point and country in each intervention group. The risk of readmission was significantly higher in the endovascular group in years one and two when pooling data from all countries. It appears that most readmissions occurred at one year follow-up and this pattern was similar across all countries. Significant differences were found between groups at two years follow-up in the United Kingdom. In addition, patients in the endovascular group in Finland underwent an additional significant 37% more readmissions than the neurosurgical group. This difference was still present at two years follow-up.

Table 5.17 shows the risk of having an angiogram after original discharge for both groups and country. In all countries, patients in the endovascular group had significantly more angiograms than neurosurgical patients. Although the statistical significance between groups differs across years and country, the largest difference occurred during the first year post-randomisation for all countries.

Table 5.16: Risk of readmission for subsequent operational procedures at each follow-up point and country

	Endovascular			Neurosurgery			
	No at risk at the beginning of the period	Patients having a readmission	Risk of readmissions	No at risk at the beginning of the period	Patients having a readmission	Risk of readmissions	Difference between groups (95% CI)
All participants							
Year 1	982	209	0.213	957	163	0.170	0.043 (0.007 to 0.077)*
Year 2	941	67	0.071	892	35	0.039	0.032 (0.011 to 0.053)*
Year 3	888	47	0.053	831	31	0.037	0.016 (-0.004 to 0.036)
Year 4	845	25	0.030	783	14	0.018	0.012 (-0.004 to 0.027)
Year 5	800	24	0.030	755	15	0.020	0.01 (-0.006 to 0.026)
United Kingdom							
Year 1	751	163	0.217	747	133	0.178	0.039 (-0.001 to 0.079)
Year 2	734	53	0.072	706	32	0.045	0.027 (0.002 to 0.052)*
Year 3	709	42	0.059	673	29	0.043	0.016 (-0.007 to 0.040)
Year 4	689	18	0.026	651	14	0.022	0.005 (-0.013 to 0.022)
Year 5	671	18	0.027	639	15	0.023	0.003 (-0.014 to 0.021)
Sweden							
Year 1	58	13	0.224	53	20	0.377	-0.153 (-0.314 to 0.017)
Year 2	58	3	0.052	52	3	0.058	-0.006 (-0.11 to 0.091)
Year 3	58	0	0.000	50	1	0.020	-0.020 (-0.105 to 0.044)
Year 4	56	2	0.036	49	0	0.000	0.036 (-0.041 to 0.121)
Year 5	56	1	0.018	48	0	0.000	0.018 (-0.058 to 0.094)
Germany							
Year 1	46	4	0.087	53	1	0.019	0.068 (-0.028 to 0.185)
Year 2	45	1	0.022	53	0	0.000	0.022 (-0.048 to 0.116)
Year 3	45	0	0.000	53	1	0.019	-0.019 (-0.099 to 0.061)
Year 4	45	1	0.022	53	0	0.000	0.022 (-0.048 to 0.116)
Year 5	45	1	0.022	52	0	0.000	0.022 (-0.049 to 0.116)
Canada							
Year 1	43	2	0.047	35	1	0.029	0.018 (-0.104 to 0.129)
Year 2	38	0	0.000	32	0	0.000	0.000 (-0.107 to 0.092)
Year 3	31	1	0.032	23	0	0.000	0.032 (-0.113 to 0.162)
Year 4	29	0	0.000	22	0	0.000	0.000 (-0.149 to 0.117)
Year 5	26	1	0.038	13	0	0.000	0.038 (-0.192 to 0.189)
Finland							
Year 1	34	20	0.588	28	6	0.214	0.374 (0.128 to 0.560)*
Year 2	34	9	0.265	28	0	0.000	0.265 (0.095 to 0.431)*
Year 3	31	2	0.065	27	0	0.000	0.065 (-0.068 to 0.207)
Year 4	30	2	0.067	27	0	0.000	0.067 (-0.067 to 0.213)
Year 5	24	3	0.125	25	0	0.000	0.125 (-0.031 to 0.31)

*$p<0.05$

Table 5.17: Risk of having an angiogram after original discharged at each follow-up point and country

	Endovascular			Neurosurgery			
	No at risk at the beginning of the period	Patients having an angiogram	Risk of angiogram	No at risk at the beginning of the period	Patients having an angiogram	Risk of angiogram	Difference between groups (95% CI)
All participants							
Year 1	982	656	0.668	957	187	0.195	0.473 (0.433 to 0.51)*
Year 2	941	219	0.233	892	54	0.061	0.172 (0.141 to 0.203)*
Year 3	888	148	0.167	831	24	0.029	0.138 (0.111 to 0.165)*
Year 4	845	76	0.090	783	15	0.019	0.071 (0.049 to 0.093)*
Year 5	800	27	0.034	755	10	0.013	0.021 (0.005 to 0.037)*
United Kingdom							
Year 1	751	529	0.704	747	161	0.216	0.489 (0.443 to 0.531)*
Year 2	734	166	0.226	706	44	0.062	0.164 (0.129 to 0.199)*
Year 3	709	129	0.182	673	22	0.033	0.149 (0.118 to 0.181)*
Year 4	689	42	0.061	651	11	0.017	0.044 (0.024 to 0.066)*
Year 5	671	26	0.039	639	10	0.016	0.023 (0.005 to 0.042)*
Sweden							
Year 1	58	33	0.569	53	13	0.245	0.324 (0.141 to 0.477)*
Year 2	58	15	0.259	52	4	0.077	0.182 (0.04 to 0.315)*
Year 3	58	7	0.121	50	0	0.000	0.121 (0.027 to 0.229)*
Year 4	56	30	0.536	49	3	0.061	0.474 (0.309 to 0.605)*
Year 5	56	0	0.000	48	0	0.000	0.00 (-0.074 to 0.064)
Germany							
Year 1	46	13	0.283	53	1	0.019	0.264 (0.128 to 0.407)*
Year 2	45	1	0.022	53	0	0.000	0.022 (-0.048 to 0.116)
Year 3	45	0	0.000	53	0	0.000	0.00 (-0.068 to 0.079)
Year 4	45	0	0.000	53	1	0.019	-0.019 (-0.099 to 0.061)
Year 5	45	0	0.000	52	0	0.000	0.00 (-0.069 to 0.079)
Canada							
Year 1	43	34	0.791	35	2	0.057	0.734 (0.541 to 0.837)*
Year 2	38	13	0.342	32	1	0.031	0.311 (0.13 to 0.472)*
Year 3	31	3	0.097	23	0	0.000	0.097 (-0.06 to 0.249)
Year 4	29	2	0.069	22	0	0.000	0.069 (-0.088 to 0.22)
Year 5	26	0	0.000	13	0	0.000	0 (-0.228 to 0.129)
Finland							
Year 1	34	23	0.676	28	4	0.143	0.534 (0.293 to 0.691)*
Year 2	34	19	0.559	28	4	0.143	0.416 (0.178 to 0.591)*
Year 3	31	7	0.226	27	2	0.074	0.152 (-0.043 to 0.332)
Year 4	30	1	0.033	27	0	0.000	0.033 (-0.094 to 0.167)
Year 5	24	1	0.042	25	0	0.000	0.042 (-0.096 to 0.202)

*p<0.05

5.4.2 EQ-5D results

Tables 5.18-5.23 show the distribution of responses to the EQ-5D across the different levels of each dimension at 2 months, 1 year and 5 years follow-up for all participants and for each country separately. The mobility and self-care dimension registered a significant difference across responses indicating better quality of life in the endovascular group at two months follow-up in the United Kingdom. The same result was observed in Canada for the self-care dimension at two months follow-up. In addition, significant differences were detected in the responses of the usual activities domain at one-year follow-up favouring the endovascular group again in Canada.

Table 5.18: Distribution (n, %) of EQ-5D responses across the dimensions for all participants at 2 months, 1 year and 5 years follow-up

	Mobility		Self-care		Usual Activities		Pain		Anxiety	
Follow-up point	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**
2 months										
1	600 (67%)	522 (60%)	725 (80%)	644 (74%)	401 (45%)	358 (41%)	386 (43%)	339 (40%)	456 (51%)	391 (46%)
2	254 (28%)	299 (35%)	120 (13%)	168 (19%)	383 (43%)	385 (45%)	480 (54%)	481 (56%)	375 (42%)	395 (46%)
3	43 (5%)	46 (5%)	52 (6%)	54 (6%)	113 (13%)	123 (14%)	26 (3%)	35 (4%)	57 (6%)	65 (8%)
χ^2 (p-value)	**8.68**	**0.013**	**12.29**	**0.002**	2.32	0.313	3.59	0.166	5.25	0.073
1 year										
1	668 (71%)	609 (68%)	794 (84%)	729 (81%)	570 (60%)	485 (54%)	504 (54%)	467 (52%)	489 (52%)	418 (47%)
2	252 (27%)	276 (31%)	117 (12%)	142 (16%)	299 (32%)	335 (37%)	397 (42%)	388 (44%)	389 (41%)	406 (46%)
3	23 (2%)	16 (2%)	33 (4%)	30 (3%)	75 (8%)	81 (9%)	40 (4%)	38 (4%)	62 (7%)	68 (8%)
χ^2 (p-value)	4.12	0.128	4.33	0.115	**8.13**	**0.017**	0.31	0.857	4.94	0.084
5 year										
1	522 (70%)	469 (68%)	623 (83%)	578 (83%)	477 (64%)	412 (60%)	474 (63%)	420 (61%)	435 (59%)	398 (58%)
2	215 (29%)	214 (31%)	102 (14%)	94 (14%)	229 (31%)	240 (35%)	237 (32%)	243 (35%)	264 (36%)	259 (38%)
3	11 (2%)	11 (2%)	24 (3%)	20 (3%)	41 (6%)	40 (6%)	36 (5%)	27 (4%)	46 (6%)	30 (4%)
χ^2 (p-value)	0.82	0.665	0.12	0.941	2.93	0.232	2.37	0.306	2.72	0.257

Bold figures indicate significant differences at 5% level on the distribution of responses

Table 5.19: Distribution (n, %) of EQ-5D responses across the dimensions for UK participants at 2 months, 1 year and 5 years follow-up

	Mobility		Self-care		Usual Activities		Pain		Anxiety	
Follow-up point	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**
2 months										
1	432 (64%)	374 (57%)	537 (79%)	476 (72%)	273 (40%)	253 (39%)	291 (43%)	242 (37%)	336 (50%)	285 (44%)
2	215 (32%)	249 (38%)	105 (16%)	147 (22%)	322 (48%)	315 (48%)	364 (54%)	382 (59%)	290 (43%)	313 (48%)
3	30 (4%)	35 (5%)	35 (5%)	34 (5%)	82 (12%)	90 (14%)	20 (3%)	29 (4%)	47 (7%)	52 (8%)
χ^2 (p-value)	**6.78**	**0.034**	**10.39**	**0.006**	0.94	0.625	6.23	0.044	4.92	0.085
1 year										
1	487 (68%)	454 (65%)	594 (83%)	550 (79%)	410 (57%)	354 (51%)	384 (50%)	347 (50%)	360 (50%)	319 (46%)
2	218 (30%)	232 (33%)	107 (15%)	129 (18%)	258 (36%)	285 (41%)	302 (42%)	313 (45%)	313 (44%)	316 (46%)
3	14 (2%)	13 (2%)	19 (3%)	20 (3%)	52 (7%)	60 (9%)	32 (5%)	31 (5%)	44 (6%)	56 (8%)
χ^2 (p-value)	1.35	0.510	3.46	0.177	5.71	0.058	1.56	0.456	3.45	0.178
5 year										
1	420 (68%)	383 (66%)	510 (82%)	478 (83%)	388 (63%)	336 (58%)	391 (63%)	343 (60%)	360 (58%)	334 (58%)
2	190 (31%)	187 (32%)	96 (16%)	86 (15%)	203 (33%)	208 (36%)	198 (32%)	207 (36%)	217 (35%)	212 (37%)
3	10 (2%)	10 (2%)	15 (2%)	14 (2%)	29 (5%)	34 (6%)	30 (5%)	26 (5%)	42 (7%)	27 (5%)
χ^2 (p-value)	0.40	0.820	0.08	0.962	2.72	0.256	2.08	0.35		

Bold figures indicate significant differences at 5% level on the distribution of responses

Table 5.20: Distribution (n, %) of EQ-5D responses across the dimensions for Swedish participants at 2 months, 1 year and 5 years follow-up

	Mobility		Self-care		Usual Activities		Pain		Anxiety	
Follow-up point	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**
2 months										
1	48 (84%)	41 (76%)	50 (88%)	45 (83%)	33 (58%)	24 (44%)	25 (45%)	22 (43%)	34 (62%)	26 (51%)
2	7 (12%)	12 (22%)	2 (4%)	3 (6%)	15 (26%)	21 (39%)	31 (56%)	28 (55%)	21 (38%)	24 (47%)
3	2 (4%)	1 (2%)	5 (9%)	6 (11%)	9 (16%)	9 (17%)	0 (0%)	1 (2%)	0 (0%)	1 (2%)
χ^2 (p-value)	2.12	0.346	0.473	0.789	2.34	0.310	1.11	0.573	2.12	0.347
1 year										
1	49 (86%)	40 (76%)	53 (93%)	47 (89%)	44 (77%)	36 (68%)	32 (56%)	28 (52%)	35 (62%)	25 (47%)
2	6 (11%)	13 (25%)	0 (0%)	2 (4%)	9 (16%)	12 (23%)	24 (42%)	24 (45%)	18 (32%)	27 (51%)
3	2 (4%)	0 (0%)	4 (7%)	4 (8%)	4 (7%)	5 (9%)	1 (2%)	1 (2%)	4 (7%)	1 (2%)
χ^2 (p-value)	5.35	0.069	2.22	0.330	1.20	0.550	0.121	0.941	5.12	0.077
5 year										
1	48 (86%)	34 (71%)	48 (86%)	41 (85%)	39 (70%)	30 (63%)	39 (70%)	32 (67%)	32 (59%)	26 (54%)
2	8 (14%)	14 (29%)	4 (7%)	4 (8%)	12 (21%)	16 (33%)	15 (27%)	15 (31%)	22 (41%)	20 (42%)
3	0 (0%)	0 (0%)	4 (7%)	3 (6%)	5 (9%)	2 (4%)	2 (4%)	1 (2%)	0 (0%)	2 (4%)
χ^2 (p-value)	3.43	0.064	0.08	0.962	2.43	0.297	0.41	0.814	2.37	0.306

Bold figures indicate significant differences at 5% level on the distribution of responses

Table 5.21: Distribution (n, %) of EQ-5D responses across the dimensions for German participants at 2 months, 1 year and 5 years follow-up

	Mobility		Self-care		Usual Activities		Pain		Anxiety	
Follow-up point	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**
2 months										
1	28 (70%)	36 (72%)	29 (73%)	38 (76%)	25 (63%)	29 (58%)	16 (42%)	28 (57%)	22 (60%)	27 (56%)
2	3 (8%)	10 (20%)	3 (8%)	7 (14%)	5 (13%)	10 (20%)	17 (45%)	18 (37%)	10 (27%)	16 (33%)
3	9 (23%)	4 (8%)	8 (20%)	5 (10%)	10 (25%)	11 (22%)	5 (13%)	3 (6%)	5 (14%)	5 (11%)
χ^2 (p-value)	5.65	0.059	2.42	0.298	0.91	0.634	2.45	0.294	0.48	0.787
1 year										
1	34 (77%)	38 (79%)	34 (77%)	38 (79%)	27 (61%)	35 (73%)	26 (61%)	32 (67%)	26 (61%)	27 (58%)
2	4 (9%)	8 (17%)	3 (7%)	6 (13%)	8 (18%)	6 (13%)	15 (35%)	14 (29%)	11 (26%)	17 (36%)
3	6 (14%)	2 (4%)	7 (16%)	4 (8%)	9 (21%)	7 (15%)	2 (5%)	2 (4%)	6 (14%)	3 (6%)
χ^2 (p-value)	3.39	0.184	1.87	0.393	1.40	0.497	0.38	0.826	2.13	0.345
5 year										
1	13 (81%)	18 (82%)	13 (81%)	20 (91%)	13 (81%)	16 (73%)	12 (75%)	17 (77%)	14 (88%)	14 (64%)
2	2 (13%)	4 (18%)	1 (6%)	1 (5%)	0 (0%)	4 (18%)	4 (25%)	5 (23%)	2 (13%)	8 (36%)
3	1 (6%)	0 (%)	2 (13%)	1 (5%)	3 (19%)	2 (9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
χ^2 (p-value)	1.55	0.457	0.89	0.640	3.65	0.161	0.03	0.871	2.72	0.099

Bold figures indicate significant differences at 5% level on the distribution of responses

Table 5.22: Distribution (n, %) of EQ-5D responses across the dimensions for Canadian participants at 2 months, 1 year and 5 years follow-up

	Mobility		Self-care		Usual Activities		Pain		Anxiety	
Follow-up point	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**
2 months										
1	33 (77%)	24 (65%)	41 (95%)	28 (76%)	24 (56%)	18 (49%)	20 (47%)	18 (51%)	24 (56%)	19 (54%)
2	10 (23%)	10 (27%)	2 (5%)	6 (16%)	17 (39%)	15 (41%)	23 (54%)	17 (49%)	17 (40%)	14 (40%)
3	0 (0%)	3 (8%)	0 (0%)	3 (8%)	2 (5%)	4 (11%)	0 (0%)	0 (%)	2 (5%)	2 (6%)
χ^2 (p-value)	3.99	0.136	**7.04**	**0.03**	1.21	0.547	0.19	0.666	0.05	0.974
1 year										
1	36 (86%)	24 (71%)	38 (91%)	31 (91%)	35 (83%)	17 (50%)	25 (60%)	15 (44%)	28 (67%)	14 (41%)
2	6 (14%)	9 (27%)	4 (10%)	2 (6%)	6 (14%)	14 (41%)	16 (38%)	18 (53%)	13 (31%)	18 (53%)
3	0 (0%)	1 (3%)	0 (0%)	1 (3%)	1 (2%)	3 (9%)	1 (2%)	1 (3%)	1 (2%)	2 (6%)
χ^2 (p-value)	3.19	0.203	1.55	0.460	**9.70**	**0.008**	1.80	0.407	5.02	0.081
5 year										
1	19 (76%)	8 (62%)	24 (96%)	11 (85%)	16 (67%)	8 (6%)	14 (56%)	7 (54%)	12 (48%)	4 (31%)
2	6 (24%)	4 (31%)	1 (4%)	1 (8%)	8 (33%)	4 (31%)	11 (44%)	6 (46%)	13 (52%)	9 (69%)
3	0 (0%)	1 (8%)	0 (0%)	1 (8%)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
χ^2 (p-value)	2.32	0.313	2.27	0.322	1.90	0.387	0.016	0.899	1.04	0.307

Bold figures indicate significant differences at 5% level on the distribution of responses

Table 5.23: Distribution (n, %) of EQ-5D responses across the dimensions for Finnish participants at 2 months, 1 year and 5 years follow-up

	Mobility		Self-care		Usual Activities		Pain		Anxiety	
Follow-up point	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**	**Endovascular**	**Neurosurgery**
2 months										
1	23 (68%)	20 (71%)	29 (85%)	24 (86%)	21 (62%)	18 (64%)	18 (53%)	17 (61%)	21 (62%)	14 (50%)
2	10 (29%)	7 (25%)	4 (12%)	2 (7%)	7 (21%)	7 (25%)	16 (47%)	11 (39%)	12 (35%)	10 (36%)
3	1 (3%)	1 (4%)	1 (3%)	2 (7%)	6 (18%)	3 (11%)	0 (0%)	0 (0%)	1 (3%)	4 (14%)
χ^2 (p-value)	0.160	0.923	0.90	0.638	0.66	0.720	0.38	0.539	2.83	0.243
1 year										
1	26 (77%)	22 (79%)	31 (91%)	25 (89%)	20 (59%)	20 (71%)	17 (50%)	20 (71%)	17 (50%)	15 (54%)
2	8 (24%)	6 (21%)	1 (3%)	2 (7%)	8 (24%)	6 (21%)	14 (41%)	7 (25%)	13 (38%)	10 (36%)
3	0 (0%)	0 (0%)	2 (6%)	1 (4%)	6 (18%)	2 (7%)	3 (9%)	1 (4%)	4 (12%)	3 (11%)
χ^2 (p-value)	0.04	0.844	0.74	0.692	1.72	0.423	3.02	0.220	0.08	0.961
5 year										
1	17 (71%)	20 (80%)	21 (88%)	22 (88%)	17 (71%)	17 (68%)	13 (54%)	18 (72%)	13 (54%)	15 (60%)
2	7 (29%)	5 (20%)	0 (0%)	2 (8%)	3 (13%)	7 (28%)	8 (33%)	7 (28%)	7 (29%)	9 (36%)
3	0 (0%)	0 (0%)	3 (12%)	1 (4%)	4 (17%)	1 (4%)	3 (13%)	0 (0%)	4 (17%)	1 (4%)
χ^2 (p-value)	0.56	0.456	3.00	0.223	3.38	0.184	3.85	0.146	2.17	0.337

Bold figures indicate significant differences at 5% level on the distribution of responses

Table 5.24 shows the EQ-5D utility values using country-specific value sets up to five years follow-up and participating country, using a fully-split analysis. In this case, the information from the EQ-5D domains was combined with each country-specific value set. The table also includes the cumulative number of deaths at each follow-up. EQ-5D utility values are showed for those patients with missing and completed EQ-5D data.

An overall consistent improvement was observed in terms of quality of life across year and country within each treatment group. A significant difference was detected between groups at 2 months follow-up with a mean (SD) of 0.68 (0.32) in the endovascular group and 0.64 (0.32) in the neurosurgical group, suggesting a better quality of life for patients in the endovascular group in the United Kingdom. No other significant differences were detected on subsequent follow-ups. None of the remainder countries presented any difference between groups at any follow-up in terms of quality of life.

Table 5.24: Utility values obtained from the EQ-5D at each follow-up point and using country-specific EQ-5D value sets for participating countries

	Endovascular				Neurosurgery			
EQ-5D	**Dead (n)**	**Missing (n)**	**Patients with completed EQ-5D (n)**	**Mean (SD)**	**Dead (n)**	**Missing (n)**	**Patients with completed EQ-5D (n)**	**Mean (SD)**
Utility values								
United Kingdom (n=1,644)								
2 months*	46	91	672	0.68 (0.32)	70	116	649	0.64 (0.32)
Year 1	56	37	716	0.73 (0.3)	83	63	689	0.7 (0.31)
Year 2	58	81	670	0.74 (0.3)	91	112	632	0.72 (0.3)
Year 3	64	101	644	0.74 (0.31)	106	133	596	0.75 (0.28)
Year 4	71	106	632	0.75 (0.31)	111	138	586	0.76 (0.28)
Year 5	76	117	616	0.75 (0.31)	116	146	573	0.75 (0.29)
Sweden (n=123)								
2 months	5	4	55	0.76 (0.26)	4	4	51	0.72 (0.27)
Year 1	5	2	57	0.77 (0.3)	5	1	53	0.75 (0.26)
Year 2	5	2	57	0.8 (0.28)	6	1	52	0.76 (0.3)
Year 3	5	1	58	0.8 (0.27)	7	2	50	0.78 (0.27)
Year 4	6	2	56	0.81 (0.26)	7	3	49	0.77 (0.27)
Year 5	6	4	54	0.83 (0.26)	7	4	48	0.77 (0.25)
Germany (n=118)								
2 months	16	9	37	0.75 (0.4)	3	5	48	0.79 (0.3)
Year 1	16	3	43	0.79 (0.34)	3	6	47	0.84 (0.27)
Year 2	17	13	32	0.82 (0.34)	3	17	36	0.87 (0.22)
Year 3	17	17	28	0.8 (0.33)	3	20	33	0.87 (0.25)
Year 4	17	24	21	0.84 (0.26)	3	29	24	0.91 (0.15)
Year 5	17	29	16	0.85 (0.28)	4	30	22	0.91 (0.15)
Canada (n=89)								
2 months	4	1	43	0.82 (0.13)	6	0	35	0.80 (0.19)
Year 1	5	1	42	0.86 (0.16)	6	1	34	0.78 (0.21)
Year 2	5	5	38	0.84 (0.18)	6	4	31	0.78 (0.21)
Year 3	5	12	31	0.85 (0.17)	6	12	23	0.83 (0.23)
Year 4	5	14	29	0.85 (0.14)	6	13	22	0.78 (0.23)
Year 5	5	19	24	0.86 (0.12)	7	21	13	0.78 (0.23)
Finland (n=63)								
2 months	0	0	34	0.71 (0.25)	1	0	28	0.7 (0.26)
Year 1	0	0	34	0.69 (0.25)	1	0	28	0.74 (0.22)
Year 2	0	0	34	0.73 (0.26)	1	2	26	0.69 (0.21)
Year 3	0	3	31	0.75 (0.28)	1	1	27	0.77 (0.23)
Year 4	1	3	30	0.76 (0.25)	2	0	27	0.79 (0.22)
Year 5	2	8	24	0.7 (0.27)	2	2	25	0.77 (0.22)
	46	91	672	0.68 (0.32)	70	116	649	0.64 (0.32)

* Statistical significant difference at $p<0.05$

Table 5.25 shows the results of using country-specific EQ-5D index value sets compared to using UK value set in the countries where a country-specific value set was available. The two months follow-up was selected as the baseline measure. To avoid the missing data in years 3, 4 and 5, year two was selected to determine whether changes over time were measured differently depending on the value set applied.

Significant baseline differences between value sets were found within each intervention group in Germany and Canada. In both cases, the country-specific value sets estimated higher utility values than the UK value set. All differences were highly significant with associated p-value $p<0.001$. No differences were found between the Finish and the UK value sets.

In terms of differences over time, the neurosurgical group experienced a significant larger improvement when the UK value set was used in German patients. The within group difference over time between value sets was estimated to be -0.03 with associated p-value $p<0.05$.

No significant differences were detected in the differences of between group mean differences in any of the countries. This result suggested that the mean difference between groups was likely to be the same independently of the value set used. However, it may be possible that the small sample size in this analysis may well lack power to detect a significant difference. Therefore, using country-specific value sets to calculate quality-adjusted life years (QALYs) has the potential to distort the interpretation of any significant result. Using a single value set on the whole sample, any differences detected are likely to be due to the interventions and not to differences in the value set estimations. To avoid this case, QALYs were calculated using the UK value set.

Table 5.25: EQ-5D mean index values comparison using country-specific value set compared to the UK value set in Germany, Canada and Finland at 2 months and 2 years follow-up

	Endovascular			Neurosurgery					
	n	Country-specific value set Mean (SD)	UK value set Mean (SD)	n	Country-specific value set Mean (SD)	UK value set Mean (SD)	Mean difference German value set (SE diff)	Mean difference UK set value (SE diff)	Differences in between group mean differences SE (diff)
Germany (n=118)									
2 months	37	0.75 (0.41)	0.63 (0.51)	48	0.79 (0.30)	0.67 (0.39)	-0.03 (0.08)	-0.05 (0.10)	0.02 (0.128)
2 year	32	0.82 (0.34)	0.74 (0.43)	36	0.87 (0.22)	0.79 (0.28)	-0.05 (0.08)	-0.05 (0.09)	0.00 (0.120)
Within group baseline differences between value sets at 2 months	37	0.12 (0.12)*		48	0.11 (0.10)*				
Within group change over time	26	0.05 (0.13)	0.03 (0.20)	35	0.09 (0.17)	0.13 (0.23)			
Within group differences over time	26	-0.028 (0.09)		35	-0.03 (0.09)**				
Canada (n=89)									
2 months	43	0.82 (0.13)	0.77 (0.20)	35	0.80 (0.20)	0.74 (0.27)	0.02 (0.04)	0.03 (0.05)	0.01 (0.086)
2 year	38	0.84 (0.19)	0.79 (0.27)	31	0.78 (0.21)	0.71 (0.29)	0.06 (0.05)	0.08 (0.07)	0.02 (0.086)
Within group baseline differences between value sets at 2 months	43	0.06 (0.07)*		35	0.06 (0.08)*				
Within group change over time	38	0.02 (0.16)	0.03 (0.23)	30	-0.012 (0.19)	-0.011 (0.28)			
Within group differences over time	38	-0.008 (0.08)		30	-0.001 (0.09)				
Finland (n=63)									
2 months	34	0.71 (0.25)	0.73 (0.29)	28	0.70 (0.26)	0.70 (0.33)	0.007 (0.06)	0.03 (0.08)	0.023 (0.10)
2 year	34	0.73 (0.27)	0.72 (0.33)	26	0.69 (0.21)	0.73 (0.26)	0.04 (0.06)	-0.014 (0.08)	0.054 (0.10)
Within group baseline differences between value sets at 2 months	34	-0.019 (0.13)		28	0.009 (0.15)				
Within group change over time	34	0.02 (0.21)	-0.01 (0.28)	26	-0.007 (0.23)	0.03 (0.27)			
Within group differences over time	34	0.04 (0.17)		26	-0.03 (0.19)				

* $p<0.001$, ** $p<0.05$

5.4.3 Quality-adjusted life years (QALYs)

Table 5.26 presents the utility values at each follow-up when applying the UK value set to the pooled dataset and imputing missing values with multiple imputation. There were no significant differences in EQ-5D utility at any follow-up point. However, QALYs gained (95% parametric CI) were estimated to be 0.112 (0.000 to 0.224) favouring the endovascular intervention using the multiple imputed data which was marginally significant at a 5% significance level. A similar result was achieved when combining patient-level QALY data across the imputed datasets.

Table 5.26: Utility values obtained from the EQ-5D at each follow-up point and using the UK value set for the pooled sample and multiple imputation for missing values.

	Endovascular (n=1073)		Neurosurgery (n=1070)		
EQ-5D	**Dead (n)**	**Mean**	**Dead (n)**	**Mean**	**Mean difference 95% parametric confidence interval**
Utility values					
2 months	75	0.678	88	0.655	0.024 (-0.008 to 0.055)
1 year	85	0.728	105	0.706	0.022 (-0.005 to 0.050)
2 year	88	0.740	115	0.725	0.015 (-0.012 to 0.043)
3 year	94	0.740	131	0.751	-0.011 (-0.038 to 0.017)
4 year	104	0.751	137	0.757	-0.006 (-0.034 to 0.021)
5 year	111	0.752	144	0.765	-0.013 (-0.040 to 0.014)
QALYs*		**2.838**		**2.726**	**0.112 (0.000 to 0.224)**
Combined QALYS from multiple imputation datasets*		**2.838**		**2.726**	**0.112 (0.001 to 0.224)**

*Statistical significance difference at $p<0.05$

5.5 Summary and conclusions

This chapter has provided details of the International Subarachnoid Aneurysm Trial case study. Resource use, unit costs and quality of life data related to endovascular and neurosurgical intervention has been described for the main contributing countries in the trial.Descriptive statistics suggested that particular resource use categories were similar across countries. For example, the time in theatre for the neurosurgical procedure was significantly

higher than the endovascular group in all countries. However, significant differences were found in the management of imaging and investigation and length of stay for the first episode of care across countries. For instance, the number of angiograms performed in the UK was similar between groups but differs between groups in Sweden, Finland and Germany.

Length of stay duration between groups also differs among countries. Participants in the endovascular group in the UK spent fewer days on wards than neurosurgical patients. A similar pattern was not shown in any of the remainder countries. In Canada, patients in the neurosurgical group spent significantly longer time in ITU than endovascular patients with none of these differences exhibited in any of the remainder countries.

The long term follow-up data suggested that patients in the endovascular group were more likely to suffer a hospital readmission due to re-rupture and re-bleedings in the UK and Finland. This result was not observed in the remainder countries but a similar pattern may be expected to occur across countries. However, the limited sample size may have well prevented observing a significant difference. Patients in the endovascular group had significantly more follow-up angiograms than the neurosurgical group and this result was observed in all countries.

A marginal significant difference in terms of QALYs was observed between the two groups at five years follow-up. A simple but illustrative example compared the different EQ-5D value sets and suggested that using a specific value set across countries provides more meaningful results. Quality of life of patients in the endovascular group was significantly better than neurosurgical patients. This estimate was calculated pooling information from all countries.

This chapter has conducted a large number of comparisons between the treatment groups across countries and follow-up period. The influence of the large sample in the UK to detect a significant difference compared to the rest of countries is clear in most of the comparisons performed.

The current chapter has described the ISAT dataset using mainly descriptive statistics on the available sample in each country. However, most sophisticated modelling techniques should be used (as indentified in Chapter 4) to determine whether differences or similarities across countries exist. The next chapter provides an in-depth overview of some of these techniques and explored resource use and outcomes data to identify differences across countries.

Chapter 6: Statistical methods to identify heterogeneity of treatment effect in multinational cost-effectiveness analyses

6.1 Introduction

Heterogeneity is defined in decision making for health economic evaluation as the proportion of the variability between patients that can be explained by differences in patient characteristics (Briggs et al. 2006a). By "characteristic" is meant a quality of the patient known for certain such as age, gender or prior history of disease. In a multinational clinical trial there may be significant differences in the characteristics of patients between the countries they have been recruited from, and these may cause the treatment effect to vary across countries. The presence of heterogeneity of treatment effect among patients from different countries is called an interaction between treatment and countries (Cook et al. 2003). Chapter 4 identified standard regression and multilevel models (MLMs) as potential alternatives to analyse and present the results from economic evaluations alongside multinational randomised trials. In addition it also identified statistical tests for heterogeneity to detect country-by-treatment interactions. The aim of this chapter is to explore these statistical tests and show that both the standard regression and MLMs are also capable of detecting heterogeneity of treatment effect. This is illustrated using data from the ISAT study.

An overview of these approaches is given, and they are then used first to detect whether heterogeneity of treatment effect is present in a set of selected resource use categories or in quality-adjusted life years (QALYs). Then, heterogeneity of treatment effect on costs is investigated using a multicountry costing approach. Finally, heterogeneity in the net benefit statistic across countries is examined. The calculation of net benefits raises separate issues

concerning cost calculations and willingness to pay for health gains, and these are also investigated in this chapter.

The use of structural equations has not been included in this chapter. Willke and colleagues (Willke et al. 1998b) were among the first to suggest the use of regression analysis to estimate country-specific cost effectiveness using data from multinational clinical trials. In their approach, the multinational characteristic of the dataset was represented using dummies for each country and interaction terms with the treatment group in a cost and an outcome equations. This method allows country-specific incremental cost-effectiveness ratios (ICERs) to be calculated, but if endogeneity is present the model would have to be estimated using instrumental variables. Evidence suggests that instruments are not straightforward to find. Moreover, the choice of a poor quality instrument in this type of regression model may distort the estimation process and provide biased results (Bound et al. 1995). As the aim of this chapter is to evaluate methods to detect heterogeneity of treatment effect, the approach by Willke and colleagues is not included in this chapter.

The work by Menzin and colleagues (Menzin et al. 1996) to inform decision makers in a particular country using data from a clinical trial conducted in a different country has not been included in this chapter either. The authors proposed the use of evidence synthesis to conduct an economic evaluation across countries. Although synthesising evidence from different sources is an important current methodological challenge (Ades et al. 2006a; Ades et al. 2006b), it deviates from the original aim of this thesis.

6.2 Statistical methods

6.2.1 Treatment effect on health outcomes, resource use, costs and net benefits

In a randomised controlled trial (RCT) where patients are allocated to two therapies A and B[15], the treatment effect is the difference between the mean effectiveness in group A and the mean effectiveness in group B. A is normally described as the new alternative which can be a new drug, a surgical intervention or a screening programme whereas B is often referred as the comparator which can be a competing alternative, existing clinical practice or doing nothing. This estimate of treatment effect is also called incremental effectiveness between these two therapies in the economic evaluation literature.

In an economic evaluation alongside an RCT, in which cost data are available, the difference between the mean cost per patient in group A and the mean cost per patient in group B is known as the incremental cost between these two therapies. This may also be called the treatment effect on costs. Similarly, the treatment effect on a particular category of resource use is the difference between the mean resource use in group A and the mean resource use in group B.

Chapter 2 introduced two measures to capture and present the results from cost-effectiveness analysis, the ICER and the net benefit approach. It was shown that a linear transformation of the ICER into net benefits can potentially solve some of the ratio statistic problems when dealing with uncertainty on the ICER (Willan and Briggs 2006). However, the use of net benefits is not a complete alternative because of the explicit assumptions on the willingness to pay for health gain needed to construct it. As a result, there is no current gold standard to present the results from cost-effectiveness analysis and analysts normally

[15] Although many RCTs compare more than two alternatives, the two-arm comparison is used here for simplicity.

present only one measure or both. In this chapter, we make use of the net-benefit approach to present the results from the cost effectiveness of endovascular and neurosurgery. Therefore, the treatment effect on net benefits is estimated as the difference in the mean net benefit between the two groups. The treatment effect on net benefits is also known as the incremental net benefit (INB) between the groups.

The following sections describe different methods to identify whether heterogeneity exists across countries in the treatment effect on health outcomes, resource use, costs and net benefits. First, formal tests for heterogeneity will be described, then standard regression analysis is used to detect country-by-treatment interaction, and finally the use of multilevel models to explain variation between countries is introduced.

6.2.2 Formal statistical tests to detect country-by-treatment interactions

Heterogeneity of treatment effect among patients from different countries can be evaluated using formal statistical tests (Cook et al. 2003). Interactions between treatment and countries are normally divided into qualitative (or cross-over interaction) and quantitative (or non-cross-over interaction) (Gail and Simon 1985). In a qualitative interaction, the treatment effect is positive for patients in some countries and negative for those in others. In a quantitative interaction, the magnitude but not the direction of treatment effect varies across countries.

In economic evaluation of health care technologies, the treatment effect is normally defined as the mean difference between the intervention and the control group of a parameter of interest. Defining treatment effect as D_i where $i = 1, \dots, k$ indicates the treatment effect for a set of k countries with mean $\delta_i$and variance σ_i^2 for each country. D_i is assumed to be

independent and normally distributed. σ_i^2 is estimated from the data using the estimated variance S_i^2 which is a consistent estimator of σ_i^2. The estimated treatment effect in each country is then δ_1 for country 1, δ_2 for country 2, and δ_k for country k.

A test for quantitative interaction $(\delta_1 = \delta_2 = \cdots = \delta_k)$ uses the test statistic

$$H = \sum_{i=1}^{k} \frac{(D_i - \overline{D})^2}{S_i^2}$$

where

$$\overline{D} = \frac{\sum_{i=1}^{k} \left[D_i / S_i^2 \right]}{\sum_{i=1}^{k} \left[1 / S_i^2 \right]}$$

is the weighted average of the individual treatment effects. H is compared to the critical values of a chi-square distribution with $k - 1$ degrees of freedom. When H is greater than the critical value, there is evidence of heterogeneity of treatment effects among the k countries.

For qualitative interaction, the null hypothesis that all δ_i are greater than or equal to 0 or less than or equal to 0 is examined against the alternative hypothesis that some of the δ_i are greater than 0 and some are less than 0. The likelihood ratio test and the range test are available to explore qualitative interaction. The likelihood ratio test is defined calculating the quantities:

$$Q^- = \sum_{i=1}^{k} \left(\frac{D_i^2}{S_i^2} \right) I(D_i > 0)$$

$$Q^{+} = \sum_{i=1}^{k} \left(\frac{D_i^2}{S_i^2} \right) I(D_i < 0)$$

where $I(D_i > 0) = 1$ if $D_i > 0$ and 0 otherwise and $I(D_i < 0) = 1$ if $D_i < 0$ and 0 otherwise. The likelihood ratio test is defined by $Q = min(Q^{+}, Q^{-}) > C_{2\alpha}$. The critical values $C_{2\alpha}$ for the likelihood ratio test can be found in the article by Gail and Simon (Gail and Simon 1985). Gail and colleagues produced critical values for significance levels of $\alpha =$ 0.001, 0.025, 0.05, 0.10 and 0.20 and a number of groups ranging from 2 to 30.

The range test for qualitative interaction from Piantadosi and Gail (Piantadosi and Gail 1993) rejects the null hypothesis of no qualitative interaction at level α if both

$$max\left\{ {}^{D_i}/_{S_i} \right\} > C'_{2\alpha} \; and \; min\left\{ {}^{D_i}/_{S_i} \right\} < -C'_{2\alpha}$$

In their article, they also provide the critical values $C'_{2\alpha}$ for significance levels of $\alpha =$ 0.001, 0.025, 0.05, 0.10 and 0.20 and a number of groups ranging from 2 to 30.

It has been shown that the likelihood ratio test has greater power relative to the range test when the number of groups with $D_i > 0$ is similar to the number of groups with $D_i < 0$. However, because this is not known *a priori*, Cook and colleagues recommended conducting both tests (Cook et al. 2003).

The treatment effect for each country (D_i) and associated variances (S_i^2) needs to be estimated separately and independently for each country and will depend on the type of data available. For instance, D_i can be the mean difference in effectiveness or mean difference in

costs, a relative risk or odds ratio, or the mean difference in net benefits between two groups in a randomised controlled trial. If a vector of covariates X is also available, D_i can be estimated using a suitable regression model as presented in the next section.

6.2.3 Standard regression analysis to detect country-by-treatment interactions

Country-by-treatment interactions can be detected using regression analysis that allows the treatment effect to vary across countries. Before formulating a multi-country model, it is very useful to specify first a single-country model, which can be expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + e_i \qquad [6.1]$$

where Y_i is the dependent variable of interest, X is a explanatory variable, T_i is a treatment indicator that, in this instance, takes value 0 if the patient belongs to the endovascular groups and 1 if the patient belongs to the neurosurgical group, and e_i is the usual stochastic error term with $i = 1, \dots, n$ indicating the individual patient observation. The parameter β_2 indicates the treatment effect, i.e. the mean difference in the variable of interest between the two groups after controlling the variable X.

The model in [6.1] can be expanded to the multivariate case, in which X is a vector of patient-specific variables instead of a single one. In this case the interpretation of β_2 is the mean difference in the variable of interest between the two groups after controlling for all variables included in the vector X.

When data from a multinational randomised controlled trial are available, the single-country model can be extended to allow the dependent variable to vary by country. This is achieved

by including a set of dummy variables to indicate the country the patient was recruited from. A model with k countries and n patients is defined then as:

$$Y_{ij} = \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_{3j} N_{ij} + e_i \qquad [6.2]$$

where N_{ij} is a vector of dummy variables indicating the countries, with the individual patient i specified as $i = 1, \dots, n$ and the country specified as j with $j = 1, \dots, k-1$.

To be able to estimate the intercept in [6.2] a reference country needs to be selected. In this case the estimated β_{3j} is the mean difference in the dependent variable between country j and the reference country category. The reference category is normally defined as the country with the largest number of observations in the sample. In this model the interpretation of β_2 indicates the mean overall treatment effect across all countries, which is constant across all settings.

Country-by-treatment interactions are included through interaction terms, extending the model [6.2] to:

$$Y_{ij} = \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_{3j} N_{ij} + \beta_{4j} T_i N_{ij} + e_i \qquad [6.3]$$

where $T_i N_{ij}$ is a vector of interaction terms that allows the treatment effect β_{4j} to vary across countries. In this case, β_2 is interpreted as the treatment effect in the reference category.

6.2.3.1 Test for heterogeneity

Country-by-treatment interactions can be evaluated testing whether all β_{4j} are zero. This is achieved using a joint F-test to determine that $\beta_{4j} = 0$. Rejecting the null hypothesis suggests there is evidence of heterogeneity of treatment effect across countries.

6.2.3.2 Model selection and regression diagnostics

Selecting an appropriate model to estimate equation [6.3] is vital for analysts dealing with this type of data. The model selected should ensure unbiased estimates of the coefficients and accurate estimation of the uncertainty around them. Excellent reviews are available in the literature that not only indentify the methods used by analysts to estimate these models but provide recommendations on the most appropriate method (Billingham et al. 1999; Mihaylova et al. 2010). In this chapter we will make use of these recommendations to select the estimation models in the next section.

Besides the selection of the estimation model, each model should be evaluated using regression diagnostics to ensure both the coefficients and the associated standard errors are accurately estimated. Particular attention should be paid to model-misspecification and heteroscedasticity. Specification errors occur when we omit relevant variables, include unnecessary covariates or adopt a wrong functional form. The consequences affect directly the estimation of coefficients and standard errors which are no longer unbiased and efficient (minimum variance). Heteroscedasticity occurs when the variance associated with each individual error term is not constant. When a model with heteroscedasticity is estimated, the associated standard errors of the coefficients are no longer efficient. Miss-specification can be evaluated formally using the link test approach (Pregibon 1980) whereas heteroscedasticity can be corrected using robust variance estimation (Manning 1998).

6.2.4 Multilevel models to model variation of treatment effect across countries

The following description of multilevel models follows the terminology used by the Centre for Multilevel Modelling at the University of Bristol in their manual for the statistical package MLwiN (Rasbash et al. 2009).

6.2.4.1 Single or multicountry multilevel models

Multilevel models (MLM) recognise that the dataset under evaluation follows a natural hierarchy or structure. In the case of a multinational randomised controlled trial, patients are normally nested within centres and centres nested within countries. This structure is represented in a multilevel model assigning level 1 to patients, level 2 to centres and level 3 to countries. This type of model assumes that units in each of the levels are a random sample from a population of units in each level. Therefore MLM allow the estimation of variation in the underlying population in each level.

Before formulating a general multilevel model, it is very useful to define a single-level model with one explanatory variable first, which is defined as:

$$y_i = a + bx_i + e_i \qquad [6.4]$$

where y_i is the variable under evaluation of interest or response variable, for example the patient-level net benefit statistic of two treatment arms in a randomised controlled trial, a is the intercept, b is the slope of the regression line and x_i the explanatory covariate. In this example we are assuming that there is only one country, such that the subscript i indicates the number of patients in this country. The term $a + bx_i$ is the fixed part of the regression and is also known as the predicted score $\widehat{y}_i$.

The error term or residual e_i is the part of the actual score (y_i) that is not predicted by the fixed part $\hat{y}_i$. It is the difference between the patient's actual score and the predicted score. In this single-country model then, the variation at level 1, i.e. the patients, is just the variance of these e_i.

To extend the single-country model to several countries and therefore to the multilevel case, we need to make the assumption that countries in our study are a random sample of a population of countries. Therefore we can define a regression relation with parallel slopes in each country as:

$$\begin{aligned} \hat{y}_{i1} &= a_1 + bx_{i1} \\ \hat{y}_{i2} &= a_2 + bx_{i2} \\ &\cdots \end{aligned}$$

This can be written as:

$$\hat{y}_{ij} = a_j + bx_{ij} \qquad [6.5]$$

With the full model more generally expressed as:

$$y_{ij} = a_j + bx_{ij} + e_{ij} \qquad [6.6]$$

In this equation, the subscript j takes values from 1 to the number of countries in the sample. To clarify the terminology and for the benefit of the reader, when an item in the regression equation has two subscripts ij, it varies from patient to patient within a country. Where an item has a j subscript only, it varies across countries but has the same value for all the patients within a country. Where an item has no subscript at all, it is constant across all patients and countries.

To simplify the exposition that follows, it is assumed that the structure of the dataset has two levels: level 1 to indicate individual patients and level 2 to indicate countries. In a multilevel analysis with a two-level hierarchy, the level 2, i.e. countries, are considered a random sample from a population of countries and a_j in equation [6.5] is expanded to $a_j = a + u_j$ yielding then

$$\hat{y}_{ij} = a + bx_{ij} + u_j \qquad [6.7]$$

Where a is a constant and u_j is the departure of the jth country intercept from the overall value a. In equation [6.7] u_j is the level 2 residual which is the same for all patients in country j. The model for actual scores y_{ij} is then:

$$y_{ij} = a + bx_{ij} + u_j + e_{ij} \qquad [6.8]$$

Where u_j and e_{ij} are random quantities with means equal to zero and form the random part of [6.8]. In this simple model, u_j and e_{ij} are uncorrelated and they follow a normal distribution with variances σ_u^2 and σ_e^2 respectively. Model [6.8] then has four items to be estimated, σ_u^2, σ_e^2, a and b. The mean intercept a and the slope b are known as the fixed parameters in this model. It is the presence of the two random variables σ_u^2 and σ_e^2 in equation [6.8] that defines this model as multilevel. When Equation [6.8] does not include any explanatory covariates, it is also called a variance component model because the overall residual variance is partitioned into the two components referring to each level of the hierarchy.

6.2.4.2 The random intercept model

To specify a more general model we adapt [6.8] to general notation. It is useful though to introduce first a special explanatory covariate x_0 which takes value 1 for all patients in all

countries. This covariate allows all terms in [6.8] to be associated with an explanatory variable. We also introduce $\beta_0, \beta_1,$ and so on for the fixed parameters to be consistent with the terminology used in the regression analysis section presented earlier in this chapter. In addition, they also match the subscript of the explanatory variable to which they are attached. Finally we incorporate the subscript 0 into the random variables and get:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + u_{0j} x_0 + e_{0ij} x_0 \qquad [6.9]$$

And this model is finally arranged as:

$$\begin{aligned} y_{ij} &= \beta_{0ij} x_{0ij} + \beta_1 x_{1ij} \\ \beta_{0ij} &= \beta_0 + u_{0j} + e_{0ij} \end{aligned} \qquad [6.10]$$

The coefficient associated with x_{0ij}, β_{0ij}, is random at both level 1 and level 2, i.e. patients and countries. This model is also known as a random intercept model.

The standard model assumes the response variable is normally distributed:

$$y \sim N(XB, \Omega)$$

where XB is a vector containing the fixed part of the model and Ω is a matrix of variances and covariances of the random terms over all the levels of the data. With the simple model we have just introduced, it indicates variances at level 1 and level 2.

Equation [6.10] can include patient level and country-specific data in the same model. This is an advantage of MLM compared with single-level regression analysis where country-level variables cannot be estimated separately. In a single-level model any country-level covariate can be expressed as a linear combination of the dummies identifying the country causing estimation problems.

Equation [6.10] allows the estimation of country-specific results through the use of shrinkage estimates and the calculation of residuals in both levels, which is more complex than the calculation of residuals in single-level models. This is described in Appendix 7.

The relative size of σ_{e0}^2 and σ_{u0}^2 provides a measure of the degree of dependency in a hierarchical dataset. This dependency occurs because patients from the same country are likely to be more similar than those from different countries and this is measured by the variance partitioning coefficient (VPC) or intraclass correlation, which is defined as:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2}$$

The intraclass correlation coefficient represents as the proportion of the total variance that is attributable to the level 2 variance. The VPC is always less than or equal to 1. The VPC is close to 1 when σ_{u0}^2 is large compared to σ_{e0}^2 and is close to 0 when σ_{u0}^2 is small compared to σ_{e0}^2.

6.2.4.3 The random slope model

Equation [6.10] described above assumes that the sole variation between countries occurs in the intercepts. A natural extension to this model is to allow the slope β_1 to vary across countries. This was shown in equation [6.3] using different interaction terms between the dummy treatment and the dummies representing each country. However, we can specify β_1 to randomly vary across countries, and hence at level 2. This can be formally written as:

$$\begin{aligned} y_{ij} &= \beta_{0j}x_{0ij} + \beta_{1j}x_{1ij} + e_{0ij}x_{0ij} \\ \beta_{0j} &= \beta_0 + u_{0j} \\ \beta_{1j} &= \beta_1 + u_{1j} \end{aligned} \qquad [6.11]$$

In model [6.11], both the intercept and the slope vary randomly across countries. Therefore β_0 and β_1 have a j subscript in the first line. The second line suggests that the intercept for the jth country (β_{0j}) equals β_0, the average intercept across all countries, plus a random difference u_{0j}. Similarly, the third line indicates that the slope for the jth country (β_{1j}) is given by β_1, the average slope across all countries, plus a random difference u_{1j}. The parameters β_0 and β_1 are the fixed part intercept and slope coefficients in this model. The terms u_{0j} and u_{1j} are random departures from β_0 and β_1, and are also called the residuals at the country level. These residuals allow the line of best fit for the jth country to differ from the average line in both its slope and its intercept. Equation [6.11] is also known as a random slope model.

The terms u_{0j} and u_{1j} in equation [6.11] are assumed to follow a bivariate normal distribution with mean vector 0 and covariance matrix Ω_u. Because we have two random variables at level 2, Ω_u is a 2 by 2 symmetric covariance matrix whose elements are:

$$\Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix}$$

Where $\sigma_{u0}^2 = var(u_{0j})$ is the variation in the intercept across the country's summary lines, $\sigma_{u1}^2 = var(u_{1j})$ is the variation in the slope across the country's summary lines, and $\sigma_{u01} = cov(u_{0j,}u_{1j,})$ is the covariance between the country's intercept and slopes. e_{0ij} is the departure of each patient's score from the line of best fit for that patient's country. As in equation [6.10], e_{0ij} is assumed to be normally distributed with mean 0 and variance σ_e^2, and is uncorrelated to Ω_u.

In this random slope model the total variance of y_{ij} conditional on x_{1ij} is:

$$var(y_{ij}|x_{ij}) = \sigma_{u0}^2 + 2\sigma_{u01}x_{ij} + \sigma_{u1}^2 x_{ij}^2 + \sigma_e^2$$

Since x_{1ij} = 0, 1 for all i and j the variances of y_{ij} in each treatment arm are

$var(y_{ij}|x_{ij} = 0) = \sigma_{u0}^2 + \sigma_e^2$ and $var(y_{ij}|x_{ij} = 1) = \sigma_{u0}^2 + 2\sigma_{u01} + \sigma_{u1}^2 + \sigma_e^2$

And hence a variance partitioning coefficient can be defined for each arm in the trial as follows:

$$\rho\,(x_{ij} = 0) = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

$$\rho\,(x_{ij} = 1) = \frac{\sigma_{u0}^2 + 2\sigma_{u01} + \sigma_{u1}^2}{\sigma_{u0}^2 + 2\sigma_{u01} + \sigma_{u1}^2 + \sigma_e^2}$$

6.2.4.4 Test for heterogeneity

A test for heterogeneity in a multilevel framework can be constructed identifying whether there is evidence of significant variation of treatment effect across countries. This is conducted therefore testing the null hypothesis that $H_0: \sigma_{u1}^2 = 0$, which is similar to test that all β_{4j} are zero in the standard regression framework or the formal statistical test described above. However, evaluating the significance of variances is not a straightforward routine because the estimated variances $\sigma_{u0}^2 = var(u_{0j})$ and $\sigma_{u1}^2 = var(u_{1j})$ are unlikely to follow a normal distribution. An alternative method is to use the likelihood ratio test (Woodward 2005). In the likelihood ratio test we compare the model [6.11] with model [6.10] where σ_{u1}^2 is constrained to zero. For each model the statistic $-2logL$ is calculated where L is the likelihood function and represents the probability of estimating the observed data if the model were true. Defining L_v as the likelihood for a model with v terms and L_{v+u} as the

likelihood for a model with the same v terms plus additional u terms, we can formulate the null hypothesis that the vector of betas for these extra u terms are zero using the statistic:

$$\Delta = -2logL_v - (-2logL_{v+u})$$

that follows a chi-square distribution with u degrees of freedom. In our case u is equal to two because the model [6.11] has the additional parameters σ_{u1}^2 and σ_{u01} to estimate compared to model [6.10].

6.2.5 Presenting treatment effect results

The tests for heterogeneity described in this chapter provide an indication of the variability of the treatment effect across countries. Once heterogeneity has been determined, the next step is to decide which estimates of treatment effect of costs and QALYs should be formally reported. Both the standard regression and the multilevel framework can calculate estimates of treatment effect in each country. The debate on which method to use has received considerable attention in the health economics community but to date no consensus has been reached (Grieve et al. 2007; Manca et al. 2005b; Manca and Willan 2006). In this chapter the treatment effect from both approaches are presented in costs and QALYs.

However, a useful approach is to select the model that provides a better fit to the data, and this can be achieved using information criterion methods. In this chapter the Akaike information criterion (AIC) (Akaike 1974) is used to indicate which regression method better fits the ISAT data. AIC is easy to construct and interpret: the model with the lowest AIC is the one that provides the best fit. AIC is defined as:

$$AIC = -2 * LogLikelihood + 2k$$

where the log likelihood can be obtained directly from the regression output and k is the number of parameters estimated in the model. AIC allows the comparison of different models with different error distribution using the same dataset.

6.2.6 Software implementation

All the methods described in this chapter can be implemented using well-known statistical software. The formal statistical tests for heterogeneity were constructed in Microsoft Excel 2007 using the formulae presented above. Standard regression analyses were conducted in STATA version 11 (StataCorp, Tx, USA), whereas the multilevel models were implemented in the statistical package MLwiN (MLwiN, Centre for Multilevel Modelling, Bristol) which provides a user-friendly interface to conduct this type of analysis.

6.3 Application of the statistical methods: an example using the ISAT dataset

6.3.1 Summary descriptive from the ISAT dataset

As described in Chapter 5, interactions may exist between treatment effect and country for several resource use items and health outcomes. It was shown that mean duration in theatre was consistently higher in the neurosurgical intervention in all countries. In addition, the mean number of additional imaging procedures and investigations, mean length of stay and mean number of readmissions, varied across countries. Quality of life using the EQ-5D favoured endovascular patients, with variation in the magnitude and direction of treatment effect in Germany. Therefore country-by-treatment interactions need to be formally evaluated to confirm the heterogeneity of treatment effect suggested in the descriptive analysis.

Using the methods described in the previous sections, heterogeneity of treatment effect is examined by looking at different categories of resource use, health outcomes, costs and net benefits. Five categories of resource use are explored: duration in the operating theatre (in minutes), number of angiograms associated with the first episode of care, length of stay at a ward facility (in days), length of stay at an ITU facility (in days) and the number of readmissions at five years follow-up. We then evaluate between country heterogeneity for QALYs over the five-year follow-up as the outcome measure. Next, any treatment effect on the total cost per patient at five years follow-up is examined using a multicountry costing method. Finally, treatment effect on net benefits is investigated using different willingness to pay for health gain. The estimation of net benefits assumes a particular willingness to pay for health gain, which may well vary across settings. However, the analysis in this chapter does not attempt to have different willingness to pay in each country. Instead, the same

willingness to pay across countries is assumed, but the impact of using different thresholds of willingness to pay levels to calculate net-benefits is also illustrated.

Tables 6.1-6.3 provide sample statistics in each randomisation group, using data from all countries combined and from each individual country. Patient-level data from all countries include information from the eleven countries participating in ISAT. However, observations from countries with a sample size of less than 20 in each arm were excluded from the analysis when using data from each individual country[16] leaving five countries in this analysis. The tables report sample size, mean and standard deviation (SD) in each group, mean and standard error of the difference and associated 95% confidence interval (95% CI). These estimates of mean differences are called unadjusted means because they are calculated using only the sample size available for each country.

Table 6.1 provides sample statistics for selected resource use categories. Patients in the neurosurgical intervention spent significantly more time in theatre than patients in the endovascular group in all countries and therefore no qualitative difference across settings seems to exist. However, the magnitude of this difference was interestingly larger in Canada compared to the remaining countries; these patients in the neurosurgical group spent an additional 114 minutes in theatre compared to endovascular patients. The following sections formally test the presence of a quantitative interaction in the duration of the operation across countries.

The number of angiograms associated with the first episode of care was similar between the two intervention groups for all countries except in Finland. In this country, the mean (SD) number of angiograms was 1.15 (0.36) in the endovascular group and 2.03 (0.42) in the

[16] A similar threshold was used in Chapter 5 to describe the ISAT dataset.

neurosurgical arm, a significant mean difference and 95% confidence interval of 0.89 (1.08 to 0.69) additional angiograms in the neurosurgical arm[17].

Table 6.1: Sample statistics for the selected resource use using all countries and participating countries data

	Endovascular		Neurosurgery			
	n	Mean resource use per patient (SD)	n	Mean resource use per patient (SD)	SE (mean difference)	Mean difference (95% confidence interval)
All countries						
Duration (in minutes)	1063	138.57 (62.78)	1060	207.04 (86.99)	3.29	-68.47 (-74.92 to -62.01)*
Number of angiograms	1071	1.22 (0.60)	1068	1.21 (0.52)	0.01	0.01 (-0.03 to 0.06)
Number of ITU days	1073	3.58 (6.75)	1067	4.24 (7.18)	0.30	-0.66 (-1.25 to -0.07)**
Number of ward days	1072	13.73 (13.50)	1067	15.91 (15.03)	0.62	-2.18 (-3.39 to -0.97)*
Number of readmissions	1073	0.35 (0.80)	1070	0.24 (0.64)	0.03	0.11 (0.05 to 0.17)*
United Kingdom						
Duration (in minutes)	803	136.91 (57.87)	829	204.09 (84.83)	3.60	-67.18 (-74.25 to -60.11)*
Number of angiograms	809	1.13 (0.44)	835	1.11 (0.40)	0.02	0.02 (-0.02 to 0.06)
Number of ITU days	809	1.95 (4.40)	835	2.41 (4.78)	0.23	-0.45 (-0.90 to -0.02)**
Number of ward days	809	14.80 (11.92)	835	17.19 (15.32)	0.68	-2.38 (-3.72 to -1.06)*
Number of readmissions	809	0.37 (0.77)	835	0.25 (0.66)	0.04	0.10 (0.03 to 0.17)**
Sweden						
Duration (in minutes)	64	135.47 (87.30)	59	217.46 (78.20)	14.99	-82 (-111.67 to -52.31)*
Number of angiograms	64	1.28 (0.58)	59	1.39 (0.67)	0.11	-0.11 (-0.33 to 0.11)
Number of ITU days	64	8.55 (7.46)	59	10 (7.38)	1.34	-1.45 (-4.10 to 1.20)
Number of ward days	64	8.72 (6.65)	59	8.39 (6.27)	1.17	0.33 (-1.98 to 2.64)
Number of readmissions	64	0.30 (1.11)	59	0.41 (0.87)	0.18	-0.11 (-0.47 to 0.24)
Germany						
Duration (in minutes)	59	154.66 (94.50)	55	223.47 (131.72)	21.36	-68.81 (-111.14 to -26.48)**
Number of angiograms	62	1.58 (1.08)	55	1.31 (0.64)	0.17	0.27 (-0.06 to 0.60)
Number of ITU days	62	13.81 (12.33)	56	17.63 (10.59)	2.13	-3.82 (-8.03 to 0.40)
Number of ward days	62	7.26 (8.80)	56	9.86 (11.36)	1.86	-2.60 (-6.28 to 1.09)
Number of readmissions	62	0.12 (0.37)	56	0.04 (0.19)	0.06	0.07 (-0.31 to 0.18)
Canada						
Duration (in minutes)	48	151.04 (59.35)	39	264.74 (73.13)	14.20	-113.70 (-141.93 to -85.47)*
Number of angiograms	48	2.06 (0.93)	41	1.95 (0.55)	0.17	0.11 (-0.22 to 0.44)
Number of ITU days	48	6.58 (5.61)	40	10.8 (12.25)	1.98	-4.22 (-8.15 to -0.29)**
Number of ward days	48	11.81 (8.85)	40	15.05 (24.24)	3.37	-3.24 (-9.93 to 3.46)
Number of readmissions	48	0.10 (0.37)	41	0.02 (0.16)	0.06	0.08 (-0.04 to 0.20)
Finland						
Duration (in minutes)	34	128.68 (57.67)	29	172.59 (49.69)	13.69	-43.91 (-71.28 to -16.54)**
Number of angiograms	34	1.15 (0.36)	29	2.03 (0.42)	0.10	-0.89 (-1.08 to -0.69)*
Number of ITU days	34	1.5 (3.53)	29	2.17 (2.62)	0.80	-0.67 (-2.26 to 0.92)
Number of ward days	34	10.77 (5.50)	29	11 (4)	1.23	-0.24 (-2.70 to 2.23)
Number of readmissions	34	1.03 (1.51)	29	0.21 (0.49)	0.29	0.82 (0.24 to 1.41)*

*p<0.001; **p<0.05

[17] The mean difference and confidence intervals has been multiplied by -1 to facilitate the interpretation of the results when the mean difference between the endovascular and neurosurgical group was negative.

Patients randomised to the endovascular group spent significantly fewer days in an intensive therapy unit (ITU) than the neurosurgery group in the UK and Canada. In the case of Canada the mean difference (95% CI) was estimated to be 4.22 (0.29 to 8.15) days whereas in the UK, the mean difference (95% CI) was estimated to be 0.45 (0.02 to 0.90) days. Patients in the endovascular group in the UK spent significantly fewer days at a ward facility than patients in the neurosurgical group, with a mean difference and 95% CI of 2.38 (1.06 to 3.72) days. In the remaining countries, the mean difference between groups varied in both directions but none of the estimates reached statistical significance.

Finally, patients in the endovascular group experienced significantly more readmissions at 5 years follow-up than neurosurgical patients in the UK and Finland. In the UK the mean difference (95% CI) was estimated to be 0.10 (0.03 to 0.17) whereas in Finland the estimated mean difference (95% CI) was estimated to be 0.82 (0.24 to 1.41).

Table 6.2 reports identical information to Table 6.1 for quality-adjusted life years and the total cost per patient at five years follow-up using country-specific unit costs. Using the overall sample and data from the eleven countries participating in ISAT, the mean (SD) number of QALYs at 5 years follow-up was estimated to be 2.84 (1.32) and 2.73 (1.31) in the endovascular and neurosurgical group respectively, a significant mean difference (95% CI) of 0.11 (0.01 to 0.22) favouring the endovascular group. Using data from the UK only, the mean (SD) number of QALYs was estimated to be 2.86 (1.29) in the endovascular and 2.70 (1.32) in the neurosurgery respectively, a significant mean difference (95% CI) favouring endovascular patients of 0.16 (0.034 to 0.287). Using data from Germany alone the mean (SD) number of QALYs was estimated to be 2.17 (1.68) in the endovascular group and 2.98 (1.21) in the neurosurgical group, a significant mean difference (95% CI) of -0.81 (-1.34 to -

0.26) favouring neurosurgical patients. No further significant differences in QALYs were detected in any of the remaining countries.

Table 6.2: Sample statistics for quality-adjusted years (QALYs) and total cost per patient at five years follow-up using all countries and participating countries data

	Endovascular		Neurosurgery			
	n	**Mean resource use per patient (SD)**	**n**	**Mean resource use per patient (SD)**	**SE (mean difference)**	**Mean difference (95% confidence interval)**
All countries						
QALYs	1073	2.84 (1.32)	1070	2.73 (1.31)	0.06	0.11 (0.001 to 0.22)*
Total cost per patient using country-specific unit costs	973	38,565 (32,631)	981	40,450 (42,500)	1,715	-1,885 (-5,249 to 1,479)
United Kingdom						
QALYs	809	2.86 (1.29)	835	2.70 (1.32)	0.06	0.16 (0.034 to 0.287)*
Total cost per patient using UK unit costs	781	37,624 (32,987)	808	39,026 (41,427)	1,883	-1,401 (-5,094 to 2,290)
Sweden						
QALYs	64	3.04 (1.38)	59	2.87 (1.34)	0.25	0.17 (-0.319 to 0.652)
Total cost per patient using Swedish unit costs	64	43,218 (50,387)	58	50,387 (36,090)	5,760	-7,168 (-18,572 to 4,236)
Germany						
QALYs	62	2.17 (1.68)	56	2.98 (1.21)	0.27	-0.81 (-1.34 to -0.26)*
Total cost per patient using German unit costs	52	47,043 (39,913)	50	47,598 (32,440)	7218	-555 (-14875 to 13,766)
Canada						
QALYs	48	2.98 (1.25)	41	2.66 (1.39)	0.28	0.32 (-0.24 to 0.87)
Total cost per patient using Canadian unit costs	46	44,494 (28,077)	38	60,045 (77,609)	12,304	-15,551 (-40,027 to 8,926)
Finland						
QALYs	34	3.01 (1.13)	29	3.14 (1.09)	0.28	-0.13 (-0.69 to 0.438)
Total cost per patient using Finnish unit costs	30	29,358 (20,898)	27	20,919 (8,200)	4,295	8,439 (-167 to 17,046)

*p<0.001; All costs expressed in $US Purchasing Power Parity from 2008

Chapter 5 described the way in which country-specific unit costs collected as part of the ISAT study were expressed in 2008 UD dollars using GDP purchasing power parities (PPPs) provided by the OECD website to eliminate any price differences and ensure comparability (Eurostat-OECD 2006). The mean (SD) total cost per patient at 5 years follow-up using country-specific unit costs was estimated to be $38,565 ($32,631)) in the endovascular group

and $40,450 ($42,500) in the neurosurgical group, a mean difference (95% CI) of -$1,885 (-$5,249 to $1,479).

Similar non-significant results were obtained in the total cost per patient at five years follow-up using country-specific unit costs and data from each individual country. However, the magnitude and direction of the mean cost difference varied across settings, and so the next section provides formal tests to determine whether treatment effect interaction exists.

Table 6.3 reports similar information to Tables 6.1 and 6.2 for net benefits in the monetary scale at five years follow-up using the same willingness to pay for health gain (λ) across all countries for different thresholds of λ. Pooling data from all countries and using country-specific unit costs to calculate net benefits, endovascular therapy has a significantly greater net benefit that neurosurgery for values of willingness to pay above $15,000 per QALY gained. Note that in all countries when λ=0, the incremental net benefit is the same as the difference in costs between neurosurgery and endovascular.

Using data from the UK alone and a willingness to pay of λ=$40,000 per QALY gained, the mean (SD) net benefit was estimated to be $76,773 ($69,131) and $69,070 ($74,058) in the endovascular and neurosurgical group respectively, a significant mean incremental net benefit (95% CI) of $7,703 ($648 to $14,758) favouring the endovascular intervention. This mean difference increases to $14,005 with associated 95% CI ($2,214 to $25,795) when λ=$80,000.

Table 6.3: Sample statistics for net monetary benefits using country-specific unit costs and different thresholds of willingness to pay (λ) in all countries and participating countries data

	Endovascular		Neurosurgery			
	n	**Mean resource use per patient (SD)**	**N**	**Mean resource use per patient (SD)**	**SE (mean difference)**	**Mean difference (95% confidence interval)**
All countries						
Net benefit (λ = 0)	973	-38,565 (32,631)	981	-40,450 (42,501)	1,715	1,885 (-1,479 to 5,249)
Net benefit (λ = 15,000)	973	4,056 (42,905)	981	712 (50,751)	2,127	3,343 (-828 to 7,515)
Net benefit (λ = 40,000)	973	75,090 (70,075)	981	69,316 (74,667)	3,277	5,774 (-652 to 12,200)
Net benefit (λ = 80,000)	973	188,745 (120,167)	981	179,082 (121,815)	5,474	9,663 (-1,073 to 20,400)
United Kingdom						
Net benefit (λ = 0)	781	-37,624 (32,987)	808	-39,026 (41,427)	1,883	1,402 (-2,291 to 5,095)
Net benefit (λ = 15,000)	781	5,275 (42,895)	808	1,510 (49,771)	2,334	3,765 (-814 to 8,344)
Net benefit (λ = 40,000)	781	76,773 (69,131)	808	69,070 (74,058)	3,597	7,703 (648 to 14,758)*
Net benefit (λ = 80,000)	781	191,169 (117,753)	808	177,165 (121,733)	6,011	14,005 (2,214 to 25,795)*
Sweden						
Net benefit (λ = 0)	64	-43,218 (27,280)	58	-50,387 (36,090)	5,760	7,168 (-4,236 to 18,575)
Net benefit (λ = 15,000)	64	2,369 (41,826)	58	-7,203 (48,144)	8,147	9,572 (-6,558 to 25,702)
Net benefit (λ = 40,000)	64	78,347 (72,914)	58	64,769 (76,615)	13,524	13,579 (-13,198 to 40,355)
Net benefit (λ = 80,000)	64	199,912 (126,263)	58	179,923 (127,394)	22,988	19,989 (-25,526 to 65,503)
Germany						
Net benefit (λ = 0)	52	-47,043 (39,913)	50	-47,598 (32,440)	7,218	555 (-13,766 to 14,875)
Net benefit (λ = 15,000)	52	-15,976 (48,183)	50	-3,557 (36,498)	8,489	-12,419 (-29,261 to 4,422)
Net benefit (λ = 40,000)	52	35,802 (79,956)	50	69,845 (58,052)	13,881	-34,043 (-61,583 to -6,503)*
Net benefit (λ = 80,000)	52	118,647 (142,603)	50	187,286 (103,523)	24,757	-68,641 (-117,757 to -19,525)*
Canada						
Net benefit (λ = 0)	46	-44,494 (28,077)	38	-60,045 (77,609)	12,304	15,551 (-8,926 to 40,028)
Net benefit (λ = 15,000)	46	-336 (37,575)	38	-17,016 (86,926)	14,180	16,680 (-11,529 to 44,888)
Net benefit (λ = 40,000)	46	73,261 (63,552)	38	54,700 (107,784)	18,932	18,561 (-19,100 to 56,223)
Net benefit (λ = 80,000)	46	191,017 (111,220)	38	169,445 (148,431)	28,354	21,572 (-34,833 to 77,977)
Finland						
Net benefit (λ = 0)	30	-29,358 (20,898)	27	-20,919 (20,898)	4,294	-8,439 (-17,046 to 167)
Net benefit (λ = 15,000)	30	17,365 (34,020)	27	26,694 (22,320)	7,715	-9,329 (-24,790 to 6,130)
Net benefit (λ = 40,000)	30	95,235 (60,748)	27	106,048 (49,595)	14,790	-10,813 (-40,453 to 18,828)
Net benefit (λ = 80,000)	30	219,829 (105,860)	27	233,015 (93,972)	26,638	-13,186 (-66,570 to 40,197)

*p<0.01; **p<0.05; net benefits and λ expressed in $US Purchasing Power Parity from 2008

Assuming λ=$40,000, patients in Germany where the treatment effect in QALYs appeared to be in the opposite direction to other countries (Table 6.2), yielded a mean (SD) net benefit of $35,802 ($79,956) in the endovascular group and $69,845 ($58,052) in the neurosurgical arm, a mean incremental net benefit (95%CI) of -$34,043 (-$61,583 to -$6,503) favouring the neurosurgical intervention. The mean (95% CI) difference rises to -$68,641 (-$117,757 to -$19,525) when λ=$80,000. However, this needs further investigation as this result uses data

from Germany only, and caution needs to be exercised because a different result was achieved in the United Kingdom with a much larger sample. In the remaining countries there was no clear evidence of a particular intervention being more cost-effective. The next section will help to clarify this issue.

The results presented in Tables 6.1-6.3 show the influence of the large UK sample size in the dataset in the calculation of confidence intervals. The standard errors associated with all the mean difference comparisons conducted were smaller for the UK, indicating more precise estimates in that country than in Sweden, Germany, Canada and Finland[18].

6.3.2 Formal statistical tests for heterogeneity

We will start our exploration of heterogeneity of treatment effect across countries with the formal tests described in Section 6.2.2.

To conduct a formal test for heterogeneity, estimates of treatment effect between two alternatives are needed. These estimates can be provided by sample statistics or by the output from a regression analysis. In this chapter, estimates of treatment effect to conduct the formal tests of heterogeneity were taken from the sample statistics analysis in Tables 5.1-5.3. A significance level of 1% was chosen due to the large number of comparisons made and to minimise significant results due to chance.

Table 6.4 presents the results of the formal quantitative and qualitative tests to detect interaction across countries on different types of resource use. The results of the quantitative

[18] The standard error of the difference between two means is defined as $SE_{diff}=\sqrt{\frac{SD_1^2}{n_1}+\frac{SD_2^2}{n_2}}$, where SD_1 is the standard deviation of group 1, SD_2 the standard deviation of group 2, n_1 the sample size in group 1 and n_2 the sample size in group 2. Smaller values of n indicate a less precise mean difference.

interaction test suggest heterogeneity of treatment effect in the number of angiograms across countries (H = 83.71, p<0.001). This result is driven by the additional number of angiograms in the neurosurgical group in Finland compared to the remaining countries. It is also shown that treatment effect differs across countries in the number of readmissions at five years follow-up (H=3,308, p<0.000). Again this result seems to be driven by the higher number of additional readmissions in the endovascular group in Finland compared with other countries. Neither the Gail and Simon and the Piantadosi and Gail tests for qualitative interaction showed cross-over interactions across countries in any of the resource use categories. This result suggests that differences across countries occurred mainly in the magnitude of the treatment effect.

Table 6.4: Quantitative and qualitative tests to detect country-by-treatment interactions on selected resource use

	Operation duration		**Number of angiograms**		**ITU days**		**Ward Days**		**Number of readmissions**	
	Treatment effect	**SE**	**Treatment effect**	**SE**	**Treatment effect**	**SE**	**Treatment effect**	**SE**	**Treatment effect**	**SE**
Countries										
UK	-67.18	3.60	0.02	0.02	-0.45	0.23	-2.38	0.68	0.1	0.04
Sweden	-82	14.99	-0.11	0.11	-1.45	1.34	0.33	1.17	-0.11	0.18
Germany	-68.81	21.36	0.27	0.17	-3.82	2.13	-2.6	1.86	0.07	0.06
Canada	-113.7	14.20	0.11	0.17	-4.22	1.98	-3.24	3.37	0.08	0.06
Finland	-43.91	13.69	-0.89	0.10	-0.67	0.80	-0.24	1.23	0.82	0.29
H	14.28		83.71		6.45		5.78		3,308	
p-value	0.046		0.000		0.489		0.566		0.000	
$Q = min(Q^+, Q^-)$	0		3.94		0		0.08		0.37	
$C_{2\alpha}$	14.95		14.95		14.95		14.95		14.95	
$max\{D_i/S_i\}$	-3.21		1.59		-0.84		0.28		2.83	
$min\{D_i/S_i\}$	-18.66		-8.90		-2.13		-3.50		-0.61	
$C_{2\alpha}$	3.48		3.48		3.48		3.48		3.48	

Table 6.5 shows the results from the quantitative and qualitative tests for QALYs. The test for quantitative interaction suggested heterogeneity of treatment effect for QALYs (H=939.69, p<0.000). This result is likely to be driven by the opposite direction of the

treatment effect in Germany compared to the remaining countries. The qualitative interaction tests failed to show cross-over interaction in the treatment effect on QALYs using a 1% significance level. However, using a 2.5% significance level both the Gail and Simon and the Piantadosi and Gail tests for qualitative interaction do detect a cross-over interaction on treatment effect on QALYs (data not shown).

Table 6.5: Quantitative and qualitative tests to detect country-by-treatment interactions quality-adjusted life years and total cost per patient at 5 years follow-up

	QALYs		**Total cost per patient at 5 years follow-up: country-specific unit costs**	
	Treatment effect	**SE**	**Treatment effect**	**SE**
Countries				
UK	0.16	0.06	-1,401	1,883
Sweden	0.17	0.25	-7,168	5,760
Germany	-0.81	0.27	-555	7,218
Canada	0.32	0.28	-15,551	12,304
Finland	-0.13	0.28	8,439	4,295
H	939.69		7.565	
p-value	0.000		0.372	
$Q = min(Q^+, Q^-)$	8.88		3.71	
$C_{2\alpha}$	14.95		14.95	
$max\{D_i/S_i\}$	2.67		1.96	
$min\{D_i/S_i\}$	-3.00		-1.26	
$C_{2\alpha}$	3.48		3.48	

Table 6.5 shows no significant quantitative or qualitative interaction detected in the total cost per patient at five years follow-up across countries at the 1% level. Finally, Table 6.6 shows the results from the formal tests of heterogeneity for net benefits on the monetary scale at five years follow-up. No significant quantitative interaction was found in net benefits for any value of λ and using a significance level of 1%. The qualitative interaction tests failed to show cross-over interaction in the treatment effect on net benefits using a 1% significance level.

Table 6.6: Quantitative and qualitative tests to detect country-by-treatment interactions of net benefits on the monetary scale using country-specific unit costs and different thresholds of willingness to pay

	Net benefit (λ=0)		Net benefit (λ=15,000)		Net benefit (λ=40,000)		Net benefit (λ=80,000)	
	Treatment effect	**SE**	**Treatment effect**	**SE**	**Treatment effect**	**SE**	**Treatment effect**	**SE**
Countries								
UK	1,402	1,883	3,765	2,334	7,703	3,597	14,005	6,011
Sweden	7,168	5,760	9,572	8,147	13,579	13,524	19,989	22,988
Germany	555	7,218	-12,419	8,489	-34,043	13,881	-68,641	24,757
Canada	15,551	12,304	16,680	14,180	18,561	18,932	21,572	28,354
Finland	-8439	4,294	-9,329	7,715	-10,813	14,790	-13,186	26,638
H	7.570		8.970		13.106		14.697	
p-value	0.372		0.255		0.070		0.040	
$Q = min(Q^+, Q^-)$	3.71		3.60		6.55		6.76	
$C_{2\alpha}$	14.95		14.95		14.95		14.95	
$max\{D_i/S_i\}$	1.26		1.61		2.14		2.33	
$min\{D_i/S_i\}$	-1.97		-1.46		-2.45		-2.77	
$C_{2\alpha}$	3.48		3.48		3.48		3.48	

6.3.3 Standard regression analysis

The second method we will explore to detect heterogeneity is the standard regression analysis approach described in Section 6.2.3. Data from the United Kingdom, Sweden, Germany, Canada and Finland only were used in the regression analysis to be consistent with the previous section and avoid estimations problems with countries with small sample sizes. Therefore, a sample size of 2,037 observations was included to estimate the models.

Figures 6.1-6.8 show the histograms by randomisation group using data from all countries to give an idea of the statistical distribution of the dependent variables included in the regression models. Therefore the figures show histograms for selected resource use, health outcomes, costs and net benefit statistic. The histograms are built using the same number of ticks in the x-axis in both intervention groups to make them visually more comparable. The major tick in the x-axis is determined by a large value for a particular patient in one of the randomisation groups. Exploring the statistical distribution of the dependent variable provides useful information to support the choice of model.

Figure 6.1: Histogram of duration of operation (in minutes) by randomisation group using all countries data

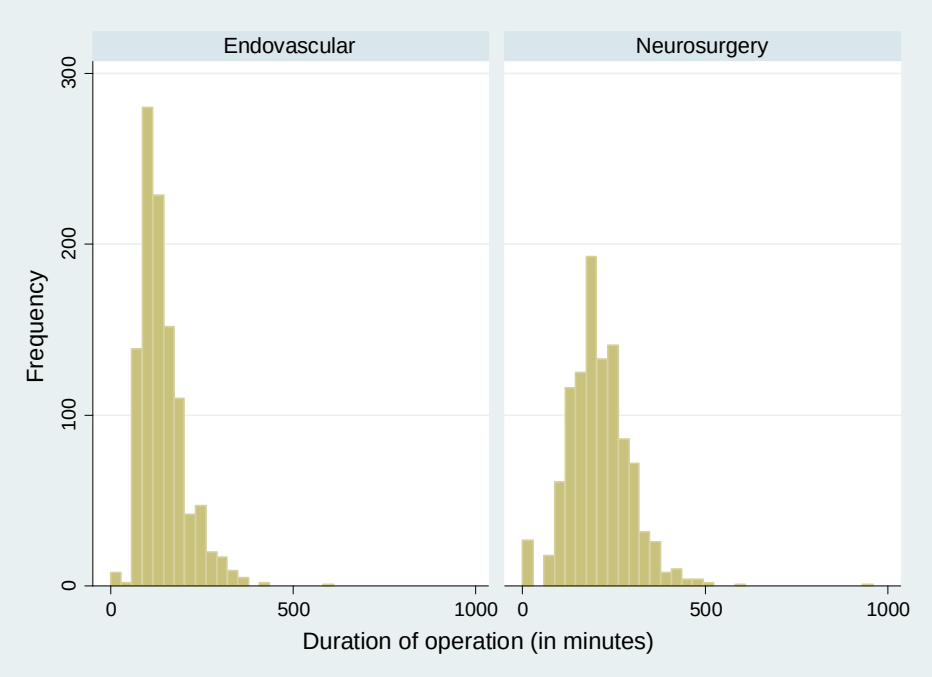


Figure 6.2: Histogram of number of angiograms associated to the first episode of care by randomisation group using all countries data

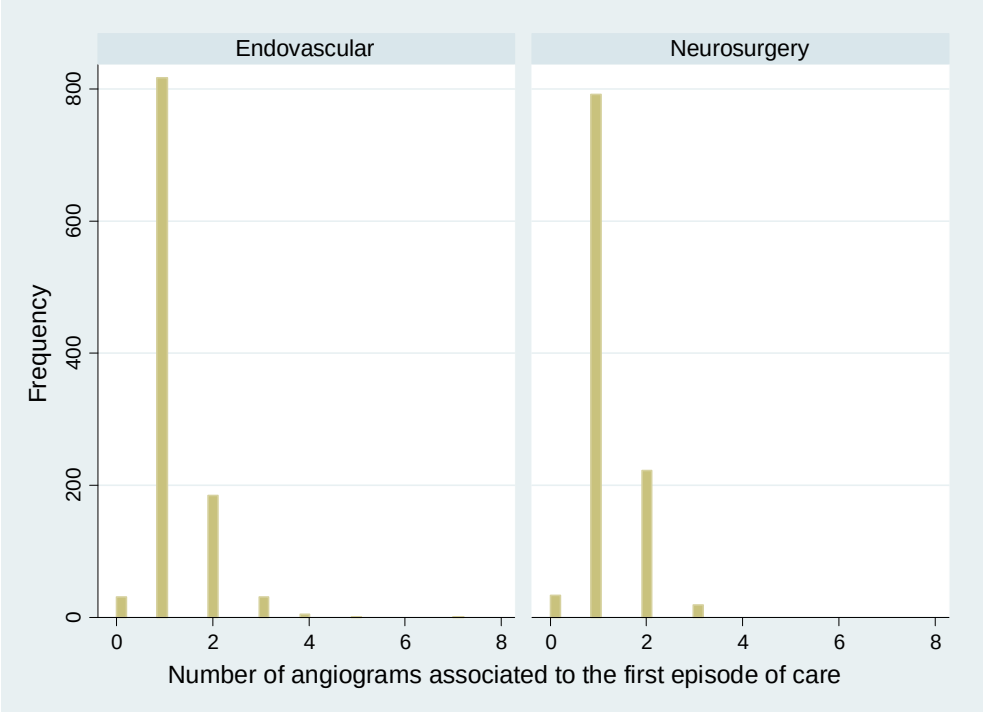


Figure 6.3: Histogram of number of days in intensive therapy unit (ITU) by randomisation group using all countries data

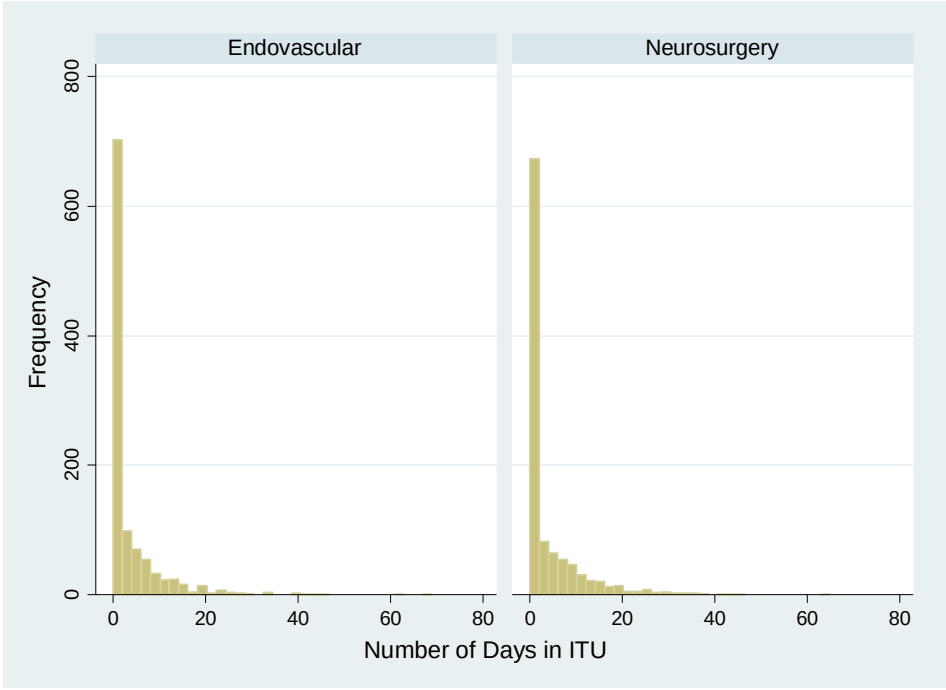


Figure 6.4: Histogram of number of days at ward facility by randomisation group using all countries data

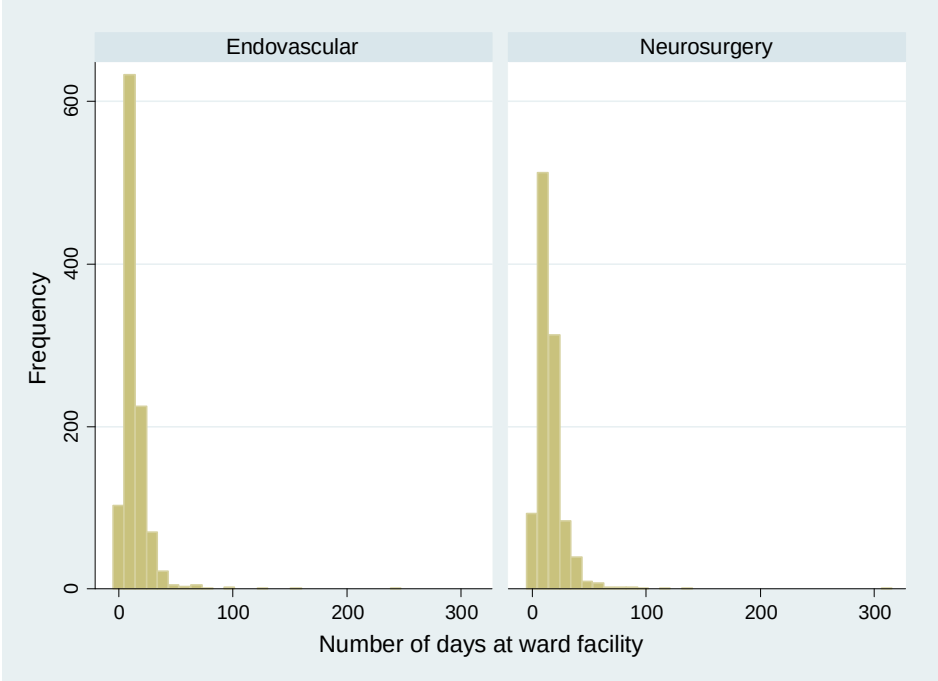


Figure 6.5: Histogram of number of readmissions at 5 years follow-up by randomisation group using all countries data

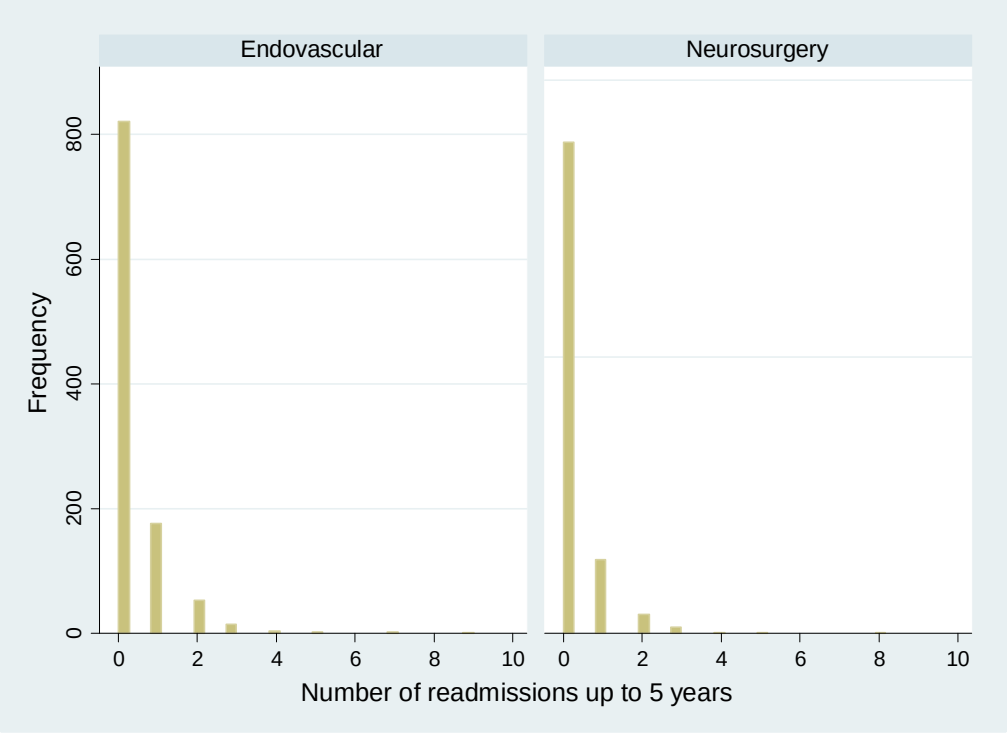


Figure 6.6: Histogram of quality-adjusted life years (QALYs) at 5 years follow-up by randomisation group using all countries data

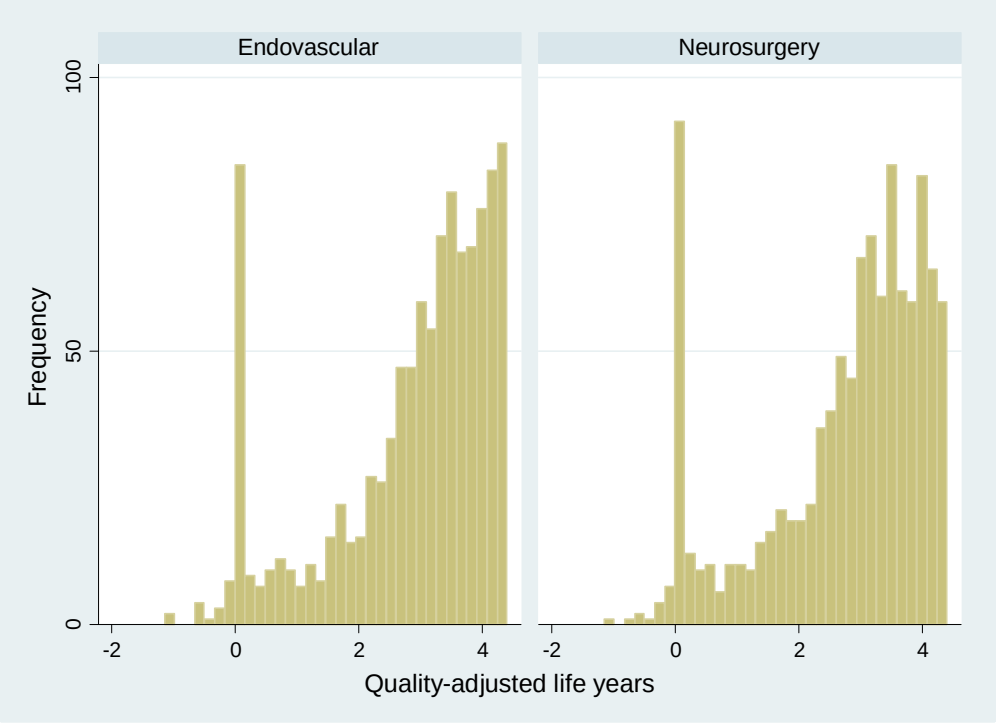


Figure 6.7: Histograms of total cost per patient using UK and country-specific unit-costs at 5 years follow-up by randomisation group using all countries data

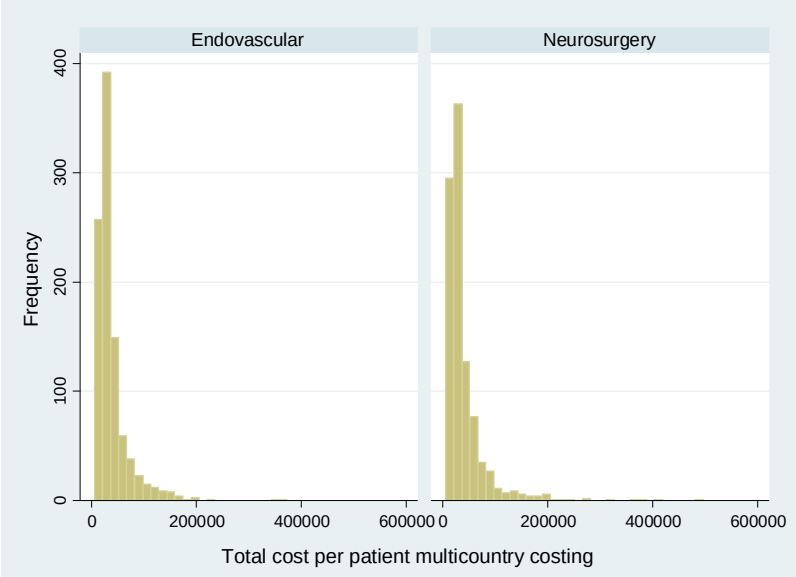
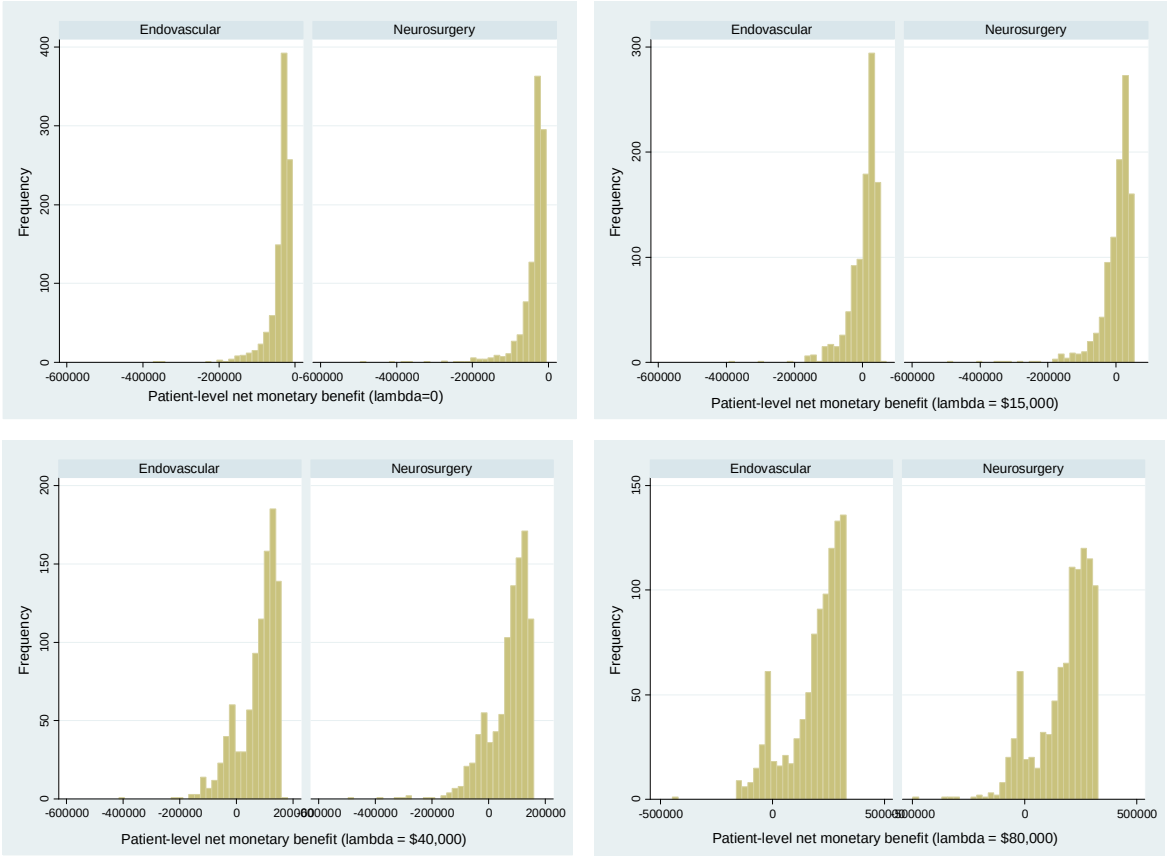


Figure 6.8: Histograms of net monetary benefits using different values of willingness to pay and country-specific unit costs by randomisation group using all countries data



Duration of operating time in theatre was estimated using a Tobit model assuming the dependent variable is truncated for values below zero because duration cannot have negative values. The Tobit model ensures the coefficients do not predict values below zero. The number of angiograms performed post randomisation and the number of readmissions at 5 years follow-up histograms look like a Poisson process. This was corroborated using a goodness-of-fit test which suggested that a Poisson regression was appropriate for these two covariates. The number of days in ITU and ward facilities was estimated applying a general gamma linear model (GLM) assuming a link identity. A similar model was used to estimate the total cost per patient using UK and country-specific unit costs. Manning and Mullahy described in a seminal article the potential of GLMs to model health care resource use and costs. (Manning and Mullahy 2001). Finally QALYs and net benefits were modelled using ordinary least squares (Hoch et al. 2002; Manca et al. 2005a).

The reference case in all models refers to an individual randomised to the endovascular group and recruited in the United Kingdom.

The interpretation of the coefficients from each of the models depends on the type of model estimated. The OLS and GLM models estimated in this chapter generate coefficients that can be easily interpreted because they are expressed in the same units as the dependent variable. Hence we can directly compare the results from OLS and GLM with the unadjusted means in Tables 6.1-6.3. The Poisson regression models the log of the expected counts as a function of the predictors and hence the coefficients are not easily interpretable. However, exponentiating the coefficients expresses the coefficients as rates, which is a more manageable statistic to interpret (Kirkwood and Sterne 2003). However, we cannot directly compare the results from the Poisson regression with the unadjusted mean difference in Table 6.1 for the number of angiograms and the number of readmissions.

Table 6.7 presents the coefficients of the models, including dummies for treatment, country and country-by-treatment interactions for the selected categories of resource use. The table also shows the results from the test for interaction to evaluate heterogeneity of treatment effect across jurisdictions.

The treatment effect on a particular resource use of endovascular coiling compared to neurosurgical intervention in the United Kingdom is given by the treatment coefficient in all models. The results indicate that, for example in the UK, patients allocated to neurosurgery had an additional 66.61 minutes in theatre compared to neurosurgery in the operation duration model, and 2.39 additional ward days in the length of stay model, with both coefficients highly significant.

The treatment coefficient associated with the number of readmissions does not have a straightforward interpretation. Exponentiating the Poisson regression coefficient $e^{-0.33}$ yields 0.72. This is interpreted as patients in the neurosurgical group in the UK have a rate of readmissions 0.72 times smaller than patients in the endovascular arm, a highly significant result.

Interpreting the results from a regression that includes interactions between different covariates is normally confusing. In the case of OLS and GLM, the country-by-treatment interaction coefficients are interpreted as the additional units of treatment effect in each country compared to the treatment effect in the United Kingdom. For instance, the treatment effect in Canada had an additional operating duration of 46.99 minutes compared to the treatment effect of 66.61 minutes in the United Kingdom, suggesting that patients in the neurosurgical group in Canada spent an additional 113.6 minutes in theatre compared to endovascular patients. This is similar to the mean duration difference between both groups

reported in Table 6.1 in Canada. Using a similar interpretation, the treatment effect for the number of ward days in Sweden had 2.72 fewer days compared to the treatment effect of 2.39 days in the United Kingdom yielding an overall treatment effect of 0.33 fewer days in the neurosurgical group compared to the endovascular patients. This is similar to the mean ward day difference between both groups reported in Table 6.1 in Sweden. Unfortunately, the interaction coefficients from the Poisson regressions do not have a straightforward interpretation but the coefficients can be used to conduct the test for interaction across countries.

Table 6.7 also presents the results from the test for interaction to determine heterogeneity of treatment effect across countries in all models. A significance level of 1% was chosen to avoid finding a spurious result given the number of comparisons conducted. The test for interaction suggests evidence of heterogeneity of treatment effect across countries in the duration of operation ($F=3.60$, $p<0.01$) and the number of angiograms ($F=82.30$, $p<0.000$). For the duration in theatre, differences across countries are caused by the longer stay in theatre of neurosurgical patients in Canada and the smaller number of angiograms amongst neurosurgical patients in Finland compared to the remaining countries. The result of the test for the number of readmissions at five years follow-up, $F=9.71$, $p<0.05$) suggested borderline significance of country-by-treatment interaction. This was mainly driven by the larger number of readmissions in the endovascular group compared to neurosurgical patients in Finland. No evidence of country-by-treatment interaction seemed to be present in the number of ITU and ward days.

Table 6.7: Country-by-treatment interaction test using regression analysis on selected resource use

	Operation duration (Tobit)			Number of angiograms (Poisson)			ITU days (GLM)			Ward days (GLM)			Number of readmissions (Poisson)		
	Coeff.	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**
Constant	136.60	2.07	0.000	0.12	0.01	0.000	1.95	0.16	0.000	14.8	0.42	0.000	-1.00	0.07	0.000
Treatment															
- *Endovascular*															
- Neurosurgery	66.61	3.66	0.000	-0.02	0.02	0.263	0.46	0.23	0.043	2.39	0.68	0.000	-0.33	0.11	0.004
Countries															
- *UK*															
- Sweden	-1.13	11.02	0.918	0.12	0.06	0.030	6.59	0.94	0.000	-6.08	0.93	0.000	-0.22	0.47	0.645
- Germany	18.06	12.38	0.145	0.33	0.09	0.000	11.85	1.56	0.000	-7.54	1.19	0.000	-1.18	0.42	0.004
- Canada	14.44	8.73	0.098	0.6	0.07	0.000	4.63	0.82	0.000	-2.99	1.33	0.025	-1.26	0.52	0.014
- Finland	-7.92	9.96	0.427	0.01	0.05	0.797	-0.45	0.62	0.463	-4.04	1.02	0.000	1.03	0.26	0.000
Country-by-treatment interactions															
- *UK*Treatment*															
- Sweden*Treatment	15.27	15.25	0.317	0.1	0.09	0.233	0.99	1.35	0.461	-2.72	1.34	0.042	0.65	0.55	0.242
- Germany*Treatment	2.10	21.74	0.923	-0.17	0.11	0.125	3.36	2.11	0.111	0.21	1.99	0.916	-0.82	0.82	0.314
- Canada*Treatment	46.99	14.80	0.002	-0.03	0.08	0.664	3.76	2.09	0.072	0.85	3.61	0.814	-1.12	1.12	0.316
- Finland*Treatment	-22.80	13.81	0.099	0.59	0.07	0.000	0.21	0.8	0.789	-2.15	1.36	0.114	-1.27	0.52	0.013
F-test	3.60			82.30			6.20			5.96			9.71		
p-value	0.01			0.00			0.18			0.20			0.05		

The reference case (in italics) indicates a patient randomised to the endovascular group and recruited in the United Kingdom. No evidence of miss-specification was found in any of the regressions using the link test approach. Robust variance estimation was used in all regressions using the vce(robust) option in STATA.

Table 6.8 shows the results of the country-by-treatment interaction test using regression analysis for QALYs and total cost per patient at 5 years follow-up. The treatment coefficient of -0.16 QALYs expresses the treatment effect of endovascular compared to neurosurgery in the United Kingdom, i.e. the mean QALYs difference between the endovascular and neurosurgical groups after controlling for the multinational nature that in this case favours endovascular patients. This is similar to the unadjusted mean QALY difference between groups of 0.16 reported in Table 6.2. The country-by-treatment interaction coefficient suggests that the treatment effect in Germany had an additional 0.96 QALYs compared to the treatment effect in the UK. This yields a treatment effect estimate in Germany of 0.80 QALYs favouring the neurosurgical group. This has driven the test for interaction which suggests heterogeneity of treatment effect for QALYs across countries ($F=3.43$, $p<0.01$).

The cost regression in Table 6.8 suggests no evidence of country-by-treatment interaction across countries. However, the country-by-treatment interaction term associated with Finland was estimated to be -9,841, a significant value ($p = 0.028$) suggesting that the treatment effect for the total cost per patient in Finland had 9,841 fewer dollars than the treatment effect of 1,402 estimated in the UK; that is, endovascular was significantly more expensive than neurosurgery in Finland. Although the coefficient is significant, no heterogeneity of treatment effect was suggested across settings using the test for interaction, using a 1% significance level.

Table 6.8: Country-by-treatment interaction test using regression analysis for quality-adjusted life years and total cost per patient at 5 years follow-up

	QALYs (OLS)			Total cost per patient at 5 years follow-up: country-specific unit costs (GLM)		
	Coeff.	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**
Constant	2.86	0.05	0.000	37,624	1,180	0.000
Treatment						
- *Endovascular*						
- Neurosurgery	-0.16	0.06	0.013	1,402	1,875	0.455
Countries						
- *UK*						
- Sweden	0.18	0.18	0.301	5,595	3,584	0.119
- Germany	-0.68	0.22	0.002	9,420	5,608	0.093
- Canada	0.12	0.18	0.500	6,871	4,262	0.107
- Finland	0.16	0.20	0.421	-8,263	3,933	0.036
Country-by-treatment interactions						
- *UK*Treatment*						
- Sweden*Treatment	-0.01	0.25	0.983	5,766	6,087	0.343
- Germany*Treatment	0.96	0.27	0.000	-847	7,363	0.908
- Canada*Treatment	-0.16	0.29	0.586	14,149	13,217	0.284
- Finland*Treatment	0.29	0.28	0.314	-9,841	4,472	0.028
F-test	3.43			7.73		
p-value	0.001			0.10		

The reference case (in italics) indicates a patient randomised to the endovascular group and recruited in the United Kingdom. No evidence of miss-specification was found in any of the regressions using the link test approach. Robust variance estimation was used in all regressions using the vce(robust) option in STATA. All costs expressed in 2008 $US Purchasing Power Parity.

Table 6.9 presents the results of the country-by-treatment interaction test using regression analysis for net benefits on the monetary scale at five years follow-up and using different willingness to pay for QALYs gained. The treatment coefficient represents the mean incremental net benefit between treatment arms in the UK. For any values of λ, the endovascular group has a greater net benefit than the neurosurgical group, suggesting that endovascular is cost-effective compared to neurosurgery. A similar pattern is observed for Sweden and Canada.

The country-by-treatment interaction coefficients in Table 6.9 suggest that in Germany and Finland, neurosurgery has a higher net benefit compared to endovascular therapy. This is driven by the treatment effect on QALYs in Germany which suggests that neurosurgery is more effective than endovascular. In Finland, neurosurgery has a greater net benefit because of the large costs of endovascular patients in this country due to the number of readmissions. The test for interaction suggests that no country-by-treatment interaction is present in net benefits across countries.

The analysis reported in Table 6.9 is based on the significance of the point estimates of incremental net benefit. Therefore, they are only an indication of the likely cost-effectiveness of endovascular and neurosurgery in each country. A full economic evaluation should incorporate a full assessment of uncertainty in each country and this is expanded in Chapter 8.

Table 6.9: Country-by-treatment interaction test using regression analysis on net benefits on the monetary scale using country-specific unit costs and different thresholds of willingness to pay

	Net benefit (λ=$0) (OLS)			Net benefit (λ=$15,000) (OLS)			Net benefit (λ=$40,000) (OLS)			Net benefit (λ=$80,000) (OLS)		
	Coeff.	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**
Constant	-37,624	1,183	0.000	5,275	1,538	0.001	76,773	2,478	0.000	191,169	4,222	0.000
Treatment												
- *Endovascular*												
- Neurosurgery	-1,402	1,879	0.456	-3,765	2,333	0.107	-7,703	3,600	0.032	-14,005	6,019	0.020
Countries												
- *UK*												
- Sweden	-5,595	3,592	0.120	-2,906	5,423	0.592	1,574	9,399	0.867	8,743	16,257	0.591
- Germany	-9,420	5,621	0.094	-21,251	6,810	0.002	-40,971	11,284	0.000	-72,522	20,083	0.000
- Canada	-6,871	4,272	0.108	-5,611	5,705	0.325	-3,511	9,616	0.715	-152	16,800	0.993
- Finland	8,266	3,943	0.036	12,090	6,313	0.056	18,463	11,210	0.100	28,660	19,513	0.142
Country-by-treatment interactions												
- *UK*Treatment*												
- Sweden*Treatment	-5,766	6,101	0.345	-5,807	8,483	0.494	-5,875	13,950	0.674	-5,984	23,645	0.800
- Germany*Treatment	848	7,380	0.909	16,184	8,700	0.063	41,746	14,161	0.003	82,645	25,157	0.001
- Canada*Treatment	-14,149	13,248	0.286	-12,915	15,173	0.395	-10,858	19,962	0.587	-7,567	29,463	0.797
- Finland*Treatment	9,841	4,482	0.028	13,094	7,797	0.093	18,516	14,855	0.213	27,191	26,754	0.310
F-test	1.92			1.91			2.69			2.98		
p-value	0.10			0.106			0.03			0.018		

The reference case (in italics) indicates a patient randomised to the endovascular group and recruited in the United Kingdom. No evidence of miss-specification was found in any of the regressions using the link test approach. Robust variance estimation was used in all regressions using the vce(robust) option in STATA. Net benefits and λ expressed in 2008 $US Purchasing Power Parity.

6.3.4 Multilevel models

So far we have explored formal statistical tests and regression analysis to detect heterogeneity across countries in a set of covariates that include resource use, QALYs, costs and net benefits. The final method we will explore is the use of multilevel models to detect heterogeneity as described in Section 6.3.4.

Multilevel models allow the treatment effect to vary randomly across countries. Because countries are considered a random effect in these models and dummies are no longer required, data from all countries in ISAT could potentially be used to estimate the parameters of the models. However, five countries provided enough data to estimate the regressions in the previous section and for comparative purposes the same dataset is used in this section. The five countries include 95% of the overall sample in the study. Therefore, the MLMs run in this section included data from largest five countries in ISAT for the selected resource use categories, total costs, QALYs, and net benefits A significance level of 1% was used to determine the significance of the coefficients to be in line with previous sections.

Table 6.10 shows the results of the multilevel random slope model for the selected categories of resource use[19]. In these models, the fixed part β_o is the mean value of the resource use across all countries and the fixed part β_1 is the mean difference between the endovascular and neurosurgical groups across all countries. For instance, the mean duration in theatre across countries was estimated to be 141.15 minutes whereas the treatment effect for this covariate was estimated to be 75.46 additional minutes in the neurosurgical procedure

[19] The number of ITU days could not be estimated in MLwiN assuming normality in the dependent variable due to the large number of zeros. Using a gamma distribution instead has the potential of correcting this problem. However, this is not currently possible in MLwiN and as a result the test for heterogeneity was not conducted for this covariate using multilevel models.

compared to the endovascular group, a highly significant result. Using another example, the treatment effect in the number of days at ward between groups across all countries was estimated to be 2.21 days, a significant result with p=0.001.

Table 6.10 includes the variance estimation for the two levels (patients and countries) and the covariance between the intercept and slope of the random part of the MLMs. In addition the variance partitioning coefficient is also reported and suggests that the total variance in each group was mostly attributable to the variance within each country. However, the VPC was large in the number of angiograms in both groups (37.7% in the endovascular treatment and 40% in the neurosurgery) indicating that a large proportion of the total variance was attributable to the between-country variance in this resource use.

Table 6.10 also presents the results of the likelihood ratio test which indicates whether treatment effect varies significantly across countries. A significance level of 1% was selected, in line with previous methods described in this chapter. The tests suggest borderline evidence of heterogeneity of treatment effect in operation duration (Δ=8.75, p =0.013) and the number of angiograms (Δ=40.6, p =0.000) but not in the number of ward days or readmissions.

Table 6.10: Modelling variation in treatment effect using a multilevel random slope model on selected resource use

	Operation duration			Number of angiograms			Ward days			Number of readmissions		
	Coeff.	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**
Fixed Part												
Constant (β_0)	141.15	4.79	0.000	1.44	0.17	0.000	10.51	1.55	0.000	0.37	0.16	0.003
Treatment effect (β_1)	75.46	9.30	0.000	0.11	0.20	0.339	2.21	0.58	0.001	-0.17	0.15	0.211
Random Part												
σ^2_{u0}	93.88	72.97		0.15	0.10		10.16	7.34		0.12	0.08	
σ^2_{u1}	372.36	270.60		0.19	0.12		0.00	0.00		0.10	0.07	
σ_{u01}	201.76	136.34		-0.09	0.09		0.00	0.00		-0.10	0.07	
σ^2_e	5709.95	179.99		0.24	0.01		172.65	5.42		0.53	0.02	
VPC												
Endovascular	1.6%			37.7%			5.6%			19.0%		
Neurosurgery	13.2%			40.0%			5.6%			3.1%		
-2LogL random slope	23203.39			2934.28			16276.01			4498.52		
-2LogL random intercept	23212.14			2974.88			16276.01			4504.89		
Δ	8.753			40.601			0			6.373		
p-value	0.013			0.000			1.000			0.041		

All models were estimated using the Restricted Iterative Generalised Least Squares (RIGLS) estimation process in MLwiN. All models assumed normality in the dependent variable.

Table 6.11 presents the results from the random slope model for QALYs and the total cost per patient at 5 years using a multicountry approach. The treatment effect parameter on QALYs (β_1) was estimated to be 0.00 with associated p=0.339 across all countries. Therefore, using the MLM suggests that there is no difference between the mean QALYs in the endovascular treatment and the mean QALYs in neurosurgical groups across all countries. The likelihood ratio tests suggest that there is no clear evidence of significant variation between countries in treatment effect (Δ=6.32, p=0.043).

Table 6.11: Modelling variation in treatment effect using a multilevel random slope model for quality-adjusted life years and total cost per patient at 5 years follow-up

	QALYs			Total cost per patient at 5 years follow-up: country-specific unit costs		
	Coeff.	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**
Fixed Part						
Constant (β_0)	2.79	0.16	0.000	41,054.22	4,374.75	0.000
Treatment effect (β_1)	0.00	0.20	0.399	2,025.97	1,705.98	0.197
Random Part						
σ^2_{u0}	0.098	0.078		79,338,336	57,934,924	
σ^2_{u1}	0.159	0.130		0	0	
σ_{u01}	-0.122	0.098		0	0	
σ^2_e	1.727	0.054		1,420,542,592	45,505,180	
VPC						
Endovascular	5.4%			5.3%		
Neurosurgery	0.7%			5.3%		
-2LogL random slope	6903.44			46734.03		
-2LogL random intercept	6909.76			46734.03		
Δ	6.32			0.00		
p-value	0.042			1.000		

All models were estimated using the Restricted Iterative Generalised Least Squares (RIGLS) estimation process in MLwiN. All models assumed normality in the dependent variable.

Table 6.10 suggests no significant difference in the total cost per patient between the interventions across countries. The likelihood ratio test suggests that there is no evidence of heterogeneity in treatment effect on costs across countries.

The VPC in QALYs and costs suggests that most of the total variance is attributable to the within-country variance in the endovascular and neurosurgical groups.

Table 6.12 shows the results from the multilevel random slope model on the net benefit statistic on the monetary scale at five years follow-up. The overall treatment effect across all countries β_1 suggests no significant differences between groups for any levels of λ. This result is different from the output from the standard regression model in Table 6.9 which showed that net benefit was significantly greater in the endovascular compared to the neurosurgical group for values of λ=£40,000 and λ=$80,000. The results from the multilevel model showed that the standard errors associated with the treatment effect were larger than those in the standard regression analysis. This is the case because in the random slope model, both the intercept and the slope are random variables and therefore more variability is included when estimating this type of model. The VPC in Table 6.12 suggests that most of the overall total variance is attributable to the level 1 variance. Finally the likelihood ratio tests suggest that the incremental net benefit does not vary significantly across countries.

Table 6.12: Modelling variation in treatment effect using a multilevel random slope model on the net benefit statistic

	Net benefit (λ=$0)			Net benefit (λ=$15,000)			Net benefit (λ=$40,000)			Net benefit (λ=$80,000)		
	Coeff.	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**	**Coeff.**	**SE**	**p-value**
Fixed Part												
Constant (β_0)	-41,054.22	4,374.75	0.000	2,364.50	5,294.06	0.361	72,297.02	9,509.70	0.000	184,157.28	13,290.83	0.000
Treatment effect (β_1)	-2,025.97	1,705.98	0.197	-3,447.43	2,117.72	0.106	-3,885.77	7,681.71	0.351	-7,372.18	12,434.65	0.335
Random Part												
σ^2_{u0}	79,338,336	57,934,924		114,991,368	84,466,368		366,556,576	282,272,960		646,801,280	453,999,168	
σ^2_{u1}	0	0		0	0		218,919,280	183,714,608		544,193,152	396,721,344	
σ_{u01}	0	0		0	0		-285,531,648	225,350,064		-602,755,968	424,012,992	
σ^2_e	1,420,542,592	45,505,180		2,189,005,056	70,121,944		5,204,370,432	166,766,240		14,528,016,384	465,505,696	
VPC												
Endovascular	5.3%			5.0%			6.6%			4.3%		
Neurosurgery	5.3%			5.0%			0.3%			-0.1%		
-2LogL random slope	46734.03			47578.70			49268.78			51271.26		
-2LogL random intercept	46734.03			47578.70			49271.71			51279.66		
Δ	0.000			0.000			2.937			8.398		
p-value	1.000			1.000			0.230			0.015		

All models were estimated using the Restricted Iterative Generalised Least Squares (RIGLS) estimation process in MLwiN. All models assumed normality in the dependent variable.

6.3.5 Summary of results from the test for heterogeneity across methods

Table 6.13 presents a summary of the results from the different tests of heterogeneity explored in this chapter. All the methods yielded similar results in most of the variables studied, except for the operation duration, the number of readmissions, and QALYs. For operation duration, the formal statistical test and the multilevel model suggest no heterogeneity whereas the standard regression detected heterogeneity across countries. In the number of readmissions, the formal statistical tests suggested heterogeneity but this was not the case when using standard regression analysis or multilevel models. The multilevel model suggests no evidence of heterogeneity of treatment effect on QALYs across countries, but this was not the case with the other two approaches. Finally, the three approaches suggested heterogeneity of treatment effect in the number of angiograms between the endovascular treatment and the neurosurgical intervention.

Table 6.13: Summary results of test for heterogeneity using formal statistical tests, regression analysis and multilevel models

Variable	Formal statistical tests	Standard regression analysis	Multilevel models
Operation duration	●	○	●
Number of angiograms	○	○	○
ITU days	●	●	
Ward days	●	●	●
Number of readmissions	○	●	●
QALYs	○	○	●
Total cost per patient using country-specific unit costs	●	●	●
Net benefit ($\lambda = 0$)	●	●	●
Net benefit ($\lambda = 15{,}000$)	●	●	●
Net benefit ($\lambda = 40{,}000$)	●	●	●
Net benefit ($\lambda = 80{,}000$)	●	●	●

○: Heterogeneity of treatment effect across countries present; ●: No heterogeneity of treatment effect across countries present; a 1% level of significance was chosen across all methods.

6.3.6 Presenting treatment effect results

The results from table 6.13 suggest evidence of heterogeneity in main categories of resource use and health outcomes in the treatment effect between endovascular and neurosurgical interventions. No evidence of heterogeneity of treatment effect was found in the total cost per patient and net benefits. However, it seems that the effect of heterogeneity diluted when adding the different cost components and hence the tests lost power. The tests of heterogeneity conducted in this chapter provide indication of the existence of heterogeneity of treatment effect across countries. This suggests that pooling data from all countries to present the results from the economic evaluation alongside the ISAT study is not appropriate as the overall treatment effect will not be a meaningful indicator in all countries.

The AIC can be used to determine which model specification fits the data best. AIC is reported for standard regression analysis and MLMs in Table 6.14. In all cases except the number of angiograms, AIC was smaller in the standard regression analysis suggesting that the MLMs did not fit the data as well as the OLS. However, MLMs assumed normality in the dependent variable for all the covariates studied whereas the standard regression analysis incorporated appropriate model specifications. This is a limitation of the current comparison between both methods. Therefore Table 6.13 also includes the results from the AIC estimating an OLS model for all covariates to improve the comparability between methods. It is clear that OLS models had smaller AIC than the MLMs in all cases.

Table 6.14: Akaike information criterion results from the standard regression analysis and the multilevel models

Variable	Standard regression analysis	Multilevel models	Standard regression analysis (OLS in all models)
Operation duration	22976.26	23224.14	23207.72
Number of angiograms	4651.54	2946.28	2908.96
ITU days	8284.47		12845.00
Ward days	14988.81	16288.01	16280.37
Number of readmissions	2982.43	4510.52	4490.64
QALYs	6906.04	6915.44	6906.04
Total cost per patient using country-specific unit costs	45264.17	46746.03	46735.57
Net benefit ($\lambda = 0$)	46735.56	46746.03	46735.56
Net benefit ($\lambda = 15{,}000$)	47578.73	47590.70	47578.73
Net benefit ($\lambda = 40{,}000$)	49269.32	49280.78	49269.32
Net benefit ($\lambda = 80{,}000$)	51276.02	51283.26	51276.02

Both the OLS and the multilevel models are able to produce country-specific estimates as described in the methods section in this chapter. Figures 6.9 and 6.10 present the results of the treatment effect (and 95% confidence interval) in QALYs and the total costs from both approaches respectively. The figures uses a forest plot with the zero indicating no difference between endovascular treatment and neurosurgery and the horizontal lines indicating the 95% confidence interval[20]. Because we are comparing endovascular treatment versus neurosurgery, estimates of treatment effect to the right indicate endovascular is more effective or more costly. The opposite occurs with estimates of treatment effect to the left.

Figure 6.9 shows the country-specific treatment effect estimate in QALYs for the OLS approach, using the observed data, and using shrinkage estimation from the MLM. It is clear that the observed data and the OLS produced similar estimates favouring endovascular treatment for all countries and across all countries except for Germany. However, the pooled

[20] Because the empirical Bayes shrinkage estimation was applied to calculate shrinkage estimates and associated standard error, the confidence interval presented here is normally referred as credible interval.

random effect from the MLM (using data from all countries) indicates no significant difference between the interventions. Table 6.11 reported a large between-country variance of 0.159 indicating variation on the treatment effect across countries driven by the German result. This has increased the uncertainty associated to the pooled random effect in the MLM. The figure also shows that some shrinkage towards the overall random pooled effect occurred in the country-specific estimates from the MLM. In this case, Germany experienced the largest shrunk.

Figure 6.9: Comparison of country-specific incremental QALY estimates from observed data, ordinary least square estimation and multilevel model

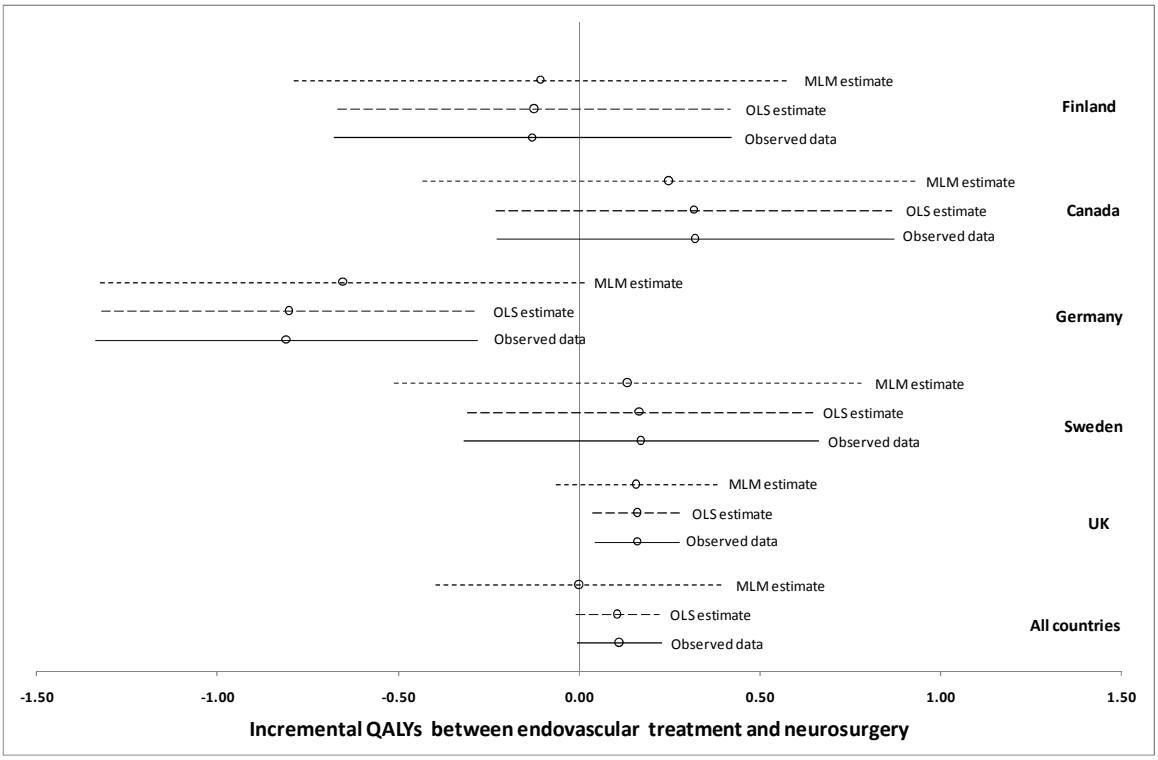
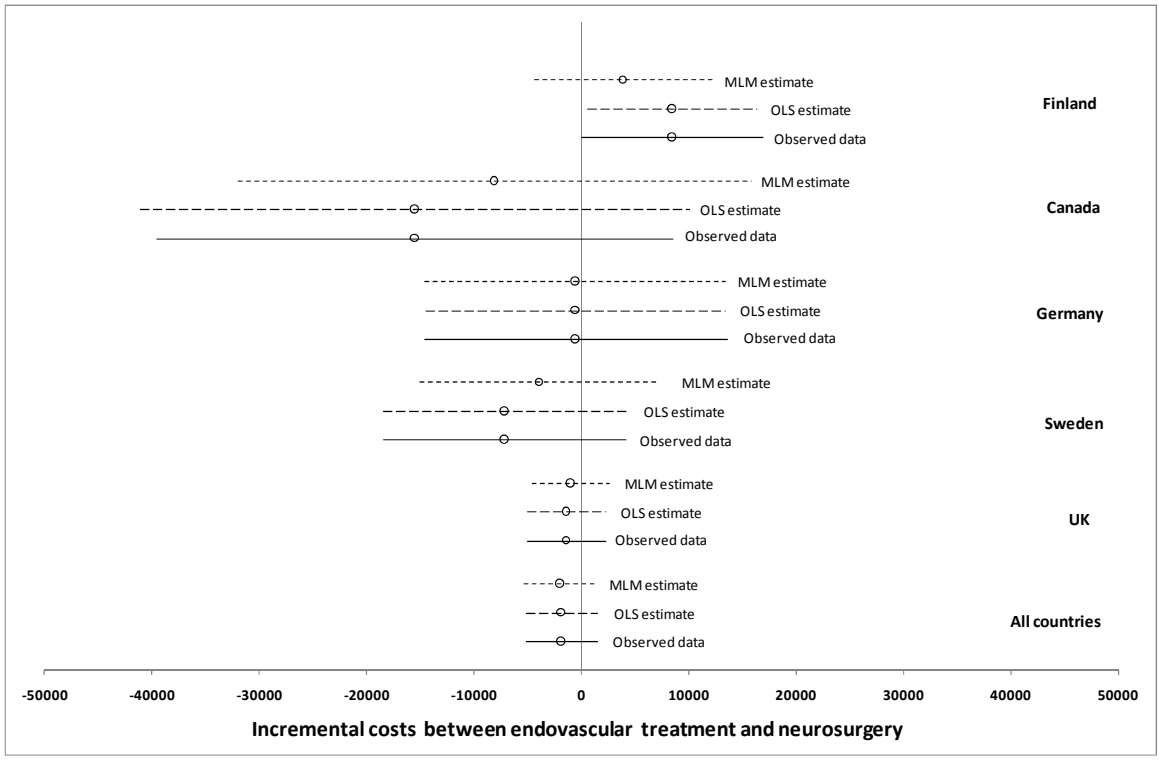


Figure 6.10 presents the results for treatment effect across countries for the total cost per patient and for the observed data, the OLS and the MLM model. Again similar results were obtained using the observed and the OLS method with no significant differences between treatment arms using data from all countries and across all countries except in Finland. In this country endovascular treatment was more expensive than neurosurgery. Using country-

specific shrinkage estimates from the MLM the mean treatment effect from all countries shrunk towards the pooled random effect. In the case of Finland the significant difference between the treatment alternative disappears. Table 6.11 suggested no variation in the between-country treatment effect in cost across countries as a result; the observed data, the OLS and the pooled random effect yielded similar mean treatment effect and associated uncertainty.

Figure 6.10: Comparison of country-specific incremental cost estimates from observed data, ordinary least square estimation and multilevel model



6.4 Summary and conclusions

The current chapter has described three methods to test for heterogeneity of treatment effect on in multinational clinical trials: formal statistical tests, standard regression analysis and multilevel models. The current chapter has investigated the impact of heterogeneity on the main drivers of costs evaluating a range of resource use categories, an aspect which has

not been explored in depth in the literature so far. The results from this chapter suggest that heterogeneity should be explored in different categories of resource use to detect important differences across countries, and proposed then that tests of heterogeneity can serve as an indicative tool to determine whether differences across countries truly exist.

There were some differences among the three methods for some of the resource use studied. For instance, formal tests suggested absence of heterogeneity in operation duration whereas this was not the case in the other two methods. Similarly formal tests suggested evidence of heterogeneity of treatment effect in the number of readmissions but standard regression and multilevel model analyses could not detect that difference. In terms of QALYs, formal tests and standard regression analysis suggested evidence of heterogeneity of treatment effect in health outcomes. This was mainly driven by the opposite sign of the treatment effect compared with the remaining countries in Germany, where it was found that quality of life of patients in the neurosurgical group was better than in the endovascular group. The three methods identified heterogeneity of treatment effect in the number of angiograms.

None of the tests observed significant heterogeneity of treatment effect in costs or net benefits. However, it is possible that none of the methods have sufficient power to detect differences across countries on these covariates and caution needs to be exercised (Cook et al. 2003; Manca et al. 2007). Given the results from the resource use analysis, it is possible that heterogeneity of treatment effect on costs and net-benefits exist, but the tests were no longer powerful enough to detect differences across countries. The results from this chapter suggest that country-by-treatment interactions in treatment effect in resource use are likely to exist. Differences across countries were found in the operation duration, the number of angiograms and the number of readmissions. These components of resource use contributed largely to the overall mean costs of endovascular and neurosurgery in a recent cost-analysis

conducted in the UK (Wolstenholme et al. 2008). As a result, it cannot be ruled out that because no differences in the treatment effect on costs and net-benefits were observed, the differences do not exist.

Exploring heterogeneity of treatment effect is an important part of the analyses of multinational clinical trials, as it provides information on whether pooled or country-specific estimates should be used to present the results of the economic evaluation. The results from this chapter suggest that heterogeneity of treatment effect is likely to be present in different components or resource use, health outcomes, and possibly costs and net benefits. Therefore, an overall mean estimate of treatment effect across all countries will not be meaningful in each jurisdiction and will probably mislead decision makers. In the ISAT study then, country-specific estimates of treatment effect are likely to provide more information than the overall pooled estimate.

This chapter presented country-specific treatment effect results in costs and QALYs using the observed data, standard regression analysis and the shrinkage estimates from the multilevel model. The AIC results suggested that standard regression analysis provided a better fit to the ISAT data than multilevel models and as a result the reporting of country-specific estimates from this method seems more appropriate. Multilevel models have received considerable attention in the literature to present the results of cost-effectiveness analyses alongside multinational clinical trials (Grieve et al. 2005; Manca et al. 2005b; Manca and Willan 2006; Thompson et al. 2006). Researchers showed enthusiasm when MLMs were introduced in this area, but recently some authors have questioned the validity of the exchangeability assumption behind these models (see Appendix 7) (Grieve et al. 2007). They argued that in international trials, systematic differences across countries in terms of resource allocation and reimbursement make the exchangeability assumption difficult to hold. They

have proposed the collection of country-specific information that could potential help to explain differences across countries in treatment effect but it is not clear what type of information should be routinely collected (Grieve et al. 2007; Manca et al. 2007).

The exchangeability assumption suggests that the country-specific mean estimates of treatment effect shrunk towards the overall pooled random effect estimate because they can benefit from this estimate. However, as discuss above, a pooled estimate for all countries, whether has been calculated using the observed data, the OLS model or from the MLM are probably meaningful to decision makers using the ISAT study. As a result, there are reasons to question the validity of the exchangeability assumption in the ISAT study.

Another aspect that has not received great attention in the literature is that one of the main potentials of multilevel models is the possibility of estimating a between-country variance of the treatment effect across countries. However, for an accurate estimation of this variance, the number of observations at the country level is an important parameter to consider. The number of countries participating in ISAT has played an important role during the estimation process. It seems that the number of observations at level 2 (countries) was not large enough to estimate an accurate between country variance, which in turn has influenced the fit of the data. The optimal number of countries participating in a multinational trial is an aspect that needs consideration in future research. This problem could potentially be solved using information from the eleven countries in ISAT. MLMs were re-estimated using information from the eleven countries in ISAT but similar results as those presented in this chapter were observed. The five countries used in the main analysis in this chapter contributed 95% of the overall sample size in the ISAT study. The remaining six countries did not include enough information to improve the estimations in the MLM models.

The current comparison of methods has assumed normality on the dependent variable in both models. This is a limitation of the current exercise and the analysis here should be expanded to include a more flexible and real selection of distribution of the variable of interest.

The above discussion of methods is based primarily on statistical grounds but Chapter 3 showed that resource use and unit costs may be systematically related. If the latter is true, having accurate unit costs from each country is important to obtain unbiased estimates of costs. However, Chapter 4 demonstrated that most studies conducted in this area have applied unit costs from one country to all resource use in the trial. The implications of using different costing methods are yet not fully understood in a multinational framework and the next chapter explores these.

Chapter 7: Economic issues in multinational cost-effectiveness analyses – a comparison of costing methods and a simulation study

7.1 Introduction

Previous chapters have identified, described and implemented three statistical methods to indentify heterogeneity across countries of treatment effects on resource use, costs and outcomes in a cost-effectiveness analysis alongside a multinational clinical trial. The outcome from these tests informs analysts whether differences in the treatment effect across countries exist. This information can then be used by analysts to decide the best way of presenting the cost-effectiveness results from the trial. The discussion so far has concentrated on statistical issues surrounding the analysis of the data, but Chapter 3 showed that if economic theory holds, the volume of resources used is likely to be closely linked to the relative costs of different inputs, and this is likely to have implications for the analysis.

If the association between resource use volumes and relative factor input prices (unit costs) truly exists, then using different costing approaches can yield different estimates of overall and country-specific mean costs. Chapter 4 identified that the collection of unit costs across countries involved in an international trial is a demanding task because of the additional economic burden associated with the gathering of data in more than one setting. As a result, it was shown that the majority of published studies in this area have applied a set of unit costs from one country across resource use in all countries (Table 4.7 in Chapter 4). However, little is known about the impact of using different costing approaches to the calculation of costs in multinational studies. For example, there is still uncertainty about whether using a one-country costing instead of a multicountry costing method produces

different results in the calculation of the country-specific incremental cost per patient in a multinational trial.

Chapter 3 introduced the concept of the production function of a health care technology and the application of the characteristics of this function to the situation of a multinational clinical trial. The chapter described the case of having a producer of health care interested in combining two inputs of a new health care technology to achieve one successfully treated patient. The two inputs that affect the production function of this technology were the number of days in an intensive therapy unit and in a ward facility. The example continued with a multinational clinical trial in which two countries participated, and resource use on the two inputs and the associated unit costs were collected. It was assumed that both countries had the same production function but faced different relative factors input prices. The amount of resource use combined in each country was different, yielding different estimates of absolute costs between the two countries.

The example in Chapter 3 described the case of a producer interested in a new health technology only. However, this scenario can be expanded to the case of a clinical trial comparing two competing alternatives, where each treatment can be represented by a health technology. In this case, the example in Chapter 3 can be extended to determine whether using different costing approaches produces different estimates of incremental costs in a multinational study comparing two treatments. Chapter 3 also introduced the concept of the elasticity of substitution, i.e. how the ratio of factor inputs changes as the slope of the isoquant varies. This parameter informs the degree of substitution between two input factors that are combined to achieve a certain level of output. Little is known about how this degree of substitution between input factors in each country influences the estimates of incremental costs in a multinational study, and the costing strategy implemented.

The aim of this chapter is to explore whether different multinational costing methods produce different estimates of incremental cost per patient in a multinational clinical trial, within the framework of the economic theory of production. The chapter begins with a formal comparison of multinational costing approaches using the ISAT data. Then using the theoretical foundations set out in Chapter 3, a simulation study is conducted to determine the relationship between incremental cost and elasticity of substitution of input factors when using different multinational costing approaches.

7.2 A comparison of multinational costing approaches using ISAT data

At least two costing approaches can be implemented in multinational economic evaluations alongside trials: a one-country or a multicountry costing approach (Reed et al. 2005a). This thesis introduced and described both approaches in Chapter 1 and then in Chapter 4 classified applied cost-effectiveness analyses alongside multinational clinical trials depending on the costing approach conducted in the study.

The aim of this section is to provide estimates of mean cost per patient in each arm of the trial for selected categories of resource use in ISAT using a one-country or a multicountry costing approach. To keep in line with previous chapters, the mean cost per patient for selected categories of resource use considered to be the main cost drivers are explored. The categories of resource use selected are angiograms, ITU and ward days, and readmissions. The mean cost per patient for each category of resource use and each intervention was calculated pooling data from all countries and separately for those countries with the largest contribution to the data, i.e. United Kingdom, Sweden, Germany, Canada and Finland.

Costs are calculated first using a one-country costing which in this case corresponds to using UK unit costs across all countries in the ISAT dataset. Costs are then re-calculated applying a multicountry costing, in which country-specific unit costs are applied to country-specific resource use. Detailed information on the country-specific unit cost data available for this analysis was presented in Chapter 5 and the reader is referred to that chapter for a complete overview. Country-specific unit cost data were available for the United Kingdom, Sweden, Germany and Finland. No country-specific unit cost data were available for Canada and so resource use in this country was weighted using UK unit costs.

Country-specific unit costs were converted to $US using purchasing power parities (PPP) from 2008 in all countries. When analysing the results using the UK unit costs only, estimates are also expressed in $US using PPP from 2008 to facilitate the comparison with the multicountry costing analysis.

7.2.1 Results of the comparison between multinational costing approaches

Tables 7.1 and 7.2 show the mean cost per patient in each treatment group and associated mean cost difference for each selected category of resource use using a one-country and a multicountry costing strategy respectively. The tables include a category called "Other", that gathers the costs of all other categories of resource use per patient in the ISAT study. The estimates of costs in the UK and Canada in both tables are the same, as UK unit costs were applied to both countries. Pooling data from all countries, the mean difference in the cost per patient between endovascular treatment and neurosurgery is similar for all resource use categories in both costing approaches. The mean difference in the cost per patient in ITU and ward suggested that endovascular was significantly cheaper than neurosurgery. In addition, the tables suggest that readmissions yielded similar mean cost difference between the costing alternatives, with endovascular more expensive than neurosurgery.

In Sweden and Germany, no significance differences in the mean cost per patient between treatments were observed in any of the categories of resource use in both costing approaches. However, different results were obtained from each costing approach in Finland. In this country, applying the set of UK unit costs to resource use in Finland yielded a significant difference in the mean cost per patient of angiograms. When applying Finnish unit costs, additional significant differences in the mean cost per patient between endovascular treatment and neurosurgery were observed in readmissions and the "other" category.

Table 7.1: Cost per patient for selected resource use using a one-country costing strategy (UK unit costs reported in $US 2008 PPP) in all countries and participating countries data

	Endovascular		Neurosurgery		
	n	Mean cost per patient (SD) $US 2008 PPP	n	Mean cost per patient (SD) $US 2008 PPP	Mean difference (95% confidence interval) $US 2008 PPP
All countries					
Angiograms	1,071	1,556 (760)	1,068	1,540 (656)	15.85 (-44.35 to 76.04)
ITU days	1,073	9,860 (18,585)	1,067	11,673 (19,792)	-1,814 (-3,441 to -186)**
Ward days	1,072	9,047 (8,895)	1,067	10,485 (9,903)	-1,438 (-2,236 to -640)*
Readmissions	1,053	2,276 (9,211)	1,062	497 (3,043)	1,779 (1,195 to 2,363)*
Other	1,019	19,232 (22,982)	1,028	19,783 (32,033)	-551 (-2,970 to 1,867)
United Kingdom					
Angiograms	809	1,436 (558)	835	1,407 (510)	29.52 (-22.16 to 81.20)
ITU days	809	5,381 (12,137)	835	6,645 (13,173)	-1,264 (-2,491 to -38.04)**
Ward days	809	9,754 (7,855)	835	11,328 (10,098)	-1,574 (-2,451 to -697)*
Readmissions	795	2,310 (9,097)	829	528 (3,238)	1,783 (1,123 to 2,442)*
Other	781	18,727 (23,200)	808	19,125 (32,225)	-398 (-3,169 to 2,373)
Sweden					
Angiograms	64	1,627 (732)	59	1,765 (851)	-138 (-421 to 145)
ITU days	64	23,547 (20,564)	59	27,550 (20,335)	-4,003 (-11,312 to 3,305)
Ward days	64	5,746 (4,385)	59	5,529 (4,130)	217 (-1,307 to 1,741)
Readmissions	64	1,636 (10,056)	59	204 (1,227)	1,432 (-1,179 to 4,042)
Other	64	20,263 (18,500)	58	25,907 (28,340)	-5,644 (-14,147 to 2,860)
Germany					
Angiograms	62	2,007 (1,371)	55	1,663 (806)	345 (-74 to 764)
ITU days	62	38,037 (33,973)	56	48,557 (29,178)	-10,520 (-22,128 to 1,087)
Ward days	62	4,783 (5,799)	56	6,496 (7,486)	-1,713 (-4,142 to 716)
Readmissions	61	1,968 (7,347)	56	151 (1,131)	1,817 (-150 to 3,784)
Other	52	29,450 (32,556)	50	27,653 (28,256)	1,797 (-10,198 to 13,792)
Canada					
Angiograms	48	2,619 (1,184)	41	2,478 (693)	141 (-277 to 559)
ITU days	48	18,137 (15,448)	40	29,754 (33,737)	-11,617 (-22,439 to -795)**
Ward days	48	7,784 (5,829)	40	9,918 (13,999)	-2,134 (-6,545 to 2,278)
Readmissions	46	2,168 (14,705)	41	0 (0)	2,168 (-2,401 to 6,737)
Other	46	13,722 (6,681)	38	21,529 (49,031)	-7,806 (-22,331 to 6,718)
Finland					
Angiograms	34	1,457 (457)	29	2,584 (535)	-1,127 (-1,377 to -877)*
ITU days	34	4,133 (9,740)	29	5,985 (7,217)	-1,853 (-6,237 to 2,532)
Ward days	34	7,094 (3,624)	29	7,249 (2,636)	-155 (-1,777 to 1,467)
Readmissions	31	4,649 (9,050)	27	1,334 (3,080)	3,315 (-349 to 6,979)
Other	30	22,775 (26,642)	27	12,905 (8,059)	9,870 (-828 to 20,568)

*p<0.001; **p<0.05, All costs expressed in $US Purchasing Power Parity from 2008

Table 7.2: Cost per patient for selected resource use using a multicountry costing strategy in all countries and participating countries data

	Endovascular		Neurosurgery		
	n	Mean cost per patient (SD) $US 2008 PPP	n	Mean cost per patient (SD) $US 2008 PPP	Mean difference (95% confidence interval) $US 2008 PPP
All countries					
Angiograms	1,071	1,565 (777)	1,068	1,546 (659)	19.38 (-41.69 to 80.45)
ITU days	1,073	8,004 (14,821)	1,067	9,543 (16,094)	-1,539 (-2851 to -228)**
Ward days	1,072	9,012 (8,894)	1,067	10,450 (9,903)	-1438 (-2236 to -640)*
Readmissions	1,053	2,183 (8,978)	1,062	475 (3,001)	1,708 (1,138 to 2,278)*
Other	973	18,621 (22,858)	981	19,262 (32,159)	-641 (-3,119 to 1,836)
United Kingdom					
Angiograms	809	1,436 (558)	835	1,407 (510)	29.52 (-22.16 to 81.20)
ITU days	809	5,381 (12,137)	835	6,645 (13,173)	-1,264 (-2,491 to -38)**
Ward days	809	9,754 (7,855)	835	11,328 (10,098)	-1,574 (-2,451 to -697)*
Readmissions	795	2,310 (9,097)	829	528 (3237)	1,783 (1,123 to 2,442)*
Other	781	18,727 (23,200)	808	19,125 (32225)	-398 (-3,169 to 2,373)
Sweden					
Angiograms	64	1,725 (776)	59	1,871 (902)	-146 (-446 to 154)
ITU days	64	18,487 (16,145)	59	21,630 (15,966)	-3143 (-8,881 to 2,595)
Ward days	64	5,789 (4,418)	59	5,571 (4,161)	218 (-1,317 to 1,754)
Readmissions	64	1,361 (8,842)	59	143 (882)	1,218 (-1,072 to 3,508)
Other	64	15,857 (15,166)	58	21,214 (27267)	-5,357 (-13,171 to 2,457)
Germany					
Angiograms	62	2,129 (1,454)	55	1,763 (855)	366 (-78 to 810)
ITU days	62	11,432 (10,210)	56	14,594 (8,769)	-3,162 (-6,650 to 327)
Ward days	62	4,587 (5,562)	56	6,229 (7,179)	-1,643 (-3,972 to 687)
Readmissions	61	1,315 (4,752)	56	81.99 (614)	1,233 (-35.15 to 2,501)
Other	52	27,230 (34,406)	50	24,579 (27588)	2,651 (-9,629 to 14,932)
Canada					
Angiograms	48	2,619 (1,184)	41	2,478 (693)	141 (-277 to 559)
ITU days	48	18,137 (15,448)	40	29,754 (33,737)	-11,617 (-22,439 to -795)**
Ward days	48	7,784 (5,829)	40	9,918 (13,999)	-2,134 (-6,545 to 2,278)
Readmissions	46	2,168 (14,705)	41	0 (0)	2,168 (-2,401 to 6,737)
Other	46	13,722 (6,681)	38	21,529 (49031)	-7,806 (-22,331 to 6,718)
Finland					
Angiograms	34	1,354 (424)	29	2,401 (497)	-1,047 (-1,279 to -815)*
ITU days	34	3,614 (8,517)	29	5,233 (6,311)	-1,620 (-5,454 to 2,214)
Ward days	34	6,276 (3,206)	29	6,413 (2,332)	-137 (-1,572 to 1,298)
Readmissions	31	3,354 (6,827)	27	742 (1,822)	2,612 (-103 to 5,327)**
Other	30	14,339 (13,332)	27	6,134 (3908)	8,205 (2,864 to 13,546)*

*p<0.001; **p<0.05, All costs expressed in $US Purchasing Power Parity from 2008

Tables 7.1 and 7.2 suggest that in terms of the incremental costs between endovascular and neurosurgery, both costing approaches estimated similar results in all countries and categories of resource use except in Finland. Therefore, the results from these tables indicate

that there may be differences in the incremental cost per patient between the interventions using a one-country or a multicountry costing approach.

To explore this issue further, Table 7.3 provides information on the distribution of each resource use category's contribution to the total cost per patient for each multinational costing approach. The table reports the mean cost in the endovascular treatment and neurosurgical group, and the associated contribution to the total cost using either UK unit costs or country-specific unit costs. The United Kingdom and Canada are not included in the table because in these countries only UK unit cost data were used. A similar distribution of resource use contribution was observed when using data from all countries and using data from Sweden alone.

However, a different pattern of distribution is observed in Germany and Finland. In Germany the cost of days in ITU comprised 50% of the total cost per patient using UK unit costs whereas that proportion falls to 24% when using German unit cost data in the endovascular group. A similar result is obtained in the neurosurgical group for that category of resource use with ITU costs contributing 57% to the total cost using UK unit costs but only 31% when using German unit cost. A difference is also observed in Germany in the category "other" in both treatment groups. In the endovascular group and using UK unit costs the "other" category contributed 39% to the total cost per patient whereas using German unit costs this estimate increased to 58%. In the neurosurgical group, a similar increase was observed. In Finland the distribution of the resource use category contribution seems to be different between the two costing methods in the neurosurgical group. Here, the "other" category contributed to 43% of the overall total cost per patient using UK unit costs and 29% using Finish unit costs.

Table 7.3: Distribution of resource use category contribution to the total cost per patient by costing approach and trial arm

	Endovascular				Neurosurgery			
	UK unit costs		Country-specific unit costs		UK unit costs		Country-specific unit costs	
	Mean cost $US 2008 PPP	%	Mean cost $US 2008 PPP	%	Mean cost $US 2008 PPP	%	Mean cost $US 2008 PPP	%
All countries								
Angiograms	1,556	4%	1,565	4%	1,540	4%	1,546	4%
ITU days	9,860	23%	8,004	20%	11,673	27%	9,543	23%
Ward days	9,047	22%	9,012	23%	10,485	24%	10,450	25%
Readmissions	2,276	5%	2,183	6%	497	1%	475	1%
Other	19,232	46%	18,621	47%	19,783	45%	19,262	47%
Total	41,971	100%	39,385	100%	43,978	100%	41,276	100%
Sweden								
Angiograms	1,627	3%	1,725	4%	1,765	3%	1,871	4%
ITU days	23,547	45%	18,487	43%	27,550	45%	21,630	43%
Ward days	5,746	11%	5,789	13%	5,529	9%	5,571	11%
Readmissions	1,636	3%	1,361	3%	204	0%	143	0%
Other	20,263	38%	15,857	37%	25,907	43%	21,214	42%
Total	52,819	100%	43,219	100%	60,955	100%	50,429	100%
Germany								
Angiograms	2,007	3%	2,129	5%	1,663	2%	1,763	4%
ITU days	38,037	50%	11,432	24%	48,557	57%	14,594	31%
Ward days	4,783	6%	4,587	10%	6,496	8%	6,229	13%
Readmissions	1,968	3%	1,315	3%	151	0%	81.99	0%
Other	29,450	39%	27,230	58%	27,653	33%	24,579	52%
Total	76,245	100%	46,693	100%	84,520	100%	47,247	100%
Finland								
Angiograms	1,457	4%	1,354	5%	2,584	9%	2,401	11%
ITU days	4,133	10%	3,614	12%	5,985	20%	5,233	25%
Ward days	7,094	18%	6,276	22%	7,249	24%	6,413	31%
Readmissions	4,649	12%	3,354	12%	1,334	4%	742	4%
Other	22,775	57%	14,339	50%	12,905	43%	6,134	29%
Total	40,108	100%	28,937	100%	30,057	100%	20,923	100%

Table 7.4 compares the mean cost per patient from Table 7.3 using each multinational costing approach in each treatment group. Only one significant difference between costing methods was obtained. This was in the cost of ITU days and was observed when using data from all countries and using data from Germany alone. Using data from all countries there was a significant mean difference in the cost of ITU days (95% confidence interval) of $1,856 ($432 to $3,279) between costing methods in the endovascular group and a significant mean difference (95% CI) of $2,130 ($599 to $3,662) in the neurosurgery group. A similar finding was obtained when using data from Germany alone. A significant mean difference between costing methods is also observed in the "Other" category in the neurosurgery group using data from Finland.

Table 7.4: Mean difference between costing approaches for selected resource use categories by trial arm

	Endovascular			Neurosurgery		
	Mean cost (SD) $US 2008 PPP		Mean difference (95% confidence interval)	Mean cost (SD) $US 2008 PPP		Mean difference (95% confidence interval)
	UK unit cost	Country-specific unit cost		UK unit cost	Country-specific unit cost	
All countries						
Angiograms	1,556	1,565	-9.59	1,540	1,546	-6.05
	(760)	(777)	(-74.68 to 55.50)	(656)	(659)	(-61.84 to 49.74)
ITU days	9,860	8,004	1,856	11,673	9,543	2,130
	(18,585)	(14,821)	(432 to 3,279)**	(19,792)	(16,094)	(599 to 3,662)*
Ward days	9,047	9,012	34.68	10,485	10,450	34.37
	(8,895)	(8,894)	(-719 to 788)	(9,903)	(9,903)	(-806 to 875)
Readmissions	2,276	2,183	92.70	497	475	22.12
	(9,211)	(8,978)	(-685 to 870)	(3,043)	(3,001)	(-235 to 279)
Other	19,232	18,621	611	19,783	19,262	521
	(22,982)	(22,858)	(-1,403 to 2,626)	(32,033)	(32,159)	(-2,288 to 3,331)
Sweden						
Angiograms	1,627	1,725	-97.38	1,765	1,871	-106
	(732)	(776)	(-361 to 167)	(851)	(902)	(-425 to 214)
ITU days	23,547	18,487	5060	27,550	21,630	5920
	(20,564)	(16,145)	(-1,410 to 11,530)	(20,335)	(15,966)	(-749 to 12,589)
Ward days	5,746	5,789	-43.59	5,529	5,571	-41.95
	(4,385)	(4,418)	(-1,583 to 1,496)	(4,130)	(4,161)	(-1,555 to 1,470)
Readmissions	1,636	1,361	275	204	143	61.10
	(10,056)	(8,842)	(-3,038 to 3,587)	(1,227)	(882)	(-329 to 452)
Other	20,263	15,857	4,407	25,907	21,214	4693
	(18,500)	(15,166)	(-1,512 to 10,326)	(28,340)	(27,267)	(-5,535 to 14,921)
Germany						
Angiograms	2,007	2,129	-122	1,663	1,763	-101
	(1,371)	(1,454)	(-624 to 381)	(806)	(855)	(-415 to 213)
ITU days	38,037	11,432	26,605	48,557	14,594	33963
	(33,973)	(10,210)	(17,625 to 35,585)*	(29,178)	(8,769)	(25,833 to 42,094)*
Ward days	4,783	4,587	196	6,496	6,229	266
	(5,799)	(5,562)	(-1,824 to 2,216)	(7,486)	(7,179)	(-2,480 to 3,012)
Readmissions	1,968	1,315	653	151	81.99	69.21
	(7,347)	(4,752)	(-1,568 to 2,875)	(1,131)	(614)	(-273 to 411)
Other	29,450	27,230	2220	27,653	24,579	3074
	(32,556)	(34,406)	(-10,807 to 15,246)	(28,256)	(27,588)	(-8,006 to 14,154)
Finland						
Angiograms	1,457	1,354	103	2,584	2,401	183
	(457)	(424)	(-110 to 317)	(535)	(497)	(-88.29 to 455)
ITU days	4,133	3,614	519	5,985	5,233	752
	(9,740)	(8,517)	(-3,910 to 4,948)	(7,217)	(6,311)	(-2,813 to 4,317)
Ward days	7,094	6,276	818	7,249	6,413	836
	(3,624)	(3,206)	(-838 to 2,475)	(2,636)	(2,332)	(-473 to 2,145)
Readmissions	4,649	3,354	1,296	1,334	742	593
	(9,050)	(6,827)	(-2,781 to 5,372)	(3,080)	(1,822)	(-796 to 1,981)
Other	22,775	14,339	8,436	12,905	6,134	6,771
	(26,642)	(13,332)	(-2,529 to 19,400)	(8,059)	(3,908)	(3,283 to 10,258)*

*p<0.001; **p<0.05, All costs expressed in $US Purchasing Power Parity from 2008

Table 7.5 reports the mean total cost per patient at five years follow-up using a one-country and a multicountry costing strategy for all countries and participating countries data. The table also reports the mean difference and associated 95% CI between endovascular and neurosurgical treatments. The total cost per patient when using UK unit costs was higher than using a multicountry approach in all countries and each country separately. No significant differences were observed in the incremental costs between endovascular treatment and neurosurgery in all countries or in each country separately.

Table 7.5: Total cost per patient at five years follow-up a one country (UK unit costs) and a multicountry costing strategy in all countries and participating countries data

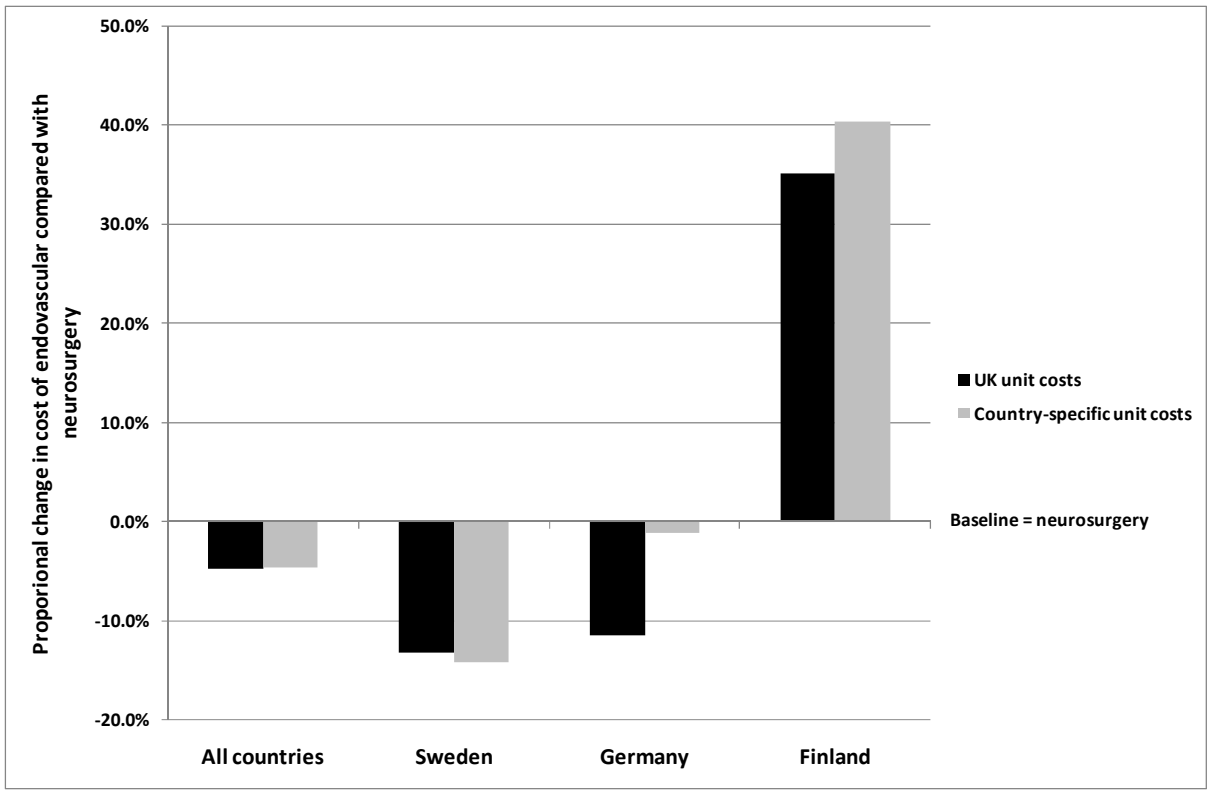
	Endovascular		Neurosurgery		
	n	**Mean cost per patient (SD) $US 2008 PPP**	**n**	**Mean cost per patient (SD) $US 2008 PPP**	**Mean difference (95% confidence interval) $US 2008 PPP**
All countries					
Total cost per patient using UK unit costs	1019	41,706 (36,812)	1028	43,783 (44,582)	-2,077 (-5,622 to 1,469)
Total cost per patient using country-specific unit costs	973	38,565 (32,631)	981	40,450 (42,500)	-1,885 (-5,249 to 1,479)
p –value difference between costing methods					0.939
Sweden					
Total cost per patient using UK unit costs	64	52,819 (33,446)	58	60,872 (40,846)	-8054 (-21,387 to 5,280)
Total cost per patient using Swedish unit costs	64	43,218 (50,387)	58	50,387 (36,090)	-7,168 (-18,572 to 4,236)
p –value difference between costing methods					0.921
Germany					
Total cost per patient using UK unit costs	52	75,688 (55,066)	50	85,487 (48,204)	-9,799 (-30,161 to 10,563)
Total cost per patient using German unit costs	52	47,043 (39,913)	50	47,598 (32,440)	-555 (-14875 to 13,766)
p –value difference between costing methods					0.462
Finland					
Total cost per patient using UK unit costs	30	40,623 (33,979)	27	30,057 (12,166)	10,566 (-3284 to 24416)
Total cost per patient using Finnish unit costs	30	29,358 (20,898)	27	20,919 (8,200)	8,439 (-167 to 17,046)
p –value difference between costing methods					0.794

*p<0.001; **p<0.05, All costs expressed in $US Purchasing Power Parity from 2008

The incremental costs using each multinational costing approach were compared parametrically and the p-value of the difference between methods is also reported in Table 7.5. The results suggest no significant differences in the incremental costs using either multinational costing approach.

This result is also observed in Figure 7.1, which indicates the proportional change in the total cost of endovascular treatment compared with neurosurgery by costing approach. The figure suggests a similar proportional change using data from Sweden, Finland, and across all countries. In Germany, however, the proportional change was much larger when using UK rather than German unit costs.

Figure 7.1: Proportional change in total cost of endovascular compared with neurosurgery by costing approach



7.2.2 Summary of the results from the comparison of multinational costing approaches

A one-country and a multicountry costing method produced similar results using the ISAT dataset when calculating incremental costs between the endovascular and neurosurgical treatments. For some categories of resource use, such as ITU days, mean costs per patient in the endovascular and neurosurgical groups were different between the multinational costing approaches. Furthermore, the distribution of each resource use category's contribution to the total cost per patient was different in Germany between costing approaches. However, all differences between methods were cancelled out when calculating the incremental cost per patient between the treatments.

The results from this comparison provide empirical data on a formal comparison of costing methods in a multinational clinical trial. Although this is valuable information, the results cannot be generalised to other multinational clinical trials in a straight forward manner. For example, it would be ideal to find out at what point multinational costing methods yield different results in the estimates of incremental costs. This and other economic aspects are explored next.

7.3 The simulation study

7.3.1 Background

The previous section has highlighted that it is difficult to assess the performance of multinational costing methods simply using available data from one international clinical trial. Little is known about when the two main costing approaches, one-country and multicountry, produce different estimates of incremental costs in a multinational clinical trial. In addition, Chapter 3 suggested that the underlying relationship of input factors and relative factor input prices across countries may well play an important role when estimating

incremental costs in each country. These issues are now explored in a simulation study that simulates mean volumes of resource use in the treatment arms of a hypothetical multinational clinical trial similar to the ISAT study.

The simulation in this chapter shares some characteristics of the experiment conducted by Raikou and colleagues in an article exploring the impact of using average or centre-specific unit costs in economic evaluations (Raikou et al. 2000). A review of this article was provided in Chapter 4. In that study, the authors simulated mean estimates of resource use of a particular health technology in a multicentre study using a production function relationship of two inputs. The simulation study in this chapter also assumes that the countries in the simulation have a production function that relates two inputs to achieve a certain level of output. In addition, instead of comparing average versus centre-specific costing strategy, the current simulation compares one-country and multicountry costing procedures. A step-by-step description of the simulation experiment is provided next.

7.3.2 Multinational costing approaches

A more formal mathematical definition of multinational costing approaches used in this section is provided as follows.

In a one-country costing method, a set of unit cost estimates from one country is applied to resource use in all countries. Defining V_a and V_b as the mean volume of resources of two inputs per patient, $C_a$and C_b the unit cost associated with each resource, and $i = 1, \dots, n$ the number of countries, country-specific mean costs in a one-country approach are calculated as follows:

$$(V_{ai} * C_a) + (V_{bi} * C_b) \qquad [7.1]$$

And the mean cost per patient across all countries in a one-country approach is:

$$\frac{1}{n}\sum_{i=1}^{n}\{(V_{ai} * C_a) + (V_{bi} * C_b)\} \qquad [7.2]$$

In [7.1] and [7.2], $C_a$and C_b do not vary across countries and the resource use in each country is multiplied by the unit cost from the available country.

However, if country-specific unit costs are available in each country then a multicountry costing analysis can be carried out and the country-specific mean cost is:

$$(V_{ai} * C_{ai}) + (V_{bi} * C_{bi}) \qquad [7.3]$$

The mean cost per patient across all countries in a multicountry approach is calculated as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\{(V_{ai} * C_{ai}) + (V_{bi} * C_{bi})\} \qquad [7.4]$$

In [7.3] and [7.4], C_{ai} and C_{bi} vary by country and the resource use in each country is multiplied by the country-specific unit cost.

7.3.3 The CES production function

The simulation exercise makes use of the theoretical framework described in Chapter 3, in particular the description in Section 3.5. A producer of health care is interested in treating patients with aneurysmal subarachnoid haemorrhage (aSAH) but is uncertain about whether a policy of endovascular treatment is more cost-effective than neurosurgical intervention. A multinational clinical trial similar to the ISAT study has been set up to compare both interventions. Endovascular treatment and neurosurgery are considered to be two different health technologies in this clinical trial.

In our experiment, one successfully treated patient in the clinical trial is achieved using a combination of intensive therapy unit days (x_1) and a number of days at a ward facility (x_2). Countries participating in this multinational clinical trial have available an endovascular and a neurosurgical technology that combine the input factors x_1 and x_2 according to a constant elasticity of substitution (CES) production function with the form.

$$V = \gamma\left[\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right]^{-1/\rho} \qquad [7.5]$$

where δ is the distribution parameter, ρ is the substitution parameter, γ the efficiency parameter, and V the desirable outcome (in this case one successfully treated patient).

The distribution parameter δ indicates the relative factor weight to achieve a level of output. It is assumed that the number of days in ITU has a higher weight than the number of days at a ward facility to represent that all patients have to go through an intensive therapy unit. The value of δ is assumed to be 0.6 in this exercise.

A recent costing analysis demonstrated that patients randomised to endovascular treatment spent less time in ITU and ward facilities overall than patients in the neurosurgical group (Wolstenholme et al. 2008). Therefore, the CES production function needs to incorporate this aspect in the health technology of each treatment. This is implemented through the efficiency parameter γ which indicates the state of the technology. Higher values of γ indicate a technology which is able to produce the level of output with fewer resources than a technology with lower values of γ. In this simulation exercise, γ equals 0.15 in the endovascular and 0.12 in the neurosurgery procedure to indicate that endovascular has an

associated shorter length of stay. These values were chosen to ensure that a difference was observed between the groups. No attempt was made to determine the magnitude of the difference, only that the difference was observable.

The substitution parameter ρ helps determine the elasticity of substitution between the number of ITU and ward days. The elasticity of substitution indicates how the producer of health is able to trade-off between the input factors to achieve the desired level of output along the isoquant. Because a CES production function is used, the elasticity of substitution is constant along the isoquant. Therefore, the experiment explores the effect of using different elasticity of substitution values in the calculation of the incremental costs in the simulated clinical trial. Empirical applications of production theory have estimated that the elasticity of substitution values in a capital-labour framework range from 0.1 to almost 2 (McFadden 1978). This is the range also used in the simulation to study the effect of the elasticity of substitution on the incremental costs between endovascular treatment and neurosurgery.

Finally V equals one indicating the combination of resources x_1 and x_2 needed to achieve one successfully treated patient.

7.3.4 Selection of number of countries

To be in line with the main dataset from this thesis, and because country-specific unit costs were available from the main contributing countries in the ISAT study, the UK, Sweden, Germany, and Finland were assumed to participate in this multinational clinical trial. In addition, Canada also participates in this trial but UK unit costs are used to calculate costs in this country. Unit cost data from each country are used to determine the level of resource use of each treatment arm and country, and this is expanded next.

7.3.5 The cost-minimisation problem

The producer of health care in each country faces the unit cost ω_1 for the cost of one day in an intensive therapy unit (ITU) and ω_2 for the unit cost of one day in a ward facility. Hence the total cost of treating one patient is given by the cost function $c(\omega_1, \omega_2, V) = \omega_1 x_1 + \omega_2 x_2$ where V is the functional form of the production function, in this case a CES. Once the relative price ($RP = \omega_1/\omega_2$) is computed, x_1 and x_2 are calculated from a cost-minimisation problem as explained in Section 3.3.1 in Chapter 3 as follows:

$$x_1 = \left[\frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta} x_2^{-\rho}\right]^{-1/\rho} \qquad [7.6]$$

$$x_2 = \delta^{1/\rho} \frac{V}{\gamma} \left[\left(\frac{(1-\delta)}{\delta} RP\right)^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta}\right]^{1/\rho} \qquad [7.7]$$

where δ is the distribution parameter, ρ is the substitution parameter, γ the efficiency parameter, V the desirable outcome and RP the value of relative factors input prices. A full mathematical description of the derivation of expressions 7.6 and 7.7 is given in Appendix 8.

Country-specific unit costs for ITU and ward days were available from the ISAT study, and this information was used to calculate the relative price in each country. This relative price is then used in equation [7.7] to calculate the number of days in a ward facility. Then, this figure is substituted in equation [7.6] to estimate the number of days in ITU. This process is repeated in each country.

This cost-minimisation process is assumed to occur in the five countries participating in this multinational clinical trial, but the producer of health care in each country faces different

relative prices and Chapter 3 showed that, as a result, the combination of ITU and ward days and the associated total cost per patient in the endovascular and neurosurgery groups are different in each country. The estimated values of x_1 and x_2 for the treatment groups in each country are considered the expected value or the mean resource use per patient in that particular country. Therefore, the outcome from the cost-minimisation problem produces an estimate of the mean number of days in ITU and ward in the endovascular and neurosurgery groups in each country. The resource use is multiplied by their respective unit costs in each country to obtain a total cost per patient in each treatment arm. The mean difference between interventions is then calculated and country-specific incremental costs are estimated. Country-specific mean costs are averaged in each treatment group to calculate the total cost per patient for each intervention across all countries. The mean difference between the endovascular and neurosurgical groups is then calculated to estimate the incremental cost across all countries.

Because one of the aims of the simulation is to understand when a one-country or a multicountry costing approach produces different results, the cost per patient in each treatment arm was computed using both approaches. For a one-country costing, the UK unit costs for ITU and ward days were used to calculate the cost per patient in each country. In a multicountry costing, country-specific unit costs for ITU and ward days were used to calculate the total cost in each treatment arm in each country. It is important to highlight that to estimate the way in which ITU and ward days are combined to produce a case in each country, the relative price is calculated using the country-specific unit costs. As a result, the "correct" total costs per patient in each arm of the trial and country are those calculated using country-specific unit costs.

For each experiment the simulation produces a set of mean number of ITU and ward days in each country. In addition, an estimate of the mean cost in each trial arm and associated incremental cost using the two multinational costing approaches are produced.

7.3.6 The simulation process

The description above explains how the volumes of ITU and ward days are calculated in each arm of the trial once the relative price is included in the cost-minimisation problem. For each country, an incremental cost between endovascular treatment and neurosurgery using either a one-country or a multicountry approach is computed. The idea of the simulation is then to create vectors of incremental costs in each country and across all countries using both costing methods, and different levels of elasticity of substitution. To achieve this, the unit costs of one day in an ITU and ward were simulated in each country imposing a normal distribution. The mean unit cost estimates for this distribution in each country were derived from the ISAT data collection.

The standard deviation around the mean unit cost estimates for both resource use in each country was estimated assuming different levels of variability defined by the coefficient of variation (Cv). Table 7.7 shows the mean unit cost estimate and the standard deviation in each country for three levels of coefficient of variation 0.2, 0.5, and 0.8 used in the simulation process. A few negative values were simulated using the normal distribution. These negative values were replaced by the observed mean unit cost estimate.

Table 7.6: Country-specific unit cost estimates used in the simulation study (unit costs reported in $US 2008 PPP)

	Mean estimate	SD (Cv=0.2)	SD (Cv=0.5)	SD (Cv=0.8)	Distribution
United Kingdom					
ITU	2,755	551	1,378	2,204	Normal
Ward	678	136	339	542	Normal
Sweden					
ITU	2,163	433	1,082	1,730	Normal
Ward	881	176	441	705	Normal
Germany					
ITU	828	166	414	662	Normal
Ward	828	166	414	662	Normal
Canada					
ITU	2,755	551	1,378	2,204	Normal
Ward	678	136	339	542	Normal
Finland					
ITU	2,409	482	1,205	1,927	Normal
Ward	583	117	292	466	Normal

Cv = coefficient of variation (Cv =SD/mean estimate)

For each random draw from the normal distribution, unit cost estimates in each country were simulated and an appropriate relative price in each country calculated. Then, estimates of incremental costs between treatment arms in each country and across all countries are computed. This simulation process was repeated 1,000 times and for different values of elasticity of substitution ranging from 0.1 to 2. For each value of elasticity of substitution, estimates of incremental costs between treatment arms in each country and across all countries were simulated 1,000 times.

For purposes of comparison with the simulation results, Table 7.7 summarises the results in terms of the total cost per patient in each arm of the trial and associated incremental cost observed in the ISAT study assuming patients only receive inpatient care to treat aSAH. The results are presented using the two multinational costing approaches when using data from all countries, and when using data from Sweden, Germany and Finland separately. The estimates of incremental costs suggests that neurosurgery is more expensive than

endovascular treatment in all cases, reaching statistical significance when using data from all countries, and using data from the United Kingdom, Germany, and Canada, separately. These estimates of incremental costs are compared later with the results from the simulation study to evaluate the accuracy of the simulation results. It is expected that if the CES production function is a reliable representation of the relationship between ITU and ward days in each country, the simulation model should replicate results similar to those observed in the ISAT study.

In terms of comparing the multinational costing approaches, no statistically significant differences were observed between the costing methods in any country or using data from all countries.

Table 7.7: Total cost per patient using only the resource use categories ITU and ward days to treat a patient with aSAH

	Endovascular		Neurosurgery		
	n	**Mean cost per patient (SD) $US 2008 PPP**	**n**	**Mean cost per patient (SD) $US 2008 PPP**	**Mean difference (95% confidence interval) $US 2008 PPP**
All countries					
Total cost per patient using UK unit costs	1,072	18,862 (21,365)	1,067	22,158 (21,462)	-3,296 (-5,112 to -1,480)*
Total cost per patient using country-specific unit costs	1,072	16,970 (18,673)	1,067	19,993 (18,718)	-3,023 (-4,609 to -1,438)*
p –value difference between costing methods					0.8249
United Kingdom					
Total cost per patient using UK unit costs	809	15,134 (14,992)	835	17,973 (16,503)	-2,838 (-4,365 to -1,312)*
Sweden					
Total cost per patient using UK unit costs	64	29,292 (20,744)	59	33,079 (18,940)	-3,787 (-10,897 to 3,324)
Total cost per patient using Swedish unit costs	64	24,276 (16,458)	59	27,201 (14,675)	-2,925 (-8,509 to 2,660)
p –value difference between costing methods					0.8504
Germany					
Total cost per patient using UK unit costs	62	42,820 (33,492)	56	55,053 (28,318)	-12,233 (-23,605 to -861)**
Total cost per patient using German unit costs	62	16,019 (10,777)	56	20,823 (9,901)	-4,804 (-8,591 to -1,018)**
p –value difference between costing methods					0.2217
Canada					
Total cost per patient using UK unit costs	48	25,922 (18,510)	40	39,672 (41,817)	-13,750 (-27,075 to -426)**
Finland					
Total cost per patient using UK unit costs	34	11,226 (10,146)	29	13,234 (7,435)	-2,008 (-6,558 to 2,543)
Total cost per patient using Finnish unit costs	34	9,889 (8,883)	29	11,646 (6,508)	-1,757 (-5,741 to 2,227)
p –value difference between costing methods					0.934

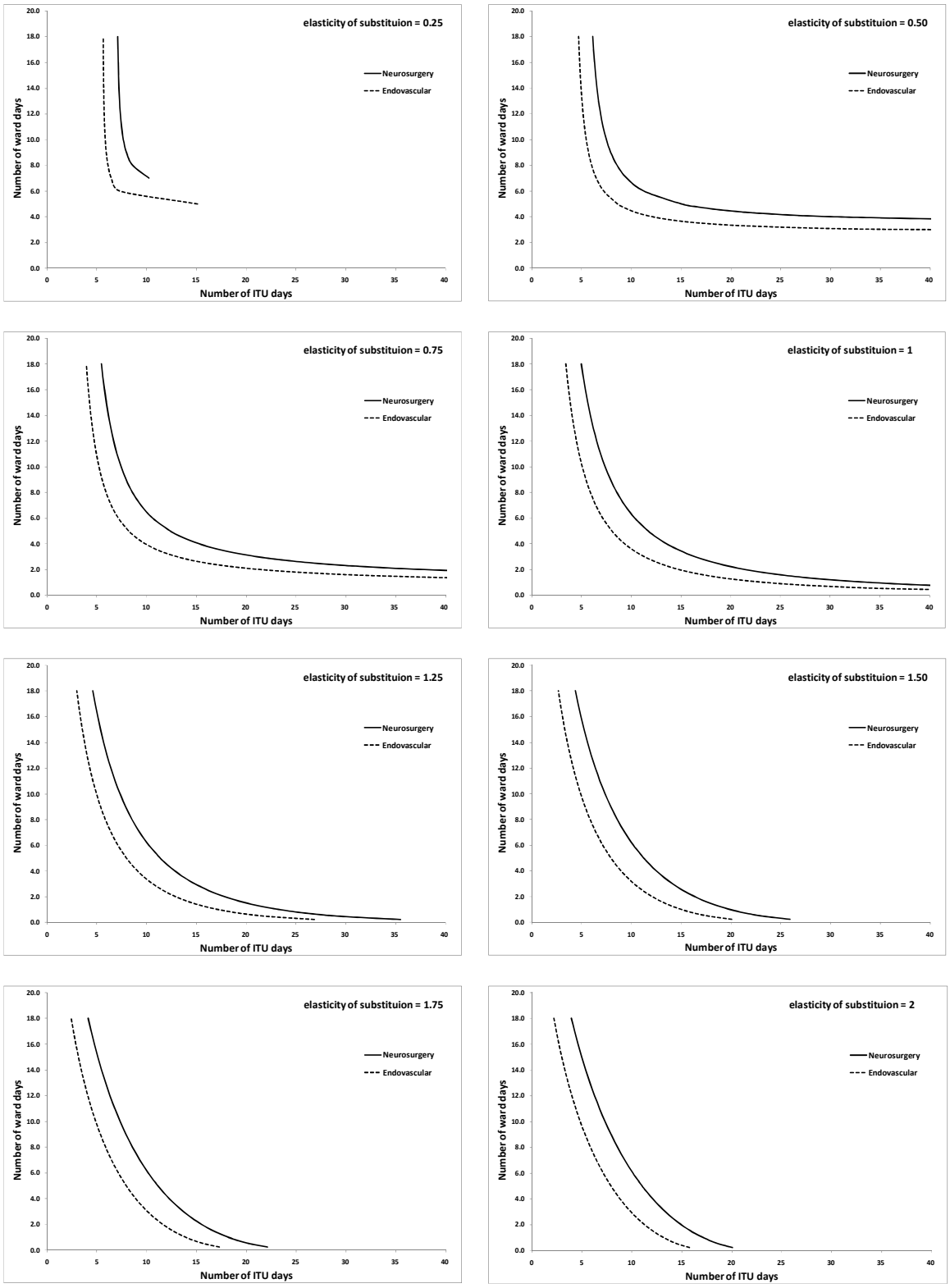
*p<0.001; **p<0.05, All costs expressed in $US Purchasing Power Parity from 2008

7.3.7 Results

Figure 7.2 provides a graphical account of the relationship between ITU and ward days in the endovascular and neurosurgical groups using the CES production function and different levels of elasticity of substitution used during the simulations. The graphs show that when the simulation assumes that the elasticity of substitution is close to zero, ITU and ward days behave similarly to perfect complements. When the elasticity of substitution equals one, ITU and ward days are combined using a Cobb-Douglas production function. Finally, for values close to 2, the relationship between the input factors is similar to the perfect substitutes case. Chapter 3 provided an overview of the CES production function and its association to common examples of isoquants in the literature.

When the relationship between ITU and ward days behaves similarly to perfect complements, a minimum number of days of both resources are needed to achieve a successfully treated patient. As a result, little substitution occurs in this case between the input factors. Moving from the perfect complement case to higher levels of elasticity of substitution, a more natural substitution between the input factors is allowed during the simulations.

Figure 7.2: Relationship between ITU and ward days in each country using a CES production for different values of elasticity of substitution



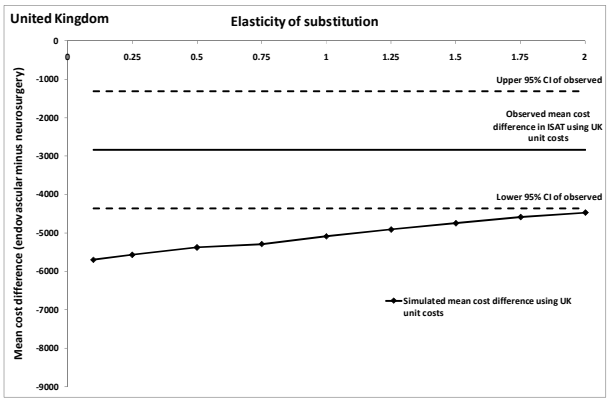
Figures 7.3 to 7.5 presents the main results from the simulation study using different levels of variation around the unit costs of ITU and ward days from 0.2 to 0.8. The graphs present the results for each of the countries separately and also for all countries together involved in the multinational clinical trial. To understand this graph it is useful to remember that in the one-country approach, costs in all countries were calculated using UK unit costs. Therefore, the UK graph only includes one line indicating how the incremental cost varies when the elasticity of substitution changes. In the remaining countries, two lines are drawn: one for the one-country costing (the solid line), and one for the multicountry costing (the dotted line). The figures also include the observed mean cost difference and associated 95% confidence interval from Table 7.7 using country-specific unit costs. In order to facilitate the interpretation of the results from the graphs, the figures have been formatted to provide a clear view of the results from the simulation but also from the observed results. In the case of Sweden, Canada and Finland it was not possible to present the observed mean estimates and confidence intervals with the output from the simulation study because a mislead graph was obtained. In those cases, partial information on the observed mean cost difference and 95% confidence interval is included to understand how observed and simulated data compare.

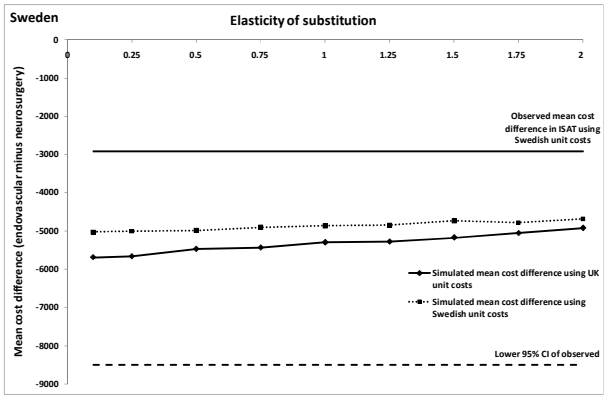
Figures 7.3 to 7.5 provide key information on the two main economic issues explored in this chapter: the effect of the elasticity of substitution on the incremental cost estimation in each country, and the effect of using different unit costs in the calculation of the incremental cost between endovascular treatment and neurosurgery.

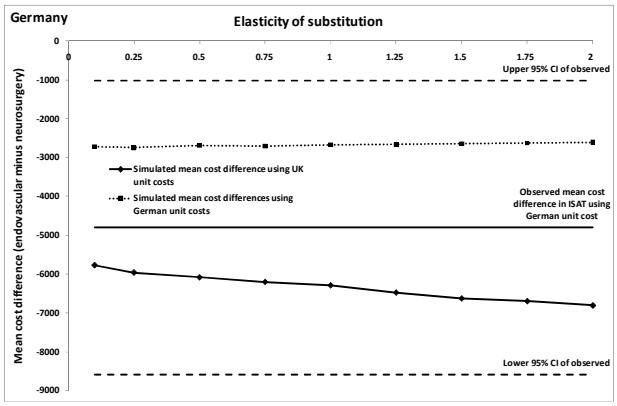
The estimation of the mean volume of days in ITU and ward in each country were simulated using the country-specific unit costs as described in the methods above. Therefore, the incremental cost calculated using the multicountry approach is considered to give the

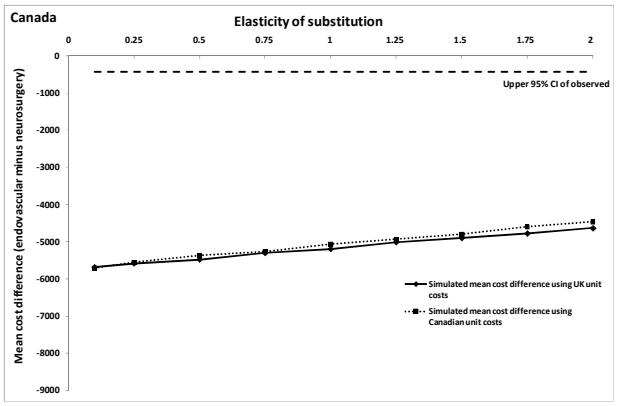
"correct" incremental costs in each country. In terms of evaluating whether using a one-country or a multicountry approach produce different results, the results from the figures suggest that in all countries and for the different values of coefficient of variation, using UK unit costs overestimates the "correct" incremental cost in each country. In Canada, when the coefficient of variation equals 0.2, little difference was observed between costing methods but this country used UK unit costs in the simulation exercise. Although the difference between costing methods was observed for the three levels of Cv, the difference was more marked when the coefficient of variation around the unit costs for ITU and ward equals 0.8. The difference between costing methods is maintained at all values of elasticity of substitution.

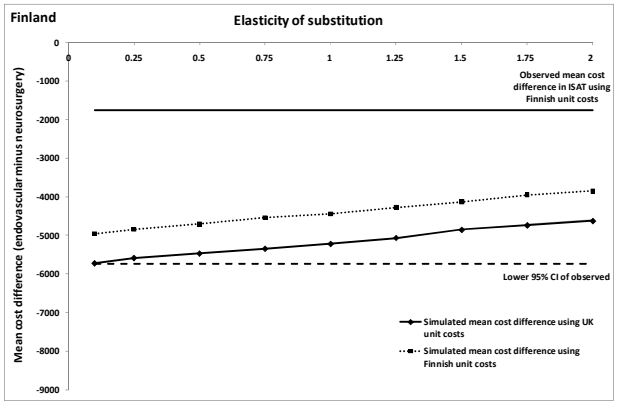
Figure 7.3: Mean cost difference between endovascular and neurosurgery for different values of elasticity of substitution by country (Cv on unit cost 0.2)











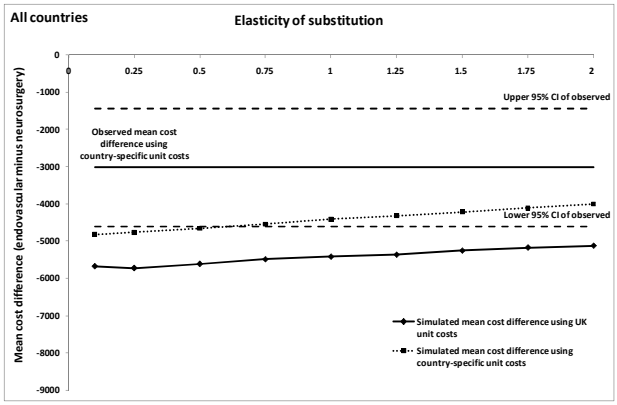


Figure 7.4: Mean cost difference between endovascular and neurosurgery for different values of elasticity of substitution by country (Cv on unit cost 0.5)

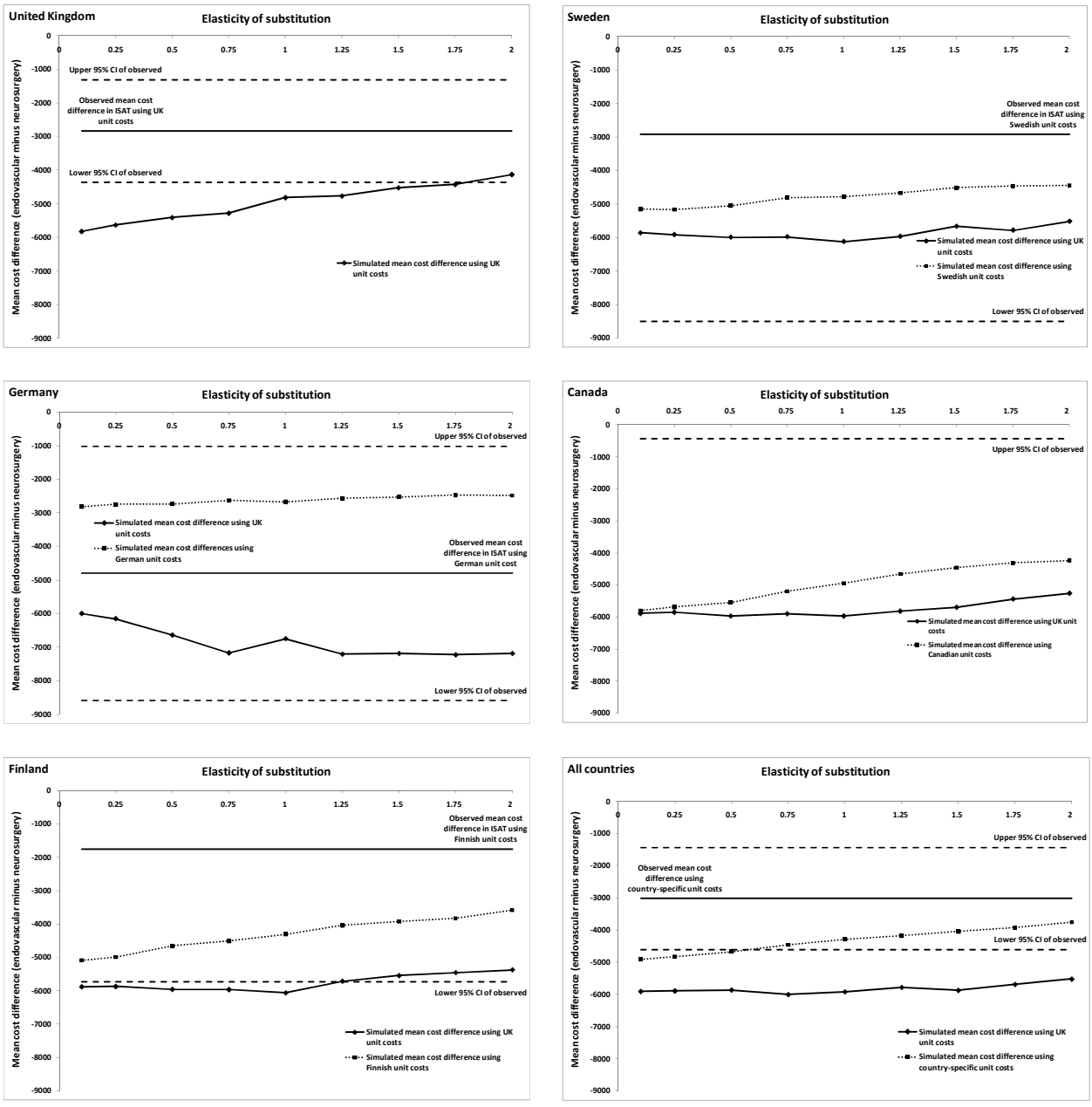
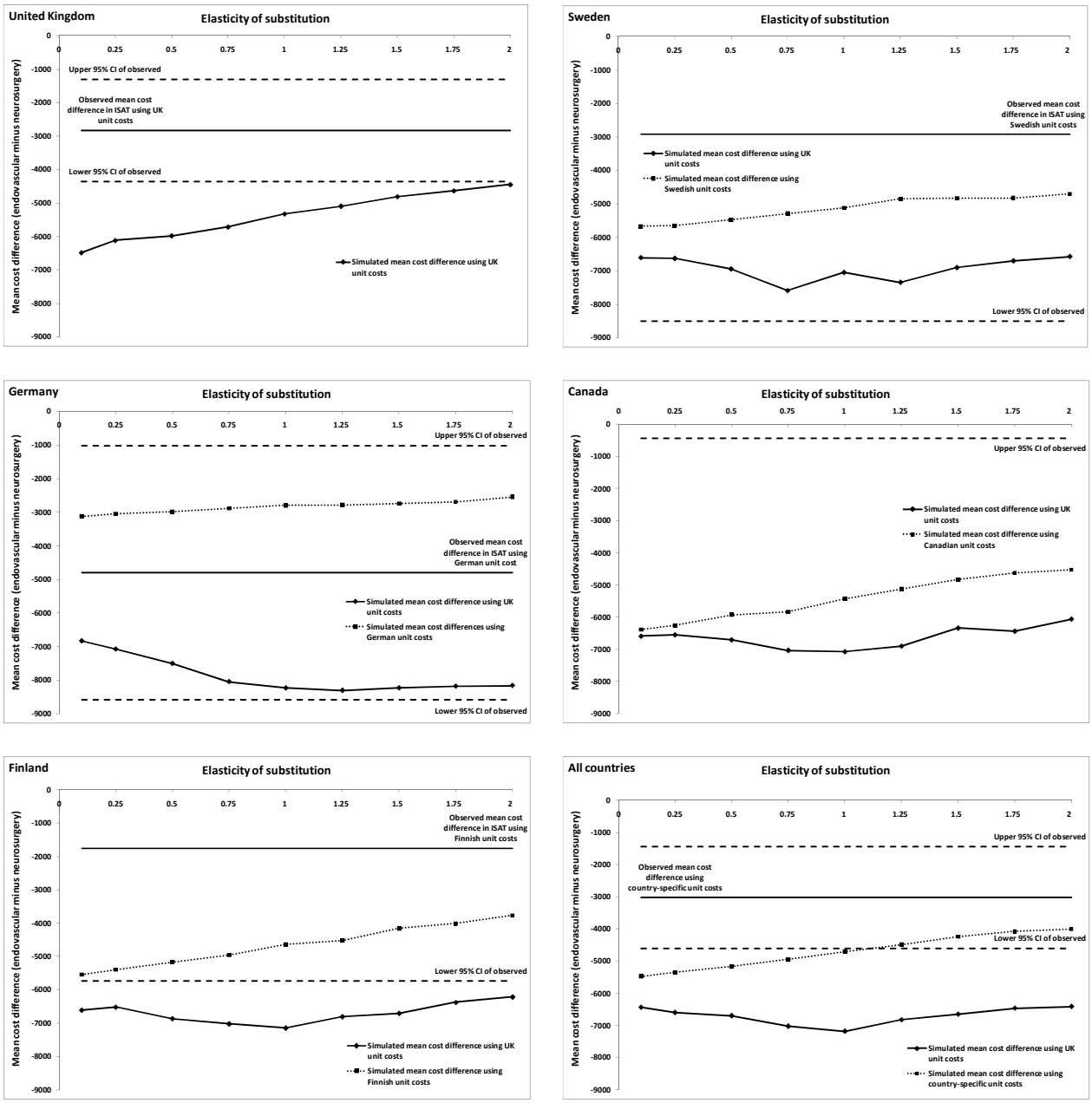


Figure 7.5: Mean cost difference between endovascular and neurosurgery for different values of elasticity of substitution by country (Cv on unit cost 0.8)



Figures 7.3 to 7.5 also suggest that there seems to be some variation in the incremental costs between endovascular treatment and neurosurgery across all countries for the different values of elasticity of substitution. A clearer picture of the variation in the incremental costs from figures 7.3 to 7.5 is presented in Table 7.8.

Table 7.8 shows the absolute difference between the incremental cost calculated for an elasticity of substitution of 2 and the incremental cost calculated using an elasticity of substitution of 0.1. These were the upper and lower limits used in the simulation study. The table also shows the minimum and maximum differences between incremental costs across all countries and costing approaches. The largest variation in the incremental cost is estimated to be $2,036 in all countries in the case of Cv=0.8 using UK unit costs. The size of the differences, reaching thousands in many cases, suggest that the substitution effect between the ITU and ward days in this simulation exercise may well have an important role in the estimation of incremental costs in the analysis.

Table 7.8: Absolute difference between incremental costs at elasticity of substitution equals to 2 and elasticity of substitution equals 0.1 (difference reported in $US 2008 PPP)

	Cv =0.2		Cv =0.5		Cv =0.8	
	UK unit costs	Country-specific unit costs	UK unit costs	Country-specific unit costs	UK unit costs	Country-specific unit costs
United Kingdom	1,230		1,687		2,036	
Sweden	767	341	344	714	35	973
Germany	1,033	117	1,183	335	1,330	574
Canada	1,059	1,243	620	1,567	533	1,861
Finland	1,097	1,113	506	1,502	396	1,782
All countries	555	811	385	1,152	21	1,468
Minimum difference	555	117	344	335	21	574
Maximum difference	1,230	1,243	1,687	1,567	2,036	1,861

Cv = coefficient of variation

Figures 7.3 to 7.5 also present the mean difference in costs using country-specific unit costs observed in the ISAT study and shown in Table 7.7. The 95% confidence interval associated with this mean difference is also presented. In the UK the simulation results of incremental costs lie clearly outside the lower confidence interval of the observed mean difference in costs in ISAT. This indicates that the simulation model overestimates the incremental costs

in the UK. When using data from all countries, a similar result is found in the simulated mean cost difference using UK unit costs. The simulated mean cost difference using country-specific unit costs is similar to the lower confidence interval of the observed mean cost difference.

In the remaining countries, the uncertainty around the observed mean difference in ISAT is wider than in the UK and using data from all countries, yielding wider confidence intervals. This indicates a considerable uncertainty in the observed mean difference in ISAT, complicating the comparison with the simulated results. Only in the case of Finland in Figure 7.5 it is clear that the simulated mean difference using UK unit costs overestimates the observed mean difference in ISAT. In the rest of cases, the results are only indicative. In Sweden the simulated results seem to have overestimated the observed mean difference in ISAT. In Germany, and Canada the simulated results seemed to have underestimated the observed mean difference in ISAT.

7.4 Summary and conclusions

This chapter has explored some important economic issues in cost-effectiveness analyses alongside multinational clinical trials using a simulation exercise. In particular, it has evaluated the impact of using different costing approaches and the result of substitution effects of input factors in the calculation of the incremental cost per patient in a hypothetical multinational clinical trial.

The chapter started by comparing a one-country and multicountry costing method using different categories of resource use and total cost per patient in the ISAT dataset. It was shown that for some categories of resource use, such as length of stay in Germany for

example, differences were observed between costing methods. However, when evaluating whether the multinational costing approaches yielded different estimates of total mean costs per patient in the endovascular and neurosurgical group and the associated incremental costs, no differences were detected. It seemed that when all categories of resources were added to calculate the total cost per patient, any differences between the costing methods in the incremental cost were cancelled out. The results from this analysis served as the starting point for a simulation study to compare the multinational costing approaches and to evaluate the impact of substitution effects in the calculation of incremental costs.

The outcome from the simulation study suggests that simulated mean differences between the endovascular treatment and neurosurgery using UK unit costs were different to the simulated results using a multicountry approach. In fact, Figures 7.3 to 7.5 showed that using UK unit costs simulated larger values of incremental cost compared to the multicountry simulated results. This result holds across all countries, for different levels of variation around unit costs and for different values of elasticity of substitution between the input factors. Because the volumes of resource use in each country were simulated using the available country-specific unit costs, the cost per patient calculated using the multicountry method was considered the "correct" estimate of total costs. Therefore, applying unit costs from one country to resource use across all countries in a trial is likely to produce biases in the country-specific incremental cost estimations, and therefore in the final cost-effectiveness estimates. If the aim of the analyst is to produce reliable country-specific cost results, an additional effort needs to be expended to ensure that reliable country-specific unit cost data are available for analysis.

The simulation study also explored the influence of substitution effects between input factors, in this case the number of days in ITU and ward, in the estimation of the

incremental costs between endovascular and neurosurgery in the simulated multinational clinical trial. The results from the experiment suggest that there was some variation in the incremental costs between treatment arms when applying a different elasticity of substitution in each country. Because it was not possible to estimate measures of uncertainty in the simulation model, the results from this part of the study should be interpreted with caution. However, the sizes of the differences among incremental costs were a good indicator of the impact of the substitution effects. These imply that the elasticity of substitution influences the estimate of country-specific incremental cost. This is an interesting result which has important implications in the current analysis of cost-effectiveness analyses alongside multinational clinical trials.

Both the standard regression and the multilevel models explored in this thesis can make use of information on substitution effects of the main categories of resources to inform the estimation of incremental costs. This can be done using the estimates of elasticity of substitution as country-specific covariates in the models. The use of country-specific covariates in the analysis of multinational clinical trials has been explored in the literature (Grieve et al. 2005; Manca et al. 2007). However, there is no evidence of using the elasticity of substitution between main input factors as a potential predictor of country-specific results. Little is known about the degree of substitution effects alongside the treatment pathway of a particular health technology in the literature. Hence, the link between clinical trials and the use of substitution effects to estimate cost-effectiveness estimates in multinational clinical trials is an area that should be expanded in future research.

The simulation study conducted here is obviously subject to certain limitations. Although one of the computational advantages of the simulations was its two-input factor world, in reality the treatment pathway of a health technology is a more complex process. Future

research should explore expanding this model to a more general case with several input-factors affecting the outcome of successfully treating a patient. This could potentially be explored using data envelopment analysis and linear programming methods (Jacobs et al. 2009).

It was not possible to estimate measures of uncertainty around the incremental costs across countries and this is also a limitation of the current study. With estimates of uncertainty such as standard errors around the incremental costs we could have formally evaluated the variation of incremental costs for different levels of elasticity of substitution. However, this was not possible and the results rely on the size of the differences across incremental costs instead. This was also a problem when comparing whether the multinational costing approaches produce different estimates of incremental costs.

The simulated model did not estimate values of incremental costs similar to the ISAT study in the UK and when using data from all countries. For the rest of countries, the results were not that clear given the uncertainty associated with country-specific mean differences in ISAT. However, comparing the results in terms of mean estimates only, the observed and simulated incremental costs were also different in Sweden, Germany, Canada, and Finland. This limitation does not invalidate the main results and message from the experiment but it is difficult to assess why these differences have occurred. It is possible that in the simulation model the CES production function may not be an accurate representation of the relationship between ITU and ward days across countries in the ISAT study. However, it is also possible that the country-specific data available from ISAT are not enough to produce a reliable estimate of the number of ITU and ward days in each particular country.

In summary, future evaluations in this area can benefit from the results from this chapter. Researchers planning to undertake a cost-effectiveness analysis alongside a multinational clinical trial need to give special consideration to the collection of unit costs across the countries participating in the study. The next chapter will make use of the conclusions so far from this thesis to report a case study in practice: a cost-effectiveness analysis alongside the International Subarachnoid Aneurysm Trial.

Chapter 8: Economic analyses of endovascular coiling versus neurosurgical clipping in patients with aneurysmal subarachnoid haemorrhage – evidence from the International Subarachnoid Aneurysm Trial (ISAT)

8.1 Introduction

This thesis has shown that the multinational nature of cost-effectiveness analysis alongside international randomised controlled trials introduces important challenges in the analysis and reporting of these studies. Chapter 6 for instance illustrated the importance of evaluating heterogeneity of treatment effect across countries and compared methods to report and present the results from these studies. Using the ISAT dataset, the chapter concluded that differences in some categories of resource use and health outcomes across countries were present, and that multilevel models were not a reliable candidate to present the results from ISAT.

In addition, Chapter 7 simulated a multinational randomised controlled trial using foundations from economic theory and information from ISAT to assess empirically the importance of collecting country-specific unit cost to conduct the cost analysis in such studies. The chapter showed that using different multinational costing approaches can lead to different estimates of incremental costs and this is likely to affect estimates of cost-effectiveness.

This chapter reports cost-effectiveness estimates of endovascular treatment versus neurosurgical clipping in patients with aneurysmal subarachnoid haemorrhage (aSAH) in the United Kingdom, Sweden, Germany, Canada and Finland. Making use of the research

output from previous chapters, the economic evaluation of the ISAT study is formally conducted.

Cost-effectiveness in this chapter is presented using the two multinational costing approaches explored in detail in the previous chapter, and therefore extends the cost comparison presented in Chapter 7.

The background to aSAH and the details of the randomised controlled trial were presented in Chapter 5, which also provides a description of the categories of resource use collected, and the strategy to collect country-specific unit costs. Quality-adjusted life years (QALYs) were used as the main outcome measure for the economic evaluation, and details of the QALY calculation are also presented in that chapter.

This chapter follows the standard structure of a scientific article (methods, results, and conclusions) to report the results from the cost-effectiveness analysis. Previous chapters have described in detailed the ISAT study and the statistical methods available to analyse and report cost-effectiveness alongside a multinational clinical trial. Therefore, this chapter does not provide extensive details of these aspects again, but a summary and guide to where more information can be found in the thesis is included.

8.2 Methods

8.2.1 Summary of the ISAT study

ISAT was a multinational randomised controlled trial across eleven countries in Europe, Canada, Australia and the USA. The aim of the study was to determine whether a policy of endovascular coiling was cost-effective compared with neurosurgical clipping in patients with

aSAH. The primary outcome of the clinical study was the change in patient's dependence or death at one year follow-up measured by the modified Rankin scale (mRS). Secondary objectives of the study included a quality of life assessment of patients and a long-term cost-effectiveness analysis.

Quality-adjusted life years (QALYs) were used as the main outcome measure in the economic evaluation. The EQ-5D instrument was used to collect quality of life, and the UK value set was applied to calculate utility values for the calculation of QALYs over the study period across countries. Detailed resource use data attributable to endovascular and neurosurgery interventions were collected using an extensive case record form. Besides, questionnaires on employment status and hospital readmissions were also included.

8.2.2 Unit cost collection

Unit cost data collection to weight resource use was primarily collected in the United Kingdom in the first instance. However, country-specific unit costs were also requested from the main contributing countries using postal questionnaires in Sweden, Germany, Canada and Finland. The country-specific unit cost questionnaires collected data on the main cost drivers in patients with aSAH. Unfortunately, centres from Canada did not provide any unit cost data. Missing unit cost items in Sweden, Germany and Finland were imputed using a market-basket method approach (Schulman et al. 1998), and UK unit cost data were used to weight Canadian resource use.

A one-country and a multicountry costing method were applied to calculate the total cost per patient in the endovascular treatment and neurosurgical groups. In a one-country method, unit cost estimates from the UK are multiplied by resource use in all countries. In a multicountry method, country-specific unit costs are attached to country-specific resource

use. Chapter 7 provided country-specific estimates using both costing methods to calculate costs in each arm of the trial and associated incremental costs. This chapter explores whether using these two multinational costing approaches produces different estimates of country-specific cost effectiveness.

8.2.3 Follow-up/perspective

Data was collected at two months and one year after randomisation and annually subsequently. Follow-up data collection still continues. This chapter presents a five-year time horizon for the cost-effectiveness analysis, and hence, costs and QALYs have been calculated over a five years period. A health care perspective approach was selected to present the results from the cost analysis. Therefore, only direct health care costs are reported.

8.2.4 Statistical methods

The incremental cost effectiveness ratio (ICER) in each country was defined as the mean estimate of differences in costs divided by the mean estimate of difference in QALYs between endovascular and neurosurgical treatments. Difference in costs and QALYs between endovascular treatment and neurosurgery in each country was derived using regression analysis. A full discussion on the choice of the statistical method to present the results from the ISAT study was presented in Chapter 6. However, a summary explanation is included in the next section.

Three regression models were conducted, two for costs and the other for QALYs. In each of the regressions the total cost per patient (using a one-country or a multicountry costing approach) and the number of QALYs for each patient were used as the dependent variable in each model. Therefore three regressions were estimated using the following model:

$$Y_{ij} = \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_{3j} N_{ij} + \beta_{4j} T_i N_{ij} + e_i \text{ [8.1]}$$

with i indicating the individual patient specified as $i = 1, \ldots, n$ and j indicating the country the patient was recruited from specified as j with $j = 1, \ldots, k$. Hence k is the total number of countries, in this case five. In order to avoid estimation problems with countries with small sample sizes, only data from the United Kingdom, Sweden, Germany, Canada and Finland were used in the regression analysis.

Y_{ij} is the dependent variable indicating the total cost per patient calculated using a one country or multicountry costing approach or the number of QALYs in each of the regression models, T_i is a treatment indicator that takes value 0 if the patient belongs to the endovascular groups and 1 if the patient belongs to the neurosurgical group, N_{ij} is a vector of dummy variables indicating the countries, and $T_i N_{ij}$ is a vector of interaction terms between the treatment arm and the countries.

β_0 is the overall mean estimate of total cost per patient or QALYs across all countries in the trial. β_1 is the treatment effect in the country of reference and it is the parameter of interest indicating the incremental cost and incremental QALYs between treatment arms in each of the regressions. In order to obtain an estimate of treatment effect (and associated standard error) for costs and QALYs in each country, the country of reference was changed to calculate these values for each country. Basically, each costs and QALYs regression was repeated five times (one for each country of ISAT) to obtain the parameter of treatment effect in each country. The associated coefficient β_{4j} of the interaction terms allows the treatment effect β_1 to vary across countries. Finally, e_i is the usual stochastic error term.

The total cost per patient in both costing approaches was estimated using a general linear model (GLM) assuming a link identity whereas the number of QALYs was estimated using ordinary least squares. A full description of model selection and full overview of results for different categories of resource use was presented in Chapter 6.

In order to be consistent with previous chapters, the treatment effect β_1 is defined as the mean estimate of costs or QALYs in the endovascular group minus the mean estimate of costs and QALYs in the neurosurgical clipping. Country-specific ICERs were calculated combining the incremental costs and QALYs estimated from the models. Because, the total cost per patient was calculated using the two multinational costing approaches, country-specific ICERs were computed using both costing methods.

Uncertainty around the country-specific ICERs was represented using confidence ellipses. Confidence ellipses indicate the joint distribution of costs and QALYs in the cost-effectiveness plane (van Hout et al. 1994). The ellipse assumes that the joint density function of differences in cost and effect follows a joint normal distribution. The correlation between costs and outcomes in each country was obtained using information from each particular country separately. Cost-effectiveness acceptability curves (CEACs) for each country and costing methods were derived to indicate the probability of endovascular treatment being cost-effective compared to neurosurgical clipping for different values of willingness to pay for QALYs gained (Briggs 2001). CEACs were derived parametrically from net benefits as described in Section 2.7 in Chapter 2.

8.3 Results

8.3.1 Overview of results from previous chapters

This section provides a summary of the main results achieved in Chapter 6. These results will help to interpret the cost-effectiveness estimates and the conclusions in this chapter.

Given the multinational nature of the study, tests for heterogeneity were conducted to detect important differences in the treatment effect between arms in the main categories of resource use, and QALYs. To identify the main categories of resource use a detailed cost analysis was conducted in the UK up to two years follow-up (Wolstenholme et al. 2008). The duration of the operation, the number of angiograms, length of stay and the number of readmissions after discharge were the categories of resource use identified. Table 6.7 in Chapter 6 shows evidence of heterogeneity of treatment effect in operation duration and the number of angiograms. In addition, the number of readmissions at five years follow-up suggested a borderline significance of country-by-treatment interaction. When evaluating country-by-treatment interaction in QALYs, there was evidence of heterogeneity of treatment effect for QALYs across countries (Table 6.8 Chapter 6). No heterogeneity was observed in the treatment effect for the total cost per patient at five years follow-up, using a multicountry costing method. Nevertheless it was decided that pooling data from all countries to present the results of the economic evaluation was not appropriate, given the results of the test for heterogeneity in resource use and QALYs. It may be that heterogeneity of treatment effect was diluted when calculating the total costs and as a result the test was no longer powerful enough to detect any difference.

Chapter 6 compared a standard regression and a multilevel approach as possible alternatives to present the results from multinational cost-effectiveness analyses. The chapter applied both approaches to the ISAT data, and concluded that the standard regression analysis

provided a better fit to the data. The comparison was made in terms of information criterion methods using the Akaike information criteria for model fit. The chapter also showed that the country-specific mean treatment effect and associated uncertainty in costs and QALYs estimated using standard regression analysis was more accurate than the multilevel alternative. The number of countries included in the analysis (in this case five) and the sample size in each country were the main limitation in the multilevel approach. Multilevel models would require a large number of countries to estimate an accurate between-country variance, and therefore country-specific standard errors through the use of shrinkage estimates. This was not possible in the ISAT dataset, where only five of the eleven participating countries provided enough information for statistical analysis. In addition, the UK contributed the majority of the data collected, and this was a problem when using the multilevel model.

For these reasons, standard regression analysis was the statistical method selected to present the cost-effectiveness results alongside the ISAT study.

8.3.2 Country-specific incremental cost and QALYs

Table 8.1 presents estimates of difference in QALYs between arms of the trial for each country from the regression models. The United Kingdom, Sweden and Canada present a point estimate of mean difference in QALYs favouring the endovascular treatment. In Germany and Finland an opposite effect was observed and the point estimate difference in QALYs suggests neurosurgery as the effective therapy. However, differences in QALYs reached a significance level only in the case of the United Kingdom with a mean difference and 95% confidence intervals of 0.161 (95% CI 0.08 to 0.25) favouring endovascular

treatment, and in Germany with a mean difference of -0.802 (95% CI -1.02 to -0.21) favouring neurosurgery.

The incremental costs per patient between the endovascular and neurosurgical arms using the two multinational costing approaches is also presented in Table 8.1. In the UK, Sweden, Germany and Canada, both costing approaches yielded similar results suggesting that endovascular was cheaper than neurosurgery although none of the differences reached a significant difference. In Finland, the incremental cost per patient between treatment arms shows a significant mean difference in costs per patient of $8,439 (95% CI $483 to $16,396) favouring the neurosurgical group between the arms of the trial using the multicountry approach.

Table 8.1: Incremental costs and QALYs between trial arm across countries, and associated incremental cost-effectiveness ratio (ICER) using different multinational costing approaches

Countries	Difference in QALYs (95% CI)	Difference in costs One-country costing (95% CI)*	ICER* One-country costing	Difference in costs Multicountry costing (95% CI)*	ICER* Multicountry costing
United Kingdom	0.161 (0.08 to 0.25)	-$1,402 (-$5,094 to $2,290)	-$8,722	-$1,402 (-$5,076 to $2,272)	-$8,722
Sweden	0.166 (-0.19 to 0.43)	-$8,054 (-$21,274 to $5,167)	-$48,529	-$7,168 (-$18,518 to $4,182)	-$43,195
Germany	-0.802 (-1.02 to -0.21)	-$9,799 (-$29,670 to $10,072)	$12,224	-$555 (-$14,510 to $13,401)	$692
Canada	0.317 (-0.25 to 0.6)	-$15,551 (-$40,027 to $8,926)	-$49,056	-$15,551 (-$41,195 to $10,093)	-$49,056
Finland	-0.126 (-0.04 to 0.37)	$10,566 (-$2,212 to $23,343)	-$84,086	$8,439** ($483 to $16,396)	-$67,165

*$US Purchasing Power Parity from 2008; **p<0.05

Table 8.1 also shows the country-specific point estimates for the incremental cost-effectiveness ratio using both costing approaches. Both costing methods yielded ICER point estimates in the same quadrant of the cost-effectiveness plane, and therefore yielded similar results. In the United Kingdom, Sweden and Canada the ICER is located in the south east quadrant suggesting that endovascular treatment is more effective and also cheaper than neurosurgical clipping. The ICER for Germany is located in the south west quadrant suggesting that endovascular treatment is less effective and cheaper than neurosurgical clipping, whereas in Finland the ICER is placed in the northwest quadrant indicating that endovascular is less effective and more expensive than neurosurgery.

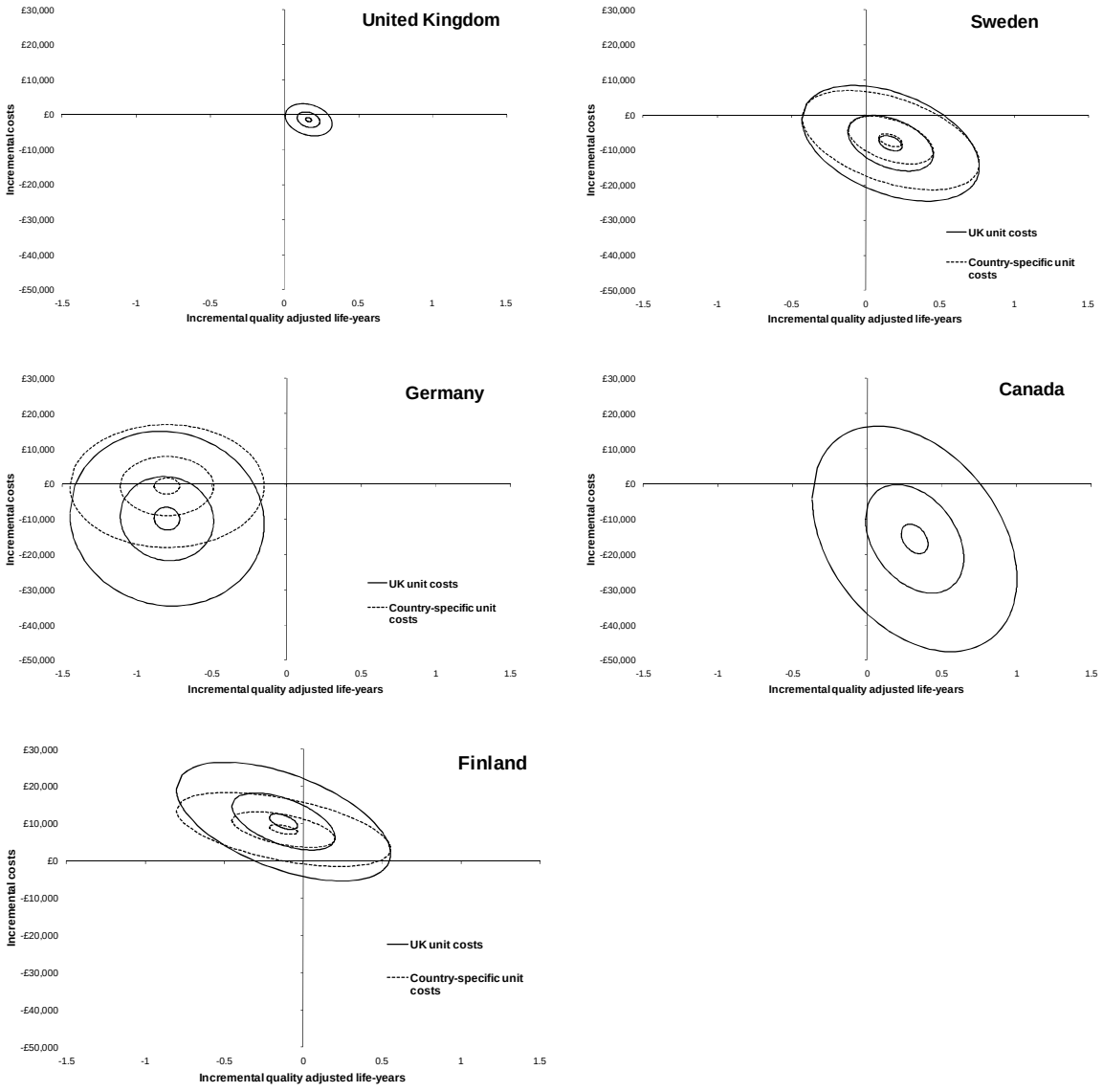
Table 8.2 reports the correlation between pairs of costs and QALYs differences by costing methods, indicating a negative correlation between costs and QALYs in most countries.

Table 8.2: Incremental costs and QALYs between trial arm across countries, and associated incremental cost-effectiveness ratio (ICER)

Countries	Correlation One-country (ΔQALY, ΔC)	Correlation Multicountry (ΔQALY, ΔC)
United Kingdom	-0.25	-0.25
Sweden	-0.41	-0.47
Germany	-0.06	-0.002
Canada	-0.36	-0.36
Finland	-0.54	-0.53

Figure 8.1 presents the joint distribution of costs and effect differences in each country by costing approach. 95%, 50% and 5% ellipses have been drawn to indicate different levels of uncertainty around the point estimate of ICER in each country and costing method.

Figure 8.1: 95%, 50% and 5% confidence ellipse of difference in costs and QALYs in the cost-effectiveness plane in each country and costing approach



The figure clearly shows the reduced uncertainty around the ICER in the case of the United Kingdom compared to the other countries. This result is mainly driven by the large sample size recruited in this country. In this case the case, the joint distribution of differences in

costs and effects lies in the south east and northeast quadrants. In the case of Sweden and Canada, the joint distribution lies across the four quadrants of the cost-effectiveness plane, although most of the distribution is placed in the northeast and southeast quadrants. Table 8.1 showed that Germany presented a significant difference in QALYs favouring the neurosurgical clipping, and the joint distribution of costs and effects in this country is located in the southwest and northwest quadrants. Finally, Finland shows most of the joint distribution of pairs of difference in costs and QALYs in the northwest and northeast quadrants. Comparing multinational costing approaches in Sweden, Germany and Finland, both approaches drew similar joint distribution in costs and effects around the ICERs.

Figures 8.2 and 8.3 present cost-effectiveness acceptability curves (CEACs) indicating the probability of endovascular treatment being cost-effective for different thresholds of willingness to pay for QALY gain in each country. Figure 8.2 presents the country-specific CEACs using a one-country costing approach and Figure 8.3 for the multicountry costing case. Endovascular treatment is cost-effective compared to neurosurgery in the UK, Canada, and Sweden even for small values of willingness to pay for health gain. For example, at a threshold of $40,000 per QALY gained the probability of endovascular being cost-effective is 99% in the United Kingdom, 92% in Canada and 85% in Sweden. The same estimates were obtained in both costing methods.

However, in Germany and Finland, the CEACs suggest that endovascular treatment is not cost-effective compared to neurosurgery. In the case of Germany and driven by the large QALY difference favouring neurosurgery, at a threshold of $40,000 per QALY gained, endovascular has a 7% probability of being cost-effective compared with neurosurgery using a one-country costing approach. This probability is reduced to 1% when using a multicountry approach. In Finland, the probability of endovascular treatment being cost-

effective tends to 20% for large values of willingness to pay for health gained using a one-country or a multicountry method.

Figure 8.2: Cost-effectiveness acceptability curves in each country indicating the probability of endovascular being cost-effective using a one-country costing approach

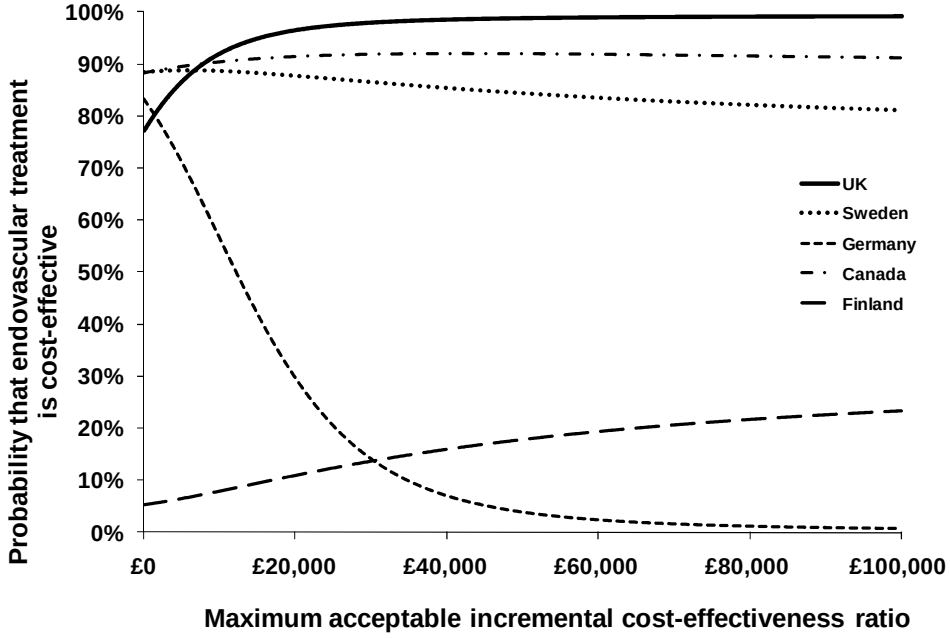
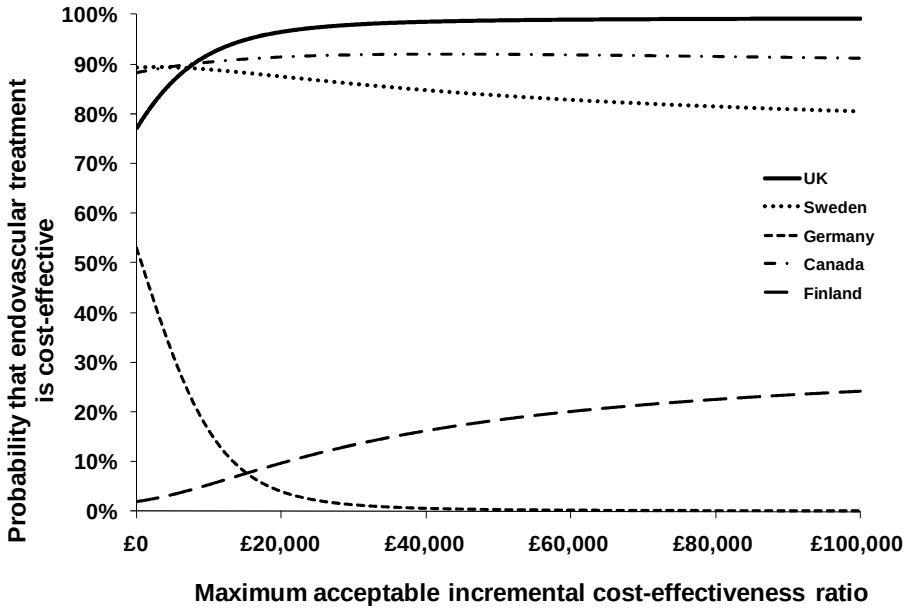


Figure 8.3: Cost-effectiveness acceptability curves in each country indicating the probability of endovascular being cost-effective using a multicountry costing approach



8.4 Summary and conclusions

This chapter has presented cost-effectiveness estimates of endovascular coiling versus neurosurgical clipping in patients with aSAH in the United Kingdom, Sweden, Germany, Canada and Finland.

Costs-effectiveness results were presented using the two multinational costing approaches explored in this thesis. The outcome from the comparison of using the multinational costing approaches suggested that both methods reached similar cost-effectiveness results using the ISAT data. To some extent, this was expected as Chapter 7 showed that both methods produced similar estimates of incremental total cost per patient when applying the methods to the ISAT dataset. Therefore, when cost and effects were combined to calculate ICERs, similar cost-effectiveness estimates of endovascular treatment versus neurosurgery were obtained. However, Chapter 7 also demonstrated in the simulation study that the two multinational costing methods can reach different estimates of incremental costs if resource use and unit costs are systematically related. As a result, it is recommended that country-specific unit costs should be collected routinely in multinational trials.

To report cost-effectiveness in each country standard regression analysis was chosen to derive the treatment effect in costs and QALYs. Although using a regression approach is similar to a fully-split analysis of the data, the approach has the advantage of incorporating more uncertainty in the standard errors because the regression recognised the multinational nature of the data through the dummies. As a result, inference on the coefficients of treatment effect in costs and QALYs is marginally improved. However, regression analysis also has some associated drawbacks that need consideration when interpreting the results reported in this chapter. Using data from each country reduces the sample size to estimate cost-effectiveness and as a result there is an immediate loss of power in the analysis. In the

case of the United Kingdom, this is probably not a serious concern given the contributing sample size in costs and outcomes in that country. However, the remaining countries may have experienced a reduction in power to detect any important differences in cost or effects. To some extent some of the uncertainty in the results is captured when combining costs and effects in the cost-effectiveness plane through ellipses and using cost-effectiveness acceptability curves. Even then, caution needs to be exercised when using the results reported in this chapter for decision-making purposes.

The alternative to the standard regression analysis is to use a multilevel model (MLM) that allows the treatment effect to vary randomly across countries. Chapter 6 showed that MLMs have several advantages compared to standard regression, in particular the estimation of the between-country variance for the treatment effect. With accurate estimation of the between-country variance, MLMs allow the calculation of shrinkage estimates. Shrinkage estimation is powerful because for those countries with large sample sizes, the estimation of a between-country variance allows a more accurate estimation of standard error associated with the country-specific treatment effect, improving the inference on these estimates. For countries with small sample size, shrinkage estimation assumes that the overall mean from the MLM is the best estimate for those countries with limited information, and in these countries, mean estimate of treatment effect is shrunk towards the overall mean. For this process to work properly and produce meaningful results, it is a requirement to have a sufficient number of countries and country-specific sample sizes. These are two limiting features in the ISAT dataset and as a result, Chapter 6 showed that MLMs did not provide a good fit to the data.

The results from the cost-effectiveness acceptability curves indicate the probability of endovascular treatment being cost-effective in each country. The CEACs for the UK, Sweden and Canada suggest endovascular treatment is a highly cost-effective strategy

compared to neurosurgical clipping. Overall, the three countries reported similar pathways of resource use and QALYs in Chapter 5 and 6, and hence similar cost-effectiveness in this chapter. Nevertheless, Canadian unit cost data was not available for analysis and UK unit cost data was used to weight Canadian resource use. This is a limitation for the cost-effectiveness estimates in Canada, and further investigation is needed to determine the generalizability of the results from this chapter.

In the case of Germany however, the CEAC suggests that endovascular treatment is not cost-effective compared to neurosurgery. This result is mainly driven by the treatment effect favouring neurosurgery from the regression analysis. One possible explanation for this result is the relative inexperience of radiologists conducting endovascular procedures in Germany at the time the trial was conducted compared to the other countries. Although a specified level of experience of the radiologist was an inclusion criterion in the original protocol, there may be a prolonged learning curve associated with the endovascular therapy. Therefore, it is possible that a new clinical trial in Germany would observe similar results to the UK, for instance. However, this needs further exploration in the future. Decision makers willing to use the information reported in this chapter need to complement the results here with further information from the literature or from clinical trials in aSAH in Germany.

The results from Finland are also subject to further exploration in the future. The CEAC for this country suggested low level of probability of endovascular being cost-effective when compared to neurosurgery. This country contributed 63 patients to the trial and therefore had the smallest contribution in the study. However, patients in the endovascular group in this country experienced significantly more readmissions at one and two years follow-up compared to neurosurgical patients. (A similar result was partially observed at two years follow-up in the United Kingdom which showed marginally more readmissions in the

endovascular group: see Chapter 5). These readmissions drove the cost-effectiveness result in Finland. Again, as in the case of Germany, further research is needed to understand better the resource use pathways in this country.

The next chapter wraps-up this thesis and summarises the output from all chapters and the main contribution of the research conducted in this research thesis. In addition, it includes recommendations for analysis of applied multinational clinical trials in the future and an agenda for future research in this area.

Chapter 9: Final summary and conclusions

The aim of this thesis was to clarify certain unanswered questions concerning the conduct of cost-effectiveness analyses alongside multinational clinical trials. Four main aspects were studied in depth during this research:

1) A structured literature review was undertaken to identify the frequency and the multinational nature of applied cost-effectiveness analysis alongside multinational clinical trials.

2) Three methods to detect heterogeneity of treatment effect across countries were evaluated, and two statistical methods to report country-specific cost-effectiveness estimates in multinational clinical trials were compared.

3) The thesis then formally linked economic evaluation methods to the economic theory of production, in particular the impact of substitution effects between input factors, and then explored the effect of this on economic evaluation of multinational clinical trials using a simulation exercise and different costing methods.

4) The cost-effectiveness of endovascular treatment compared with neurosurgical intervention for the treatment of aneurysmal subarachnoid haemorrhage was then estimated using data from the UK, Sweden, Germany, Canada and Finland.

This final chapter provides a brief summary and describes the main contributions of each of the chapters included in the thesis. Suggestions for directions for future research in this area are also included.

Chapter 1 introduced the context in which this thesis was undertaken and current research activity. It showed that economic evaluation alongside international trials has received

particular attention from the health economics community. Researchers have spent time identifying the main characteristics of this type of study, and have suggested statistical methods to analyse and report them. When this thesis was started, little was known about the multinational characteristics of the clinical trials involved in these studies. For example, estimates of the average number of countries participating in these studies and their contribution to the sample size had not been estimated. The chapter also showed that the research drawing on traditional economic theory to explain whether different multinational costing methods produce different estimates of cost-effectiveness was very limited. As a result, the aim of this thesis was to answer some of these questions.

In **Chapter 2** the background to economic evaluation of health care technologies in multinational clinical trials was explained. The chapter aimed to provide a clear overview of the main components of economic analysis, but was not meant to be an exhaustive exposition. Economic evaluation was defined and the measurement of costs and outcomes in cost-effectiveness analyses were described. The material in this chapter was intended in particular to aid the interpretation of the country-specific cost-effectiveness estimates and associated uncertainty reported in Chapter 8.

Chapter 3 explained the concept of a production function and its relationship to economic evaluation alongside multinational clinical trials. The chapter described how the behaviour of a potential decision maker in a particular country can be explained using aspects of the theory of production. If economic theory holds, a cost-minimisation problem can provide the optimal solution as to how to allocate resources in a particular setting. The solution to this problem depends on how input factors relate to one another, and the associated relative prices of these input factors. The relationship of input factors indicates the type of substitution that could occur for a particular health technology. This substitution parameter

played an important role in the simulation study conducted in Chapter 7 to determine whether different costing approaches produce different estimates of incremental costs across countries.

Chapters 1, 2 and 3 therefore provided the appropriate descriptive and theoretical background to understand this thesis, and the work carried out in the remaining of chapters.

Chapter 4 reported the structured review conducted to identify applied and methodological studies of cost-effectiveness analyses alongside multinational clinical trials. The review identified 65 applied and 20 methodological studies. The 20 methodological articles identified statistical methods and made suggestions concerning the collection of country-specific unit costs. A full exploration of the multinational nature of the 65 applied articles was conducted. The chapter presented extracted data on the number and identity of countries participating in economic evaluations of this type of study. In addition, information on the unit cost data collection conducted and the type of multinational cost-effectiveness analyses performed was also retrieved during the review.

The main outcome from the chapter was the clear picture that the reporting of the multinational nature of cost-effectiveness analyses alongside multinational clinical trials is very poor. The review identified that 57% of articles failed to report the country contribution to the overall sample size in the study. Moreover, 20% of articles did not acknowledge or recognise the multinational nature of the study in any section of the manuscript. These results are a concern because they suggest that most of the studies published to date are of limited value to health care decision makers. If decision makers cannot even identify which countries patients have been recruited from, then it is difficult to assess whether the results from the study can be applicable to their settings. Future guidelines on this topic must

recommend that authors report clearly the country contribution to the overall sample size of the trial (Drummond et al. 2009).

Chapter 5 described in detail the case study which has provided the data used during this research. The International Subarachnoid Aneurysm Trial (ISAT) was a large multinational clinical trial which had suitable characteristics to explore the study questions in this thesis. ISAT was multinational with eleven countries mainly from Europe and Canada that recruited 2,143 patients. However, the UK contributed with 1,644 (77%) of the overall sample and as a result drove the main cost and effectiveness results when pooling data from all countries. In addition, only five countries provided enough data for statistical analysis. The large proportion of patients contributed by one country is quite common in such studies, as shown in the review in Chapter 4. As a result, future multinational trials with an economic component similar to the ISAT study can benefit from the work conducted in this thesis.

Chapter 6 put into practice the statistical methods identified in the structured review on how to report and analyse cost-effectiveness estimates alongside multinational clinical trials. The chapter explored different methods to detect heterogeneity of treatment effect in resource use, QALYs, costs and net benefits. Three statistical methods were reviewed in depth: formal tests of heterogeneity, standard regression analysis, and multilevel models. This chapter formally evaluated these statistical methods using the ISAT dataset. The review in Chapter 4 suggested that the majority of papers in this area have explored heterogeneity of treatment effect across countries in costs and cost-effectiveness estimates, but not much attention has been paid to the different components of resource use. The tests for heterogeneity were therefore conducted here on the main cost drivers of patients with aneurysmal subarachnoid haemorrhage. The results from this analysis suggested that heterogeneity of treatment effect may be present in important cost components across

countries. Nevertheless, the tests were not able to detect any heterogeneity of treatment effect when adding up resource use to calculate the total cost per patient. It seems that any detectable heterogeneity of treatment effect was diluted when computing the total cost per patient. This outcome from Chapter 6 leads to the recommendation that analysts should look for important differences in the treatment effect on categories of resource use across countries. The results from the chapter also suggested that tests of heterogeneity can serve as an indicative tool to identify potential differences across countries. These indications will help the analyst to decide whether splitting or pooling the data is the appropriate option.

In Chapter 6, the comparison of statistical methods using the ISAT study suggested that MLMs did not provide the best fit to the data compared with the standard regression alternative. MLMs have been showed to have interesting characteristics for the analysis of country-specific estimates. For instance, MLMs allow the estimation of the between-country variance in treatment effect and the use of shrinkage estimates to reduce the uncertainty associated with the standard error of the estimates, thereby improving inference. In addition, MLMs allow the use of country-specific covariates to control for any differences across countries. These are characteristics in which standard regression analysis cannot compete and where MLMs have a huge potential to improve the reporting of cost-effectiveness analysis alongside multinational clinical trial. However, for MLMs to work properly, certain characteristics of the data are needed, in particular an adequate number of units at the country level and that the assumption of exchangeability of treatment effect across countries is valid. These are features that most economic evaluations alongside multinational clinical trials lack, as was shown in Chapter 4 and Chapter 6 respectively. This was also clearly the case in the ISAT study.

In the majority of cases, centres from different countries are recruited to trials with the aim of improving the overall sample size and hence the power of study. The selection of countries is normally an opportunity for the lead investigators to achieve a larger sample size in a shorter period of time. This approach may be an efficient and robust solution in a multinational trial of a drug with a solid protocol and a simple patient treatment pathway across countries. In this case, the main assumption that effectiveness results are applicable across countries may be valid.

However, in a surgical trial (such as ISAT) the treatment pathway of patients is going to be a more complex process. In this case the health care labour component, the length of stay and the management of adverse events and readmissions are going to be different across countries. Differences in the absolute and relative level of costs may also exist across countries, because health care systems vary, the funding and reimbursement of new technologies tend to be country-specific and the level of remuneration for health care labour is different across countries. These differences may lead to significantly different levels of absolute costs of the interventions. In addition, it is possible that differences in the treatment effect across country are also found. Chapter 6 suggested that this was the case in the ISAT study with evidence of heterogeneity across countries of treatment effect on certain categories of resource use. This suggests that in the economic analysis, the overall results from a multinational clinical trial may well not be representative of each individual country in the trial.

To ensure accurate reporting of country-specific cost-effectiveness results, the multinational nature of the study should be considered at the early stages of the design. In particular, this design should take into account the fact that to produce reliable country-specific results will require an adequate number of patients recruited from that country.

The number and identity of countries participating in the trial should also be discussed at early stages of the design. Lead investigators designing these studies need to understand that it may be better to have a small number of countries participating in the trial with a large sample size in each country, than a large number of centres each contributing a small proportion of the overall sample. In the case of ISAT, for instance, it is clear that enough information is available to inform UK decision-makers. In addition, Sweden and Canada which seem to have the same treatment effect in costs and effects as in the UK, have the potential to benefit from ISAT. However, this thesis has shown the difficulties of providing reliable cost-effectiveness information for those countries with opposite treatment effect results and/or a small number of patients recruited.

In **Chapter 7** the two main multinational costing methods, one-country and multicountry, were compared and assessed using a simulation study with input from the ISAT trial. The chapter explored the impact of the substitution effect, measured by the elasticity of substitution, in both methods. The simulation assumed that resource use was determined using the economic theory foundations from Chapter 3. The output from the simulation suggested that different estimates of incremental costs were obtained depending on the method used. This result holds for all values of elasticity of substitution studied. Because country-specific unit costs were used to simulate resource use, the estimates of incremental costs using country-specific unit costs were considered the "correct" estimates. Therefore, the use of a one-country costing to report country-specific estimates is likely to cause bias in the results of the analyses.

Chapter 7 also reported that the incremental cost between endovascular treatment and neurosurgery for different levels of elasticity of substitution was subject to some variation. This is an interesting result that needs further exploration in future research. If the parameter

representing the substitution effect can inform estimates of incremental costs, it may be possible to include this information in the regression models explored in Chapter 6. However, little is known about the substitution effect of input factors in trials in general, and this may need to be carefully re-examined in the future.

The outcomes from Chapters 4-7 suggest that there are multinational characteristics that need special attention when designing the cost-effectiveness plan alongside a multinational clinical trial. Table 9.1 summarises these areas and provide a brief description. Future studies should discuss these areas during the development of the economic protocol.

Table 9.1: Multinational characteristics to consider when developing the economic evaluation protocol alongside a multinational clinical trial

Item	Description
The nature of the technology	Researchers should pay attention to whether the intervention is a medicine, or a labour intensive procedure. The treatment pathways in both type of technologies tend to differ and as a result, the number of cost categories that may be different across countries.
The number of countries participating in the study	Researchers should discuss the countries that will take part in the study. Ideally, the treatment pathway of a patient participating in the trial should be similar across countries to ease the analysis of the data. To clarify this aspect, country specific data collection is needed.
The contributing sample size from each country	The sample size of the overall trial is determined through power calculations, but countries contributing with small samples area likely to rely on data from countries with more contribution to derive cost-effectiveness. This is a limitation of the current methodology and should be acknowledged at earlier stages of the design.
Collection of country-specific data	A template with country-specific categories resource use from the treatment pathway of the technology under evaluation should be prepared. This exercise will provide valuable data that could potentially identify differences and inform the modelling section during the analysis.
The type of costing analysis to implement	A decision should be made at early stages specifying the countries where cost-effectiveness will be derived. This will drive the type of costing study needed and will inform whether additional country-specific unit cost data is needed.

The clinical results from the ISAT study had an important impact in decision making, changing clinical practice worldwide. Endovascular treatment has become the first line of

action when treating aSAHs similar to those in ISAT. However, when the results were originally published, the authors suggested that they observed a larger number of readmissions in the endovascular group that needed further research and concluded that evidence was also required on the cost-effectiveness of both therapies. This was explored in **Chapter 8** which produced country-specific cost-effectiveness estimates in the countries participating in the ISAT study. The results suggested that endovascular treatment is highly cost-effective in the UK, Canada and Sweden at five years follow-up. However, in Germany and Finland, the information from ISAT needs to be complemented with more evidence from each of these countries before reaching a final decision on the cost-effectiveness of the interventions in these settings.

Since the literature review was completed, the author of this thesis has found one more useful piece of methodological work, published in December 2009. The article by Gauthier and colleagues (Gauthier et al. 2009) explored the use of shrinkage estimates to calculate treatment effects in resource use of a multinational clinical trial. Once shrinkage estimates are available for each item of resource use, the authors suggested multiplying these shrinkage estimates of resource use by country-specific unit costs to calculate country-specific incremental costs. This approach is in line with the suggestions made in Chapter 6 from this thesis; however, the shortcomings associated with limited number of countries in MLMs still apply.

Research from multinational clinical trials will continue to increase in the literature and these studies will likely include a health economic component. The burgeoning need for cost-effectiveness evidence worldwide will ensure that this will be the case. However, this thesis has demonstrated that there is potential for improvement in the conducting of economic evaluation alongside multinational clinical trials. In particular, this thesis advocates that

studies carefully consider their multinational nature at the design stage, the appropriate exploration of heterogeneity of treatment effect in resource use across countries, and the importance of collecting country-specific unit costs in these studies.

References

Anon. (1999) St. Anthony's DRG Guidebook 2000. Reston, VA: Saint Anthony Pub.

Anon. (2002) Towards a better understanding of multinational economic evaluations. Drugs and Therapy Perspectives; **18**: 20-24.

Ades, A.E., Claxton, K., and Sculpher, M. (2006a) Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Economics; **15**: 373-381.

Ades, A.E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N., and Lu, G. (2006b) Bayesian methods for evidence synthesis in cost-effectiveness analysis. PharmacoEconomics; **24**: 1-19.

Aiyar, S. and Dalgaard, C.J. (2009) Accounting for productivity: Is it OK to assume that the world is Cobb-Douglas? Journal of Macroeconomics; **31**: 290-303.

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control; **19**: 716-723.

Altman, D.G., Machin, D., Bryant, T.N., and Gardner, M.J. (2000) Statistics with confidence. confidence intervals and statistical guidelines. 2nd Edition, Bristol: BMJ.

Alvord, E.C., Loeser, J.D., Bailey, W.L., and Copass, M.K. (1972) Subarachnoid hemorrhage due to ruptured aneurysms. A simple method of estimating prognosis. Archives of Neurology; **27**: 273-284.

Amerini, G. (2000) Chapter 6: Purchasing Power Parities for Medical Care and Health Expenses. International and Interarea Comparisons of Income, Output and Prices. In Heston, A. and Lipsey, R.E. (Eds.). Chicago: Chicago University Press.

Antonanzas, F., Rodriguez-Ibeas, C., Juarez, C., Hutter, F., Lorente, R., and Pinillos, M. (2009) Transferability indices for health economic evaluations: methods and applications. Health Economics; **18**: 629-643.

Armstrong, S.H. and Ruckley, C.V. (1997) Use of a fibrous dressing in exuding leg ulcers. Journal of Wound Care; **6**: 322-324.

Arrow, K.J., Chenery, H.B., Minhas, B.S., and Solow, R.M. (1961) Capital-Labor Substitution and Economic Efficiency. The Review of Economics and Statistics; **43**: 225-250.

Aspelin, P., Aubry, P., Fransson, S.G., Strasser, R., Willenbrock, R., and Lundkvist, J. (2005) Cost-effectiveness of iodixanol in patients at high risk of contrast-induced nephropathy. American Heart Journal; **149**: 298-303.

Assessment of the Safety and Efficacy of Tenecteplase Group (2001) Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: the ASSENT-3 randomised trial in acute myocardial infarction. Lancet; **358**: 605-613.

Bachert, C., Vestenbaek, U., Christensen, J., Griffiths, U.K., and Poulsen, P.B. (2007) Cost-effectiveness of grass allergen tablet (GRAZAX) for the prevention of seasonal grass pollen induced rhinoconjunctivitis - a Northern European perspective. Clinical & Experimental Allergy; **37**: 772-779.

Barbieri, M., Drummond, M., Willke, R., Chancellor, J., Jolain, B., and Towse, A. (2005) Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: lessons for inferring generalizability. Value in Health; **8**: 10-23.

Bennett-Guerrero, E., Sorohan, J.G., Gurevich, M.L., Kazanjian, P.E., Levy, R.R., Barbera, A.V., White, W.D., Slaughter, T.F., Sladen, R.N., Smith, P.K., and Newman, M.F. (1997) Cost-benefit and efficacy of aprotinin compared with epsilon-aminocaproic acid in patients having repeated cardiac operations: a randomized, blinded clinical trial. Anesthesiology; **87**: 1373-1380.

Billingham, L.J., Abrams, K.R., and Jones, D.R. (1999) Methods for the analysis of quality-of-life and survival data in health technology assessment. Health Technology Assessment; **3**: 1-152.

Black, W.C. (1990) The CE plane: a graphic representation of cost-effectiveness. Medical Decision Making; **10**: 212-214.

Boersma, C., Radeva, J.I., Koopmanschap, M.A., Voors, A.A., and Postma, M.J. (2006) Economic evaluation of valsartan in patients with chronic heart failure: Results from Val-HeFT adapted to The Netherlands. Journal of Medical Economics; **9**: 121-131.

Bound, J., Jaeger, D.A., and Baker, R.M. (1995) Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak. Journal of the American Statistical Association; **90**: 443-450.

Briggs, A. and Gray, A. (1999) Handling uncertainty when performing economic evaluation of health care interventions. Health Technology Assessment; **3**: No-2.

Briggs, A., Sculpher, M., and Claxton, K. (2006a) Decision Modelling for Health Economic Evaluation. 1st Edition, Oxford: Oxford University Press.

Briggs, A.H. (2001) Chapter 8: Handling uncertainty in economic evaluation and presenting the results. Economic evaluation in health care: merging theory with practice. In Drummond, M. and McGuire, A. (Eds.). Oxford: Oxford University Press.

Briggs, A.H., Bousquet, J., Wallace, M.V., Busse, W.W., Clark, T.J., Pedersen, S.E., Bateman, E.D., and GOAL Investigators Group (2006b) Cost-effectiveness of asthma control: an economic appraisal of the GOAL study. Allergy; **61**: 531-536.

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

Briggs, A.H. and O'Brien, B.J. (2001) The death of cost-minimization analysis? Health Economics; **10**: 179-184.

Briggs, A.H., Wonderling, D.E., and Mooney, C.Z. (1997) Pulling Cost-Effectiveness Analysis Up by Its Bootstraps: A Non-parametric Approach to Confidence Interval Estimation. Health Economics; **6**: 327-340.

Britz, G.W., Newell, D.W., West, G.A., and Lam, A. (2003) The ISAT trial. The Lancet; **361**: 431-432.

Brouwer, W., Rutten, F., and Koopmanschap, M. (2001) Chapter 4: Costing in economic evaluations. Economic evaluation in health care: merging theory with practice. In Drummond, M. and McGuire, A. (Eds.). Oxford: Oxford University Press.

Brown, J., Nichols, D.A., and Meyer, F.B. (2003) The ISAT trial. The Lancet; **361**: 433-433.

Brown, R. and Armstrong, P. (2003) Cost effectiveness in Canada of eptifibatide treatment for acute coronary syndrome patients using PURSUIT subgroup analysis. Canadian Journal of Cardiology; **19**: 161-166.

Burstrom, K., Johannesson, M., and Diderichsen, F. (2006) A comparison of individual and social time trade-off values for health states in the general population. Health Policy; **76**: 359-370.

Cairns, J. (2001) Chapter 10: Discounting in economic evaluation. Economic evaluation in health care: merging theory with practice. In Drummond, M. and McGuire, A. (Eds.). Oxford: Oxford University Press.

Camps, J. (2004) Conducting efficient multi-country clinical studies. American Pharmaceutical Outsourcing; **5**: 39-42.

Canonica, G.W., Poulsen, P.B., and Vestenbaek, U. (2007) Cost-effectiveness of GRAZAX for prevention of grass pollen induced rhinoconjunctivitis in Southern Europe. Respiratory Medicine; **101**: 1885-1894.

Chaikin, P., Durrleman, S., Berger, F., Reinhoudt, J., Farr, I., and Ford, M. (2000) Insourcing international clinical trials to affiliates: A business-driven and medically sound strategy. Drug Information Journal; **34**: 1293-1301.

Chiang A.C. and Wainwright K. (2005) Fundamental Methods of Mathematical Economics. 4th Edition, London: Mc Graw-Hill.

Clinical Effectiveness Unit. (2006) National Study of Subarachnoid Haemorrhage. London: The Royal College of Surgeons of England and the London School of Hygiene and Tropical Medicine.

Cobb, C.W. and Douglas, P.H. (1928) A Theory of Production. The American Economic Review; **18**: 139-165.

Cook, J.R., Drummond, M., Glick, H., and Heyse, J.F. (2003) Assessing the appropriateness of combining economic data from multinational clinical trials. Statistics in Medicine; **22**: 1955-1976.

Curtis, L. (2008) Unit Costs of Health and Social Care. University of Kent at Canterbury: Personal Social Services Research Unit.

Curtis, L. (2009) Unit Costs of Health and Social Care. University of Kent at Canterbury: Personal Social Services Research Unit.

Dasbach, E.J., Rich, M.W., Segal, R., Gerth, W.C., Carides, G.W., Cook, J.R., Murray, J.F., Snavely, D.B., and Pitt, B. (1999) The cost-effectiveness of losartan versus captopril in patients with symptomatic heart failure. Cardiology; **91**: 189-194.

Davey, P., Craig, A.M., Hau, C., and Malek, M. (1999) Cost-effectiveness of prophylactic nasal mupirocin in patients undergoing peritoneal dialysis based on a randomized, placebo-controlled trial. Journal of Antimicrobial Chemotherapy; **43**: 105-112.

de Silva, J. (1997) Current practical issues in Phase III clinical trials. Drug Information Journal; **31**: 1155-1156.

Department of Health (2006) NHS Reference costs 2006. Department of Health .URL: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_062884 (Last Accessed: 8 January 2010)

Department of Health (2008) NHS Finance Manual. Financial Matters; Newsletter. Department of Health .URL: http://www.info.doh.gov.uk/doh/finman.nsf/Newsletters?OpenView (Last Accessed: 8 January 2010)

Detournay, B., Huet, X., Fagnani, F., and Montalescot, G. (2000) Economic evaluation of enoxaparin sodium versus heparin in unstable angina. A French sub-study of the ESSENCE trial. PharmacoEconomics; **18**: 83-89.

Dolan, P. (1997) Modeling valuations for EuroQol health states. Medical Care; **35**: 1095-1108.

Dolan, P. (2000) : The measurement of health-related quality of life. Handbook of Health Economics. In Culyer, A.J. and Newhouse, J.P. (Eds.). Amsterdam: Elsevier.

Donaldson, C. and Gerard, K. (2005) Economics of health care financing: The Visible Hand. 2nd Edition, New York: Palgrave Macmillan.

Drummond, M.F., Barbieri, M., Cook, J., Glick, H.A., Lis, J., Farzana, M., Reed, S.D., Rutten, F., Sculpher, M., and Severens, J. (2009) Transferability of Economic Evaluations Across Jurisdictions: ISPOR Good Research Practices Task Force Report. Value in Health; **12**: 409-418.

Drummond, M.F., Becker, D.L., Hux, M., Chancellor, J.V., Duprat-Lomon, I., Kubin, R., and Sagnier, P.P. (2003) An economic evaluation of sequential i.v./po moxifloxacin therapy compared to i.v./po co-amoxiclav with or without clarithromycin in the treatment of community-acquired pneumonia. Chest; **124**: 526-535.

Drummond, M.F., Bloom, B.S., Carrin, G., Hillman, A.L., Hutchings, H.C., Knill-Jones, R.P., de Pouvourville, G., and Torfs, K. (1992) Issues in the cross-national assessment of health technology. International Journal of Technology Assessment in Health Care; **8**: 671-682.

Drummond, M.F., O'Brien, B., Stoddart, G.L., and Torrance, G.W. (1997) Methods for the economic evaluation of health care programmes. 2nd Edition Edition, Oxford: Oxford University Press.

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J., Stoddart, G.L., and . (2005) Methods for the Economic Evaluation of Health Care Programmes. 3rd Edition, Oxford: Oxford University Press.

Egger, M. et al.. (2001) Systematic Reviews in Health Care: Meta-analysis in context. In Egger, M. et al. (Eds.). 2nd Edition, London: BMJ Publishing Group.

Elixhauser, A., Halpern, M., Schmier, J., and Luce, B.R. (1998) Health care CBA and CEA from 1991 to 1996: an updated bibliography. Medical Care; **36**: MS1-147.

EuroQol Group (1990) EuroQol - a new facility for the measurement of health-related quality of life. Health Policy; **16**: 199-208.

EUROSTAT (2009) EUROSTAT: Data Explorer. Eurostat .URL: http://ec.europa.eu/eurostat (Last Accessed: 20 April 2009)

Eurostat-OECD. (2006) Methodological Manual on Purchasing Power Parities. Paris: OECD Publishing.

Fernandez, J.L., Montgomery, S., and Francois, C. (2005) Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. PharmacoEconomics; **23**: 155-167.

Folland, S., Goodman, A.C., and Stano, M. (2001) The economics of health and health care. 3rd Edition, New Jersey: Prentice Hall.

Forbes, R.B., Lees, A., Waugh, N., and Swingler, R.J. (1999) Population based cost utility study of interferon beta-1b in secondary progressive multiple sclerosis. British Medical Journal; **319**: 1529-1533.

Friedman, L.M., Furberg, C.D., and DeMets, D.L. (1998) Fundamentals of clinical trials. 3rd Edition, New York: Springer.

Gafni, A., Goeree, R., Myhr, T.L., Hannah, M.E., Blackhouse, G., Willan, A.R., Weston, J.A., Wang, E.E., Hodnett, E.D., Hewson, S.A., Farine, D., and Ohlsson, A. (1997) Induction of labour versus expectant management for prelabour rupture of the membranes at term: an economic evaluation. TERMPROM Study Group. Term Prelabour Rupture of the Membranes. CMAJ Canadian Medical Association Journal; **157**: 1519-1525.

Gail, M. and Simon, R. (1985) Testing for Qualitative Interactions Between Treatment Effects and Patient Subsets. Biometrics; **41**: 361-372.

Gauthier, A., Manca, A., and Anton, S. (2009) Bayesian modelling of healthcare resource use in multinational randomized clinical trials. PharmacoEconomics; **27**: 1017-1029.

Glick, H., Willke, R., Polsky, D., Llana, T., Alves, W.M., Kassell, N., and Schulman, K. (1998) Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage. Economic evaluation of a phase III clinical trial in Europe and Australia. International Journal of Technology Assessment in Health Care; **14**: 145-160.

Glick, H.A., Orzol, S.M., Tooley, J.F., Polsky, D., and Mauskopf, J.O. (2003) Design and analysis of unit cost estimation studies: How many hospital diagnoses? How many countries? Health Economics; **12**: 517-527.

Goeree, R., Burke, N., O'Reilly, D., Manca, A., Blackhouse, G., and Tarride, J.E. (2007) Transferability of economic evaluations: approaches and factors to consider when using

results from one geographic area for another. Current Medical Research & Opinion; **23**: 671-682.

Goeree, R., Gafni, A., Hannah, M., Myhr, T., and Blackhouse, G. (1999) Hospital selection for unit cost estimates in multicentre economic evaluations. Does the choice of hospitals make a difference? PharmacoEconomics; **15**: 561-572.

Gold, M.R., Siegel, J.E., Russell, L.B., and Weinstein, M.C. (1996) Cost-effectiveness in health and medicine. New York: Oxford University Press.

Gray, A., Clarke, P.M., Wolstenholme, J., and Wordsworth, S. (2010) Applied Methods of Cost-Effectiveness Analysis in Health Care. In Gray, A. and Briggs, A.H. (Eds.). Oxford: Oxford University Press.

Grieve, R., Cairns, J., and Thompson, S.G. (2010) Improving costing methods in multicentre economic evaluation: the use of multiple imputation for unit costs. Health Economics; **19**: 939-954.

Grieve, R., Hutton, J., Bhalla, A., Rastenyte, D., Ryglewicz, D., Sarti, C., Lamassa, M., Giroud, M., Dundas, R., and Wolfe, C.D. (2001) A comparison of the costs and survival of hospital-admitted stroke patients across Europe. Stroke; **32**: 1684-1691.

Grieve, R., Nixon, R., Thompson, S.G., and Cairns, J. (2007) Multilevel models for estimating incremental net benefits in multinational studies. Health Economics; **16**: 815-826.

Grieve, R., Nixon, R., Thompson, S.G., and Normand, C. (2005) Using multilevel models for assessing the variability of multinational resource use and cost data. Health Economics; **14**: 185-196.

Griffin, A.D., Perry, A.S., and Fleming, D.M. (2001) Cost-effectiveness analysis of inhaled zanamivir in the treatment of influenza A and B in high-risk patients. PharmacoEconomics; **19**: 293-301.

Groener, M.G., van Ineveld, B.M., Byttebier, G., van Hout, B.A., and Rutten, F.F. (1999) An economic evaluation of Tomudex (raltitrexed) and 5-fluorouracil plus leucovorin in advanced colorectal cancer. Anti-Cancer Drugs; **10**: 283-288.

Halliday, R.G. and Darba, J. (2003) Cost data assessment in multinational economic evaluations: some theory and review of published studies. Applied Health Economics and Health Policy; **2**: 149-155.

Henderson, R.A. and Brown, R. (1999) The costs of routine eptifibatide use in acute coronary syndromes in Western Europe: An economic substudy of the PURSUIT trial. European Heart Journal, Supplement; **1**: N35-N41.

Herman, W.H., Shahinfar, S., Carides, G.W., Dasbach, E.J., Gerth, W.C., Alexander, C.M., Cook, J.R., Keane, W.F., and Brenner, B.M. (2003) Losartan reduces the costs associated with diabetic end-stage renal disease: the RENAAL study economic evaluation. Diabetes Care; **26**: 683-687.

Hilbrich, L. and Sleight, P. (2006) Progress and problems for randomized clinical trials: from streptomycin to the era of megatrials. European Heart Journal; **27**: 2158-2164.

Hillner, B.E., Agarwala, S., and Middleton, M.R. (2000) Post hoc economic analysis of temozolomide versus dacarbazine in the treatment of advanced metastatic melanoma. Journal of Clinical Oncology; **18**: 1474-1480.

HM Treasury. (2008) The Green Book: Appraisal and Evaluation in Central Government. London: The Stationary Office.

Hoch, J.S., Briggs, A.H., and Willan, A.R. (2002) Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Economics; **11**: 415-430.

Hsiao, W.C., Braun, P., Dunn, D., and Becker, E.R. (1988) Resource-based relative values. An overview. Journal of the American Medical Association; **260**: 2347-2353.

Huang, I.C., Willke, R.J., Atkinson, M.J., Lenderking, W.R., Frangakis, C., and Wu, A.W. (2007) US and UK versions of the EQ-5D preference weights: does choice of preference weights make a difference? Quality of Life Research; **16**: 1065-1072.

Hull, R.D., Raskob, G.E., Rosenbloom, D., Pineo, G.F., Lerner, R.G., Gafni, A., Trowbridge, A.A., Elliott, C.G., Green, D., and Feinglass, J. (1997) Treatment of proximal vein thrombosis with subcutaneous low-molecular-weight heparin vs intravenous heparin. An economic perspective. Archives of Internal Medicine; **157**: 289-294.

Hutton, G., Fox-Rushby, J., Mugford, M., Thinkhamrop, J., Thinkhamrop, B., Galvez, A.M., and Alvarez, M. (2004) Examining within-country variation of maternity costs in the context of a multicountry, multicentre randomised controlled trial. Applied Health Economics and Health Policy; **3**: 161-170.

Institute for Scientific Information. (2005) Reference Manager Professional Edition: Release 11. USA.

Jacobs, R., Smith, P.C., and Street, A. (2009) Measuring efficiency in health care. Cambridge: Cambridge University Press.

Jonsson, B., Cook, J.R., and Pedersen, T.R. (1999) The cost-effectiveness of lipid lowering in patients with diabetes: results from the 4S trial. Diabetologia; **42**: 1293-1301.

Kirkwood, B.R. and Sterne, J.A.C. (2003) Essential Medical Statistics. 2nd Edition, Oxford: Blackwell Science.

Koopmanschap, M.A., Touw, K.C., and Rutten, F.F. (2001) Analysis of costs and cost-effectiveness in multinational trials. Health Policy; **58**: 175-186.

Leung, C.H.S., Poon, W.S., and Yu, L.M. (2003) The ISAT trial. The Lancet; **361**: 430-431.

Levy, P., Lechat, P., Leizorovicz, A., and Levy, E. (1998) A cost-minimization of heart failure therapy with bisoprolol in the French setting: an analysis from CIBIS trial data. Cardiac Insufficiency Bisoprolol Study. Cardiovascular Drugs & Therapy; **12**: 301-305.

Light, R.J. and Pillemer, D.B. (1984) Summing up: the science of reviewing research. Cambridge: Harvard University Press.

Linn, F.H.H., Rinkel, G.J.E., Algra, A., and van Gijn, J. (1996) Incidence of Subarachnoid Hemorrhage : Role of Region, Year, and Rate of Computed Tomography: A Meta-Analysis. Stroke; **27**: 625-629.

Lofdahl, C.G., Ericsson, A., Svensson, K., and Andreasson, E. (2005) Cost effectiveness of budesonide/formoterol in a single inhaler for COPD compared with each monocomponent used alone. PharmacoEconomics; **23**: 365-375.

Lothgren, M. and Zethraeus, N. (2000) Definition, interpretation and calculation of cost-effectiveness acceptability curves. Health Economics; **9**: 623-630.

Manca, A., Hawkins, N., and Sculpher, M.J. (2005a) Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Economics; **14**: 487-496.

Manca, A., Lambert, P.C., Sculpher, M., and Rice, N. (2007) Cost-effectiveness analysis using data from multinational trials: the use of bivariate hierarchical modeling. Medical Decision Making; **27**: 471-490.

Manca, A., Rice, N., Sculpher, M.J., and Briggs, A.H. (2005b) Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models.[erratum appears in Health Econ. 2005 May;14(5):486]. Health Economics; **14**: 471-485.

Manca, A. and Willan, A.R. (2006) 'Lost in translation': accounting for between-country differences in the analysis of multinational cost-effectiveness data. PharmacoEconomics; **24**: 1101-1119.

Manning, W.G. (1998) The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of Health Economics; **17**: 283-295.

Manning, W.G. and Mullahy, J. (2001) Estimating log models: to transform or not to transform? Journal of Health Economics; **20**: 461-494.

Mano, M.S., Rosa, D.D., and Dal, L.L. (2006) Multinational clinical trials in oncology and post-trial benefits for host countries: where do we stand? European Journal of Cancer; **42**: 2675-2677.

Mark, D.B., Harrington, R.A., Lincoff, A.M., Califf, R.M., Nelson, C.L., Tsiatis, A.A., Buell, H., Mahaffey, K.W., vidson-Ray, L., and Topol, E.J. (2000) Cost-effectiveness of platelet glycoprotein IIb/IIIa inhibition with eptifibatide in patients with non-ST-elevation acute coronary syndromes. Circulation; **101**: 366-371.

McFadden, D. (1978) Chapter 4: Estimation Techniques for the Elasticity of Substitution and Other Production Parameters. Production Economics: A Dual Approach to Theory and Applications. In Melvyn, F. and McFadden, D. (Eds.). Holland: Elsevier North.

McIntosh, E., Louviere, J.J., Frew, E., and Clarke, P.M. (2010) Applied Methods of Cost-Benefit Analysis in Health Care. In Gray, A. and Briggs, A.H. (Eds.). Oxford: Oxford University Press.

Menzin, J., Oster, G., Davies, L., Drummond, M.F., Greiner, W., Lucioni, C., Merot, J.L., Rossi, F., vd Schulenburg, J.G., and Souetre, E. (1996) A multinational economic evaluation of rhDNase in the treatment of cystic fibrosis. International Journal of Technology Assessment in Health Care; **12**: 52-61.

Mihaylova, B., Briggs, A., O'Hagan, A., and Thompson, S.G. (2010) Review of statistical methods for analysing healthcare resources and costs. Health Economics Early View.

Mohr, J.P. (2003) The ISAT trial. The Lancet; **361**: 431-431.

Molyneux, A. and Kerr, R. (2003) The ISAT trial. The Lancet; **361**: 432-432.

Molyneux, A., Kerr, R., Stratton, I., Sandercock, P., Clarke, M., Shrimpton, J., and Holman, R. (2002) International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. Lancet; **360**: 1267-1274.

Molyneux, A.J., Kerr, R.S.C., Yu, L.M., Clarke, M., Sneade, M., Yarnold, J.A., and Sandercock, P. (2005) International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. Lancet; **366**: 809-817.

Nathwani, D., Li, J.Z., Balan, D.A., Willke, R.J., Rittenhouse, B.E., Mozaffari, E., Tavakoli, M., and Tang, T. (2004) An economic evaluation of a European cohort from a multinational trial of linezolid versus teicoplanin in serious Gram-positive bacterial infections: the importance of treatment setting in evaluating treatment effects. International Journal of Antimicrobial Agents; **23**: 315-324.

National Institute for Health and Clinical Excellence. (2008) Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence.

NHS Centre for Reviews and Dissemination. (2001) Improving access to cost-effectiveness information for health care decision-making: the NHS Economic Evaluation Database. 2nd Edition, York: NHS CRD, University of York.

Nord, E., Wisloff, F., Hjorth, M., and Westin, J. (1997) Cost-utility analysis of melphalan plus prednisone with or without interferon-alpha 2b in newly diagnosed multiple myeloma. Results from a randomised controlled trial. PharmacoEconomics; **12**: 89-103.

O'Brien, B.J. (1997) A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. American Journal of Managed Care; **Suppl 3**: S33-S39.

Organisation for Economic Co-Operation and Development. (2009) OECD Health Data. Paris: Organisation for Economic Co-Operation and Development.

Pang, F. (2002) Design, analysis and presentation of multinational economic studies: the need for guidance. PharmacoEconomics; **20**: 75-90.

Pauly, M.V. (1995) Chapter 6: Valuing health care benefits in money terms. Valuing health care: costs, benefits, and effectiveness of pharmaceuticals and other medical technologies. In Sloan, F.A. (Eds.). Cambridge: Cambridge University Press.

Peeters, P., Ortonne, J.P., Sitbon, R., and Guignard, E. (2005) Cost-effectiveness of once-daily treatment with calcipotriol/betamethasone dipropionate followed by calcipotriol alone compared with tacalcitol in the treatment of psoriasis vulgaris. Dermatology; **211**: 139-145.

Piantadosi, S. and Gail, M.H. (1993) A comparison of the power of two tests for qualitative interactions. Statistics in Medicine; **12**: 1239-1248.

Pinto, E.M., Willan, A.R., and O'Brien, B.J. (2005) Cost-effectiveness analysis for multinational clinical trials. Statistics in Medicine; **24**: 1965-1982.

Pregibon, D. (1980) Goodness of Link Tests for Generalized Linear Models. Journal of the Royal Statistical Society.Series C (Applied Statistics); **29**: 15-14.

Radeva, J., I, Reed, S.D., Kalo, Z., Kauf, T.L., Cantu, E., Cretin, N., and Schulman, K.A. (2005) Economic evaluation of everolimus vs azathioprine at one year after de novo heart transplantation. Clinical Transplantation; **19**: 122-129.

Raikou, M., Briggs, A., Gray, A., and McGuire, A. (2000) Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. Health Economics; **9**: 191-198.

Rasbash, J., Steele, F., Browne, W., and Goldstein, H. (2009) A User's Guide to MLwiN. Version 2.10 Edition, Centre for Multilevel Modelling, University of Bristol.

Rawlins, M. and Culyer, A.J. (2004) National Institute for Clinical Excellence and its value judgments. British Medical Journal; **329**: 224-227.

Reed, S.D., Anstrom, K.J., Bakhai, A., Briggs, A.H., Califf, R.M., Cohen, D.J., Drummond, M.F., Glick, H.A., Gnanasakthy, A., Hlatky, M.A., O'Brien, B.J., Torti, F.M., Jr., Tsiatis, A.A., Willan, A.R., Mark, D.B., and Schulman, K.A. (2005a) Conducting economic evaluations alongside multinational clinical trials: toward a research consensus. American Heart Journal; **149**: 434-443.

Reed, S.D., Friedman, J.Y., Gnanasakthy, A., and Schulman, K.A. (2003) Comparison of hospital costing methods in an economic evaluation of a multinational clinical trial. International Journal of Technology Assessment in Health Care; **19**: 396-406.

Reed, S.D., Friedman, J.Y., Velazquez, E.J., Gnanasakthy, A., Califf, R.M., and Schulman, K.A. (2004a) Multinational economic evaluation of valsartan in patients with chronic heart failure: results from the Valsartan Heart Failure Trial (Val-HeFT). American Heart Journal; **148**: 122-128.

Reed, S.D., Radeva, J.I., Glendenning, G.A., Saad, F., and Schulman, K.A. (2004b) Cost-effectiveness of zoledronic acid for the prevention of skeletal complications in patients with prostate cancer. Journal of Urology; **171**: 1537-1542.

Reed, S.D., Radeva, J.I., Weinfurt, K.P., McMurray, J.J., Pfeffer, M.A., Velazquez, E.J., Allsbrook, J.S., Masselink, L.E., Sellers, M.A., Califf, R.M., Schulman, K.A., and VALIANT, I. (2005b) Resource use, costs, and quality of life among patients in the multinational Valsartan in Acute Myocardial Infarction Trial (VALIANT). American Heart Journal; **150**: 323-329.

Rivero-Arias, O., Gray, A., and Wolstenholme, J. (2010) Burden of disease and costs of aneurysmal subarachnoid haemorrhage (aSAH) in the United Kingdom. Cost Effectiveness and Resource Allocation; **8**: 6-

Rogoff, K. (1996) The Purchasing Power Parity Puzzle. Journal of Economic Literature; **34**: 647-668.

Rubin, D.B. (1996) Multiple imputation after 18+ years. Journal of the American Statistical Association; **91**: 473-489.

Scandinavian Simvastatin Survival Study Group (1994) Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet; **344**: 1383-1389.

Schulman, K., Burke, J., Drummond, M., Davies, L., Carlsson, P., Gruger, J., Harris, A., Lucioni, C., Gisbert, R., Llana, T., Tom, E., Bloom, B., Willke, R., and Glick, H. (1998) Resource costing for multinational neurologic clinical trials: methods and results. Health Economics; **7**: 629-638.

Sculpher, M. (2001) Chapter 5: The role and estimation of productivity costs in economic evaluation. Economic evaluation in health care: merging theory with practice. In Drummond, M. and McGuire, A. (Eds.). Oxford: Oxford University Press.

Sculpher, M., Manca, A., Abbott, J., Fountain, J., Mason, S., and Garry, R. (2004a) Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. British Medical Journal; **328**: 134-137.

Sculpher, M.J., Pang, F.S., Manca, A., Drummond, M.F., Golder, S., Urdahl, H., Davies, L.M., and Eastwood, A. (2004b) Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technology Assessment; **8**: No-49.

Sellar, R. and Whittle, I. (2003) The ISAT trial. The Lancet; **361**: 432-433.

Shaw, J.W., Johnson, J.A., and Coons, S.J. (2005) US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care; **43**: 203-220.

Shearer, J. and Shanahan, M. (2006) Cost effectiveness analysis of smoking cessation interventions. Australian & New Zealand Journal of Public Health; **30**: 428-434.

Simon, J., Gray, A., Duley, L., and Magpie Trial Collaborative Group (2006) Cost-effectiveness of prophylactic magnesium sulphate for 9996 women with pre-eclampsia from 33 countries: economic evaluation of the Magpie Trial. BJOG: An International Journal of Obstetrics & Gynaecology; **113**: 144-151.

Smith, S.R. (2002) Multinational economic evaluations of pharmaceuticals. Lancet; **360**: 816-816.

Spiegelhalter, D.J., Abrams, K.R., and Myles, J.P. (2003) Bayesian Approaches to Clinical Trials and Health-care Evaluation. Chichester: Wiley.

StataCorp. (2007) Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.

Szende, A., Oppe, M., and Devlin, N. (2007) EQ-5D value sets: inventory, comparative review and user guide. In Szende, A. et al. (Eds.). Dordrecht: Springer.

Thompson, S.G., Nixon, R.M., and Grieve, R. (2006) Addressing the issues that arise in analysing multicentre cost data, with application to a multinational study. Journal of Health Economics; **25**: 1015-1028.

Torti, F.M., Jr., Reed, S.D., and Schulman, K.A. (2006) Analytic considerations in economic evaluations of multinational cardiovascular clinical trials. Value in Health; **9**: 281-291.

van Hout, B., Al, M.J., Gordon, G.S., and Rutten, F.F. (1994) Costs, effects and C/E-ratios alongside a clinical trial. Health Economics; **3**: 309-319.

Varian, H.R. (1992) Microeconomic Analysis. 3rd Edition, New York: Norton.

Varian, H.R. (1999) Intermediate Microeconomics. 5th Edition, New York: Norton.

Villar, J., Ba'aqeel, H., Piaggio, G., Lumbiganon, P., Miguel, B.J., Farnot, U., Al-Mazrou, Y., Carroli, G., Pinol, A., Donner, A., Langer, A., Nigenda, G., Mugford, M., Fox-Rushby, J., Hutton, G., Bergsjo, P., Bakketeig, L., Berendes, H., and Garcia, J. (2001) WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care [erratum appears in Lancet 2001;358(9292):1556]. Lancet; **357**: 1551-1564.

Weaver, W.D., Reisman, M.A., Griffin, J.J., Buller, C.E., Leimgruber, P.P., Henry, T., D'Haem, C., Clark, V.L., Martin, J.S., Cohen, D.J., Neil, N., and Every, N.R. (2000) Optimum percutaneous transluminal coronary angioplasty compared with routine stent strategy trial (OPUS-1): a randomised trial. Lancet; **355**: 2199-2203.

Wermer, M.J.H., van der Schaaf, I.C., Velthuis, B.K., Algra, A., Buskens, E., and Rinkel, G.J.E. (2005) Follow-up screening after subarachnoid haemorrhage: frequency and determinants of new aneurysms and enlargement of existing aneurysms. Brain; **128**: 2421-2429.

Willan, A.R. and Briggs, A.H. (2006) Statistical Analysis of Cost-Effectiveness Data. Chichester: Wiley.

Willan, A.R., Goeree, R., Pullenayegum, E.M., McBurney, C., and Blackhouse, G. (2006) Economic evaluation of rivastigmine in patients with Parkinson's disease dementia. PharmacoEconomics; **24**: 93-106.

Willan, A.R. and Kowgier, M.E. (2008) Cost-effectiveness analysis of a multinational RCT with a binary measure of effectiveness and an interacting covariate. Health Economics; **17**: 777-791.

Willan, A.R. and O'Brien, B.J. (1996) Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. Health Economics; **5**: 297-305.

Willan, A.R., Pinto, E.M., O'Brien, B.J., Kaul, P., Goeree, R., Lynd, L., and Armstrong, P.W. (2005) Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. Health Economics; **14**: 327-338.

Willke, R.J., Glick, H.A., Polsky, D., and Schulman, K. (1998a) Estimating country-specific cost-effectiveness from multinational clinical trials. Health Economics; **7**: 481-493.

Willke, R.J., Glick, H.A., Polsky, D., and Schulman, K. (1998b) Estimating country-specific cost-effectiveness from multinational clinical trials. Health Economics; **7**: 481-493.

Wimo, A., Winblad, B., Engedal, K., Soininen, H., Verhey, F., Waldemar, G., Wetterholm, A.L., Mastey, V., Haglund, A., Zhang, R., Miceli, R., Chin, W., Subbiah, P., and Donepezil Nordic Study Group (2003) An economic evaluation of donepezil in mild to moderate Alzheimer's disease: results of a 1-year, double-blind, randomized trial.[erratum appears in

Dement Geriatr Cogn Disord. 2003;16(2):102]. Dementia & Geriatric Cognitive Disorders; **15**: 44-54.

Wolfe, C.D., Taub, N.A., Woodrow, E.J., and Burney, P.G. (1991) Assessment of scales of disability and handicap for stroke patients. Stroke; **22**: 1242-1244.

Wolstenholme, J., Rivero-Arias, O., Gray, A., Molyneux, A.J., Kerr, R.S., Yarnold, J.A., Sneade, M., and International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group (2008) Treatment pathways, resource use, and costs of endovascular coiling versus surgical clipping after aSAH. Stroke; **39**: 111-119.

Woodward, M. (2005) Epidemiology: Study design and analysis. 2nd Edition, Florida: Chapman & Hall/CRC.

Wordsworth, S. and Ludbrook, A. (2005) Comparing costing results in across country economic evaluations: the use of technology specific purchasing power parities. Health Economics; **14**: 93-99.

World Health Organisation. (2003) The World Health Report 2003: shaping the future. Geneva: World Health Organisation.

Appendix 1: Properties of a health care production function

In order for the production function to make sense from an economic point of view, we have to formally assume the following set of properties: monotonic, convexity and regularity. A formal description of each is given next.

In a two-input world, if we increase the amount of one of the inputs, we should be able to produce at least as much output as we were producing before the input was increased. Technologies that fulfil this requirement are known as monotonic. More formally, if we have a set of all possible combination of inputs V and x is one of the feasible ways to produce a level of output y, if x' is an alternative feasible way such us $x' \geq x$ that also produce y, then x' is also included in V.

The second property of a production function is convexity. This suggests that if we have two feasible ways of producing y units of output, $A = (x_1, x_2)$ and $B = (z_1, z_2)$, then their weighted average will also produce at least y units of output. Convexity is better understood with a numerical example and some graphical input. Let's assume that we have two input sets A= (100, 200) and B= (200, 100) that produces 100 units of output. A convex technology would allow a combination of the two input sets to produce at least 100 units. For example, we could use 50 processes of the activity A and 50 processes of activity B or 25 processes of activity A and 75 processes of activity B. This would imply that

0.50(100,200) + 0.50(200,100) = (150,150) for the first example

0.25(100,200) + 0.75(200,100) = (175,125) for the second example.

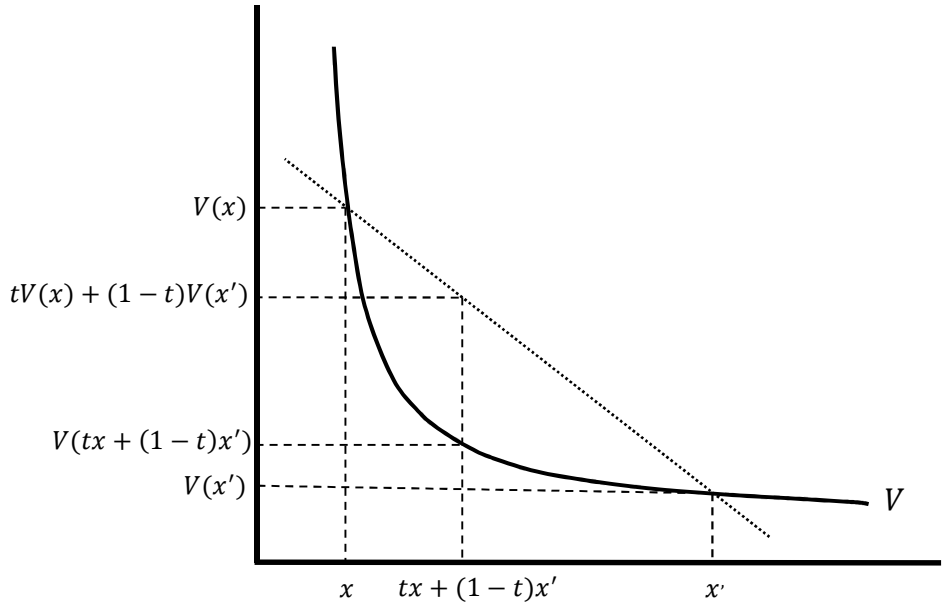
Therefore for any $t \in (0,1)$, 100 units of output can be generated as follows:

$$t(100,200) + (1-t)(200,100) = (100t + 200(1-t), 200t + (1-t)100)$$

In general, if x and x' are feasible ways of generating a particular output and belong to the same set of possible combinations V, then $tx + (1-t)x'$ is also part of V for all $0 \leq t \leq 1$. Figure A1 illustrates the idea of convexity graphically. The isoquant represents a production function V which is said to be convex if for any two points x and x' and any $t \in (0,1)$:

$$V(tx + (1-t)x') \leq tV(x) + (1-t)V(x')$$

Figure A1: The convexity property of the production function



The final property is called regularity and it is a simple characteristic that ensures that there is always a feasible way of producing a level of output. This means that the set of possible combinations of inputs V is non empty for all $y \geq 0$.

Appendix 2: Keyword proforma used in the structured review

Area	Keyword (when applicable)
Main author	
Year of publication	
Journal of publication	
Disease area	
Interventions evaluated	
Overall sample size of the trial	
Trial Acronym	
Number of countries included and sample size	
Country specific information provided	No information reported Limited information reported Detailed information reported
Categories of resource use included. Perspective	
Unit cost data sources	One-country costing Multi-country costing Not clear
Average versus centre-specific unit costs in each country	Single centre Centre-specific Average Mixed sources Not clear
Handling of cost analysis	
Cost results currency	
Conversion adopted	One-country costing no need for conversion Exchange rate GDP PPPs Average conversion rate Conversion rate not reported
Health outcomes included in the economic analysis	Disease-specific HRQoL (SF-36) HRQoL (EQ-5D) HRQoL (other utilities) Life-years
Handling of health outcomes analysis	Sub-sample of cases Trial-wide effectiveness data
Multinational classification	Fully pooled one-country costing Fully pooled multicountry costing Partially split one country costing Partially split multicountry costing Fully split one-country costing
Time horizon and extrapolation	Less than 1 month Between 1 month and 12 months More than 12 months Long-term extrapolation Not clear

Method to combine costs and health outcomes	
Country-specific ICER reported	One estimate of ICER for a particular country One estimate of ICER reported for all countries Country-specific ICER reported for each country
Handling uncertainty of cost-effectiveness results	No uncertainty Sensitivity analysis Statistical methods
Acknowledgement of multinational nature	
Funding source	Industry Non-industry Not stated
Database consulted	OvidSP- MEDLINE OvidSP-EMBASE NHS-EED Cross-reference

Appendix 3: Full list of references included in the structured review

Armstrong, S.H. and Ruckley, C.V. (1997) Use of a fibrous dressing in exuding leg ulcers. Journal of Wound Care; **6**: 322-324.

Aspelin, P., Aubry, P., Fransson, S.G., Strasser, R., Willenbrock, R., and Lundkvist, J. (2005) Cost-effectiveness of iodixanol in patients at high risk of contrast-induced nephropathy. American Heart Journal; **149**: 298-303.

Bachert, C., Bousquet, J., Canonica, G.W., Durham, S.R., Klimek, L., Mullol, J., Van Cauwenberge, P.B., Van, H.G., and XPERT Study Group (2004) Levocetirizine improves quality of life and reduces costs in long-term management of persistent allergic rhinitis. Journal of Allergy & Clinical Immunology; **114**: 838-844.

Bachert, C., Vestenbaek, U., Christensen, J., Griffiths, U.K., and Poulsen, P.B. (2007) Cost-effectiveness of grass allergen tablet (GRAZAX) for the prevention of seasonal grass pollen induced rhinoconjunctivitis - a Northern European perspective. Clinical & Experimental Allergy; **37**: 772-779.

Baran, R., Sigurgeirsson, B., Berker, D.D., Kaufmann, R., Lecha, M., Faergemann, J., Kerrouche, N., and Sidou, F. (2007) A multicentre, randomized, controlled study of the efficacy, safety and cost-effectiveness of a combination therapy with amorolfine nail lacquer and oral terbinafine compared with oral terbinafine alone for the treatment of onychomycosis with matrix involvement. British Journal of Dermatology; **157**: 149-157.

Bennett-Guerrero, E., Sorohan, J.G., Gurevich, M.L., Kazanjian, P.E., Levy, R.R., Barbera, A.V., White, W.D., Slaughter, T.F., Sladen, R.N., Smith, P.K., and Newman, M.F. (1997) Cost-benefit and efficacy of aprotinin compared with epsilon-aminocaproic acid in patients having repeated cardiac operations: a randomized, blinded clinical trial. Anesthesiology; **87** : 1373-1380.

Bjorholt, I., Andersson, F.L., Kahan, T., and Ostergren, J. (2002) The cost-effectiveness of ramipril in the treatment of patients at high risk of cardiovascular events: a Swedish sub-study to the HOPE study. Journal of Internal Medicine; **251**: 508-517.

Boersma, C., Radeva, J.I., Koopmanschap, M.A., Voors, A.A., and Postma, M.J. (2006) Economic evaluation of valsartan in patients with chronic heart failure: Results from Val-HeFT adapted to The Netherlands. Journal of Medical Economics; **9**: 121-131.

Bracco, A., Jonsson, B., Ricci, J.F., Drummond, M., and Nyhlin, H. (2007) Economic evaluation of tegaserod vs placebo in the treatment of patients with irritable bowel syndrome: an analysis of the TENOR study. Value in Health; **10**: 238-246.

Briggs, A.H., Bousquet, J., Wallace, M.V., Busse, W.W., Clark, T.J., Pedersen, S.E., Bateman, E.D., and GOAL Investigators Group (2006) Cost-effectiveness of asthma control: an economic appraisal of the GOAL study. Allergy; **61**: 531-536.

Brown, R. and Armstrong, P. (2003) Cost effectiveness in Canada of eptifibatide treatment for acute coronary syndrome patients using PURSUIT subgroup analysis. Canadian Journal of Cardiology; **19**: 161-166.

Calvert, M.J., Freemantle, N., Yao, G., Cleland, J.G., Billingham, L., Daubert, J.C., and Bryan, S. (2005) Cost-effectiveness of cardiac resynchronization therapy: results from the CARE-HF trial. European Heart Journal; **26**: 2681-2688.

Canonica, G.W., Poulsen, P.B., and Vestenbaek, U. (2007) Cost-effectiveness of GRAZAX for prevention of grass pollen induced rhinoconjunctivitis in Southern Europe. Respiratory Medicine; **101**: 1885-1894.

Charnas, R., Luthi, A.R., and Ruch, W. (1997) Once daily ceftriaxone plus amikacin vs. three times daily ceftazidime plus amikacin for treatment of febrile neutropenic children with cancer. Pediatric Infectious Disease Journal; **16**: 346-353.

Cook, J.R., Drummond, M., Glick, H., and Heyse, J.F. (2003) Assessing the appropriateness of combining economic data from multinational clinical trials. Statistics in Medicine; **22**: 1955-1976.

Dasbach, E.J., Rich, M.W., Segal, R., Gerth, W.C., Carides, G.W., Cook, J.R., Murray, J.F., Snavely, D.B., and Pitt, B. (1999) The cost-effectiveness of losartan versus captopril in patients with symptomatic heart failure. Cardiology; **91**: 189-194.

Davey, P., Craig, A.M., Hau, C., and Malek, M. (1999) Cost-effectiveness of prophylactic nasal mupirocin in patients undergoing peritoneal dialysis based on a randomized, placebo-controlled trial. Journal of Antimicrobial Chemotherapy; **43**: 105-112.

Detournay, B., Huet, X., Fagnani, F., and Montalescot, G. (2000) Economic evaluation of enoxaparin sodium versus heparin in unstable angina. A French sub-study of the ESSENCE trial. Pharmacoeconomics; **18**: 83-89.

Drummond, M.F., Becker, D.L., Hux, M., Chancellor, J.V., Duprat-Lomon, I., Kubin, R., and Sagnier, P.P. (2003) An economic evaluation of sequential i.v./po moxifloxacin therapy compared to i.v./po co-amoxiclav with or without clarithromycin in the treatment of community-acquired pneumonia. Chest; **124**: 526-535.

Ekman, M., Zethraeus, N., and Jonsson, B. (2001) Cost effectiveness of bisoprolol in the treatment of chronic congestive heart failure in Sweden: analysis using data from the Cardiac Insufficiency Bisoprolol Study II trial. Pharmacoeconomics; **19**: 901-916.

Erhardt, L., Ball, S., Andersson, F., Bergentoft, P., and Martinez, C. (1997) Cost effectiveness in the treatment of heart failure with ramipril: A swedish substudy of the AIRE study [erratum appears in Pharmacoeconomics; 12(6):706]. Pharmacoeconomics; **12**: 256-266.

Fernandez, J.L., Montgomery, S., and Francois, C. (2005) Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. Pharmacoeconomics; **23**: 155-167.

Gafni, A., Goeree, R., Myhr, T.L., Hannah, M.E., Blackhouse, G., Willan, A.R., Weston, J.A., Wang, E.E., Hodnett, E.D., Hewson, S.A., Farine, D., and Ohlsson, A. (1997) Induction of labour versus expectant management for prelabour rupture of the membranes at term: an economic evaluation. TERMPROM Study Group. Term Prelabour Rupture of the Membranes. CMAJ Canadian Medical Association Journal; **157**: 1519-1525.

Glick, H., Willke, R., Polsky, D., Llana, T., Alves, W.M., Kassell, N., and Schulman, K. (1998) Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage. Economic evaluation of a phase III clinical trial in Europe and Australia. International Journal of Technology Assessment in Health Care; **14**: 145-160.

Glick, H.A., Orzol, S.M., Tooley, J.F., Polsky, D., and Mauskopf, J.O. (2003) Design and analysis of unit cost estimation studies: How many hospital diagnoses? How many countries? Health Economics; **12**: 517-527.

Goeree, R., Gafni, A., Hannah, M., Myhr, T., and Blackhouse, G. (1999) Hospital selection for unit cost estimates in multicentre economic evaluations. Does the choice of hospitals make a difference? Pharmacoeconomics; **15**: 561-572.

Grieve, R., Nixon, R., Thompson, S.G., and Normand, C. (2005) Using multilevel models for assessing the variability of multinational resource use and cost data. Health Economics; **14**: 185-196.

Grieve, R., Nixon, R., Thompson, S.G., and Cairns, J. (2007) Multilevel models for estimating incremental net benefits in multinational studies. Health Economics; **16**: 815-826.

Griffin, A.D., Perry, A.S., and Fleming, D.M. (2001) Cost-effectiveness analysis of inhaled zanamivir in the treatment of influenza A and B in high-risk patients. Pharmacoeconomics; **19**: 293-301.

Groener, M.G., van Ineveld, B.M., Byttebier, G., van Hout, B.A., and Rutten, F.F. (1999) An economic evaluation of Tomudex (raltitrexed) and 5-fluorouracil plus leucovorin in advanced colorectal cancer. Anti-Cancer Drugs; **10**: 283-288.

Hansson, C. (1998) The effects of cadexomer iodine paste in the treatment of venous leg ulcers compared with hydrocolloid dressing and paraffin gauze dressing. Cadexomer Iodine Study Group. International Journal of Dermatology; **37**: 390-396.

Henderson, R.A. and Brown, R. (1999) The costs of routine eptifibatide use in acute coronary syndromes in Western Europe: An economic substudy of the PURSUIT trial. European Heart Journal; **20**: N35-N41.

Herman, W.H., Shahinfar, S., Carides, G.W., Dasbach, E.J., Gerth, W.C., Alexander, C.M., Cook, J.R., Keane, W.F., and Brenner, B.M. (2003) Losartan reduces the costs associated with diabetic end-stage renal disease: the RENAAL study economic evaluation. Diabetes Care; **26**: 683-687.

Hillner, B.E., Agarwala, S., and Middleton, M.R. (2000) Post hoc economic analysis of temozolomide versus dacarbazine in the treatment of advanced metastatic melanoma. Journal of Clinical Oncology; **18**: 1474-1480.

Hull, R.D., Raskob, G.E., Rosenbloom, D., Pineo, G.F., Lerner, R.G., Gafni, A., Trowbridge, A.A., Elliott, C.G., Green, D., and Feinglass, J. (1997) Treatment of proximal vein thrombosis with subcutaneous low-molecular-weight heparin vs intravenous heparin. An economic perspective. Archives of Internal Medicine; **157**: 289-294.

Hutton, G., Fox-Rushby, J., Mugford, M., Thinkhamrop, J., Thinkhamrop, B., Galvez, A.M., and Alvarez, M. (2004) Examining within-country variation of maternity costs in the context of a multicountry, multicentre randomised controlled trial. Applied Health Economics and Health Policy; **3**: 161-170.

Jonsson, B., Johannesson, M., Kjekshus, J., Olsson, A.G., Pedersen, T.R., and Wedel, H. (1996) Cost-effectiveness of cholesterol lowering. Results from the Scandinavian Simvastatin Survival Study (4S). European Heart Journal; **17**: 1001-1007.

Jonsson, B., Cook, J.R., and Pedersen, T.R. (1999) The cost-effectiveness of lipid lowering in patients with diabetes: results from the 4S trial. Diabetologia; **42**: 1293-1301.

Koopmanschap, M.A., Touw, K.C., and Rutten, F.F. (2001) Analysis of costs and cost-effectiveness in multinational trials. Health Policy; **58**: 175-186.

Lamy, A., Yusuf, S., Pogue, J., Gafni, A., and Heart Outcomes Prevention Evaluation Investigators (2003) Cost implications of the use of ramipril in high-risk patients based on the Heart Outcomes Prevention Evaluation (HOPE) study. Circulation; **107**: 960-965.

Levy, P., Lechat, P., Leizorovicz, A., and Levy, E. (1998) A cost-minimization of heart failure therapy with bisoprolol in the French setting: an analysis from CIBIS trial data. Cardiac Insufficiency Bisoprolol Study. Cardiovascular Drugs & Therapy; **12**: 301-305.

Liljas, B., Stadhl, E., and Pauwels, R.A. (1997) Cost-effectiveness analysis of a dry powder inhaler (Turbuhaler) versus a pressurised metered dose inhaler in patients with asthma. Pharmacoeconomics; **12**: 267-277.

Lindgren, P., Kahan, T., Poulter, N., Buxton, M., Svarvar, P., Dahlof, B., and Jonsson, B. (2007) Utility loss and indirect costs following cardiovascular events in hypertensive patients: The ASCOT health economic substudy. European Journal of Health Economics; **8**: 25-30.

Lofdahl, C.G., Ericsson, A., Svensson, K., and Andreasson, E. (2005) Cost effectiveness of budesonide/formoterol in a single inhaler for COPD compared with each monocomponent used alone. Pharmacoeconomics; **23**: 365-375.

Manca, A., Rice, N., Sculpher, M.J., and Briggs, A.H. (2005) Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models.[erratum appears in Health Econ. 2005 May;14(5):486]. Health Economics; **14**: 471-485.

Manca, A. and Willan, A.R. (2006) 'Lost in translation': accounting for between-country differences in the analysis of multinational cost-effectiveness data. Pharmacoeconomics; **24**: 1101-1119.

Manca, A., Lambert, P.C., Sculpher, M., and Rice, N. (2007) Cost-effectiveness analysis using data from multinational trials: the use of bivariate hierarchical modeling. Medical Decision Making; **27**: 471-490.

Mark, D.B., Cowper, P.A., Berkowitz, S.D., vidson-Ray, L., DeLong, E.R., Turpie, A.G., Califf, R.M., Weatherley, B., and Cohen, M. (1998) Economic assessment of low-molecular-weight heparin (enoxaparin) versus unfractionated heparin in acute coronary syndrome patients: results from the ESSENCE randomized trial. Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q wave Coronary Events [unstable angina or non-Q-wave myocardial infarction]. Circulation; **97**: 1702-1707.

Mark, D.B., Harrington, R.A., Lincoff, A.M., Califf, R.M., Nelson, C.L., Tsiatis, A.A., Buell, H., Mahaffey, K.W., vidson-Ray, L., and Topol, E.J. (2000) Cost-effectiveness of platelet glycoprotein IIb/IIIa inhibition with eptifibatide in patients with non-ST-elevation acute coronary syndromes. Circulation; **101**: 366-371.

McKinnon, P.S., Sorensen, S.V., Liu, L.Z., and Itani, K.M. (2006) Impact of linezolid on economic outcomes and determinants of cost in a clinical trial evaluating patients with MRSA complicated skin and soft-tissue infections. Annals of Pharmacotherapy; **40**: 1017-1023.

Menzin, J., Oster, G., Davies, L., Drummond, M.F., Greiner, W., Lucioni, C., Merot, J.L., Rossi, F., vd Schulenburg, J.G., and Souetre, E. (1996) A multinational economic evaluation of rhDNase in the treatment of cystic fibrosis. International Journal of Technology Assessment in Health Care; **12**: 52-61.

Nathwani, D., Li, J.Z., Balan, D.A., Willke, R.J., Rittenhouse, B.E., Mozaffari, E., Tavakoli, M., and Tang, T. (2004) An economic evaluation of a European cohort from a multinational trial of linezolid versus teicoplanin in serious Gram-positive bacterial infections: the importance of treatment setting in evaluating treatment effects. International Journal of Antimicrobial Agents; **23**: 315-324.

Nord, E., Wisloff, F., Hjorth, M., and Westin, J. (1997) Cost-utility analysis of melphalan plus prednisone with or without interferon-alpha 2b in newly diagnosed multiple myeloma. Results from a randomised controlled trial. Pharmacoeconomics; **12**: 89-103.

Peeters, P., Ortonne, J.P., Sitbon, R., and Guignard, E. (2005) Cost-effectiveness of once-daily treatment with calcipotriol/betamethasone dipropionate followed by calcipotriol alone compared with tacalcitol in the treatment of psoriasis vulgaris. Dermatology; **211**: 139-145.

Pinto, E.M., Willan, A.R., and O'Brien, B.J. (2005) Cost-effectiveness analysis for multinational clinical trials. Statistics in Medicine; **24**: 1965-1982.

Radeva, J., I, Reed, S.D., Kalo, Z., Kauf, T.L., Cantu, E., Cretin, N., and Schulman, K.A. (2005) Economic evaluation of everolimus vs azathioprine at one year after de novo heart transplantation. Clinical Transplantation; **19**: 122-129.

Raikou, M., Briggs, A., Gray, A., and McGuire, A. (2000) Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. Health Econ.; **9**: 191-198.

Reed, S.D., Friedman, J.Y., Gnanasakthy, A., and Schulman, K.A. (2003) Comparison of hospital costing methods in an economic evaluation of a multinational clinical trial. International Journal of Technology Assessment in Health Care; **19**: 396-406.

Reed, S.D., Radeva, J.I., Glendenning, G.A., Saad, F., and Schulman, K.A. (2004) Cost-effectiveness of zoledronic acid for the prevention of skeletal complications in patients with prostate cancer. Journal of Urology; **171**: 1537-1542.

Reed, S.D., Friedman, J.Y., Velazquez, E.J., Gnanasakthy, A., Califf, R.M., and Schulman, K.A. (2004) Multinational economic evaluation of valsartan in patients with chronic heart failure: results from the Valsartan Heart Failure Trial (Val-HeFT). American Heart Journal; **148**: 122-128.

Reed, S.D., Radeva, J.I., Weinfurt, K.P., McMurray, J.J., Pfeffer, M.A., Velazquez, E.J., Allsbrook, J.S., Masselink, L.E., Sellers, M.A., Califf, R.M., Schulman, K.A., and VALIANT, I. (2005) Resource use, costs, and quality of life among patients in the multinational Valsartan in Acute Myocardial Infarction Trial (VALIANT). American Heart Journal; **150**: 323-329.

Rutten-van Molken, M.P., van Doorslaer, E.K., and Till, M.D. (1998) Cost-effectiveness analysis of formoterol versus salmeterol in patients with asthma. Pharmacoeconomics; **14**: 671-684.

Rutten-van Molken, M.P., Van Nooten, F.E., Lindemann, M., Caeser, M., and Calverley, P.M. (2007) A 1-year prospective cost-effectiveness analysis of roflumilast for the treatment of patients with severe chronic obstructive pulmonary disease. Pharmacoeconomics; **25**: 695-711.

Schulman, K., Burke, J., Drummond, M., Davies, L., Carlsson, P., Gruger, J., Harris, A., Lucioni, C., Gisbert, R., Llana, T., Tom, E., Bloom, B., Willke, R., and Glick, H. (1998) Resource costing for multinational neurologic clinical trials: methods and results. Health Economics; **7**: 629-638.

Schulman, K.A., Buxton, M., Click, H., Sculpher, M., Guzman, G., Kong, J., Backhouse, M., Mauskopf, J., Bell, L., and Eisenberg, J.M. (1996) Results of the economic evaluation of the first study: A multinational prospective economic evaluation. International Journal of Technology Assessment in Health Care; **12**: 698-713.

Sculpher, M.J., Poole, L., Cleland, J., Drummond, M., Armstrong, P.W., Horowitz, J.D., Massie, B.M., Poole-Wilson, P.A., and Ryden, L. (2000) Low doses vs. high doses of the angiotensin converting-enzyme inhibitor lisinopril in chronic heart failure: a cost-effectiveness analysis based on the Assessment of Treatment with Lisinopril and Survival (ATLAS) study. European Journal of Heart Failure; **2**: 447-454.

Serruys, P.W., van, H.B., Bonnier, H., Legrand, V., Garcia, E., Macaya, C., Sousa, E., van der, G.W., Colombo, A., Seabra-Gomes, R., Kiemeneij, F., Ruygrok, P., Ormiston, J., Emanuelsson, H., Fajadet, J., Haude, M., Klugmann, S., and Morel, M.A. (1998) Randomised comparison of implantation of heparin-coated stents with balloon angioplasty in selected patients with coronary artery disease (Benestent II)[erratum appears in Lancet 1998 Oct 31;352(9138):1478]. Lancet; **352**: 673-681.

Simon, J., Gray, A., Duley, L., and Magpie Trial Collaborative Group (2006) Cost-effectiveness of prophylactic magnesium sulphate for 9996 women with pre-eclampsia from

33 countries: economic evaluation of the Magpie Trial. BJOG: An International Journal of Obstetrics & Gynaecology; **113**: 144-151.

Stalhammar, N., Carlsson, J., Peacock, R., Muller-Lissner, S., Bigard, M., Porro, G.B., Ponce, J., Hosie, J., Scott, M., Weir, D.G., Fulton, C., Gillon, K., and Bardhan, K.D. (1999) Cost effectiveness of omeprazole and ranitidine in intermittent treatment of symptomatic gastro-oesophageal reflux disease. Pharmacoeconomics; **16**: 483-497.

Strandell, A., Lindhard, A., and Eckerlund, I. (2005) Cost-effectiveness analysis of salpingectomy prior to IVF, based on a randomized controlled trial. Human Reproduction; **20**: 3284-3292.

Tellier, G., Chang, J.R., Asche, C.V., Lavin, B., Stewart, J., and Sullivan, S.D. (2004) Comparison of hospitalization rates in patients with community-acquired pneumonia treated with telithromycin for 5 or 7 days or clarithromycin for 10 days[erratum appears in Curr Med Res Opin. 2004 Aug;20(8):1331]. Current Medical Research & Opinion; **20**: 739-747.

Thompson, S.G., Nixon, R.M., and Grieve, R. (2006) Addressing the issues that arise in analysing multicentre cost data, with application to a multinational study. Journal of Health Economics; **25**: 1015-1028.

Tonkin, A.M., Eckermann, S., White, H., Friedlander, D., Glasziou, P., Magnus, P., Kirby, A., Mulray, S., Denton, M., Sallaberger, M., Hunt, D., Simes, J., and LIPID Study Group (2006) Cost-effectiveness of cholesterol-lowering therapy with pravastatin in patients with previous acute coronary syndromes aged 65 to 74 years compared with younger patients: results from the LIPID study. American Heart Journal; **151**: 1305-1312.

Topol, E.J., Mark, D.B., Lincoff, A.M., Cohen, E., Burton, J., Kleiman, N., Talley, D., Sapp, S., Booth, J., Cabot, C.F., Anderson, K.M., and Califf, R.M. (1999) Outcomes at 1 year and economic implications of platelet glycoprotein IIb/IIIa blockade in patients undergoing coronary stenting: results from a multicentre randomised trial. EPISTENT Investigators. Evaluation of Platelet IIb/IIIa Inhibitor for Stenting. Lancet; **354**: 2019-2024.

Tunis, S.L., Johnstone, B.M., Gibson, P.J., Loosbrock, D.L., and Dulisse, B.K. (1999) Changes in perceived health and functioning as a cost-effectiveness measure for olanzapine versus haloperidol treatment of schizophrenia. Journal of Clinical Psychiatry; **60**: 38-45.

Weaver, W.D., Reisman, M.A., Griffin, J.J., Buller, C.E., Leimgruber, P.P., Henry, T., D'Haem, C., Clark, V.L., Martin, J.S., Cohen, D.J., Neil, N., and Every, N.R. (2000) Optimum percutaneous transluminal coronary angioplasty compared with routine stent strategy trial (OPUS-1): a randomised trial. Lancet; **355**: 2199-2203.

Weintraub, W.S., Culler, S., Boccuzzi, S.J., Cook, J.R., Kosinski, A.S., Cohen, D.J., and Burnette, J. (1999) Economic impact of GPIIB/IIIA blockade after high-risk angioplasty:

Results from the RESTORE trial. Journal of the American College of Cardiology; **34**: 1061-1066.

Weintraub, W.S., Zhang, Z., Mahoney, E.M., Kolm, P., Spertus, J.A., Caro, J., Ishak, J., Goldberg, R., Tooley, J., Willke, R., and Pitt, B. (2005) Cost-effectiveness of eplerenone compared with placebo in patients with myocardial infarction complicated by left ventricular dysfunction and heart failure. Circulation; **111**: 1106-1113.

Weintraub, W.S., Mahoney, E.M., Lamy, A., Culler, S., Yuan, Y., Caro, J., Gabriel, S., and Yusuf, S. (2005) Long-term cost-effectiveness of clopidogrel given for up to one year in patients with acute coronary syndromes without ST-segment elevation. Journal of the American College of Cardiology; **45**: 838-845.

Willan, A.R., Pinto, E.M., O'Brien, B.J., Kaul, P., Goeree, R., Lynd, L., and Armstrong, P.W. (2005) Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. Health Economics; **14**: 327-338.

Willan, A.R., Goeree, R., Pullenayegum, E.M., McBurney, C., and Blackhouse, G. (2006) Economic evaluation of rivastigmine in patients with Parkinson's disease dementia. Pharmacoeconomics; **24**: 93-106.

Willan, A.R. and Kowgier, M.E. (2008) Cost-effectiveness analysis of a multinational RCT with a binary measure of effectiveness and an interacting covariate. Health Econ.; **17**: 777-791.

Willke, R.J., Glick, H.A., Polsky, D., and Schulman, K. (1998) Estimating country-specific cost-effectiveness from multinational clinical trials. Health Economics; **7**: 481-493.

Wimo, A., Winblad, B., Engedal, K., Soininen, H., Verhey, F., Waldemar, G., Wetterholm, A.L., Mastey, V., Haglund, A., Zhang, R., Miceli, R., Chin, W., Subbiah, P., and Donepezil Nordic Study Group (2003) An economic evaluation of donepezil in mild to moderate Alzheimer's disease: results of a 1-year, double-blind, randomized trial.[erratum appears in Dement Geriatr Cogn Disord. 2003;16(2):102]. Dementia & Geriatric Cognitive Disorders; **15**: 44-54.

Wordsworth, S. and Ludbrook, A. (2005) Comparing costing results in across country economic evaluations: the use of technology specific purchasing power parities. Health Economics; **14**: 93-99.

Appendix 4: Detailed results of the cost-effectiveness analysis performed in multi-country costing studies (11 studies)

Study ID	Notes	Measure of effectiveness	Conversion method used	Currency	ΔC (CIs)	ΔE (CIs)	ICER (CIs)
Fully pooled multi-country costing							
Baran et al 2007	2 alternatives compared	Clinical cure	Exchange rates	2006 EURO	-	14.2%, p=0.03	Average analysis
Fernandez et al 2005	2 alternatives compared	Differences in EQ-5D scores	Not reported	2003 EURO	-51.7 (-192.7, -3.9)	0.004 (-0.4, 0.052)	-12925
Radeva et al 2005	3 alternatives compared	Patient free of efficacy failure	GDP PPPs	2001 US Dollar	1) 1250 (-6598, 9314) 2) 1815 (-5601, 9424)	1) 0.1118 2) 0.2057	1) 11181 (-98836, 149528) 2) 8823 (-27894, 52215)
Reed et al 2005	3 alternative compared	Survival	GDP PPPs	2002 US Dollar	1) 1065 (365, 1875) 2) 553 (-102, 1216)	1) n/a 2) n/a	Joint distribution of mean differences in costs and effects on the CEP
Simon et al 2005	2 alternatives compared. ICERs reported by GNI	Number of cases of eclampsia prevented	Exchange rates	2001 US Dollar	High GNI: 65 (26,86) Middle GNI: 13 (7,17) Low GNI: 11 (9,12)	High GNI: 0.003 (0, 0.008) Middle GNI: 0.005 (0.0001, 0.011) Low GNI: 0.024 (0.015, 0.033)	High GNI: 21202 (3407, ∞) Middle GNI: 2473 (402, 21015) Low GNI: 456 (301, 779)

Study ID	Notes	Measure of effectiveness	Conversion method used	Currency	ΔC (CIs)	ΔE (CIs)	ICER (CIs)
	regions						
Nathwani et al 2004	2 alternatives compared	Antibiotic treatment duration (days)	Exchange rates	2001 US Dollar	691, p=0.67	3.2, p=0.001	Not reported
Reed et al 2004a	2 alternatives compared. ICERs reported in a 3 subgroup analysis	Life-years	GDP PPPs	1999 US Dollar	1) -929 (-3243, -1034) 2) 334 (-497, 1199) 3) 1246 (54 ,2230)	1) 0.19 (0.02, 0.35) 2) 0.02 (-0.03, 0.08) 3) -0.07 (-0.14, -0.01)	1) -4889 2) 16700 3) -17800
Reed et al 2004b	2 alternatives compared	QALYs	GDP PPPs	2000 US Dollar	-324 (-1781, 1146)	0.002	159200 (88500, 786600)
Glick et al 1998	2 alternatives compared	Number of deaths averted	GDP PPPs	1993 US Dollar	2695, p=0.04	0.091 (0.214, 0.123)	29615
Rutten-van Molken et al 1998	2 alternatives compared	Episode-free days	-	-	-	-	Average analysis
Partially split multi-country costing							

Study ID	Notes	Measure of effectiveness	Conversion method used	Currency	ΔC (CIs)	ΔE (CIs)	ICER (CIs)
Hansson et al 1998	3 alternatives compared	Percentage reduction in ulcer size	GDP PPPs	1994 US Dollar	-	-	Average analysis

Appendix 5: Detailed results of the cost-effectiveness analysis reporting country-specific ICERs (10 studies)

Study ID	Alternatives compared and handling of cost analysis	Measure of effectiveness	Currency	ΔC (CIs)	ΔE (CIs)	ICER results by country
Fully pooled, one country costing						
Bachert et al 2007	This study compared Grass allergen tablet and symptomatic medication for the treatment of Allergic rhinoconjuntivitis. Trial-wide resource use was multiplied for each country-specific unit costs to derive costs in each country	QALYs	€, £	-	-	Graphical representation of ICERs per country as a function of annual cost of Grass allergen tablet
Canonica et al 2007	This study compared Grass allergen tablet and symptomatic medication for the treatment of Allergic rhinoconjuntivitis. Trial-wide resource use was multiplied for each country-specific unit costs to derive costs in each country	QALYs	€, £	-	-	Graphical representation of ICERs per country as a function of annual cost of Grass allergen tablet
Willan et al 2006	This study compared Rivastigmine and	QALYs	£, $Can	UK: -26.18 (SE: 1447.36)	2.81 QALDs (-0.243, 5.86)	UK: -£3403 per QALYs

Study ID	Alternatives compared and handling of cost analysis	Measure of effectiveness	Currency	ΔC (CIs)	ΔE (CIs)	ICER results by country
	placebo for the treatment of Parkinson's Disease Dementia. Trial-wide resource use multiplied by each vector separately. Hierarchical models using country as random effect was adopted to report costs and costs-effectiveness			Canada: 55.76 (SE: 2120)		Canada: 7249$Can per QALYs
Aspelin et al 2005	This study compared iodixanol and iohexol for the treatment of acute renal failure. Trial-wide resource use multiplied by each vector separately to calculate costs in each country	Rate of adverse drug reaction	€	Sweden: -371 (-967, -21) Germany: -399 (-1062, -18) France: -445 (-1206, -30)	-0.08 adverse drug reaction per patient	Sweden: 91% of bootstrapped ICERs showed iodixanol less expensive and more effective Germany: 71% of bootstrapped ICERs showed iodixanol less France: 92% of bootstrapped ICERs showed iodixanol less
Lindgren et al 2005	This study compared atorvastatin and placebo for the prevention of coronary and stroke events. Trial-wide resource use multiplied by each unit cost vector	Number of cardiovascular events and procedures avoided	€, £, SEK	Sweden: 449 UK: 414	0.035 (0.018, 0.053)	Sweden: 12673 (3679, 36228) UK: 11693 (190, 35486)

Study ID	Alternatives compared and handling of cost analysis	Measure of effectiveness	Currency	ΔC (CIs)	ΔE (CIs)	ICER results by country
	separately					
Drummond et al 2003	This study compared Moxiflxacin and Co-amoxiclav with or without clarithromycin for the treatment of Community-acquired pneumonia. Trial wide resource use multiplied by each unit cost vector separately. Country by treatment interactions tested using ANOVA to ensure pooling the data from all countries	Rate of additional patient cured	€	Germany: -266 (-549, 17) France: -381 (-835, 73)	0.053 (0.012, 0.118)	Germany: -10019 per additional patient cured France: -13732 per additional patient cured
Lamy et al 2003	This study compared Ramipril and placebo for the treatment of cardiovascular events in high risk patients. Trial-wide resource use was multiplied for each vector of unit costs to derive costs in each country.	Absolute risk reduction for primary events	$US, $Can	US: -111 (-734, 512) Canada: 114 (-269, 497)	3.8% absolute risk reduction for primary event saved	US: 94% of bootstrapped ICERs <$US 10000 Canada: 91% of bootstrapped ICERs <$Can 10000
Jonsson et al 1999	This study examined the effectiveness of Simvastatin and placebo in diabetic	Life-years gained	€	Incremental costs reported for 5 subgroups of patients	Incremental effectiveness reported for 5 subgroups of	Sweden: 5422 Denmark: 5673

Study ID	Alternatives compared and handling of cost analysis	Measure of effectiveness	Currency	ΔC (CIs)	ΔE (CIs)	ICER results by country
	patients. Trial-wide resource use was multiplied for each country-specific unit cost to derive costs in each country.			using Swedish unit costs	patients	Norway: 3556 Finland: 8566 UK: 6476 Germany: 6928 France: 4243 Italy: 6002 Portugal: 6047 Belgium: 6743 Spain: 5504
Jonsson et al 1996	This study compared Simvastatin and placebo for the treatment of cholesterol. Trial-wide resource use from all countries multiplied by each country-specific unit costs separately.	Life-years gained	£	Only reported for Sweden 738	0.240 life years gained	Sweden: 5502 Norway: 6361 Belgium: 5165 France: 4137 Germany: 7827 Italy: 5869 Portugal: 8312 Spain: 6148

Study ID	Alternatives compared and handling of cost analysis	Measure of effectiveness	Currency	ΔC (CIs)	ΔE (CIs)	ICER results by country
						UK: 6983 Australia: 5970 New Zealand: 8824
Fully pooled, multi-country costing						
Simon et al 2005	This study compared magnesium sulphate and placebo for the treatment of pre-eclampsia. Country-specific resource use was combined with country-specific unit costs in each of the GNI groups. Independent analyses were performed in each of the GNI group as sample sizes seemed large enough to detect differences	Number of cases of eclampsia prevented	$US	High GNI: 65 (26,86) Middle GNI: 13 (7,17) Low GNI: 11 (9,12)	High GNI: 0.003 (0, 0.008) Middle GNI: 0.005 (0.0001, 0.011) Low GNI: 0.024 (0.015, 0.033)	High GNI: 21202 (3407, ∞) Middle GNI: 2473 (402, 21015) Low GNI: 456 (301, 779)

Appendix 6: Covering letter and unit cost survey to collect country-specific unit costs

Dear Centre Coordinator

The ISAT team in Oxford is now working on a full economic evaluation of the ISAT data. We aim to estimate the cost-effectiveness of coiling in comparison with neurosurgery.

To do this, we need to obtain information on the unit costs of the resources used during the ISAT study in different centres. We have tried to identify the items that are most important, in particular staff in theatre, cost of coils and the cost per day in hospital, and would be very grateful if you could provide this information in the attached questionnaire.

The questionnaire describes in detail the information we would like you to collect and send to us. In particular, as staff grades and titles are likely to vary, we have provided a brief description of different staff types included in the questionnaire.

If you have any questions please contact Oliver Rivero-Arias, (Health Economist for ISAT) preferably by email (**oliver.rivero@dphpc.ox.ac.uk**), and we will try to respond as quickly as possible.

It would be very helpful if you could complete this questionnaire in electronic format, save it as a word file, and forward it as an email attachment to the Oxford centre at julia.yarnold@nds.ox.ac.uk .

If there are any problems with this, please print it off and send it by fax to +44 (0) 1865 224490.

We would like to thank you for taking the time to complete this questionnaire.

Yours sincerely,

Andy Molyneux & Richard Kerr
(ISAT Chief Investigators)

UNIT COST RECORD SHEET

Completed by	
Contact phone number	
Contact email address	
ISAT Centre & number	
Country	

First, please write down the currency in which you are reporting these costs, and the year they refer to:

Currency:	
Year:	

In this table, please tell us how many staff are normally present during endovascular and surgical procedures and their average gross salary, including any employer costs such as insurance and pension contributions.

Staff type	Annual average Salary, including employer contributions	Number in theatre for endovascular procedure	Number in theatre for surgical procedure
Consultant Radiologist/Surgeon			
Assistant Radiologist/Surgeon			
Other Radiologist/Surgeon Specify: _____			
Consultant Anaesthetist			
Assistant Anaesthetist			
Other Anaesthetic staff Specify: _____			
Radiographer			
Nurse Specify: _____			
Nurse Specify: _____			
Nurse Specify: _____			
Nurse Specify: _____			
Other Specify: _____			
Other Specify: _____			

The number of coils used will be an important influence on the cost of the endovascular procedure. Please tell us the price your hospital typically pays per coil, and the coil manufacturers used.

Cost per coil:	
Manufacturers	

Finally, we would like an estimate of the average cost per patient per day in the different wards detailed below. This unit cost should if possible be based on the total running costs, including the cost of all staff involved, and the cost of equipment and consumables used during a period of 24 hours.

	Cost per inpatient day
Intensive Therapy Unit (ITU)	
Neurology Ward	
Neurosurgery Ward	
Rehabilitation Unit	
District General Hospital or patient's local hospital	

If you have any queries please contact Oliver Rivero-Arias at oliver.rivero@dphpc.ox.ac.uk or telephone number +44 (0) 1865 289266.

Thank you very much for completing this unit costs record sheet.

The ISAT Team

Appendix 7: Calculation of residuals and country-specific shrinkage estimates in multilevel models (MLMs)

The raw residual for a particular patient in a MLM is defined as $r_{ij} = y_{ij} - \hat{y}_{ij}$, where y_{ij} is the observed or actual score for this patient, and $\hat{y}_{ij}$ is the predicted score from the regression. The raw residual for the jth country is the mean of the r_{ij} for the patients in the country, r_{+j} . The estimated level 2 residuals for this country are obtained by multiplying r_{+j} by a shrinkage factor defined as:

$$\hat{u}_{0j} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \frac{\sigma_{e0}^2}{n_j}} * r_{+j}$$

Where n_j is the number of patients in country j. The level 1 residual is the observed or actual score, minus the predicted score from the regression, minus the level 2 residual, $\hat{e}_{0ij} = y_{ij} - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1ij}\right) - \hat{u}_{0j}$. This shrinkage factor is always less than or equal to 1 and it will be less than 1 when σ_{e0}^2 is large compared to σ_{u0}^2 , when n_j is small or both. In these cases, we have relatively little information about the country and the level 2 residual is pulled towards zero. Therefore the predicted score for patients in this country is similar to the predicted score using the pool random effect estimate, and substantial shrinkage occurs. This is the procedure implemented by the software MLwiN to calculate the country-specific shrinkage estimates. However, another method to shrinkage estimation in cost-effectiveness analyses alongside multinational clinical trials has been proposed (Pinto et al. 2005; Willan et al. 2005). This approach produces similar shrinkage estimates as the procedure used by MLwiN, but also allows the estimation of the variance associated to the shrinkage estimates of the treatment effect. This is the approach used in Chapter 6 to calculate country-specific results from the MLMs in costs and QALYs and is described briefly below. The method is

called empirical Bayes shrinkage estimation and the reader is referred to the original publications for full statistical derivation (Pinto et al. 2005; Willan et al. 2005).

Defining $\hat{\theta}_i$ as the observed treatment effect for country i, and $i = 1,2,\dots,M$; $\hat{\theta}$ the pool random effect estimate from the MLM; $\hat{\sigma}_i^2$ the within-country variance and $\hat{\sigma}^2$ the between-country variance, the shrinkage estimate is given by:

$$\tilde{\theta}_i = \hat{\theta}_i + \left(\frac{M-3}{M-1}\right)\hat{\sigma}_i^2(\hat{\sigma}_i^2 + \hat{\sigma}^2)^{-1}(\hat{\theta} - \hat{\theta}_i)$$

The variance of $\tilde{\theta}_i$ can be approximately estimated by:

$$Var(\tilde{\theta}_i) = \hat{\sigma}_i^2 - \left(\frac{M-3}{M-1}\right)\hat{\sigma}_i^2(\hat{\sigma}_i^2 + \hat{\sigma}^2)^{-1} \times \left\{1 - \left(\frac{M-3}{M-1}\right)\hat{\sigma}_i^2(\hat{\sigma}_i^2 + \hat{\sigma}^2)^{-1}W^{*-1}\right\}$$

with $W^* = \sum_i W_i^*$ and $W_i^* = (\hat{\sigma}_i^2 + \hat{\sigma}^2)^{-1}$. In this case $\hat{\sigma}_i^2$ is estimated directly from the first and second moments of the patient-level, country-specific data and $\hat{\sigma}^2$ is obtained from the MLM.

When calculating shrinkage estimates, the countries are said to "borrow strength" from countries with more precise estimates (Grieve et al. 2007; Pinto et al. 2005) and assume that the parameter of inference of interest (in our case the treatment effect in costs and outcomes) is exchangeable across countries. The exchangeability assumption suggests that when using MLMs the analyst is making the implicit assumption that the knowledge of the country may not be enough to inform that particular country. The exchangeability assumption states that the parameter of inference of interest in each country, for example treatment effect on health outcomes, is drawn randomly from an unknown parametric population distribution. In a clinical trial setting, the treatment effect on health outcomes

becomes a random variable that follows that distribution as specified in the multilevel model specified in equation [6.11]. This implies that the treatment effects across countries are similar but not identical. However, the parameters of this population distribution are always unknown *a priori* and the justification of assuming exchangeability is based on our knowledge on the topic.

Appendix 8: Algebra to calculate x_1 and x_2 in the cost-minimisation problem in the simulation study

In Chapter 3 section 3.2.8, a CES production function was defined as

$$V = \gamma[\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}]^{-1/\rho} \qquad [1]$$

The technical rate of substitution (TRS) of [1] can be expressed as (Chapter 3 section 3.2.8):

$$|TRS| = \frac{\delta}{(1-\delta)}\left(\frac{x_1}{x_2}\right)^{-\rho-1} \qquad [2]$$

In addition, the cost minimisation problem in Section 3.3.1 in Chapter 3 demonstrated that the optimal solution of x_1 and x_2 is given when the TRS equals the relative factor prices (RP) such as:

$$RP = \frac{\omega_1}{\omega_2} = TRS$$

Hence for a CES production function we can write:

$$RP = \frac{\omega_1}{\omega_2} = \frac{\delta}{(1-\delta)}\left(\frac{x_1}{x_2}\right)^{-\rho-1} \qquad [3]$$

We will work with expression [1] and [3] to determine the optimal values of x_1 and x_2 used in the simulation study. From [1], we can express x_1 as a function of the remaining parameters $x_1 = f(x_2, V, \gamma, \delta, \rho)$. We divide first by γ in both sides of expression [1]:

$$\frac{V}{\gamma} = \left[\delta x_1^{-\rho} + (1-\delta)x_2^{-\rho}\right]^{-1/\rho};$$

Then exponentiating by $-\rho$ in each side:

$$\left[\frac{V}{\gamma}\right]^{-\rho} = \delta x_1^{-\rho} + (1-\delta)x_2^{-\rho};$$

Re-arranging:

$$\delta x_1^{-\rho} = \left[\frac{V}{\gamma}\right]^{-\rho} - (1-\delta)x_2^{-\rho};$$

Dividing by δ in each side:

$$x_1^{-\rho} = \frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho};$$

And finally exponentiating both sides by $-1/\rho$:

$$x_1 = \left[\frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho}\right]^{-1/\rho} \qquad [4]$$

We substitute expression [4] in [3] and work out x_2 as a function of the remaining parameters $x_2 = f(V, \delta, \gamma, \rho, RP)$. We then first substitute [4] in [3]:

$$RP = \frac{\delta}{(1-\delta)}\left(\frac{1}{x_2}\right)^{-\rho-1}\left(\left[\frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho}\right]^{-1/\rho}\right)^{-\rho-1};$$

Multiplying the exponents $\frac{-1}{\rho} * (-\rho - 1)$ we get:

$$RP = \frac{\delta}{(1-\delta)}\left(\frac{1}{x_2}\right)^{-\rho-1}\left[\frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho}\right]^{\frac{1+\rho}{\rho}};$$

Re-arranging:

$$\frac{RP(1-\delta)x_2^{-\rho-1}}{\delta} = \left[\frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho}\right]^{\frac{1+\rho}{\rho}};$$

Exponentiating both sides by $\rho/1+\rho$:

$$\left[\frac{RP(1-\delta)x_2^{-\rho-1}}{\delta}\right]^{\frac{\rho}{1+\rho}} = \frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho};$$

Working with the expression within [ ] and the exponents:

$$\frac{RP^{\left(\frac{\rho}{1+\rho}\right)}(1-\delta)^{\left(\frac{\rho}{1+\rho}\right)}x_2^{-\rho}}{\delta^{\left(\frac{\rho}{1+\rho}\right)}} = \frac{V^{-\rho}}{\delta\gamma^{-\rho}} - \frac{(1-\delta)}{\delta}x_2^{-\rho};$$

Working and re-arrange the whole expression:

$$\left[\frac{(1-\delta)}{\delta}\right]^{\left(\frac{\rho}{1+\rho}\right)} RP^{\left(\frac{\rho}{1+\rho}\right)} x_2^{-\rho} + \frac{(1-\delta)}{\delta} x_2^{-\rho} = \frac{V^{-\rho}}{\delta\gamma^{-\rho}};$$

Taking $x_2^{-\rho}$ out from the expression within [ ]:

$$x_2^{-\rho}\left[\left(\frac{1-\delta}{\delta}\right)^{\left(\frac{\rho}{1+\rho}\right)} RP^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta}\right] = \frac{V^{-\rho}}{\delta\gamma^{-\rho}};$$

Arranging now the expression within [ ] and working with the exponents:

$$x_2^{-\rho}\left[\left(\frac{(1-\delta)}{\delta} RP\right)^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta}\right] = \frac{V^{-\rho}}{\delta\gamma^{-\rho}};$$

Dividing both sides by $x_2^{-\rho}$:

$$\left(\frac{(1-\delta)}{\delta} RP\right)^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta} = \frac{V^{-\rho}}{\delta\gamma^{-\rho}} x_2^{\rho};$$

Re-arranging:

$$x_2^{\rho} = \frac{\delta\gamma^{-\rho}}{V^{-\rho}}\left[\left(\frac{(1-\delta)}{\delta} RP\right)^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta}\right];$$

Working with the expression $\frac{\delta\gamma^{-\rho}}{V^{-\rho}}$:

$$x_2^{\rho} = \frac{\delta V^{\rho}}{\gamma^{\rho}}\left[\left(\frac{(1-\delta)}{\delta} RP\right)^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta}\right];$$

Finally exponentiating both sides by $1/\rho$:

$$x_2 = \delta^{1/\rho}\frac{V}{\gamma}\left[\left(\frac{(1-\delta)}{\delta} RP\right)^{\left(\frac{\rho}{1+\rho}\right)} + \frac{(1-\delta)}{\delta}\right]^{1/\rho}; \qquad [5]$$

The parameters $V, \delta, \gamma, \rho, RP$ are exogenous in the simulation study conducted and therefore once they are selected, x_2 can be easily calculated from expression [5] using a spreadsheet. That value is then substituted in expression [4] to calculate x_1.