Methodological issues in the analysis of individual- and aggregate-participant level data for cost effectiveness analysis

Pedro Saramago Goncalves

PhD Thesis

University of York

Department of Economics

June 2012

ABSTRACT

Health care economic evaluations assess the costs and consequences of competing interventions, programmes or services. Such assessments use a decision model, with parameters informed by available evidence. Evidence, however, is rarely derived from a single source, in which case researchers are expected to combine information on multiple sources. This thesis contributes to the methodological debate on the use of evidence, particularly, the use of individual level data (IPD), for cost effectiveness analysis.

This thesis defines a taxonomy which summarises the methodological and analytical issues in the use and synthesis of evidence for cost effectiveness modelling. For alternative parameter types (e.g. relative effectiveness, costs) the taxonomy offers guidance on appropriate synthesis methodologies to use and identifies areas where further methodological contributions are needed. The thesis also explores methods of synthesis of IPD and develops novel frameworks which allow both IPD and AD to be jointly modelled, specifically in estimating relative effectiveness. The use of IPD from studies is found desirable, particularly when the estimation of subgroup effects is of interest.

An applied decision model of the cost effectiveness of smoke alarm equipment in households with pre-school children is developed within this thesis. This application offers a means to evaluate the impact of using IPD on the cost effectiveness outcomes, compared to the use of AD. The thesis examines the advantages of having access to IPD when quantifying decision uncertainty. Additionally, it discusses the use of IPD in estimating the value of further research. Specifically, a framework is used which allows considering population subgroups. It is argued that the use of IPD allows a more suitable characterisation of decision uncertainty, appropriately allowing for subgroup value of information analysis.

LIST OF CONTENTS

Abstract 2

List of Contents 3

List of Tables 7

List of Figures 11

Acknowledgements 14

Publications 15

Author's declaration 15

1. Introduction 16

1.1 Background 16

1.2 Thesis aims and objectives 22

1.2.1 Objective 1: Exploring the use of evidence for cost effectiveness analysis – enhancing the use of individual level data 22

1.2.2 Objective 2: Assess the added value of having access to and using individual patient level data for cost effectiveness decision making 22

1.3 Case study 23

1.4 Structure of the thesis 25

2. Deriving Input Parameters for Cost effectiveness Modelling: Taxonomy of Data Types and Approaches to their Statistical Synthesis 28

2.1 Introduction 28

2.2 A taxonomy for the use of evidence in cost effectiveness models: application to clinical effectiveness evidence 31

2.2.1 Single source of evidence 34

2.2.2 Multiple sources of evidence 36

2.3 Use of evidence for other model input parameters 46

2.3.1 Disease natural progression data 47

2.3.2 Cost / resource use data 49

2.3.3 Health-state utility data 50

2.4 Discussion 53

3. Synthesizing Evidence using Aggregate- and Individual-participant Level Data . 57

3.1 Introduction 57

3.2 Motivating example: The effectiveness of home safety education and the provision of safety equipment for the prevention of accidents in pre-school children and the impact of socioeconomic characteristics 60
3.3 Methods for estimating a pooled treatment effect – direct comparisons 63
3.3.1 Synthesizing aggregate data only 63
3.3.2 Synthesizing individual participant-level data only 64
3.3.3 Synthesizing individual and aggregate level data 65
3.3.4 Meta-regression models: the inclusion of covariates 68
3.3.5 Application 73
3.4 Methods for estimating a pooled treatment effect – indirect and mixed treatment comparisons 82
3.4.1 Synthesizing aggregate data only 84
3.4.2 Synthesizing individual participant-level data only 85
3.4.3 Synthesizing individual and aggregate level data 85
3.4.4 Mixed treatment comparisons models: including covariates 88
3.4.5 Application 92
3.5 Challenges for the synthesis of (little) evidence of different formats: extending developed synthesis models 100
3.5.1 Extensions to mixed treatment comparisons models including covariate(s): the synthesis of aggregate and individual level data 101
3.5.2 Application 104
3.6 Discussion 108
4. Economic Evaluation of Smoke Alarm Programs for Preventing Fire-related Injuries of Pre-school Children in the Home 112
4.1 Defining the context: Economic Evaluation of Public Health interventions 112
4.2 Review of the existing economic evidence in accident prevention 115
4.3 Methods 117
4.3.1 The decision problem 117
4.3.2 Decision model structure 117
4.3.3 Decision analytic model assumptions/simplifications 121
4.3.4 Model implementation 122
4.3.5 Identifying, combining and analysing existing relevant evidence 122
4.3.6 Base case and scenario analysis 138
4.4 Cost effectiveness results 141
4.4.1 Overall population results (no covariates) 141
4.4.2 Subgroup cost effectiveness analysis results 145

4.4.3 Results of analysis under alternative scenarios ... *148*
4.5 Discussion ... 149
5. Using Aggregate- and Individual-participant Level Data for Cost Effectiveness Modelling ... 154
5.1 Background ... 154
5.2 Framework of analyses ... 155
5.3 Results in the absence of subgroups ... 158
5.3.1 Effectiveness ... *158*
5.3.2 Cost effectiveness ... *159*
5.4 Results in the presence of mutually exclusive subgroups ... 161
5.4.1 Subgroup analyses: number of parents in the household (two vs single) ... *161*
5.4.2 Subgroup analyses: household employment status (employed vs at least one parent unemployed) ... *165*
5.4.3 Subgroup analyses: number of parents in the household and their employment status *168*
5.5 Discussion ... 172
6. The Value of Further Research: the Added Value of Individual-participant Level Data ... 175
6.1 Introduction ... 175
6.1.1 Background ... *175*
6.1.2 Aims and objectives ... *178*
6.2 Value of additional research in the absence of subgroups ... 181
6.2.1 Definitions and methods ... *181*
6.2.2 Available evidence and the value of further research ... *182*
6.2.3 Application: value of additional research in the context of having access to individual participant level data ... *187*
6.3 Value of additional research in the presence of mutually exclusive subgroups ... 191
6.3.1 Definitions and methods ... *191*
6.3.2 Available evidence and the value of further research in the presence of subgroups ... *194*
6.3.3 Application: value of additional research in the presence of subgroups and in the context of having access to individual participant level data ... *199*
6.4 Discussion ... 205
7. Discussion ... 210
7.1 Summary of the main thesis findings ... 210
7.2 What could have been done differently? ... 215
7.3 Recommendations for accessing and analysing individual-level evidence ... 216

7.4 Recommendations for future research 217
7.5 Conclusions 219
Appendices 220
Appendix 1 220
Appendix 2 224
Appendix 3 228
Appendix 4 233
Appendix 5 234
Appendix 6 252
Appendix 7 256
Appendix 8 263
List of abbreviations (in alphabetic order) 266
References 270

LIST OF TABLES

Table 2.1 – A gallery of scenarios arising when using clinical evidence in cost effectiveness models 33

Table 2.2 – Scenarios and corresponding current methods literature for when using clinical evidence to inform cost effectiveness modelling. 45

Table 3.1 – Available evidence on interventions seeking to increase the ownership of functioning smoke alarm safety equipment to prevent fire injuries in children..... 62

Table 3.2 - Parameter estimates from fitting the pair-wise meta-analytic model to AD without covariates to the functioning smoke alarm outcome..... 74

Table 3.3 - Parameter estimates from fitting the pair-wise meta-analytic models to IPD and to AD and IPD without covariates to the functioning smoke alarm outcome..... 76

Table 3.4 - Parameter estimates from fitting the pair-wise meta-regression model to AD with a binary covariate (single parent status) to the functioning smoke alarm outcome. 78

Table 3.5 - Parameter estimates from fitting pair-wise meta-regression models to IPD and to AD plus IPD with covariate 'single parent status' to the functioning smoke alarm outcome data considering exchangeable treatment interactions and modelling separately within and between study associations. 79

Table 3.6 - Parameter estimates from fitting the pair-wise model to IPD and to AD plus IPD with information on the covariate 'single parent status' to the binary functioning smoke alarm outcome data considering exchangeable and independent treatment interactions, not separating between- and within-study interactions. 81

Table 3.7 - Parameter estimates from fitting different MTC synthesis models without including covariates to the functioning smoke alarm outcome data 95

Table 3.8 - Parameter estimates from fitting a MTC model of AD and IPD with covariate 'single parent status' to the functioning smoke alarm outcome data, considering exchangeable treatment interactions and modelling separately within and between associations. 97

Table 3.9 - Parameter estimates from fitting the MTC model of AD, IPD only and AD and IPD with covariate 'single parent status' to the functioning smoke alarm outcome data;

considering exchangeable treatment interactions, estimating subgroup baseline effects and imputing missing covariate values. 107

Table 4.1 – ORs and absolute probabilities of success estimates for each intervention result of fitting MTC model for AD without including covariates to the functioning smoke alarm outcome data. 127

Table 4.2 - Relative intervention effect estimates and absolute probabilities of success estimates for each intervention, result of fitting MTC models for AD including: (a) a covariate relating to the number of parents (i.e. 1P *vs.* 2P) in the household; and (b) a covariate parents' employment status (i.e. 2U *vs.* 1U), to the functioning smoke alarm outcome data. 129

Table 4.3 – Absolute probabilities of success estimates for each intervention result of fitting MTC model for AD including covariates 'single parent status' and 'parents' employment status' to the functioning smoke alarm outcome data. 130

Table 4.4 - List of model input parameters used within part 1 (intervention) of the decision model for functioning smoke alarms. Parameter descriptions and sources of evidence used to inform the parameter are shown. 131

Table 4.5 - All cause mortality. 134

Table 4.6 – List of model input parameters used within part 2 (5 year Markov structure) of the decision model for functioning smoke alarms. Parameter descriptions and sources of evidence used to inform the parameter are shown. 136

Table 4.7 – List of model input parameters used within the decision model for functioning smoke alarms. Sources of evidence used to inform the parameter and parametric assumption used to model parameter uncertainty is also shown. 138

Table 4.8 - Summary of the base case 139

Table 4.9 – List of scenarios considered in current analysis (each one assessed as a specific scenario, includes base case) showing how these are interconnected and their inherent hierarchical structure. 140

Table 4.10 – Results of the base case scenario for all interventions (for when effectiveness AD were synthesised and used to populate the decision model) 142

Table 4.11 - ICERs and probabilistic sensitivity analysis results for each intervention for the base case scenario and for the subgroup of (a) 2Us and of (b) 1Us. 146

Table 4.12 - Cost effectiveness results for the functioning smoke alarms decision model for four subgroups (a) two employed parent household (2EP); (b) employed single parent household (1EP); (c) two parent household with at least one unemployed (2UP); and (d) unemployed single parent household (1UP). 147

Table 4.13 – Cost effectiveness results of the 4 scenarios for all interventions and all participant households and for when AD on effectiveness was synthesised and used to populate the decision model, all the rest remaining constant..... 148

Table 5.1 - Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) for each intervention result of fitting different MTC models (AD and AD plus IPD) without including covariates, to the functioning smoke alarm outcome data. 159

Table 5.2 – Cost effectiveness results for all interventions and for when AD and AD plus IPD on effectiveness was synthesised and used to populate the decision model. Both analyses use base case characteristics and, in both, evidence informing all other economic model parameters remained the same. 160

Table 5.3 - Cost effectiveness results for all interventions, for the 2P and 1P subgroups and for when AD and AD plus IPD effectiveness evidence was synthesised and used to populate the decision model. 164

Table 5.4 – Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) from fitting the MTC model for AD and for AD plus IPD including covariate 'single parent status' to the functioning smoke alarm outcome data. .. 165

Table 5.5 – Cost effectiveness results for all interventions, for the 2U and 1P subgroups and for when AD and AD plus IPD effectiveness evidence was used. 167

Table 5.6 - Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) from fitting the MTC model for AD and for AD plus IPD including covariates 'single parent status' and 'parents' employment status' to the functioning smoke alarm outcome data. 169

Table 5.7 – Cost effectiveness results for all interventions, for 4 subgroups of families: (a) 2 employed (2EP); (b) employed single (1EP); (c) 2 with at least one unemployed (2UP); and (d) unemployed single parents (1UP). Results are shown for when using AD and AD plus IPD to populate the decision model. 171

Table 6.1 – Expected cost effectiveness of functioning smoke alarms interventions per participant for when using AD plus IPD effectiveness evidence to inform the economic model. 189

Table 6.2 – Population EVPI at a threshold value of £20,000 per QALY (2009 values) derived from the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD for when decisions are made for: 1 subgroup (the entire targeted population); 2 subgroups – single parent status specification; 2 subgroups – parent's employment status specification; and 4 subgroups. 203

LIST OF FIGURES

Figure 3.1 – Diagrams of evidence structures: (a) indirect comparison of intervention A and B given studies on the comparisons of CA and CB, and (b) network of studies reflecting MTCs of CA, CB, AB and AD trials. 83

Figure 3.2 - Network diagram for the functioning smoke alarm outcome with information on the number and format of evidence available for each treatment comparison (continuous line) and on the number and format of evidence available for single parent status (dashed line). 93

Figure 3.3 - Parameter estimates from fitting the MTC of AD and the MTC of AD plus IPD models with information on covariate 'single parent status' to the functioning smoke alarm outcome data, considering exchangeable and independent treatment interactions, not separating between- and within-study interactions. 99

Figure 4.1 - Decision analytical model structure, part 1: model for households receiving interventions; part 2: Markov state transition model for pre-school children aged 1 to 5; and part 3: Markov state transition model for rest of life (5 years onwards)..... 120

Figure 4.2 – (a) Cost effectiveness plane and (b) acceptability curves for the functioning smoke alarms decision model. 143

Figure 4.3 – Cost effectiveness acceptability frontier for the functioning smoke alarms decision model. 144

Figure 5.1 – Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) for two parent families (2P) and single parent families (1P) from fitting the MTC model for AD (i.e. MTC AD RE) and for AD plus IPD (i.e. MTC AD+IPD RE)..... 162

Figure 6.1 - Representation of (a) an increase in estimates precision when IPD is considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 1.1). 185

Figure 6.2 - Representation of (a) a decrease (removal) of bias in obtained estimates when IPD is considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 1.2)..... 186

Figure 6.3 - Representation of (a) an increase in precision and a decrease (removal) of bias in obtained estimates when IPD is considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 1.3)..... 187

Figure 6.4 – Cost effectiveness results for the smoke alarms decision model: NMBs estimates at £30,000 threshold ratio versus the probability of the intervention(s) being cost effective. Results shown are for when AD only was used to inform the effectiveness decision model parameters and when no subgroups are considered. 188

Figure 6.5 – Population expected NMBs at a threshold value of £20,000 per QALY (2009 values) derived from the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD effectiveness evidence. 190

Figure 6.6 – Two forms of VoH: (i) the *static* value (represented by gain A and representing the NMB obtained with existing information for the average population, equivalent to the gains obtained when considering 2 subgroups); and (ii) the dynamic value (represented by the vertical distance B, where $EVPI_{B1} > EVPI_{A}$). The *x-axis* reflects the number of subgroups and the *y-axis* the NMBs. Reproduced from Espinoza *et al.* (2011) and Claxton (2011), with permission from the authors..... 194

Figure 6.7 – Representation of (a) considering existing heterogeneity in the identification of 2 subgroups and possible impact in the distribution of subgroup effects when IPD is considered and the inability of performing this analysis when using AD; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 2.1). 196

Figure 6.8 - Representation of (a) considering existing heterogeneity in the identification of 4 subgroups and possible impact in the distribution of subgroup effects when IPD is considered and the inability of performing this analysis when using AD, although possible for 2 subgroups; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 2.2)..... 197

Figure 6.9 – Representation of (a) considering existing heterogeneity in the identification of 4 subgroups and possible impact in the distribution of subgroup effects when AD and IPD are considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 2.3)..... 198

Figure 6.10 - Population expected NMBs at a threshold value of £20,000 per QALY (2009 values) derived from the cost effectiveness decision model for functioning smoke alarms.

Results are shown for the use of AD and AD plus IPD for when decisions are made for (a) the average population (repeated from above for completeness); (b) 2 subgroups, specification 1 – single parent status specification; (c) 2 subgroups, specification 2 – parent's employment status specification; and (d) 4 subgroups. 201

Figure 6.11 – Expected (individual level) NMBs with perfect information (for £20,000 threshold values and 2009 values) for the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD effectiveness evidence to inform the decision model when heterogeneity is considered. That is, no subgroups, weighted average for 2 subgroups (2 specifications – number of parents in the family and their employment status) and the weighted average for the 4 subgroups are considered. ... 204

ACKNOWLEDGEMENTS

The completion of this thesis was made possible by the financial support of the Medical Research Council and the Fundação para a Ciência e Tecnologia (Portugal). I would like to thank them for this assistance.

I am eternally grateful to a number of people for their help and support throughout my studies. In particular, my supervisors Dr Andrea Manca and Professor Alex Sutton, for their time, expertise, constant support, encouragement and guidance. Members of the Thesis Advisory Group (TAG), Professor Mark Sculpher and Dr Cynthia Iglesias also provided useful input, for which I am thankful. Special thanks goes to Professor Karl Claxton, a precious late addition to the TAG, and Professor Nicola Cooper, who provided me with helpful guidance during the initial stages of the thesis.

Members of the academic community within the Centre for Health Economics are also acknowledged, especially all members of the Team for Economic Evaluation and Health Technology Assessment for welcoming me into their group. In particular, my friend, office colleague and fellow PhD student Manuel Espinoza who provided me with continued support throughout the PhD program.

An immense thank you is due to my beautiful and gifted partner Marta Soares for her constant colossal support and for making my existence so joyful. My last word of thanks goes to my father, mother, sister and my departed grandmother for their love and support.

PUBLICATIONS

I have made efforts to disseminate the research presented in this thesis. The following papers have been submitted to peer-review journals and were accepted for publication:

Saramago P., Manca A., Sutton A. J., Deriving input parameters for cost-effectiveness modelling: taxonomy of data types and approaches to their statistical synthesis. Value in Health, 2012, 15 (5), 639-649, July 2012, DOI:10.1016/j.jval.2012.02.009.

Saramago P., Sutton A. J., Cooper N. J., Manca A., Mixed treatment comparisons using aggregate- and individual-participant level data. Statistics in Medicine, 2012, 31 (28), DOI: 10.1002/sim.5442.

AUTHOR'S DECLARATION

I declare that this thesis is my original work and that none of the material contained in this thesis has previously been submitted for a degree in this, or any other, awarding institution. The contents and views expressed reflect the best of my own knowledge, investigation, and belief.

Pedro Saramago, 2012

CHAPTER 1

1. INTRODUCTION

1.1 Background

The rising costs of health and demand pressures associated with demographic alterations, together with rising consumer expectations, have led to an escalating burden on health care systems. To guide health care policy decisions in the efficient allocation of available health resources, an increasing emphasis on the use of economic evidence has been observed in recent years. This emphasis on economic evidence goes hand in hand with a move towards evidence-based health care, which has raised health service researchers' awareness concerning a number of methodological issues. This thesis contributes to some of the recent methodological debates, aiming to explore and structure how evidence of different disaggregation levels, and in particular, evidence at the individual level can be used to best inform economic evaluations of health care technologies.

Economic evaluation in health care

Economic evaluation is defined as the "*…comparative analysis of alternative courses of action in terms of both their costs and consequences.*" (Drummond et al., 2005). Its main purpose is to assess the economic and health consequences of health care interventions, programmes or services with the aim of informing policy decisions regarding resource provision within health care systems operating under a fixed budget. Economic evaluation provides a way of systematically analysing the relevant alternatives, without which it would be impossible to explicitly identify the interventions that should be made available to maximize benefits from the available budget. This involves making difficult judgements regarding the value for money of alternative health interventions (e.g. drugs, medical devices and surgical techniques). The use of economic evaluation in health care decision making appears to have increased over the last couple of decades. This tool is seen by many health systems as a helpful instrument in controlling costs and improving efficiency in an evidence-based decision-making environment[1].

In the United Kingdom (UK), the National Institute for Health and Clinical Excellence (NICE) was set up to take the lead in the provision of clinical and cost effectiveness evidence - issues considered central to its mission (Government, 2005)[2]. The NICE is seen as an independent organisation responsible for providing

---

[1] In Australia [Department of Health and Ageing's Health Technology Assessment (DoHA - HTA)], Canada [Canadian Agency for Drugs and Technologies in Health (CADTH)], and other European countries such as Finland [Finnish Office for Health Technology Assessment (FINOHTA)], the Netherlands [Medicines Evaluation Board (CBG-MEB)], Portugal [National Authority of Medicines and Health Products (INFARMED)] and Sweden [Swedish Council on Health Technology Assessment (SBU) and the Medical Products Agency (MPA)] have introduced economic evaluation guidelines at the end of the 90's and beginning of 2000's. The United States of America (USA) – a country of, in many ways, different health care system – also considered the need to guarantee an efficient use of collective health care resources, making some Health Maintenance Organisations (HMOs) use formal economic criteria in performing decisions about which interventions to subsidize.

[2] The National Health System (NHS) Research & Dissemination Health Technology Assessment programme was setup in 1996 to evaluate health care technologies. Its main objective was to guarantee that information on costs, effectiveness and, more generally, the broader impact of health

national guidance on promoting good health and preventing and treating ill health. The NICE publishes guidelines on how it selects interventions for review and on the methods and types of analysis performed to assess them[3]. These guidelines are comprised of a set of documents that describe the processes and methods the NICE uses to undertake technology appraisals, providing guidance for the organisations invited to contribute to these appraisals.

Cost effectiveness analysis (CEA) is the type of economic evaluation that, by adhering to the principles of resource allocation, is promoted in the above mentioned guidance and has been widely used in health care. In CEA, effectiveness is commonly measured in terms of Quality Adjusted Life Years (QALYs), a composite measure combining mortality and morbidity (Williams, 1985). The theoretical and methodological strengths and weaknesses of the QALY approach have been discussed elsewhere (Weinstein et al., 2009, Drummond et al., 2005) and will be revisited later in this thesis. The main application of CEA is to support reimbursement decisions made by health care providers regarding health technologies. This tool evaluates technologies to find the one minimizing the cost of generating a given level of health, or maximizing the level of health within a specified budget (Garber & Phelps, 1997). Intrinsically, CEA is a comparative tool, requiring the contrast of costs and consequences of at least two alternative options. When making the choice between alternatives, the potential health outcomes gained

technologies, was provided by quality research. Moreover, this information was to be produced in an efficient way, aimed at those who use, manage and work in the NHS (Burns, 1998). In 1999, the NICE was established in an attempt to resolve the so-called 'postcode lottery' of health care in England and Wales, where treatments that were available depended upon the NHS primary care trust area in which the patient happened to live. In 2005 it was amalgamated with the Health Development Agency to become the new National Institute for Health and Clinical Excellence.

[3] In 2001, the NICE published a first set of technical guidance for manufacturers' and sponsors' submissions. The aim was to harmonize submissions for technology appraisal and maintain a transparency format in policy decisions. In 2004, an updated guidance was issued (NICE, 2004) which attempted to establish a clear role for economic evaluation within the NICE appraisal process and incorporate a range of methodological developments, which, by then, were raised in the cost effectiveness literature. An updated version of 2004 guidance was released in 2008 (NICE, 2008) and a new updated version is being prepared with the help of technical reports produced by the Decision Support Unit (DSU).

must be compared to those lost from interventions displaced by reallocating resources to fund this new technology. In CEA, the summary measures of interest to the decision maker are the expected values of both costs and effectiveness outcomes for each treatment strategy. These are commonly aggregated in a distinctive cost effectiveness outcome measure, as the incremental cost effectiveness ratio (ICER) $=\Delta C / \Delta E$ (where ΔC are the mean differential costs, and ΔE the mean differential effects), or its reformulation, the net benefit[4] (NB) measure. When a trade-off situation is raised, decision rules should be applied (Drummond & McGuire, 2001). If the ICER is used, it is interesting to assess the probability of its estimates being smaller than predefined fixed threshold values[5], $ICER < \lambda$ (with $\Delta E > 0$). In these circumstances, the intervention is cost effective in relation to the comparator (NICE, 2008). Cost effectiveness analysis has been undertaken in the literature using either individual participant level data (IPD) collected alongside primary studies, such as randomised controlled clinical trials (RCTs), or using decision analytic models. The latter, as discussed below, are mathematical models used to combine information from various sources (Sculpher, M. J. et al., 2006).

Decision analytic modelling for economic evaluation

In the presence of multiple sources of information, mathematical relationships need to be established to synthesise or gather data on both costs and effectiveness components of interest (Drummond et al., 2005, Gold, 1996). Cost effectiveness analysis combining information sources is denominated decision analytic modelling

[4] The Net Health Benefit (NHB) of an intervention, as defined by Stinnett and Mullahy (1998), is interpreted as *"... the net benefit (measured in units of health) of investing resources in intervention T_1, compared with T_0, rather than investing those resources in a marginally cost effective program."*. Within the NB framework, the new technology is accepted if: $NHB = \Delta E - \Delta C / \lambda > 0$, or equivalently, $NMB = \lambda \bullet \Delta E - \Delta C > 0$ (net monetary benefit (NMB)), where λ is the predefined threshold value.

[5] Current guidance (NICE, 2008) considers the existence of a cost effectiveness threshold in the form of a range of empirically plausible values (i.e. between £20,000 and £30,000/QALY). Issues surrounding the existence or not of a cost effectiveness threshold as well as its (approximate) value have received extensive debate in recent literature (Raftery, 2009, Towse, 2009, McCabe et al., 2008, Appleby et al., 2007, Culyer et al., 2007, Birch & Gafni, 2006).

(Briggs, A. H. et al., 2006). Decision analysis provides a systematic approach to decision making under uncertainty (Briggs, A. H. et al., 2006), by allowing: *(i)* a clear definition of the decision problem; *(ii)* the choice of an appropriate time horizon for the analysis; *(iii)* consistency in costs and benefits perspective; *(iv)* comparison of the new technology judged against all relevant comparators and consideration of all relevant evidence; and *(v)* an appropriate understanding of existing uncertainty and assessment of the value of acquiring additional research (Claxton, K. et al., 2007, Sculpher, M. J. et al., 2006). The majority of these items are briefly discussed in Appendix 1.

Use of evidence in cost effectiveness analysis

The current state of the art of the use of evidence in the economic assessment of health care interventions is reflected in the methods guidance for technology appraisal from the NICE (NICE, 2008). As with the 2001 and 2004 versions, more recent guidance acknowledges the need to assemble an analytical framework to synthesise available evidence for the estimation of clinical and cost effectiveness results, central to the clinical decision making context. These guidelines advise the identification, assessment and use of 'relevant' available evidence. The usefulness of the results obtained from decision models directly depends on the source and quality of the estimates informing the model (Cooper, N. J. et al., 2007, Cooper, N. et al., 2005, Briggs, A., 2000).

Currently, decision model parameter estimates are mainly obtained from diverse sources of evidence ranging from more reliable sources as RCTs to less robust ones as the estimates obtained from eliciting expert opinion. The range of relevant study designs depends on the nature of the parameter to be informed. For instance, RCT data is likely to be the preferred to inform estimates of treatment efficacy, while for costs and health related quality of life associated to an intervention, administrative sources or observational studies may be more appropriate. Techniques for systematic reviews are often used to identify the range of evidence sources available and their quality. These techniques are known for their transparency and replicable

characteristics. The formal synthesis of (effectiveness) evidence tends to be limited to RCT data and usually applied using standard meta-analytic techniques. Pair-wise and network meta-analyses, together with indirect treatment comparisons, are now commonly used to summarize evidence on clinical effectiveness in the NICE technology appraisals, and the estimates of effect they generate are frequently used to inform the economic analyses.

As discussed by Cooper *et al.* (2007, 2005), there are a number of methodological issues relating to the use of evidence in the economic assessment of heath care technologies. Problems that threaten the validity of study findings, may, for instance, be related to: *(i)* the suitability of methods to analyse/synthesise evidence – irrespective of its level of disaggregation; *(ii)* failure to consider all relevant evidence – through discarding or reducing evidence; *(iii)* failure to adequately model/adjust for effect modifiers; *(iv)* failure to appropriately reduce/eliminate existing confounding effects and/or bias; *(v)* failure to correctly reflect existing decision uncertainty. Ultimately, all these potential issues may undermine decisions to approve/reject particular options and to undertake further research. These and other methodological issues are explored throughout this thesis.

In relation to point (ii) above, and although no inclination is shown to exist in the NICE methods guidance (NICE, 2008), the use/synthesis of IPD is usually preferred by analysts/modellers because it offers several advantages over using/synthesizing published aggregated data (Simmonds et al., 2005). The idea behind this preference is that IPD better informs model input parameter estimates and correctly reveal their uncertainty by illustrating consequences on the cost effectiveness outputs of interest, helping the decision making process (Drummond & McGuire, 2001) – and responding to issues (iii) and (iv) above. These arguments are the starting points of this thesis.

1.2 Thesis aims and objectives

The overall aim of this thesis is to develop the methods on the use and synthesis of evidence for CEA for health care decision making. To achieve this, two main objectives are defined.

1.2.1 Objective 1: Exploring the use of evidence for cost effectiveness analysis – enhancing the use of individual level data

The first objective is to explore how evidence, and, in particular, how aggregate data (AD) and IPD are used to populate decision model input parameters. The use and relevance of a particular source of data varies depending on its characteristics, the type of model parameter it seeks to inform and the number of parameters it may inform, among other issues.

Two specific research targets have been set. The first target is to review the literature on the use of evidence for economic modelling of heath care interventions. This review aims to further the reader's appreciation of: *(i)* the diversity of (evidence type) scenarios that the analyst/modeller may face when wanting to inform a decision model and consequently answer a particular research question; *(ii)* the variety of modelling options available in the current literature to synthesise evidence; *(iii)* the gaps in the methods literature; and *(iv)* to provide key references where such methods have been used in practice. The second target, and linked to aim (iii) above, is to develop novel methodology to address identified gaps and fully substantiate the advantages of these with the support of a case study (see section 1.3).

1.2.2 Objective 2: Assess the added value of having access to and using individual patient level data for cost effectiveness decision making

The second objective of this thesis is to investigate the use of evidence at the individual level, compared to aggregate level, when: *(a)* assessing the cost effectiveness of alternative options (thus informing the suitability of the provision of

health care, given available information); *(b)* judging the feasibility of funding supplementary research (thus diminishing existing decision uncertainty); and *(c)* understanding heterogeneity by undertaking subgroup analysis (and therefore checking for whom (a) and (b) is true/successful). Again, to work towards this second objective of the thesis, two immediate research goals have been set.

Firstly, the thesis aims to use a case study that provides the grounds for a comparison of performances between alternative, new and existing methods of synthesis. Additionally, the case study will be a vehicle for the investigation of the impact on the estimation of CEA outcomes and decision uncertainty across a set of scenarios. Among other issues, these scenarios will evaluate decisions for the overall population and for subgroups of the population.

Secondly, as decisions based on available information are inherently uncertain, it is important to evaluate the opportunity costs of getting them wrong, conditional to the format of evidence used. Therefore, quantification of the upper boundary of the value of conducting further research is required and will be estimated. Additionally, it is crucial to have an understanding of heterogeneity in order to guide decisions about further research for different population strata. Thus, the final aim of this thesis is to evaluate the value of further research and the added value of individual level evidence in the presence and in the absence of population subgroups.

1.3 Case study

Throughout this thesis, a case study is used to illustrate the methodological issues being considered. This case study is based on a Public Health (PH) accident prevention scheme, focusing on the evaluation of interventions that promote the provision of functioning smoke alarm safety devices for the prevention of accidents in the home in pre-school children.

In 2008-9, local authority fire services attended over 700,000 fires of false alarms in the UK, almost 50,000 of which were domestic fires. In 2008-9 in the UK, 335 fatalities were estimated to have happened as a result of fires in the home, and approximately 10,000 people suffered non-fatal injuries (Government, 2009). The majority of childhood injuries are found to occur in the home (Unicef, 2001) and within the range of possible causes, fire-related injuries are considered one of the most relevant in terms of resultant disabilities, deaths and costs incurred (Government, 2009, 2004a). For several years, the UK government has conducted publicity campaigns in order to increase the number of households which have smoke alarms fitted and fully operational. Fires detected by smoke alarms tend to be discovered more rapidly and are associated with a reduced risk of death and less property damage (DiGuiseppi & Higgins, 2001). Although some parents of children aged 0-4 have smoke alarms installed, fewer take other safety measures inside the home. Generally, there is higher incidence of safety measures being adopted inside the home when children reach one year and older (Government, 2004b).

Despite several interventions being available to improve the uptake of security measures in this context, few attempts have been made to systematically review and subsequently synthesise evidence in this area (DiGuiseppi & Higgins, 2001, DiGuiseppi & Roberts, 2000, Elkan et al., 2000). These have found that interventions which were based in counselling and education did not have a significant impact on the increased ownership of smoke alarms. Nonetheless, interventions delivered with counselling as part of primary care child health surveillance have revealed an effect on smoke alarm ownership (DiGuiseppi & Higgins, 2001).

A systematic review (Kendrick et al., 2007) and its update (Wynn et al., 2010) were found to play a key role in the identification of evidence and in performing an initial synthesis of findings from a variety of sources. The primary intention of these systematic reviews was to obtain IPD from all relevant studies and to subsequently synthesise these in a meta-analysis. Unfortunately, the investigators were only successful in obtaining IPD for a proportion of the studies. The reviews included non-randomised and RCTs, as well as controlled before-and-after studies. While the initial

Cochrane review (Kendrick et al., 2007) identified much of the relevant literature base, comparative studies that did not have usual care as a comparator (e.g. studies comparing 'smoke alarms education' *vs.* 'smoke alarms education plus low-cost/free fitted smoke alarm', etc) were not considered. Therefore, a supplementary systematic review of existing reviews was conducted (Kendrick et al., 2010) to identify further relevant "head-to-head" primary studies that could be included in a network analysis. The exploration of participant-level socioeconomic characteristics was of primary interest for both reviews because there were concerns that the effectiveness of such interventions was dependent on socioeconomic characteristics.

The evidence base used for this thesis was identified by both these review studies. Details on the included trials will be provided later in this thesis but it also can be found in the systematic reviews by Kendrick *et al.* (2007) and Wynn *et al.* (2010), and also in Sutton *et al.* (2008) and Cooper *et al.* (2012).

1.4 Structure of the thesis

This thesis is structured as follows.

Chapter 2 develops a taxonomy based on possible scenarios typically faced by the analyst when dealing with the evidence base. This provides guidance to modellers on the appropriateness of certain methodologies which may enable the use/synthesis of available data to inform a given model parameter. Although its main focus is on effectiveness type parameters, this chapter also briefly considers available methods for the use of evidence in other key economic model parameters. Advantages and disadvantages of using evidence at the individual level, compared to aggregate level, are discussed throughout and gaps in the methods literature identified.

Chapter 3 begins by revising available methodologies for the estimation of combined statistics when direct head-to-head comparisons are at stake. Revised

methods consider the synthesis of study summary estimates, of evidence available at individual level and of the mixture of AD and IPD – all in a binary outcome setting. It illustrates how to incorporate treatment-effect modifiers in all the above mentioned modelling scenarios. These are extended to the indirect and mixed treatment comparisons (MTCs) framework, with the development of novel synthesis methodologies. All model implementation, including that for novel models, is supported by the above mentioned motivating example.

Chapter 4 focuses mainly on the case study decision problem relating to the assessment of 'functioning' smoke alarm programs for the prevention of fire-related injuries of pre-school children in the home. A decision analytic model that addresses the decision problem is described and cost effectiveness results are discussed. Summary evidence is the source used to populate the effectiveness model input parameters. Different viewpoints of the analysis are evaluated and subgroup cost effectiveness analysis is implemented in order to evaluate whether suboptimal intervention decisions are being made for different subgroups of patients.

Chapter 5 evaluates the impact of using different effectiveness model inputs (from alternative synthesis models for AD and/or IPD) over the cost effectiveness outcomes. Chapter 4 and Chapter 5 are interrelated in the sense that both follow from Chapter 3 by using the results from the relevant (novel and existing) synthesis models, and both make use of the same (case study) decision analytic model.

Chapter 6 deals with the issue of whether and for whom it is worthwhile funding additional research. For the whole population of interest or for subsets thereof, this chapter highlights the advantages of having access to IPD, compared to having summary data only, when quantifying decision uncertainty and estimating the expected cost of uncertainty.

Finally, **Chapter 7** brings together the conclusions of the thesis, focusing on its contributions to the methods in evidence synthesis and CEA. This chapter concludes

with a discussion of future research topics that emerges from the work that has been produced.

CHAPTER 2

2. DERIVING INPUT PARAMETERS FOR COST EFFECTIVENESS MODELLING: TAXONOMY OF DATA TYPES AND APPROACHES TO THEIR STATISTICAL SYNTHESIS

2.1 Introduction

Economic evaluations assess the costs and health consequences of competing health care interventions, programme or services. Their aim is to inform policy decisions regarding resource provision within health care systems operating under a limited or fixed budget.

The information required to carry out an economic evaluation rarely comes from a single study (Sculpher, M. J. et al., 2006). More commonly, the evidence base informing the model parameters is represented by one or more data sources, including individual patient level datasets (e.g. RCTs and observational studies), expert opinions, and secondary data analyses (e.g. meta-analysis). Decision analytic models represent an ideal vehicle to structure the decision problem, combine all available data and characterise the various sources of uncertainty associated with the

decision problem (Sculpher, M. & Claxton, 2005). As with any modelling framework, the results of the analysis depend on the suitability of the model structure, the quality of the data inputs and the methods used to derive these (Cooper, N. J. et al., 2007).

The NICE (or the Institute) for England and Wales is one of the many national agencies worldwide that recognise the value of decision models to inform the assessment of whether or not technologies represent value for money. The Institute's guideline for methods of technology appraisal (NICE, 2008) recommends that after defining *'…explicit criteria by which studies are included and excluded…'*(page 14) *'…all relevant evidence...'* should be *'…identified, quality assessed and, when appropriate, pooled using explicit criteria...'* by means of *'...justifiable and reproducible methods* (page 27)'.

One issue typically faced by health economics modellers is how to proceed when multiple sources of evidence are available to inform the same model input (e.g. relative effectiveness). In the last decade, at least for effectiveness parameters, there has been a shift towards recognising the need for a more systematic identification and utilisation of statistical evidence synthesis in decision models (Cooper, N. J. et al., 2007), with approaches such as meta-analysis or MTCs increasingly being used in CEA (Ades, A. E. et al., 2006b).

Parameters used in decision analytic models, for instance, are increasingly being estimated from AD available from published literature. There are, however, several examples where the model parameters have been derived almost exclusively from a single individual patient level trial dataset (Epstein, D. M. et al., 2008, Henriksson et al., 2008, Briggs, A. et al., 2007, Mihaylova et al., 2006, 2005). Advantages of the latter approach, compared to using AD only, include more accurate modelling of the disease's natural history and the possibility of exploring heterogeneity in baseline risk (and/or relative treatment effect) across patient groups. In this case, the challenge is how to integrate IPD with any other component of the evidence-base that may be available in aggregate or summary measures format. Methods for

combining multiple individual level (Higgins et al., 2001), or IPD and AD, are rapidly developing (Riley, Richard D. & Steyerberg, 2010, Riley, Richard D. et al., 2008, Riley, R. D. et al., 2008, Sutton, A. J. et al., 2008, Riley, R. D. et al., 2007), although many applied health economics modellers are currently unaware of these.

This chapter develops a taxonomy based on possible scenarios typically faced by the analyst when dealing with the evidence base. Since most of the methods development took place in the area of statistical synthesis of clinical effectiveness measures from RCTs, the proposed taxonomy is structured around examples concerning such parameters (section 2.2). Statistical approaches available to synthesise the evidence base under different scenarios are briefly reported and discussed together with key references to full explanations of the methodologies and examples of where such methods have been used in practice. This chapter makes no claim to be exhaustive with respect to reviewing the various applications, as this is not its objective. Instead, the aim is to use these examples to illustrate and to provide recommendations regarding which techniques are most appropriate in order to use synthesise available information depending on its *format*, *number of data sources* and *number of parameters to be derived*.

In addition to applying this taxonomy to clinical effectiveness parameters, its application is considered in relation to other key economic model input parameters (in section 2.3) of an economic model including disease natural history, resource use / costs, and preferences, with a view to discussing issues with the application of the taxonomy to these other parameters. In doing so, it is hoped to encourage a fuller application of this taxonomy to non-effectiveness parameters in future modelling. Finally, section 2.4 summarises the main points of the manuscript and includes suggestions for future research.

2.2 A taxonomy for the use of evidence in cost effectiveness models: application to clinical effectiveness evidence

Good practice in health economic evaluation suggests that decision models should be structured in a way that appropriately reflects the decision problem at hand (Weinstein et al., 2003). The evidence used to inform the model inputs often comes from different sources, with potentially heterogeneous designs (e.g. RCTs, observational and expert panels). International HTA standards in systematic reviews and meta-analysis (Higgins & Green, 2008) indicate that good quality RCT evidence is the preferred data source for estimating the main clinical effect(s) of interest.[6] Country-specific HTA guidance documents provide more heterogeneous indications as to whether or not it is acceptable to use non-randomised evidence to inform the main clinical effectiveness part of the model in the absence of evidence from good quality randomised studies (McGhan et al., 2009). In the case of the NICE for instance, its methods guidance states that any limitations of the methods used, potential biases in obtained parameter estimates, caveats about the interpretation of results and appropriate reflection of parameter uncertainty should be extensively reported in the analyses submitted for consideration of the Institute (NICE, 2008). For simplicity, this chapter deals with situations in which the main body of evidence for effectiveness comes from randomised studies[7].

It is argued here that the selection of an appropriate method for the analysis and synthesis of clinical effectiveness data for use in a decision model does depend on three dimensions of the evidence base (listed below), the combination of which gives rise to a taxonomy of possible scenarios the analyst may face, as illustrated in Table

[6] There are of course many features of the evidence base (e.g. characteristics of target population, use of intermediate outcomes rather than final ones) that may complicate its use for informing a particular economic analysis. However these are not specific to randomised data alone.

[7] The added level of complication deriving from the inclusion of non-randomised data is discussed when relevant to the argument.

2.1. These are briefly introduced here and discussed more fully in the following sections.

Number of available sources of evidence Depending on the research question, there may be multiple sources of evidence from RCTs from which to derive an estimate of clinical effectiveness, although there are examples where a single RCT provides the only evidence available.

Formats in which data are available The above evidence may be available in *(a)* aggregate form only (sometimes referred to as summary data), *(b)* at the individual level or, when multiple sources of data are available, *(c)* a combination of AD and IPD[8].

Number of (effectiveness) parameters to be derived It is important to distinguish between the need to synthesise the evidence to inform a single parameter versus the need to estimate multiple parameters for use in the decision model[9].

[8] Care must be taken when classifying these data formats, since in some contexts IPD may not contain any extra information beyond what is conveyed by available summary statistics. For example, basic IPD can be reconstructed from a summary 2 x 2 table recording numbers of individuals at risk and those who experienced a binary outcome in a 2-arm trial. Either approach will give the same estimate of the odds ratio (OR) of effect (Lambert et al., 2002). In this case, the OR is a sufficient statistic, in the sense that *"...no other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter"* (Fisher, 1922).

[9] One example may be the synthesis of the sensitivity and specificity of a diagnostic test, another one may be the need to synthesise (one or more) clinical outcome(s) reported at different time-points - most often these requiring different analytical and evidence synthesis strategies.

Table 2.1 – A gallery of scenarios arising when using clinical evidence in cost effectiveness models

		Single source of evidence		Multiple sources of evidence	
		Parameter		Parameter	
		Single	Multiple	Single	Multiple
Format of data available	**Aggregate level data**	Scenario A1	Scenario B1	Scenario A2	Scenario B2
	Individual level data	Scenario C1	Scenario D1	Scenario C2	Scenario D2
	Mixture of aggregate and individual level data	---	---	Scenario E2	Scenario F2

2.2.1 Single source of evidence

Let's start with the simplest scenario of all, that is, where there is only one single source of evidence from which to derive the parameter(s) of interest. In this case the problem is not how to synthesise the available evidence but how to make the best of this single source to inform parameter estimate(s) for use in the economic model.

2.2.1.1 Aggregate data to inform the estimation of a single parameter (A1)

If all the available evidence is in the form of published (summary) results of a single study, the simplest option is to use these data in the model 'as they are' to inform the derivation of the relevant parameter estimate in the decision model. For a probabilistic representation of these parameters the analyst will need to have access to multiple statistics from the source of evidence (e.g. mean and standard error). Also, plausibility and sample characteristics (e.g. skewness) may be used to define an appropriate distribution.

Clearly, exploration of any statistical heterogeneity[10] relating to a parameter in this circumstance is unfeasible and usually no further appraisals of the evidence are possible, other than a simple sensitivity or threshold analysis. At this stage and in the absence of other source of evidence, the analyst may want to explore whether attempting to acquire further evidence, through other techniques (e.g. expert elicitation) is worth the effort.

2.2.1.2 Aggregate data to inform the estimation of multiple parameters (B1)

It is possible for a single published study to provide several inputs that may be used to derive model parameters. For instance, a single (three-arm) trial may provide effectiveness data for a decision model evaluating the same three alternative

[10] Statistical heterogeneity refers to variability between effect sizes from studies than would be expected from chance only.

treatments. Since the resulting measures of relative effectiveness (such as ORs or log-ORs) between these three arms are inherently correlated, it has been recommended that correlation between parameters should be explicitly modelled where possible. Failure to do so would produce not just an incorrect assessment of the uncertainty[11] in the model but also result in an incorrect estimation of each treatment's expected costs and benefits (Epstein, D. & Sutton, 2011, Ades, A. E. et al., 2006a, Ades, A. E. & Lu, 2003). Methods such as indirect or MTC models can be successfully used to address the above problem (see section 2.2.2 and references therein) (NICE, 2008).

Another situation where it is possible to derive multiple parameters from a single study occurs when the interest of the modeller lies in estimating multiple outcomes/multiple time points on the same treatment comparison. The range of possible analytical options here may be limited by the lack of information on the correlation between outcomes. In some cases, approximate or *ad hoc* methods may be available to take into account the correlated nature of the outcomes (e.g. the phi coefficient, Yule's Q or Yule's Y – see Epstein and Sutton (2011) for further details), and this will usually be preferable to assuming the outcomes are independent (but less desirable than obtaining the IPD and estimating the correlations directly).

2.2.1.3 Individual patient data to inform the estimation of a single parameter (C1)

Access to IPD, especially when there is only one relevant study forming the evidence base, is particularly advantageous since it allows re-analysis of the data (e.g. inclusion of further explanatory covariates, conduct of more in-depth analyses than is possible from summary evidence extracted from published reports (Stewart & Clarke, 1995)) aimed to derive appropriate model input parameters. Indeed, compared to the use of AD, the analysis of IPD may be considered the most flexible

[11] In such instances, propagation of correlations is automatic if parameter estimation is conducted in the same program as the decision model (sometimes called one-step comprehensive decision modelling (Cooper, N. J. et al., 2004)), or can be achieved by specifying the full multivariate distributions for the correlated parameters (Ades, A. E. & Lu, 2003).

way to explore and answer clinical and economic research questions. In this case, data can be analysed using the range of statistical models developed to analyse trial data to estimate the decision model parameter of interest (Glick, 2007, Bland, 2000). Given that appropriate methods in this context are extensively documented elsewhere, this section is kept brief.

2.2.1.4 Individual patient data to inform the estimation of multiple parameters (D1)

The availability of IPD often enables the estimation of multiple model parameters. The economic model constructed around the third Randomised Intervention Trial of unstable Angina (RITA 3) (Henriksson et al., 2008) is an example where the trial data were used to derive estimate rates of cardiovascular death and myocardial infarction (as well as costs and health-related quality of life) through regression models applied to a single individual patient-level dataset. As noted in section 2.2.1.2, the possibility to estimate correlations between correlated input model parameters based on summary measures is often limited by the data being reported. This is no longer an issue when one has access to the original study IPD. Another important area where access to IPD facilitates estimation of multiple parameters relates to the analysis of time-to-event data. Since trials' follow up are short in duration (Sculpher, M. J. et al., 2006) to produce long term estimates of cost effectiveness, most models need to extrapolate the observed trial results (e.g. fatal and non-fatal events) beyond the trial follow up. This can be achieved employing parametric distributions to model the outcome of interest, which are typically governed by a combination of two or more correlated ancillary parameters. Popular examples of parametric distributions include the Weibull, the Log-Logistic and the Generalised Gamma (Collett, 2003) for the analysis of time-to-event outcomes.

2.2.2 Multiple sources of evidence

There are situations where the evidence base is represented by multiple studies. Depending on the format in which they are available and the number of parameters needed to estimate, these give rise to six possible scenarios (Table 2.1).

2.2.2.1 Multiple aggregate data to inform the estimation of a single parameter (A2)

A typical case occurs when there are several studies reporting results on the same parameter of interest, and the researcher needs to combine these into a single quantitative estimate. The statistical methods most commonly used to achieve such a synthesis fall within the meta-analytic family (Whitehead, 2002). In standard meta-analysis of clinical trials, the parameter of interest is usually some measure of comparative effectiveness between treatment arms.

A fixed effect[12] (FE) meta-analysis is carried out under the assumption that a single common (or 'fixed') effect underlies every study in the meta-analysis (Higgins & Green, 2008). It is common, though, to observe between-study variation in treatment effect estimates (heterogeneous treatment effects). In such a case, it is customary to use a random effects[13] (RE) model. For an up to date comprehensive review of recent developments in meta-analysis, the reader can refer to the paper by Sutton and Higgins (2008).

Some authors have argued that one of the prime weaknesses of meta-analysis is a possible failure to control for sources of bias, and that a good meta-analysis of badly designed studies will still result in a biased combined statistic. In addition to the inclusion of evidence of sub-optimal quality, publication and other related biases may be present (Sutton, A. J. et al., 2000b).

Study level features, such as participants' characteristics, which may lead to between-study heterogeneity, can be investigated by adopting a meta-regression approach in which study-level covariates are included in the analysis. Some

[12] In the FEs approach, if, for instance, a meta-analysis of odds ratios is being done, it is assumed that every study is estimating the same odds ratio. Therefore, only within-study variation is taken to influence the uncertainty in the results.

[13] Random effects meta-analysis makes the assumption that while individual studies are estimating different treatment effects, these come from a common distribution with some measure of central tendency and some measure of dispersion (Higgins & Green, 2008).

researchers would prefer to include 'weaker'/'low quality' studies in the meta-analysis, and add a study-level covariate reflecting the methodological quality of the trials in order to assess the impact of trial quality on the effect size. Unfortunately, meta-regression methods also have a number of weaknesses (Thompson & Higgins, 2002). The analyst should be aware of the fact that the use of such mean study-level covariate values has low power (over IPD methods) and, more importantly, carries the risk of 'ecological fallacy'[14] (Piantadosi et al., 1988) if these average patient level characteristics are considered (Berlin et al., 2002, Lambert et al., 2002). In this sense, as it shall be seen in section 2.2.2.3, access to IPD can be used to disentangle the relationship between the parameter of interest and baseline covariates (Wakefield, 2008). Ades *et al.* (2005) provide an extensive discussion of how between-study heterogeneity can be incorporated into the parameters of a decision analytic model[15].

In CEA, the use of estimates of relative treatment effects derived from a meta-analysis is common (Gold, 1996). An example of its use within an economic model is the prevention and treatment of influenza A and B (Turner, D. et al., 2003), where a separate meta-analysis was conducted to evaluate time to symptoms alleviated and time to return to normal activities for different baseline risk groups. In another study, McKenna *et al.* (2010) recently carried out a systematic review and economic evaluation of the clinical effectiveness and cost effectiveness of aldosterone antagonists for post-myocardial infarction heart failure. The authors estimated the effectiveness parameter to inform their cost effectiveness model using a Bayesian meta-regression model.

[14] Ecological fallacy refers to situations in which relationships observed at the aggregate variable level are incorrectly inferred to exist also at the individual level.

[15] Notice that if the heterogeneity parameter is used to derive parameters for a decision analytic model – technically the analysis is estimating multiple parameters (i.e. a RE estimate of the treatment effect, its variability and a measure of heterogeneity) and thus belongs to the B2 category.

2.2.2.2 Multiple aggregate data to inform the estimation of multiple parameters (B2)

Meta-analysis can also be used to achieve more complex forms of evidence synthesis, to address issues related to multiple (indirect and mixed treatment) comparisons and combinations of evidence on multiple or surrogate/intermediate endpoints (Baker, 2006). Much of the published work on these complex methods of synthesis has been undertaken within a Bayesian framework, mainly for computational reasons but also because of its coherent link to decision making (Ades, A. E. et al., 2006b). The term 'multi-parameter evidence synthesis' (MPES) adapted from Hasselblad and McCrory (1995) has been coined to designate these extended methods of synthesis.

When multiple outcomes are of interest, a multivariate meta-analysis model facilitates the joint estimation of these endpoints, thus estimating possible correlation between them. Often the advantage of a multivariate REs meta-analysis lies in its ability to use the within-study and between-study correlation of the multiple endpoints of interest. For example, Reitsma *et al.* (2005) have suggested applying a bivariate REs meta-analysis to jointly synthesise logit-sensitivity and logit-specificity values from diagnostic accuracy studies.

More generally, a common feature of the evidence base used to inform health care funding decisions is the absence of head-to-head trials comparing all relevant treatment strategies. When more than two treatments are to be compared and the evidence base contains different randomised pair-wise or multi-arm comparisons, the appropriate techniques to use in the decision making context are indirect treatment comparisons and network meta-analysis (or MTCs), which are simple extensions of the pair-wise meta-analysis method (Lu, G. & Ades, 2004, Lumley, 2002). MTCs can be recognized as an example of MPES, in which parameters are related to one another by a definable structure (Ades, A. E., 2003).

In a MTC the modeller may choose between a FEs and a REs analysis, depending on the assumptions made about any between-trial heterogeneity, as discussed in A2

(Ades, A. E., 2003, Higgins & Whitehead, 1996). MTC relies on exactly the same assumptions as standard pair-wise meta-analysis (i.e. choice and quality of the studies), although now these are applicable to the full set of interlinked trials. Therefore, the similarity between trials included in the network will also be a determinant of the internal validity of the analyses, at the risk of having high confounding bias (Dias et al., 2010a). In the instances where direct and indirect evidence are combined for a particular comparison, it is also vital that there are no disagreements between the direct and indirect comparisons[16] (Dias et al., 2010b, Salanti, Georgia et al., 2008, Lu, G. B. & Ades, 2006). As for standard meta-analysis, in network meta-analysis it is important to allow for between-study heterogeneity (Cooper, Nicola J. et al., 2009). An extension of this family of techniques, allowing for the incorporation of study-level covariates to explain between-study heterogeneity and reduce synthesis model inconsistency, is also available (Salanti, G. et al., 2010, Cooper, Nicola J. et al., 2009, Salanti, G. et al., 2009). Additional details on the use of indirect and mixed treatment comparisons can be found elsewhere (Sutton, A. et al., 2008).

A good example illustrating the use of the MTC framework when multiple follow-up times are available is the paper by Lu *et al.* (2007). For an application of the MPES approach, the reader is referred to the recent work by Welton *et al.* (2008), which was originally developed using data from the earlier economic appraisal of antiviral treatment by Turner *et al.* (2003) referred to in A2. Another example where MTC was used in an economic analysis can be found in Woolacott *et al.* (2006). The authors synthesised clinical effectiveness data from several published trials in epilepsy to estimate the transition probabilities needed to populate a state-transition model developed to assess the cost effectiveness of alternative medications for epilepsy.

[16] For instance, in a MTC model comparing three treatments (e.g. A, B, C) consistency is achieved when, for each pair-wise comparison, no discrepancies can be found between the direct and indirect estimates of the parameter of interest (e.g. OR) derived from the model. The issue here is in defining how big a difference is considered a discrepancy – although this is arbitrary to define, there are statistical tests (potentially with reduced power) for discrepancies found (Dias et al., 2010b).

2.2.2.3 Multiple individual patient datasets to inform the estimation of a single parameter (C2)

Meta-analysis of IPD or 'mega-analysis', where raw data from each study is obtained and synthesised to inform the estimation of a single parameter of interest, is considered the 'gold-standard' in evidence synthesis (Higgins et al., 2001, Sutton, A. J. et al., 2000a). This approach has a series of advantages, which are summarised by Stewart and Parmar (1993), Stewart and Tierney (2002) and Simmonds *et al*. (2005). Access to multiple individual level datasets avoids the risk of bias associated with published AD; it allows one to obtain information possibly not available from published reports (or not available in the format required for the meta-analysis and cost effectiveness model); and it facilitates consistent inclusion/exclusion criteria to be used across studies (Jeng et al., 1995, 1993). An increase in statistical power to detect true patient-treatment relationships is gained when compared to meta-regression of AD (Smith, C. T. et al., 2005, Lambert et al., 2002), which only assesses treatment in relation to group-level summary data (Cooper, N. J. et al., 2007). It should be highlighted that, however, in most situations, access to IPD may be difficult due to issues such as confidentiality, sponsors' or investigators' rigidity in releasing this data.

Surprisingly, there is a paucity of published literature concerning methodologies for meta-analysis of IPD. Simmonds *et al.* (2005) recently published a review of meta-analysis using trial-based IPD suggesting that most methods used in practice are straightforward. The review shows that the majority of applications use a 'two-stage' process where initially each dataset is analysed separately, AD is drawn for each study (stage one) and subsequently combined using a 'standard' meta-analytic model for aggregate evidence (stage two). This approach may be considered a simplification of the techniques discussed in scenario A2. Alternative and more robust approaches for dealing with binary (Turner, R. M. et al., 2000), ordinal (Whitehead et al., 2001), continuous (Higgins et al., 2001, Goldstein et al., 2000) and longitudinal outcomes (Jones et al., 2009, Farlow et al., 2005) based on REs

generalised hierarchical models exist. Unfortunately, these approaches appear to be rarely used in practice.

It should be noted that if the outcome of interest is binary and the analyst does not need to control for covariates, the information from the AD will report the *sufficient statistics*, that is, no additional benefit is obtained from access to IPD (as in scenario A2).

2.2.2.4 Multiple individual patient data to inform the estimation of multiple parameters (D2)

As outlined in section 2.2.2.2, there are clear theoretical and practical benefits (besides an obvious policy rationale) that justify why it would be desirable to carry out an MTC (and MPES models in general) when deriving parameters for use in CEA. Many of these benefits will also apply when the purpose is to analyse multiple datasets from which to derive multiple parameters for use in decision modelling. Nevertheless, some authors believe that this need is exacerbated by the fact that an MTC is essentially an observational study comparing several treatment strategies. For instance, Salanti *et al.* (2009) point out that while each individual trial may have high internal validity, studies included in an MTC will almost inevitably display between-study variability in study-level characteristics that can affect the relative effectiveness of the strategies being compared. One example of this is the definition of 'Placebo' or 'Standard Care' in many MTCs (Salanti, G. et al., 2009) and cost effectiveness models (Hawkins & Scott, 2010), which has been found to vary enormously between studies. Another example is the work by Nixon *et al.* (2007) and subsequent two cost effectiveness models (Wailoo et al., 2008, Brennan et al., 2007), in which the authors conducted a covariate adjusted AD MTC of trial evidence in drugs for rheumatoid arthritis. While this approach is better than an unadjusted MTC, it still suffers the same limitations as standard meta-regression, making it essential to carry out a series of tests to assess the consistency of the evidence in the evidence base network.

2.2.2.5 Mixture of individual and aggregate level data to inform a single parameter (E2)

When the analyst opts for trying to acquire IPD from each relevant study forming the evidence-base, the most frequent scenario s/he encounters is that these data will be made available only for a subset of the evidence base. In such situations, analysts have traditionally taken two alternative routes: *(a)* include only the studies for which IPD was available; or *(b)* for the studies where IPD was available, collapse these to summary evidence and use only the latter. Neither solution makes optimal use of the available data. The first option throws away important information and the second ignores all the advantages that IPD may bring towards an improved estimation of the effect size. A better approach would be to *jointly model* the IPD and AD (Riley, R. D. et al., 2007). As yet, this issue has not received attention in the health economics literature; although a series of meta-analysis models have been recently developed (Riley, Richard D. & Steyerberg, 2010, Jackson et al., 2008, Riley, Richard D. et al., 2008, Riley, R. D. et al., 2008, Sutton, A. J. et al., 2008, Jackson et al., 2006) in the statistical literature specifically for this purpose. These models are valuable in reducing between study heterogeneity, or in identifying patient subgroups with differential treatment effects.

2.2.2.6 Mixture of individual and aggregate level data to inform multiple parameters (F2)

Access to IPD (alongside existing AD) is particularly important when the objective of the evidence synthesis model is the estimation of multiple input parameters to populate a cost effectiveness model. This is clearly the most technically challenging scenario the analyst may face as the existence of such models have not been encountered, despite the scenario discussed in this section being quite common. In this sense further methodological research aimed to develop models appropriate to deal with these situations are welcome.

An interesting application does exist for diagnostic test evaluation (Riley, Richard D. et al., 2008), with a bivariate meta-analysis model being used to model outcomes of

diagnostic studies, although this was not used to inform a cost effectiveness model. By reconstructing IPD from available AD, the authors manage to take into account the mixture of IPD and AD which allows for all evidence set to be simultaneously considered in estimating the parameters of interest (i.e. sensitivity and specificity of the test). Both IPD and AD studies contribute to the estimation of the impact of study-level covariates and the across-study effects.

For guidance, a summary of the scenarios with corresponding recommend methods and related methodological and applied literature is shown in Table 2.2.

Table 2.2 – Scenarios and corresponding current methods literature for when using clinical evidence to inform cost effectiveness modelling.

Scenario	Source of evidence	Format of data	Parameter	Methods *	Relevant references
A1	Single	AD	Single	Direct inclusion of reported estimate (or transformation of it) in the model.	----
B1	Single	AD	Multiple	Direct inclusion of estimates or transformations of them in the model - correlation should be included if reported.	(Epstein, D. & Sutton, 2011, Rodgers et al., 2011, Ades, A. E. et al., 2006a, Ades, A. E. & Lu, 2003)
C1	Single	IPD	Single	Using standard analysis procedures relevant for the primary study.	(Glick, 2007, Stewart & Clarke, 1995)
D1	Single	IPD	Multiple	Using standard multivariate estimation procedures - including correlations.	(Henriksson et al., 2008, Stewart & Tierney, 2002)
A2	Multiple	AD	Single	Meta-analysis	(Sutton, A. J. & Higgins, 2008, Ades, A. E. et al., 2005, Turner, D. et al., 2003, Berlin et al., 2002, Lambert et al., 2002, Thompson & Higgins, 2002, Whitehead, 2002, Sutton, A. J. et al., 2000b, Stewart & Parmar, 1993)
				Meta-regression	
B2	Multiple	AD	Multiple	Multivariate meta-analysis (e.g. Bivariate random-effects meta-analysis)	(Dias et al., 2010b, Salanti, G. et al., 2010, Cooper, Nicola J. et al., 2009, Salanti, G. et al., 2009, Salanti, Georgia et al., 2008, Sutton, A. et al., 2008, Welton et al., 2008, Lu, G. et al., 2007, Lu, G. B. & Ades, 2006, Woolacott et al., 2006, Reitsma et al., 2005, Lu, G. & Ades, 2004, Ades, A. E., 2003, Lumley, 2002, Hasselblad & McCrory, 1995)
				Mixed treatment comparison	
C2	Multiple	IPD	Single	'Mega-analysis' (meta-analysis using IPD)	(Jones et al., 2009, Farlow et al., 2005, Simmonds et al., 2005, Smith, C. T. et al., 2005, Whitehead et al., 2001, Turner, R. M. et al., 2000)
D2	Multiple	IPD	Multiple	Multivariate meta-analysis using IPD (e.g. Bivariate random effects meta-analysis)	(Nixon, R. M. et al., 2007)
E2	Multiple	Mixture	Single	Two-stage - reduce IPD to AD or reconstruct IPD from AD	(Riley, Richard D. & Steyerberg, 2010, Jackson et al., 2008, Riley, Richard D. et al., 2008, Riley, R. D. et al., 2008, Sutton, A. J. et al., 2008, Riley, R. D. et al., 2007, Jackson et al., 2006)
				One-stage - hierarchical modelling (e.g. meta-analysis using AD and IPD)	
F2	Multiple	Mixture	Multiple	Extensions of previous synthesis models to the hierarchical framework (e.g. Bivariate random-effects meta-analysis of IPD)	(Riley, Richard D. et al., 2008)

* Where there is the need to define uncertain parameters for inclusion in a decision model and evidence is available as in scenarios A1 and B1, it may be necessary that multiple statistics are reported. Also, it may be of interest to have some information on the characteristics of the distribution best representing the parameter. All subsequent scenarios allow for expressions of uncertainty.

2.3 Use of evidence for other model input parameters

While quantitative evidence synthesis methods typically focus on clinical effectiveness (including adverse events), cost effectiveness models require information on many other input parameters, the most important being disease natural progression, cost/resource use, and (health state) utility data. There are some examples in the literature where evidence synthesis techniques have been applied to estimate these parameters. In this section the specific characteristics of these parameters are described and the implications for methods of synthesis highlighted – aiming to illustrate taxonomy's applicability.

While clinical effectiveness data used to populate the model typically come from RCTs, the evidence base used to estimate disease natural progression, resource use/cost and utility data parameters is often derived from observational evidence (e.g. registries, administrative claim data) (Briggs, A. H. et al., 2006). There are various reasons for this, beyond the fact that RCT evidence may not be available to populate these model parameters.

First, due to their intrinsic design, randomised data on resource use/cost are often considered to have low external validity. This may be due to the fact that the trial evidence does not reflect true clinical practice (Drummond et al., 2005), or the evidence may not be relevant to the decision-maker for whom the model is being developed (Urdahl et al., 2006). In this case, observational evidence may provide an opportunity to calibrate the model parameters and assess the extent to which trial evidence reflects real-world situations. Methods developed in the generalised evidence synthesis framework, which facilitate the synthesis of both randomised and non-randomised data while accounting for the different study designs (Spiegelhalter, David J. & Best, 2003, Prevost et al., 2000) and methods developed for cross-design synthesis (Ades, A. E. & Sutton, 2006) may be useful here. Applications of these methods outside the analysis of clinical effectiveness have not been encountered, and

further research in this direction, focussing on parameters such as disease natural history, could be of great interest. Second, resource use / cost (and to some extent utility) data are country specific (Manca & Willan, 2006), and, in many cases, it is almost impossible to find jurisdiction-specific RCT evidence on these parameters (Augustovski et al., 2009, Schulman et al., 1998). Third, most RCTs have a short follow up duration. To model disease natural history as well as long term costs and utilities, (large) long term, observational studies are often the only solution. Other specific issues for each type of parameter will be discussed in turn in the next subsections.

2.3.1 Disease natural progression data

One of the initial and most important phases in building any decision model is to explicitly define its structure. This entails, among other things, giving an appropriate representation of the key health states that the population of interest may experience over time, and reflecting what is known about the natural history of the particular health condition being modelled as well as the impact of alternative treatment options on the disease process. These procedures should be performed in collaboration with both clinical and non-clinical experts from the field(s) of interest. Evidence about the natural history of a disease is crucial for a good understanding of possible clinically-defined states and, in view of the complexity of the task, long-term IPD are ideal for this.[17] Furthermore, given the concerns about the external validity of trial-based data evidence, a favoured source for baseline risk data is often case series or high-quality individual level administrative or epidemiological datasets (Cooper, N. et al., 2005). Despite relaxing the evidence base inclusion criteria to model disease natural history, IPD are still very often unavailable, leaving published summary evidence as the only feasible option to inform the model parameters.

[17] Detailed individual level natural history data are particularly important for modelling the impact that baseline characteristics may have on parameters in the model that capture the occurrence of clinical events beyond follow up, and associated resource use and health-related quality of life.

With respect to the synthesis of data informing the baseline history part of the model, it is worth highlighting the publication of a recent useful document from the NICE DSU (Dias et al., 2011c). This report reviews evidence synthesis issues which may arise when dealing with baseline natural history modelling. A discussion of the source of evidence to use for baseline outcomes, the simultaneous versus separate modelling of baseline and treatment effects and the inclusion of covariates in baseline models is provided by the authors.

Although not presented within an evidence synthesis framework, Isaman and colleagues (Ye et al., Isaman, D. J. M. et al., 2009, Isaman, D. J. et al., 2006) proposed an approach that allows the use of published regression data to populate a multi-state model describing disease natural history, even when the published study may have ignored intermediary states in the multi-state model (taxonomy section B1). The authors applied their proposed methodology to model several chronic conditions, including heart disease and diabetes. Welton and Ades (2005), however, use evidence synthesis methods applied to AD to estimate transition probabilities from transition rates, with the objective of using these to model disease progression. The authors illustrate how to statistically combine data from multiple sources, including partially observed data at several follow-up times, to inform an epidemiologically realistic model (taxonomy section B2). Chao and Chen (2009) recently used a similar approach and developed a multi-state Markov model to predict the progression of age-related hearing loss, by synthesising partially observed AD from four studies from which they derived progression rates (taxonomy section B2).

Modelling disease natural history becomes a lot easier when the analyst has access to IPD, and there are many examples in the literature that show how one can proceed in this case. Marshall and Jones (1995), for instance, developed a multi-state model to describe disease progression in diabetic patients with retinopathy, and used patient-level covariates in the model to capture the natural course of the disease and identify the factors associated with progression and regression between disease stages

(taxonomy section D1). Applications in the medical field, which, in the presence of multiple patient-level data, carried out the synthesis of these for the purpose of modelling the natural history of a disease have not been encountered.

2.3.2 Cost / resource use data

The quantification of the cost of each alternative strategy being compared is essential for any economic assessment. The best study design for quantifying health care resource utilisation includes prospective data collection within a long term naturalistic trial setting. In the absence of these, retrospective analysis of existing data sets, complemented with examination of administrative databases, can be an alternative solution. In fact, it is not uncommon for model parameters associated with health care resource use to be estimated by reviewing routine data (e.g. hospital records) (Cookson et al., 2005, Tumeh et al., 2005). Elicitation of expert opinion (Connock et al., 2006, Wu et al., 2006) may also play a role in informing the estimates of some resource use model parameters, although this source of evidence is considered the least preferred, as it usually carries considerable levels of subjectivity.

An article by Bower *et al.* (2003) is one of the few examples in which meta-analytic techniques were employed to synthesise cost data. Using data on costs from trials of counselling in primary care, the authors attempted to overcome sample size limitations in their economic analysis by pooling short and long term resource use data from four different studies using a FEs meta-analysis (taxonomy sections A2 and B2). This approach has a number of limitations, as the authors pointed out (e.g. the significant variation between trials in standard deviations of costs and the difficulty in identifying comparable data and consequently of standardising cost means), reinforcing the fact that under no circumstances will the performed analysis approach the precision of primary data collection.

Further to these conclusions it is argued here that while it is possible to carry out quantitative evidence synthesis of health care resource use data, there are some real concerns that limit its validity, over and above the issues mentioned in the

introductory part of section 2.3. The first one is a technical issue. Since resource use and costs are non-normally distributed, their statistical synthesis is particularly challenging from the analytical point of view, a problem that is exacerbated when the information is only available at aggregate study level. Second, there is typically a large methodological heterogeneity that affects costing studies, which is often impossible to characterise statistically. Some of the study-level features that may be responsible for it are differences in data collection strategies, methods for measuring resource use and costs, follow up duration, methods of analysis and reporting. A third issue relates to time. Technological innovation, relative price changes and many other factors that may affect resource use and costs are difficult to capture in a synthesis of secondary data.

2.3.3 Health-state utility data

Quality-adjusted life years are used extensively as a measure of health benefits in CEA for policy decisions (Kind et al., 2009). Their advantage stems from the fact that QALYs combine morbidity and mortality into a single *numeraire*. Morbidity, for the purpose of QALY calculation, is measured in terms of its impact on a preference-based generic measure of health related quality of life. Instruments that can be used to estimate preference-based generic health-related quality of life include, the EuroQol five-dimensional (or EQ-5D) (Brooks, 1996), the Health Utility Index (HUI) (Feeny et al., 2004), the Quality of Well-Being (QWB) (Kaplan et al., 1998) and the Short Form six-dimentional (SF-6D) (Brazier et al., 2002). Several country-specific preference weights exist for the EQ-5D and SF6D. The abundance of alternative instruments means that researchers and policy makers are often unclear as to which of these should be used and accepted in a given country or jurisdiction. Differences in the descriptive systems used by each of these instruments generate a comparability problem (Wee et al., 2007, Conner-Spady & Suarez-Almazor, 2003), which is compounded when the analyst intends to synthesise the available evidence to produce one parameter estimate of utility for use in the model. In addition,

disagreement in the literature as to whose preferences should be used to value health states (Gandjour, 2010) makes the synthesis even more complex.

Despite this ongoing debate, the EQ-5D has become the most widely used preference-based generic measure of health-related quality of life in recent years. In the UK, the third edition of the NICE methods guidance for technology assessment (NICE, 2008) recommend the use of the EQ 5D for the reference case analyses. Since then, there has been an increasing interest in this instrument.

Publicly available repositories of health state utility values for a variety of health conditions are potentially a very useful data source. Tengs and Wallace (2000) were the first to publish a national repository of one thousand utility values gathered from 154 published reports. More recently, Sullivan and Ghushchyan (2006) and Sullivan *et al.* (2009) used Medical Expenditure Panel Survey data to develop a prediction tool for preference-based EQ-5D index scores for chronic conditions in the USA and the UK, respectively, based on responders ICD-9 codes. These repositories are particularly useful in the absence of (primary or secondary) preference-based generic health-related quality of life data. When these are available and depending on the data's format and number of parameters to inform, one of the scenarios described in Table 2.1 is considered.

To date, little work has been undertaken on the methods for statistical synthesis of preference-based health related quality of life data. This may be due to the fact that its synthesis is not stated as a requirement by national bodies such as the NICE (NICE, 2008). Recently, the NICE DSU released a technical support document giving guidance on the identification, review and synthesis of health state utility values (Papaioannou et al., 2011). Among a series of recommendations, this document emphasises the importance of selecting a main set of relevant utility values, or, in the presence of multiple relevant values, the pooling of these is suggested in order to improve precision of both mean and variance estimates. Meta-regression is one of the synthesis methods proposed to account for variability and to provide support to the choice of values used.

Perhaps one of the first manuscripts to apply evidence synthesis methods to health related quality of life data was published by Kinney *et al.* (1996), who conducted a meta-analysis of 84 studies reporting summary quality of life data in a cardiac patient population (taxonomy section A2). Tengs and Lin (2003, 2002) published meta-regressions in two different clinical areas (i.e. HIV/AIDS and cardiovascular) with the objective of estimating utility values associated with specific health states while controlling for specific study- and instrument- related features (taxonomy section A2). Using the same methodology Sturza (2010) recently published a meta-regression of utility values in lung cancer, finding a great deal of heterogeneity in the data even after applying strict inclusion criteria, and concluded that analysts should avoid direct comparisons of lung cancer utility values elicited with dissimilar methods (taxonomy section A2). Donnan *et al.* (2009) and Cheng and Niparko (1999) found similar problems with respect to combining utility values from a variety of assessment methods. For Cheng and Niparko (1999) it was found to be problematic to do so and *"…to some extent, this heterogeneity limits the meaningfulness of statistical pooling…"* (page 1217). Other examples of quantitative evidence synthesis of utility estimates can be found in the literature (Peasgood et al., 2010, Peasgood et al., 2009, McLernon et al., 2008, Bremner et al., 2007, Post et al., 2001, Dijkers, 1997) (all examples lay within categories A2 and B2 of the taxonomy).

These findings suggest that quantitative synthesis of aggregate preference-based values is limited by: *(i)* the between-study heterogeneity in the instruments used; *(ii)* the value set used to quantify utilities; *(iii)* the models used to approximate scores for health states; and over and above *(iv)* the typical issues related to standard meta-regression of summary binary outcome data. It has therefore been argued that, particularly in this context, the use of IPD would be essential (Ara & Brazier, 2010).

Further work is required in this area, both with regard to methods of quantitative synthesis of heterogeneous preference-based outcomes when these are available at

aggregate study level, and the need to control for between-study heterogeneity induced by the use of different preference-based instruments.

2.4 Discussion

The information required to carry out economic evaluation studies for policy decisions often comes from several different data sources, which often provide multiple estimates of the parameter(s) of interest. Statistical evidence synthesis techniques and decision analytic models represent an ideal vehicle to structure the decision problem, combine all available data and characterise the various sources of uncertainty associated with the decision problem. Using the synthesis of clinical effectiveness data as a conceptual framework, a taxonomy of possible scenarios that the analyst may face was developed (and appropriate methodologies to use discussed) based on a combination of three factors: *(a)* the number of data sources; *(b)* their format(s); and, *(c)* whether the analyst wishes to derive single or multiple parameters from the synthesis. Recommendations concerning appropriate methods to use under different scenarios were provided throughout. This chapter also reviewed the way in which evidence has been used to inform decision model parameters related to the disease natural history, costs and utilities. Areas where further methodological research may be needed are also identified.

The proposed taxonomy is designed to be used by health economics modellers as an instrument to support the development of their analysis plan, help them to fulfil methodological requirements and adequately address the research question at hand. The three dimensions on which the taxonomy is based provide a simple method of characterising and categorising the evidence base available (i.e. in terms of its quantity and format) linking this to the (type and number of) decision model parameter(s) to be derived. Following this 'checklist', the analyst can easily identify within the relevant taxonomy cell (or cells) methods that are available and those that are recommended. This list of approaches and methods has been (wherever

possible) supported by references to key methods literature and case studies where these have been put into practice. The references can then be consulted by analysts in search of further methodological and/or practical details of the subject. In this sense, the taxonomy helps to ensure consistency and completeness when carrying out the task of using evidence to inform decision models, standardising approaches and the adequate use of methods to analyse/synthesise evidence. Additionally, it provides a useful reference on the more recent methodological developments in the context of evidence synthesis for health care CEA.

The current taxonomy foundations are not, however, without limitations. First, the evidence synthesis methodologies and applied studies described throughout the chapter are not the result of a comprehensive systematic review (i.e. not exhaustive). It is believed, however, that these are representative of the methodologies found in the methods and applied literature in this area of research. Second, despite the efforts to make this taxonomy easy to generalise, the three dimensions (number and format of data sources and number of parameters to inform) may still not capture all possible scenarios. For instance, the taxonomy could be extended to include extra dimensions – e.g. extrapolation of model estimates – or detailed to cover other aspects – e.g. role of covariates within each taxonomy section. It was felt, however, that such an extension would unnecessarily increase its complexity without adding substantial benefits. Finally, the taxonomy is applied to clinical effectiveness but not to other key economic model parameters (i.e. disease natural history, resource use / costs, and preferences). Nonetheless, issues relating to the application of the taxonomy to these other parameters are discussed and its fuller application encouraged in future research. Methodological and applied literature is scarce regarding the quantitative synthesis of evidence to inform these – it is believed that further research is required despite recent relevant contributions in this area (Dias et al., 2011c, Papaioannou et al., 2011). Moreover, the specific characteristics of these parameters and of the evidence used to inform them may pose further challenges. Some of these evidence characteristics are highlighted and discussed next.

In practice, studies included in a certain synthesis may vary in their degree of rigour and possibly in their relevance towards the research question. This is particularly important when 'relevant' available evidence comes from observational data. Flaws in the design or conduct of a study can result in bias, and in some cases this can have as much influence on observed effects as that of treatments. Important intervention effects, or lack thereof, can be obscured by bias. Assessment of study quality gives an indication of the strength of evidence provided by the pooled result and, ultimately, quality assessment helps to answer the question of whether included studies are sufficiently robust to guide treatment, prevention, diagnostic or policy decisions. Most of the bias adjustment proposals published so far are reweighting schemes, usually attributing lower weight to evidence with a high risk of bias. More information on this topic can be found in Spiegelhalter *et al.* (2003), Turner *et al.* (2009) and Welton *et al.* (2009).

Common to all evidence identified, to potentially inform decision model parameters, is the case of partial reporting of information. If, for instance, mean differences without a measure of variance are reported, difficulties may arise when attempting to parameterize data for probabilistic modelling and strong assumptions may have to be imposed. A variety of methods for imputing variances have been proposed – see Abrams *et al.* (2005), Wiebe *et al.* (2006) and Furukawa *et al.* (2006) for further details. Another example occurs when different studies report different (multiple) outcome measures, at different time points and possibly on different scales. All these issues raise important obstacles for the synthesis of evidence.

A number of authors have recently published papers relating to the synthesis of cost effectiveness model outputs (Anderson, 2010, Sculpher, M. J. & Drummond, 2006, Pignone et al., 2005, Sculpher, M. J. et al., 2004, Welte et al., 2004, Birch & Gafni, 2003, Nixon, J. et al., 2001). One of the key questions here is whether or not it is appropriate to do so. I argue that there is no apparent rationale for assuming that the costs of a particular health care intervention (or their health-utilities) estimated in different studies carried out in different countries and health care settings, probably

using different measurement and assessment instruments, should converge towards a common value to be estimated using evidence synthesis techniques (Pignone et al., 2005).

In conclusion, this chapter brings recent developments in quantitative evidence synthesis to the attention of the health economics modelling community, encouraging a broader and more explicit consideration of these methods in the future. Several of the techniques presented here fall in the spheres of epidemiology, statistics and operational research, which in some cases are not directly accessible (due to lack of exposure and increased complexity of methods) to health economics modellers.

The taxonomy should be viewed by readers/analysts as a supplement to the guidelines on methods for technology assessment published by the NICE (NICE, 2008), and increase the users' confidence surrounding the validity of decision model inputs and subsequent outputs.

CHAPTER 3

3. SYNTHESIZING EVIDENCE USING AGGREGATE- AND INDIVIDUAL-PARTICIPANT LEVEL DATA

3.1 Introduction

As a consequence of the move towards evidence-based health care, with its underlying principle that evidence synthesis must be seen as the key to more coherent and efficient research (Sutton, Alexander J. et al., 2009), it is necessary to systematically identify and consider evidence from all the relevant studies (Higgins & Green, 2008). The lack of clear guidelines on what data can be used and how to effectively synthesise it led to the development of a taxonomy in Chapter 2. This taxonomy is based on possible scenarios faced by the analyst when dealing with all relevant available evidence. It aims to help analysts to identify the most appropriate method(s) to use when synthesizing the available data for a given model parameter, working towards the standardisation of approaches. Moreover, Chapter 2 pushes forward the methodological agenda in the synthesis field by highlighting several existing gaps in the methods literature. The current chapter aims to fill in some of these gaps.

It has been noted that in many cases the synthesis of evidence is conducted using pair-wise meta-analysis (Sutton, A. J. et al., 2000a, Hasselblad & McCrory, 1995). This method is described extensively in the literature (Sutton, A. J. et al., 2000a). It was also acknowledged that meta-regression techniques can be used to explore any apparent between-study heterogeneity, with the aim of estimating *treatment x covariate* interactions (Simmonds & Higgins, 2007, Berlin et al., 2002). When these methods rely only on patient-level covariates at aggregate level, they have been shown to have low statistical power (Lambert et al., 2002) and to be highly susceptible to ecological fallacy biases (Berlin et al., 2002). This thesis also acknowledges that another way of obtaining combined statistics is to perform meta-analyses over IPD (Stewart & Clarke, 1995). Individual participant data may, however, be available for only a (small) proportion of all relevant studies. Recent extensions to this modelling framework have been developed, which allow IPD and AD to be used jointly to estimate the effects of a treatment (Sutton, A. J. et al., 2008, Kendrick et al., 2007).

In most medical conditions, however, multiple interventions are available, and clinicians and policy makers need to decide on the optimal strategy among all relevant alternatives. It was discussed in Chapter 2 that MTCs (Lumley, 2002, Tudur et al., 2001) may be used to combine evidence on multiple alternative interventions, informing treatment comparisons that may not have been trialled head-to-head and without breaking randomization. Nonetheless, several assumptions are imposed by the MTC approach, such as: *(i)* such methods can only be applied to connected networks of studies; *(ii)* the treatment effects are thought to be generalisable across patients from trials included in the network; and *(iii)* in the presence of evidence loops, consistency across the evidence base must exist (Dias et al., 2011b, Dias et al., 2010b, Cooper, Nicola J. et al., 2009, Lu, G. & Ades, 2009, Song et al., 2009). As with any meta-analysis, in the case of MTCs it is desirable to account for heterogeneity/inconsistency, otherwise results may be biased (Cooper, Nicola J. et al., 2009). Also, the identification of factors contributing to these two factors may be valuable clinically, as the optimal treatment strategy may vary across different

patient groups. Thus, if treatment recommendations are made for subgroups of patients, this will lead to efficiency gains when compared to the suboptimal framework of decisions based on overall (mean) effectiveness. Despite the advantages of IPD for exploring heterogeneity/inconsistency, IPD has rarely been used in the context of MTCs.

This chapter considers synthesis models for binary outcomes where AD and IPD are available (although direct simplifications of the models allow the analysis of just AD or just IPD), and where patient and/or study level information on covariates may be of interest. Models are fitted to a motivating dataset on uptake of smoke alarms to prevent accidents in pre-school children.

After describing the motivating example dataset in section 3.2, existing and novel models for the synthesis of AD, IPD, and AD and IPD simultaneously are outlined in sections 3.3 and 3.4, considering direct and indirect comparisons, respectively. In both sections, models which estimate mean intervention effects, ignoring the influence of covariates, are considered first. These are followed by models which incorporate AD and IPD data, allowing for both individual and aggregate study level covariate information to be included, enabling discussions over estimates of *treatment x covariate* interactions and their assumptions. Results of applying the described methods to the motivating datasets are subsequently discussed. These sections are followed by a modelling extension to consider all available evidence in section 3.5 and some discussion topics and concluding remarks in section 3.6.

3.2 Motivating example: The effectiveness of home safety education and the provision of safety equipment for the prevention of accidents in pre-school children and the impact of socioeconomic characteristics

The motivating example comes from an evaluation of the effectiveness of home safety education and the provision of functioning smoke alarm safety devices for the prevention of accidents in pre-school children. Studies were identified as part of a more general and updated Cochrane systematic review of safety equipment (Wynn et al., 2010, Kendrick et al., 2007) – see Chapter 1 for further detail.

As a case study for this chapter, the outcome measure assessed is the provision of functioning smoke alarms (binary – Yes/No) given different interventions designed to increase their prevalence in households with children. The following relevant evidence is used for the smoke alarm outcome: 9 studies available in IPD format (Bulzacchelli et al., 2009, Phelan et al., 2009, Gielen et al., 2007, Hendrickson, 2005, Watson et al., 2005, Sznajder et al., 2003, DiGuiseppi et al., 2002, Kendrick et al., 1999, Clamp & Kendrick, 1998) and 11 available in AD format (Sangvai et al., 2007, Harvey et al., 2004, Mock et al., 2003, Gielen et al., 2002, Gielen et al., 2001, King et al., 2001, Johnston et al., 2000, Schwarz, D. F. et al., 1993, Barone, 1988, Matthews, 1988, Miller et al., 1982), summing up to approximately 11,500 participants. Seven out of 9 available studies in IPD format and 7 out of 11 in AD format are RCTs. Also, 2 of the IPD studies and 6 of the AD studies are cluster-allocated trials[18], but in none of these had the cluster design been accounted for in the original analysis. In all subsequent analyses, the effect of the clustering is modelled for the IPD. For the summary study estimates, an approximate adjustment is made through the inflation of treatment effect variances prior to all modelling described below (Note, Sutton *et al.* (2008) made

[18] The allocation level for clustered RCTs varied across studies and ranged from paediatricians to general practices and electoral wards. The allocation level for the clustered non-RCTs included one study allocated at the GP practice level and one where allocation was based on time periods. The allocation level for the clustered CBA was the child health clinic.

cluster adjustments to the AD within the modelling, this has the advantage of allowing for the uncertainty in the estimation of the Intra Class Correlation Coefficient (ICC) and could be pursued within a MTC framework if desired. See also Kendrick *et al.* (2007) for further details). Information for these studies is provided in Table 3.1. Note that values are here presented in summary form, but where IPD is available it is used as such in the analysis – exceptions are made for the models where AD is synthesised by reducing available IPD to AD.

Seven implementation strategies are defined across the available evidence base, namely:

1) usual care (UC);
2) education (E);
3) education plus low cost/free safety equipment (E + FE);
4) education plus low cost/free safety equipment plus home inspection (E + FE + HI);
5) education plus low cost/free safety equipment plus fitting (E + FE + F);
6) education plus home inspection (E + HI);
7) education plus low cost/free safety equipment plus fitting plus home inspection (E + FE + F + HI).

The exploration of (binary) participant-level socioeconomic characteristics is of interest for this chapter. Thus, in the last column of Table 3.1 information on the binary covariate relating to the number of parents in the family (i.e. single *vs.* two parents households) can be found.

The following section will focus on obtaining a combined statistic of the intervention effect in a pair-wise framework.

Table 3.1 – Available evidence on interventions seeking to increase the ownership of functioning smoke alarm safety equipment to prevent fire injuries in children.

Study lead author, year	Study design [2]	Allocation type [3]	Data available [4]	Strategies [1] [number of participants with functioning smoke alarms / total number of participants (numbers adjusting for clustering in parentheses)]							
				(1) UC	(2) E	(3) E + FE	(4) E + FE + HI	(5) E + FE + F	(6) E + HI	(7) E + FE + F + HI	Single-parent household (by treatment arm) - %
Matthews, 1988	NRT	IA	AD				6/12		6/12		NA
Schwarz *et al.*, 1993	CBA	IA	AD	816/1060						866/902	NA
King *et al.*, 2001	RCT	IA	AD				394/469		406/482		NA
Sangvai *et al.*, 2007	RCT	IA	AD	5/10			16/17				NA
Gielen *et al.*, 2001	RCT	CA	AD	54/56 (52.02/53.95)	77/80 (74.18/77.07)						0.848 (0.830; 0.860)
Miller *et al.*, 1982	NRT	CA	AD	46/105 (9.34/21.31)		61/108 (12.38/21.92)					NA
Barone, 1988	RCT	CA	AD		34/38 (20.08/22.45)	39/41 (23.04/24.22)					NA
Johnston *et al.*, 2000	RCT	CA	AD				211/211 (20.05/21.15)		136/143 (31.07/31.14)		0.573 (0.486; 0.638)
Gielen *et al.*, 2002	RCT	CA	AD			47/56 (44.2/52.66)	47/58 (44.2/54.54)				NA
Mock *et al.*, 2003	CBA	CA	AD	10/297 (2.33/69.18)	18/308 (3.03/71.74)						NA
Harvey *et al.*, 2004	RCT	CA	AD				997/1545 (781.6/1211.2)			1421/1583 (1114.0/1241.0)	NA
Clamp *et al.*, 1998	RCT	IA	IPD	71/82		81/83					0.103 (0.122; 0.084)
Hendrickson, 2002	RCT	IA	IPD	26/40			37/38				0.244 (0.132; 0.350)
Sznajder *et al.*, 2003	RCT	IA	IPD		6/50			27/47			0.135 (0.120; 0.152)
Watson *et al.*, 2005	RCT	IA	IPD	619/737				692/764			0.264 (0.255; 0.273)
Gielen *et al.*, 2007	RCT	IA	IPD	325/375	345/384						0.695 (0.720; 0.669)
Bulzacchelli *et al.*, 2009	NRT	IA	IPD	55/71	109/139						0.719 (0.761; 0.698)
Phelan *et al.*, 2010	RCT	IA	IPD	112/138						130/140	NA
Kendrick *et al.*, 1999	NRT	CA	IPD	305/339 (233.4/259.4)			341/385 (260.9/294.6)				0.108 (0.104; 0.112)
DiGuiseppi *et al.*, 2002	RCT	CA	IPD	5/30 (5/30)				8/44 (8/44)			NA

Notes:

1) UC - usual care; E - education; E + FE - education plus low cost / free equipment; E + FE + HI - education plus low cost / free equipment plus home inspection; E + FE + F - education plus low cost / free equipment plus fitting; E + HI - education plus home inspection; E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

2) NRT - non-randomised trial; CBA - controlled before and after trial; RCT - randomised controlled trial.

3) IA - participants are individually allocated; CA - participants are cluster allocated.

4) AD - aggregate data; IPD - individually participant data.

3.3 Methods for estimating a pooled treatment effect – direct comparisons

This section is divided in two subsections. Initially, a description of existing RE meta-analytic models for binary outcomes is given for the synthesis of AD, of IPD only and for the simultaneous synthesis of both AD and IPD. Extensions to the inclusion of information on covariates are also discussed. In a second part, results of the application to the motivating example are shown and thoroughly discussed. It is important to note that distinctions between control and intervention groups, as outlined in Table 3.1, are ignored in this section with all trials combined in a single 'pair-wise' comparison.

3.3.1 Synthesizing aggregate data only

A REs meta-analysis assumes that the true effects in each two-arm trial are not equal, but are random observations drawn from a common distribution. The Bayesian REs model for binary outcomes can then be written as (Smith, T. C. et al., 1995):

$$\begin{aligned} r_{Aj} &\sim Bin\left(p_{Aj}, n_{Aj}\right) & r_{Bj} &\sim Bin\left(p_{Bj}, n_{Bj}\right) \\ \text{logit}\left(p_{Aj}\right) &= \mu_j & \text{logit}\left(p_{Bj}\right) &= \mu_j + \delta_j \\ \mu_j &\sim N\left(0,10^6\right) \quad \delta_j \sim N\left(d,\tau^2\right) \quad d \sim N\left(0,10^6\right) \quad \tau \sim Unif\left(0,10\right) \end{aligned} \tag{3.1}$$

Where n_{Aj} and n_{Bj} denote the total number of individuals in the two arms A and B of the j^{th} study, r_{Aj} and r_{Bj} denote the number of events in these two arms, with the underlying probabilities of an event represented by p_{Aj} and p_{Bj}, respectively. The estimated unconstrained log-odds of an event in group A in the j^{th} trial is denoted by μ_j, and requires a prior distribution. The true treatment effect (on a log-odds scale)

in trial *j* is represented by δ_j, allowed to be different from each other and assumed to be sampled from a Normal distribution with mean *d* and variance τ^2. Prior distributions are required for these parameters. The parameter τ^2 indicates how much variability there is between estimates from the different studies. For the parameters requiring a prior distribution, these were specified above (model (3.1)), and are intended to be vague.

If one wanted to describe a FEs model, the above could be modified so that the δ_j's are fixed across studies and a (vague) prior distribution assigned to this overall treatment effect.

3.3.2 Synthesizing individual participant-level data only

The Bayesian REs model for binary outcomes using IPD data, while controlling for participant allocation, can be written as (notice that the model structure is similar to that described by Sutton *et al.* (2008) and Turner *et al.* (2000)):

$$\begin{gathered} Y_{ij} \sim Bernoulli\left(p_{ij}\right) \\ \text{logit}\left(p_{ij}\right) = \mu_j + \delta_j treat_{ij} \\ \mu_j \sim N\left(0,10^6\right) \qquad \delta_j \sim N\left(d,\tau^2\right) \qquad d \sim N\left(0,10^6\right) \qquad \tau \sim Unif\left(0,10\right) \end{gathered} \tag{3.2}$$

The binary response of the i^{th} participant in the j^{th} study, Y_{ij} (i.e. 1 = event, 0 = no event), is assumed Bernoulli distributed, with probability of the event of interest described by p_{ij}. A standard logistic regression is fitted to each of the *j* trials and the linear predictor considers an independent term, μ_j, which estimates the log-odds of an event in the control group, and a term for the treatment difference, δ_j, multiplied by a treatment group indicator, $treat_{ij}$ (i.e. *0* if in control group, and *1* if in

intervention group). As in (3.1), all estimated parameters require specification of prior distributions which are intended to be vague.

3.3.3 Synthesizing individual and aggregate level data

One practical limitation of carrying out an IPD only meta-analysis, is that it relies on the availability of data sets for all studies. Therefore, one may be faced with the difficulty of having to statistically synthesise evidence in two different formats (i.e. IPD and AD). The model described here follows the approach used by Sutton *et al.* (2008) which, in addition to IPD and AD, allows also for the additional level of complexity introduced by different randomisation procedures (i.e. cluster- and individual-allocation). This model is described in five interrelated parts and can be viewed as an integration of the previous two models, (3.1) and (3.2), with all notation conventions remaining the same. The Bayesian REs model for the pair-wise combination of IPD and AD for a binary outcome can be written as**:**

(3.3) *Part I - Model for individually allocated IPD studies*

$$\begin{aligned} Y_{ij} &\sim Bernoulli\left(p_{ij}\right) \\ \text{logit}\left(p_{ij}\right) &= \mu_j^{IPD} + \delta_j treat_{ij} \\ \mu_j^{IPD} &\sim N\left(0,10^6\right) \end{aligned} \tag{3.3.1}$$

For: i = 1, 2, ... , number of participants in the j^{th} individually allocated IPD study; and j = 1, 2, . . . , number of individually allocated IPD studies.

Part II - Model for cluster allocated IPD studies

$$\begin{aligned} Y_{imj} &\sim Bernoulli\left(p_{imj}\right) \\ \text{logit}\left(p_{imj}\right) &= \mu_{mj}^{c.IPD} + \delta_j treat_{imj} \end{aligned} \tag{3.3.2}$$

$$\mu_{mj}^{c.IPD} \sim N\left(\theta_j, \tau.c_j^2\right) \qquad \theta_j \sim N\left(0,10^6\right) \qquad \tau.c_j \sim N\left(0,10^6\right)$$

For: i = 1, 2, ... , number of participants in the m[th] cluster of the j[th] cluster allocated IPD study; m = 1, 2, . . . , number of clusters in the j[th] study; and j = (number of individually allocated IPD studies + 1),…,(number of individually allocated IPD studies + number of cluster allocated IPD studies)

While equation (3.3.1) is used for individually allocated IPD studies, equation (3.3.2) allows for clustering effects within studies. A separate unconstrained control group odds for each cluster is estimated on the logit scale within each study, $\mu_{mj}^{c.IPD}$, assuming that these are exchangeable[19] within each study. Cluster effects between studies are assumed independent.

Part III - Model for individually allocated AD studies

$$r_{Aj} \sim Bin\left(p_{Aj}, n_{Aj}\right) \qquad r_{Bj} \sim Bin\left(p_{Bj}, n_{Bj}\right)$$

$$\text{logit}\left(p_{Aj}\right) = \mu_j^{AD} \qquad \text{logit}\left(p_{Bj}\right) = \mu_j^{AD} + \delta_j \qquad (3.3.3)$$

$$\mu_j^{AD} \sim N\left(0,10^6\right)$$

For: j = (number of individually allocated IPD studies + number of cluster allocated IPD studies + 1), . . . , (number of individually allocated IPD studies + number of cluster-allocated IPD studies + number of individually allocated AD studies)

Part IV - Model for cluster allocated AD studies

$$design.effect_j = 1 + (ave.c.size_j - 1) \cdot ICC_j \qquad (3.3.4)$$

[19] As described in Higgins *et al.* (2009): "*exchangeability represents a judgement that the treatment effects may be non-identical but their magnitudes cannot be differentiated a priori*", that is, "*exchangeability describes the a priori position of expecting underlying effects to be similar, yet non-identical*" and it "*reflects a degree of prior ignorance in that the magnitudes of the effects cannot be differentiated*".

$$\sigma.adj^2_j = \sigma^2_j \cdot design.effect_j$$

$$T_j \sim N\left(\delta_j, \sigma.\text{adj}^2_j\right) \qquad ICC_j \sim \textit{as described in the text}$$

For: j = (number of individually allocated IPD studies + number of cluster-allocated IPD studies + number of individually allocated AD studies + 1), . . . , (total number of studies)

Equation (3.3.4) combines cluster-allocated studies in AD format over the assumption that the original analysis ignored the clustering effect. If this effect is ignored, cluster allocated studies will benefit from being allocated more weight than they should. Therefore, adjustments were made to inflate the treatment effect variances to take this fact into account. These were inflated based on the design effect, function of the average cluster size in the j^{th} study and the ICC in the j^{th} study[20] (Donner & Klar, 2002). Further details on how ICCs are calculated can be found in Kendrick *et al.* (2007) and Sutton *et al.* (2008).

Part V: Combining estimates of the effect of the intervention from the 4 data sources

$$\delta_j \sim N\left(d, \tau^2\right) \qquad d \sim N\left(0, 10^6\right) \qquad \tau \sim Unif\left(0, 10\right) \qquad (3.3.5)$$

For: j = 1 . . . , total number of studies

Equation (3.3.5) specifies a RE to be placed across all treatment effect estimates from the IPD and AD, δ_j's, imposing the exchangeability property. In this way

[20] It is unlikely that the ICCs will be reported in published studies. For the studies which do not provide this information it is necessary to estimate them from other sources. Estimates of ICC may be derived, for instance, from the IPD used in part 3.3.2 of the model or based on external evidence. If distributions are specified for the unknown ICCs then these should represent the uncertainty in its value.

synthesis across both types of data is achieved, since equations (3.3.1) and (3.3.2) "share" parameters. As before, specified prior distributions are intended to be vague. It is noted that for an analysis such as this, which does not consider patient level covariates, no loss of information is to be expected when collapsing IPD to AD and thus the model presented here might seem over-elaborate in this specific circumstance, though this is not the case if covariates are considered, as is outlined in the next section.

3.3.4 Meta-regression models: the inclusion of covariates

Random-effects models take into consideration the possible heterogeneity between studies, although they do not explain the reasons study results vary. While associations of treatment effect with patient characteristics may be explored using AD and average study level covariates, as outlined in this chapter's introduction, this approach is problematic since possible associations can occur purely by chance, or due to the presence of confounding factors, and are susceptible to low power and aggregation/ecological bias (Berlin et al., 2002). Use of IPD is recommended to explore patient characteristics, and avoids the bias introduced by group-level analyses.

Covariates can be included in a Bayesian meta-analysis model in a straightforward way. For example, in the previously described Bayesian REs meta-analytic model of AD for binary outcomes on a log- OR scale (section 3.3.1), the model can be simply extended to include a covariate, *X*, measured at trial level as means in the two arms:

$$\begin{aligned} r_{Aj} &\sim Bin\left(p_{Aj}, n_{Aj}\right) & r_{Bj} &\sim Bin\left(p_{Bj}, n_{Bj}\right) \\ \text{logit}\left(p_{Aj}\right) &= \mu_j & \text{logit}\left(p_{Bj}\right) &= \mu_j + \delta_j + \beta \cdot X_j \\ \mu_j &\sim N\left(0,10^6\right) \quad \delta_j \sim N\left(d,\tau^2\right) \quad d \sim N\left(0,10^6\right) \end{aligned} \tag{3.4}$$

$$\tau \sim Unif(0,10) \quad \beta \sim N(0,10^6)$$

The estimated odds (on the log-scale), denoted by μ_j, can be interpreted as the model intercept (i.e. the effect where the covariate takes the value *0*), and β, an unconstrained model coefficient which estimates the effect of the covariate on the treatment effect, and requires a prior distribution – model (3.4). Extension to multiple regression coefficients is straightforward. The IPD model without covariates (3.2) can be extended in the same manner, although it is important to note that in such an analysis the covariate data is at the patient, not the study, level (not shown but demonstrated below).

Riley and Steyerberg (Riley, Richard D. & Steyerberg, 2010) have shown that IPD allows for the modelling of *treatment x covariate* interactions using both within and between-study variability, and such variability can be partitioned to produce a meta-regression and a "pure" IPD estimate of the interaction of interest. These estimates can subsequently be merged into an overall interaction estimate if deemed appropriate. The following model is described in (interrelated) parts and extends model (3.3) for the combination of IPD and AD. It is an adaptation of the model described in Sutton *et al.* (2008), now including binary covariates and considering the partitioning of the variability regarding the interactions (extending the framework used by Riley (2010)). Alternative assumptions could be accommodated with relatively straightforward modifications to the model specification.

(3.5) *Part I – Model for individually allocated IPD studies including covariates*

$$Y_{ij} \sim Bernoulli(p_{ij}) \tag{3.5.1}$$

$$\text{logit}\left(p_{ij}\right) = \mu_j^{IPD} + \delta_j \cdot treat_{ij} + \beta_{0j} \cdot x_{ij} + \beta^B \cdot \bar{x}_j \cdot treat_{ij} + \beta^W \cdot \left(x_{ij} - \bar{x}_j\right) \cdot treat_{ij}$$

$$\mu_j^{IPD} \sim N\left(0,10^6\right) \qquad \beta_{0j} \sim N\left(0,10^6\right)$$

For: i = 1, 2, ... , number of participants in the individually allocated j^{th} IPD study; and j = 1, 2, . . . , number of individually allocated IPD studies

Part II - Model for cluster allocated IPD studies including covariates

$$Y_{imj} \sim Bernoulli\left(p_{imj}\right)$$

$$\text{logit}\left(p_{imj}\right) = \mu_{mj}^{c.IPD} + \delta_j \cdot treat_{imj} + \beta_{0j} \cdot x_{imj} + \beta^B \cdot \bar{x}_j \cdot treat_{imj} + \beta^W \cdot \left(x_{imj} - \bar{x}_j\right) \cdot treat_{imj} \qquad (3.5.2)$$

$$\mu_{mj}^{c.IPD} \sim N\left(\theta_j, \tau.c_j^2\right) \qquad \theta_j \sim N\left(0,10^6\right) \qquad \tau.c_j \sim Unif\left(0,10\right)$$

$$\beta_{0j} \sim N\left(0,10^6\right)$$

For: i = 1, 2, ... , number of participants in the m^{th} cluster of the j^{th} cluster allocated IPD study; m = 1, 2, . . . , number of clusters in the j^{th} study; and j = (number of individually allocated IPD studies + 1),…,(number of individually allocated IPD studies + number of cluster allocated IPD studies)

Compared to model (3.3), the changes applied are the addition of three extra terms: *(i)* a study-specific individual level covariate regression term, $\beta_{0j} \cdot x_{ij}$ (or $\beta_{0j} \cdot x_{imj}$ for cluster-allocated studies), where β_{0j} is the main covariate effect and x_{ij} (x_{imj}) refers to the value for the binary covariate in the i^{th} participant (in the m^{th} cluster) of the j^{th} study; *(ii)* an interaction term to account for the within-study association, $\beta^W \cdot \left(x_{ij} - \bar{x}_j\right) \cdot treat_{ij}$ (or $\beta^W \cdot \left(x_{imj} - \bar{x}_j\right) \cdot treat_{imj}$ for cluster-allocated studies), where within-study relationship is modelled by centring x_{ij} (x_{imj}) about the mean covariate value, $\bar{x}_j$, in each study and β^W is assumed the same throughout all IPD

studies; and *(iii)* an interaction term to model the between-study relationship, $\beta^B \cdot \bar{x}_j \cdot treat_{ij}$ (or $\beta^B \cdot \bar{x}_j \cdot treat_{imj}$ for cluster-allocated studies), by interacting with the mean covariate value, where, like β^W, β^B is assumed the same throughout all IPD studies, but also equivalent to the estimated slope for the covariate in model equations (3.5.3) and (3.5.4) below, the β^B in $\beta^B \cdot x.agg_j$ (where $x.agg_j$ represents the proportion of participants with the characteristic of interest in the meta-regression of aggregate results).

Part III - Model for individually allocated AD studies including covariates

$$r_{Aj} \sim Bin\left(p_{Aj}, n_{Aj}\right) \qquad r_{Bj} \sim Bin\left(p_{Bj}, n_{Bj}\right)$$
$$\text{logit}\left(p_{Aj}\right) = \mu_j^{AD} \qquad \text{logit}\left(p_{Bj}\right) = \mu_j^{AD} + \delta_j + \beta^B \cdot x.agg_j \tag{3.5.3}$$
$$\mu_j^{AD} \sim N\left(0, 10^6\right)$$

For: j = (number of individually allocated IPD studies + number of cluster-allocated IPD studies + 1), . . . , (number of individually allocated IPD studies + number of cluster-allocated IPD studies + number of individually allocated AD studies)

Part IV - Model for cluster allocated AD studies including covariates

$$design.effect_j = 1 + (ave.c.size_j - 1) \cdot ICC_j$$
$$\sigma.adj^2_j = \sigma^2_j \cdot design.effect_j \qquad \delta^*_j = \delta_j + \beta \cdot x.agg_j \tag{3.5.4}$$
$$T_j \sim N\left(\delta^*_j, \sigma.\text{adj}^2_j\right) \qquad ICC_j \sim \textit{as described in the text}$$

For: j = (number of individually allocated IPD studies + number of cluster-allocated IPD studies + number of individually allocated AD studies + 1), . . . , (total number of studies)

Both individually- and cluster-allocated IPD studies are contributing to the estimation of the regression coefficient representing the within-study associations, β^W.Vague prior distributions are attributed to the participant-level regression parameters, β_{0j}. In equation (3.5.5) below the specification of vague priors for β^W and β^B is required.

V - Combining estimates of the effect of the intervention from the 4 data sources including covariates

$$\delta_j \sim N(d, \tau^2) \qquad d \sim N(0,10^6) \qquad \tau \sim Unif(0,10)$$
$$\beta^B \sim N(0,10^6) \qquad \beta^W \sim N(0,10^6) \tag{3.5.5}$$

For j = 1 . . . , total number of studies

In this model, it is necessary that the aggregate covariate (i.e. proportion of people in the study with the characteristic of interest) is expressed as a decimal. This way, estimated slopes from both the AD and the IPD have a comparable interpretation (Lambert et al., 2002). The regression coefficients from the interaction intervention – mean covariate value from the IPD (model equations (3.5.1) and (3.5.2)) – where the covariate indicates the presence or absence of a certain characteristic, and the slope from the AD meta-regressions, where the covariate value is the proportion of participants in a study with a certain characteristic, both estimate the change in outcome for a one unit increase in the proportion (i.e. from *0* to *1*, or the difference in effect of the intervention in the presence and absence of the covariate). Therefore, both the IPD and the AD, whether individually- or cluster-allocated, can contribute to the estimation of a single regression coefficient, β^B, representing the between-study associations.

3.3.5 Application

As for previous publications in this field, because of the flexibility of the modelling allowed, all models described below are fitted using Bayesian Markov chain Monte Carlo (MCMC) methods, as implemented in the software WinBUGS (Spiegelhalter, D. et al., 2003). Unless stated otherwise, for all models applied in this chapter, the MCMC sampler was run for 10,000 iterations, discarded as 'burn-in', and further 5,000 iterations were done on which inferences are based. Chain convergence was checked on all presented posterior sample summaries, including checking stability across distinct sets of initial values. All unknown parameters require prior distributions within a Bayesian paradigm, and are given prior distributions which are intended to be vague throughout. See Appendix 2 for more details on Bayesian methods.

For each subsequent set of results, medians of the MCMC posterior sample are presented and 95% credible intervals (CrI) of the posterior distribution as a Bayesian measure of the uncertainty of estimated parameter. The models considered in sections 3.3.1 to 3.3.4 have been fit sequentially, "building up" to the full complexity of model (3.5) to the outcome of interest, possession of a 'functioning smoke alarm'. Initially, it is compared the results of the pair-wise models without covariates for AD (all evidence base), AD only (55% of the evidence base) and IPD (reduced to AD – 45% of the evidence base). This is followed by the synthesis of IPD and of AD plus IPD, including the IPD where available, ignoring covariates. It is then considered the binary covariate on 'number of parent status' as a potential treatment effect modifier. This covariate is *binary* (0 = two parent household (2P); 1 = single parent household (1P)) for each individual in the IPD and a *percentage* (mean) when referring to the proportion of 1P in a study for the AD studies. Initially, two meta-regression models are specified, using summary evidence (i.e. the entire evidence base reduced to AD). Given the relatively small number of trials available, the model has been specified separating the variability regarding interaction effects. This is followed by a model with overall interaction coefficients. The latter models are slight modifications of

model (3.5) in which within- and between-study associations are merged into an overall parameter.

3.3.5.1 Analysis of models without covariates

Table 3.2 shows the results of meta-analytic models applied to the functioning smoke-alarm outcome without considering any covariates. The 3 columns of results relate to the application of model (3.1) to different sets of evidence.

Table 3.2 - Parameter estimates from fitting the pair-wise meta-analytic model to AD without covariates to the functioning smoke alarm outcome.

		Model (3.1) - 20 studies included		Model (3.1) - 11 studies included		Model (3.1) - 9 studies included	
		Random effects MA of AD		**Random effects MA of AD only**		**Random effects MA of IPD reduced to AD**	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter						
Log odds ratios for intervention effect	d	0.884	0.414 to 1.387	0.862	0.063 to 1.685	0.926	0.169 to 1.859
Between-study variance	τ^2	0.708	0.285 to 1.893	0.919	0.267 to 3.936	0.926	0.18 to 4.677
	Function of parameter						
Odds ratios for intervention effect	e^d	2.420	1.513 to 4.004	2.368	1.065 to 5.391	2.524	1.184 to 6.420
Deviance Information criteria	*DIC*	235.97		122.41		115.28	

Notes:

i) available IPD was reduced to AD, and this was combined with the existing AD.

The synthesis of all available evidence in AD (i.e. 20 studies) using a REs model, returns an intervention effect which favours the treatment (OR 2.4, 95% CrI 1.51 –

4.0). Between-study heterogeneity, $\hat{\tau}^2$, is taken into account and estimated as 0.7 (95% CrI 0.29 – 1.89), revealing a fair amount of variability across studies. Similar point estimates are obtained when splitting the study sample in the group for which evidence at aggregate level is available and the group for which evidence at the individual level (reduced to AD) is available. It is noted that for these models, uncertainty of estimated parameters is found to be higher, most certainly a consequence of the smaller number of studies included.

Table 3.3 refers to the case of meta-analyzing IPD studies only (i.e. 9 studies) and of synthesizing the entire evidence base whether in aggregate or in individual level format (i.e. 20 studies). If evidence at the individual level is the only format of evidence available, the statistical combination of these could be implemented by using model (3.2). These consider an extension to this model in order to consider the fact that 2 of the studies are cluster-allocated and 7 are individually-allocated (Table 3.3 column 1 of results). This model estimates almost equivalent results, compared to the ones obtained when applying model (3.1) to the same set of studies reduced to AD.

In Table 3.3 column 2, results of applying model (3.3) to the evidence base are shown. This model considers the simultaneous synthesis of AD and IPD. It is interesting to note the resemblance of estimated parameter posterior distributions of applying model (3.1) and model (3.3), to all studies (i.e. 20) – results columns 1 of Table 3.2 and results column 2 of Table 3.3. These similarities were expected and small differences can be observed. These differences can be justified by: *(i)* the use of different approaches in the cluster adjustment as adjustments were made prior to the modelling in the AD model and both outside (for AD) and within (for IPD) modelling in the IPD + AD model; *(ii)* the slight influence of the prior distributions; and by *(iii)* simulation error.

Table 3.3 - Parameter estimates from fitting the pair-wise meta-analytic models to IPD and to AD and IPD without covariates to the functioning smoke alarm outcome.

		Model (3.2)[i)] - 9 studies included		Model (3.3) - 20 studies included	
		Random effects MA of IPD only		**Random effects MA of AD and IPD**	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter				
Log odds ratios for intervention effect	d	0.920	0.084 to 1.867	0.850	0.311 to 1.408
Between-study variance	τ^2	1.070	0.231 to 4.948	0.823	0.3 to 2.458
	Function of parameter				
Odds ratios for intervention effect	e^d	2.510	1.088 to 6.472	2.339	1.364 to 4.087
Deviance Information criteria	*DIC*	2939.13		3099.71	

Notes:

i) Results shown relate to an extension to model (3.2) which takes into consideration the fact that in some studies participants were cluster-allocated and in some others individually-allocated.

3.3.5.2 Analysis of models considering a binary covariate

It is well documented that children suffering socio-economic disadvantages are at higher risk of unintentional injury at home than the more fortunate (Edwards et al., 2006). In fact, parents in lower income households and/or those in lower social grade households are less likely to have taken certain preventive actions (Kendrick et al., 2007, Reimers & Laflamme, 2005, Haynes et al., 2003). Literature is also available which has explored the relationship between maternal age and child injury risk (Reading et al., 1999, Scholer et al., 1999). The majority of these studies found that children are at greater risk of injury in families where the mother is younger or in single parent families (O'Connor et al., 2000). Information on the risk factor related to

the number of parents in the household (i.e. two or single parent families) is considered in the next section.

In nine of the 20 studies, information was available for the binary covariate related to whether the family was composed of a single parent or two parents and hence whether the uptake of smoke alarms was different for these two participating subgroups. Meta-regression models described consider *treatment x covariate* interaction terms which result in non-linear intervention effects. In addition to this, potential baseline covariate imbalances between study-arms and across studies may contribute to differences between the log-OR of the intervention effect estimates between models which consider and do not consider the covariate of interest.

Applying model (3.4) to the set of 9 studies, all at aggregate level, and considering the overall study covariate proportion, the OR for intervention effect in 2Ps is 2.5 (95% CrI 0.93 – 8.56) and 1Ps is 0.4 (95% CrI 0.01 – 19.59) – results reported in Table 3.4.

Though for both subgroups, CrI include 1, the results suggest that the intervention may be more effective for one subgroup than for the other. A *treatment x covariate* interaction (β) of -1.9 (95% CrI -5.94 – 1.73) indicates that 1Ps do not benefit from strategies aimed at increasing safety equipment ownership as those investigated in the current case study. The point estimate for the between-study variance parameter is reasonably large (approximately 1.4) and that the 95% CrI is quite wide (0.29 to 8.76) reflecting the uncertainty in its estimation.

If, instead of considering all evidence formats with information on the covariate, the focus is on the evidence in IPD format, reducing it to AD (i.e. 7 studies), and using the overall covariate proportion for each of the studies, the same conclusions can be obtained. Nevertheless, a slight increase in uncertainty is observed in all estimates, particularly in the treatment interaction term, β, and in the between-study variance.

Table 3.4 - Parameter estimates from fitting the pair-wise meta-regression model to AD with a binary covariate (single parent status) to the functioning smoke alarm outcome.

		Model (3.4) - 9 studies included		Model (3.4) - 7 studies included	
		Random effects MR of AD, including a binary covariate		**Random effects MR of IPD reduced to AD, including a binary covariate**	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter				
Log odds ratios for intervention effect	d	0.926	-0.069 to 2.147	1.049	-0.309 to 2.619
Regression coefficient for treatment interaction with single-parent status	β	-1.853	-5.941 to 1.725	-2.014	-7.783 to 3.864
Between-study variance	τ^2	1.433	0.287 to 8.76	1.925	0.354 to 15.591
	Function of parameter				
Odds ratios for intervention effect	e^d	2.525	0.933 to 8.563	2.854	0734 to 13.724
Ratio of the odds ratios for single-parent *vs* two-parent households	e^β	0.157	0.003 to 5.611	0.133	<0.001 to 47.6661
Odds ratio for intervention effect in single-parent households	$e^{d+\beta}$	0.395	0.007 to 19.59	0.383	0.001 to 162.74
Deviance Information criteria	*DIC*	104.844		90.334	

It is worth highlighting that estimated *treatment x covariate* associations through the use of model (3.4) for AD may not reflect the true relationship at the individual level – these associations may have occurred purely by chance, due to the presence of confounding factors, or ecological bias (Berlin et al., 2002). Therefore, all the meta-regression results described above have to take these factors into consideration.

Single parent status information was available in 7 out of the 9 studies in IPD format. The results of estimating treatment interactions terms, separating the between- and the within-study associations are shown in Table 3.5.

Table 3.5 - Parameter estimates from fitting pair-wise meta-regression models to IPD and to AD plus IPD with covariate 'single parent status' to the functioning smoke alarm outcome data considering exchangeable treatment interactions and modelling separately within and between study associations.

		Model (3.5) [i] - 7 studies included		Model (3.5) - 9 studies included	
		Random effects MA of IPD only		**Random effects MA of AD and IPD**	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter				
Regression coefficient for within study association with single-parent (*vs* two parent families)	β^{W}	0.035	-0.483 to 0.556	0.054	-0.508 to 0.585
Regression coefficient for between study association with single-parent (*vs* two parent families)	β^{B}	-1.957	-7.881 to 3.238	-1.918	-5.951 to 1.611
Difference between regression coefficient for within and between study associations with single-parent (*vs* two parent families)	β^{diff}	-1.985	-7.937 to 3.247	-1.999	-6.032 to 1.616
Deviance Information criteria	*DIC*	2436.89		2451.52	

Notes:

i) This model is a short version of model (3.5) which considers only the 2 initial parts of this model, i.e. the ones relating to the synthesis of IPD.

The first set of results refers to using IPD only and a shorter version of model (3.5). The second column of results refers to the direct use of model (3.5) with IPD included where possible (7 studies) in addition to AD (2 studies), and interactions are assumed to be exchangeable. To assess whether ecological bias exists, the difference between the association terms is estimated, β^{diff}'s. The CrIs of the difference between within- and between-study association estimates do include 0. Hence, there is little evidence to suggest systematic differences between the covariate treatment interactions estimated by the within-study and between-study variation. In the subsequent analysis presented, these two sets of coefficients are replaced with a single one combining the between and within study association information.

The meta-regression of IPD only and of AD + IPD, without separating within- and between-study association effects is shown in Table 3.6. In both synthesis models, results favour the intervention in the two-parent households (OR 2.7, 95% CrI 0.84 – 11.05 and OR 2.5, 95% CrI 0.87 – 8.27, respectively), nonetheless this effect is somewhat 'weak' as CrIs include 1. The *treatment x covariate* interaction estimate, $\hat{\beta}$, is approximately 0, suggesting an almost null impact. This indicates a similar intervention effect between both subgroups, as ORs for the IPD only model and for AD + IPD are approximately 2.7 (95% CrI 0.8 – 11.64) and 2.5 (95% CrI 0.88 – 8.67).

Table 3.6 - Parameter estimates from fitting the pair-wise model to IPD and to AD plus IPD with information on the covariate 'single parent status' to the binary functioning smoke alarm outcome data considering exchangeable and independent treatment interactions, not separating between- and within-study interactions.

		Model (3.5) [i) and ii)] - 7 studies included		Model (3.5) [ii)] - 9 studies included	
		Random effects MA of IPD only		**Random effects MA of AD and IPD**	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter				
Log odds ratios for intervention effect	d	0.992	-0.177 to 2.404	0.922	-0.136 to 2.112
Regression coefficient for treatment interaction with single-parent status	β	-0.001	-0.519 to 0.521	-0.018	-0.529 to 0.495
Between-study variance	τ^2	1.695	0.335 to 11.33	1.422	0.299 to 7.702
	Function of parameter				
Odds ratios for intervention effect	e^d	2.697	0.838 to 11.045	2.515	0.873 to 8.269
Ratio of the odds ratios for single-parent *vs* two-parent households	e^β	0.999	0.595 to 1.684	0.983	0.589 to 1.641
Odds ratio for intervention effect in single-parent households	$e^{d+\beta}$	2.696	0.799 to 11.64	2.473	0.884 to 8.666

Notes:

i) This model is a short version of model (3.5), considering only the 2 initial parts of this model relating to the synthesis of IPD.

ii) Models were simplified by aggregating in a single term both within and between study associations.

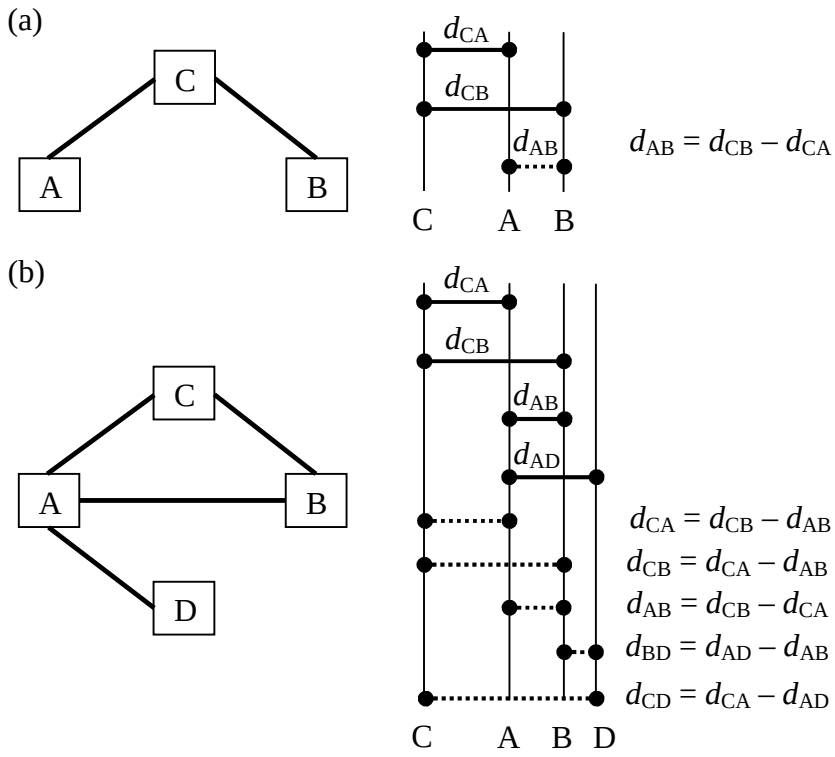
The following section considers the fact that control and intervention groups differ in different studies. This extends and restructures evidence in order to have not only pair-wise comparisons (A *vs.* B) but a whole set of comparisons among the relevant set of interventions using direct and indirect evidence.

3.4 Methods for estimating a pooled treatment effect – indirect and mixed treatment comparisons

This section reviews existing MTC methodology, which I extend in future sections. First, two simple hypothetical cases will be used to exemplify an indirect treatment comparison and a MTC. Figure 3.1a shows a network formed by pair-wise comparisons between the three interventions – A, B and C. Direct head-to-head evidence is available for comparisons of A *vs.* C and B *vs.* C. There is no direct evidence between A and B (no solid line connects these two treatments). Treatments A and B are *indirectly* linked through C, which is the common comparator in this diagram. An indirect comparison estimate of A and B can be derived using evidence from A *vs.* C and B *vs.* C trials. On the log-OR scale, an estimate of A *vs.* B comparison, d_{AB}, can be derived from existing evidence (Ades, A. E. et al., 2006b): $d_{AB} = d_{CB} - d_{CA}$.

Figure 3.1b represents the case when four interventions, A, B, C and D, are of interest. Not only is direct evidence available for the comparisons C *vs.* A and C *vs.* B, but also for A *vs.* B and A *vs.* D. Given the network of existing direct evidence, indirect estimates can be derived for C *vs.* A, C *vs.* B, A *vs.* B and B *vs.* D, using the same rationale as before. Except for some cases (i.e. B *vs.* D and C *vs.* D), Figure 3.1b shows that both types of evidence (i.e. direct and indirect) are available for most pair-wise comparisons. When analysing data in this way, as for standard meta-analysis, a decision needs to be made as to whether each trial is assumed to provide an estimate of exactly the same quantity (fixed effect) or if the studies included in the meta-analysis provide estimates that are realisations from a common distribution of possible, exchangeable (random effect) outcomes . Such an approach generalises to networks of any complexity; see elsewhere for further details (Caldwell et al., 2005, Lu, G. & Ades, 2004).

Figure 3.1 – Diagrams of evidence structures: (a) indirect comparison of intervention A and B given studies on the comparisons of CA and CB, and (b) network of studies reflecting MTCs of CA, CB, AB and AD trials.



Section 3.4.1 below outlines a standard REs MTC model for a binary outcome (Ades, A. E. et al., 2006b) which is expanded upon in future sections – FEs approaches are not pursued here, but could be attained with straightforward simplifications of the model. Once again, all models described below are fitted using Bayesian MCMC methods and all unknown parameters requiring prior distributions are given priors which are intended to be vague throughout.

3.4.1 Synthesizing aggregate data only

Following the approach used by Ades *et al.* (2006b) a REs MTC model for binary outcome data, using only AD, can be written as:

$$r_{jk} \sim Bin\left(p_{jk}, n_{jk}\right)$$

$$\text{logit}\left(p_{jk}\right) = \begin{cases} \mu_{jb} & b = A, B, C, ... \quad \text{if } k = b \\ \mu_{jb} + \delta_{jbk} & \text{if } k \text{ alphabetically after } b \end{cases} \tag{3.6}$$

$$\delta_{jbk} \sim N\left(d_{bk}, \sigma^2\right) \sim N\left(d_{Ak} - d_{Ab}, \sigma^2\right)$$

where r_{jk}, denotes the number of observed events, and n_{jk} the total number of individuals in the k^{th} treatment arm of the j^{th} trial. The underlying probabilities of an event for each arm in each trial are represented by p_{jk}. The quantity μ_{jb} represents the log-odds of an event for treatment *b* in study *j*, and d_{bk} is the log-OR for treatment *k* relative to the study-specific baseline treatment *b*. Each δ_{jbk}, the log-OR for treatment *k* relative to treatment *b* in trial *j*, is assumed to be Normally distributed with mean d_{bk} and variance σ^2. Prior distributions need to be specified for μ_{jb} and *d* $\left(\sim N\left(0, 10^6\right)\right)$ and for σ $\left(\sim Unif\ (0,2)\right)$. Note that $d_{AA} = 0$.

This synthesis model requires modifications in order to incorporate trials with 3 or more arms, as the model must take into account the correlation structure between arms of the same trials. These alterations rely on the use of the multivariate normal distribution for the intervention effects (see elsewhere for details of implementation (Lu, G. & Ades, 2004)).

3.4.2 Synthesizing individual participant-level data only

A REs model for MTC using IPD only can be written as:

$$Y_{ijk} \sim Bernoulli\left(p_{ijk}\right)$$

$$\text{logit}\left(p_{ijk}\right)=\begin{cases} \mu_{jb} & b = A,B,C,.... \text{ if } k = b \\ \mu_{jb}+\delta_{jbk} & \text{if } k \text{ alphabetically after } b \end{cases} \tag{3.7}$$

$$\delta_{jbk} \sim N\left(d_{bk},\sigma^2\right) \sim N\left(d_{Ak}-d_{Ab},\sigma^2\right)$$

For each study, the binary response of the i^{th} participant, in the k^{th} treatment arm of the j^{th} study, Y_{ijk} (i.e. 1 = event, 0 = no event), is assumed to follow a Bernoulli distribution with the probability of the event of interest occurring of p_{ijk} . As above, a standard logistic regression is fitted to each participant *i* of the j^{th} trial, with μ_{jb}, representing the log-odds for the control group (baseline *b*) in study *j*. The δ_{jbk}, derived from the treatment group indicator for each participant *i*, is the log-OR for treatment *k* relative to the study-specific (*j*) baseline treatment *b*. Prior distributions need to be specified for μ_{jb}, *d* and σ, as specified in model (3.6). A similar modification to allow for studies with more than 2 arms, akin to that in section 3.4.1, is straightforward and can also be applied to all subsequent models described. Applications of the IPD MTC model as in (3.7) have not been encountered.

3.4.3 Synthesizing individual and aggregate level data

Similar to the pair-wise case discussed in Section 3.3.3, an IPD only network meta-analysis would rely on the availability of IPD datasets for all studies. Since the availability of IPD for all studies is fairly unlikely, one may have to statistically synthesise evidence in IPD and AD simultaneously. The model described below can

be viewed as an integration of the previous two models, (3.6) and (3.7), with all notation conventions remaining the same. Similarly to model (3.3) and (3.5), the following model is described in 4 interrelated parts. Thus, a REs MTC model for the combination of IPD and AD for binary outcomes, allowing for different allocation procedures, can be written as:

(3.8) *Part I – Model for individually allocated IPD studies*

$$Y_{ijk} \sim Bernoulli\left(p_{ijk}\right)$$

$$\text{logit}\left(p_{ijk}\right) = \begin{cases} \mu_{jb}^{IPD} & b = A, B, C, \ldots \quad \text{if } k = b \\ \mu_{jb}^{IPD} + \delta_{jbk} & \text{if } k \text{ alphabetic ally after } b \end{cases} \qquad (3.8.1)$$

$$\mu_{jb}^{IPD} \sim N\left(0, 10^6\right)$$

For i = 1, 2, ... , number of participants in the j^{th} individually allocated IPD study; j =1, 2, . . . , number of individually allocated IPD studies; and k = 1, 2, … , number of treatments for which participants were allocated to.

Part II - Model for cluster allocated IPD studies

$$Y_{imjk} \sim Bernoulli\left(p_{imjk}\right)$$

$$\text{logit}\left(p_{imjk}\right) = \begin{cases} \mu_{mjb}^{c.IPD} & b = A, B, C, \ldots \quad \text{if } k = b \\ \mu_{mjb}^{c.IPD} + \delta_{jbk} & \text{if } k \text{ alphabetic ally after } b \end{cases} \qquad (3.8.2)$$

$$\mu_{mjb}^{c.IPD} \sim N\left(\theta_j, \tau.c_j^2\right)$$

For i = 1, 2, ... , number of participants in the m^{th} cluster of the j^{th} cluster-allocated IPD study; m = 1, 2, . . . , number of clusters in the j^{th} study; j = (number of individually allocated IPD studies + 1),…,(number of individually allocated IPD studies + number of cluster allocated IPD studies) ; and k = 1, 2, … , number of treatments for which participants were allocated to.

While equation (3.8.1) is used for individually allocated IPD studies, equation (3.8.2) allows for clustering effects within studies. In equation (3.8.2) a separate unconstrained control group odds for each cluster is estimated on the *logit* scale within each study, $\mu_{mjb}^{c.IPD}$, assuming that these are exchangeable within each study, with mean θ_j and variance $\tau.c_j^2$ - both these parameters requiring the specification of (vague) prior distributions ($N(0,10^6)$ and $Unif\ (0,10)$, respectively). Cluster effects between studies are assumed to be independent.

Part III - Model for both cluster and individually allocated AD studies

$$r_{jk} \sim Bin(p_{jk}, n_{jk})$$

$$\text{logit}(p_{jk}) = \begin{cases} \mu_{jb}^{AD} & b = A, B, C, ... \quad \text{if } k = b \\ \mu_{jb}^{AD} + \delta_{jbk} & \text{if } k \text{ alphabetically after } b \end{cases} \qquad (3.8.3)$$

For j = (number of individually allocated IPD studies + number of cluster allocated IPD studies + 1)... (total number of studies).

Equation (3.8.3) combines both individually- and cluster-allocated studies in AD since pre-model data adjustments were made to appropriately inflate the treatment effect variances for the effects of clustering in the cluster-allocated trials. This inflation may be made by estimating the ICC, as previously demonstrated in related literature (Donner & Klar, 2002). A prior distribution needs to be specified for μ_{jb}^{AD} $\left(\sim N(0,10^6)\right)$.

Part IV: Combination of estimates of the intervention effect

$$\delta_{jbk} \sim N\left(d_{bk}, \sigma^2\right) \sim N\left(d_{Ak} - d_{Ab}, \sigma^2\right) \quad (3.8.4)$$

For k = 1 . . . , total number of treatments.

Equation (3.8.4) specifies a RE to be placed across all treatment effect estimates from the IPD and AD, δ_{jb}'s, imposing the exchangeability property. In this way, synthesis across both types of data is achieved since equations (3.8.1), (3.8.2), and (3.8.3) "share" parameters. Prior distributions need to be specified for d $\left(\sim N\left(0,10^6\right)\right)$ and σ $\left(\sim Unif\ (0,2)\right)$. It is notable that for an analysis such as this, which does not consider participant level covariates, no loss of information is expected when using AD compared to IPD. As for model 3.3, model 3.8 is expected to provide the same results as the synthesis of AD and may seem over-elaborated. Nonetheless, this is no longer the case when information on covariates is included, as it is shown in the next section.

3.4.4 Mixed treatment comparisons models: including covariates

The inclusion of study-level covariates may explain some of the between-study heterogeneity and reduce inconsistency in the network (Cooper, Nicola J. et al., 2009). Of course, treatment effect associations may be explored using AD and average study level covariates. Meta-regression techniques can be translated into the MTC situation in quite a straightforward way, although the disadvantages previously highlighted still persist (Dias et al., 2011a, Nixon, R. M. et al., 2007, Jansen, 2006). Cooper *et al.* (2009), consider three different assumptions that can be made about *treatment x covariate* interactions, namely: *(i)* they are independent for every treatment in the network; *(ii)* they are the same for all treatments in the network; and *(iii)* they are assumed exchangeable for all treatments in the network. Option (i) is the least stringent but requires more data; option (ii) makes the strongest assumption, while option (iii) is a 'half-way-house' between the first two, where interactions can be

different across treatments but they borrow strength from one another. When data availability is expected to be limited - as it is in the motivating example - option (iii) may be most appealing and is pursued below.

The following MTC model is described in parts and extends model (3.8) for the combination of IPD and AD, considering *(a)* individual level covariate values for the IPD, *(b)* study-level covariate information for AD, *(c)* exchangeable *treatment x covariate* interactions, and *(d)* the partitioning of the variability regarding the interactions (extending the framework used by Riley and Steyerberg (2010) to the MTC setting). Alternative assumptions could be accommodated with relatively straightforward modifications to the model specification.

(3.9) *Part I - Model for individually allocated IPD studies including covariates*

$$Y_{ijk} \sim Bernoulli\left(p_{ijk}\right)$$

$$\text{logit}\left(p_{ijk}\right) = \begin{cases} \mu_{jb}^{IPD} + \beta_{0j} \cdot x_{ij} & b = A, B, C, \ldots \quad \text{if } k = b \\ \mu_{jb}^{IPD} + \delta_{jbk} + \beta_{0j} \cdot x_{ij} + & \text{if } k \text{ alphabetically after } b \\ + \beta_{bk}^{B} \cdot \bar{x}_j + \beta_{bk}^{W} \cdot \left(x_{ij} - \bar{x}_j\right) & \end{cases} \tag{3.9.1}$$

$$\mu_{jb}^{IPD} \sim N\left(0, 10^6\right)$$

For i =1, 2, ... , number of participants in the individually allocated j^{th} IPD study; and j =1, 2, . . . , number of individually allocated IPD studies

Part II - Model for cluster allocated IPD studies including covariates

$$Y_{imjk} \sim Bernoulli\left(p_{imjk}\right) \tag{3.9.2}$$

$$\text{logit}\left(p_{imjk}\right)=\begin{cases}\mu_{mjb}^{c.IPD}+\beta_{0j}\cdot x_{imj} & b=A,B,C,\ldots \text{ if } k=b\\ \mu_{mjb}^{c.IPD}+\delta_{jbk}+\beta_{0j}\cdot x_{imj}+ & \text{if } k \text{ alphabetically after } b\\ +\beta_{bk}^{B}\cdot\bar{x}_j+\beta_{bk}^{W}\cdot\left(x_{imj}-\bar{x}_j\right) & \end{cases}$$

$$\mu_{mjb}^{c.IPD}\sim N\left(\theta_j,\tau.c_j^2\right)$$

For i =1, 2, ... , number of participants in the m^{th} cluster of the j^{th} cluster allocated IPD study; m = 1, 2, . . . , number of clusters in the j^{th} study; and j = (number of individually allocated IPD studies + 1),…,(number of individually allocated IPD studies + number of cluster allocated IPD studies).

There are three extra terms in (3.9.1) and (3.9.2) compared to (3.8.1) and (3.8.2), namely: *(i)* a study-specific individual level covariate regression term, $\beta_{0j}\cdot x_{ij}$ (or $\beta_{0j}\cdot x_{imj}$ for cluster-allocated studies), where β_{0j} is the main covariate effect and x_{ij} (x_{imj}) refers to the value for the binary covariate in the i^{th} participant (in the m^{th} cluster) of the j^{th} study; *(ii)* an interaction term to account for the within-study interaction, $\beta_{bk}^{W}\cdot\left(x_{ij}-\bar{x}_j\right)$ (or $\beta_{bk}^{W}\cdot\left(x_{imj}-\bar{x}_j\right)$ for cluster-allocated studies), where within-study relationship is modelled by centring x_{ij} (x_{imj}) about the mean covariate value, $\bar{x}_j$, in each study and β_{bk}^{W} is assumed different for each (active) treatment *vs.* control comparator but exchangeable throughout all IPD studies (i.e. $\beta_{bk}^{W}\sim N\left(\beta_{Ak}^{W}-\beta_{Ab}^{W},\sigma_{B^w}^2\right)$, needing a prior distribution to be specified for σ_{B^w} $\left(\sim Unif\ (0,2)\right)$; and *(iii)* an interaction term to model the between-study relationship, $\beta_{bk}^{B}\cdot\bar{x}_j$, by interacting with the mean covariate value, which, like β_{bk}^{W}, β_{bk}^{B} is also assumed different but exchangeable. While within-study relationships may only be estimated through IPD and are captured by β_{bk}^{W}, indicating variations in an individual's event risk for a change in x_{ij} (x_{imj}), the between-study relationships may be estimated by both IPD and AD and are captured by β_{bk}^{B}, denoting the variations in underlying mean event risk for a change in $\bar{x}_j$ (the mean covariate value for the j^{th}

study). The difference between these two terms represents an estimate of the ecological bias (Riley, Richard D. & Steyerberg, 2010). In (3.9.2) both θ_j and $\tau.c_j$ require the specification of vague prior distributions $\left(\sim N\left(0,10^6\right)\right)$.

Part III - Model for both cluster and individually allocated AD studies including covariates

$$r_{jk} \sim Bin\left(p_{jk}, n_{jk}\right)$$

$$\text{logit}\left(p_{jk}\right) = \begin{cases} \mu_{jb}^{AD} & b = A, B, C, \ldots \quad \text{if } k = b \\ \mu_{jb}^{AD} + \delta_{jbk} + \beta_{bk}^{B} \cdot X_j & \text{if } k \text{ alphabetic ally after } b \end{cases} \tag{3.9.3}$$

j = (number of individually allocated IPD studies + number of cluster allocated IPD studies + 1), . . . , (total number of studies)

As in model (3.6), the likelihood contribution, r_{jk}, is described in the usual way. One additional term is included, $\beta_{bk}^{B} \cdot X_j$, which represents a study-level specific covariate regression term for treatment *k* relative to the study-specific baseline treatment *b* for each *j* trial. This term is equivalent to the exchangeable interaction term estimated in the two IPD statistical models above, which model the between-study relationship, $\beta_{bk}^{B} \cdot \overline{x}_j$. Once again, a vague prior distribution is specified for μ_{jb}^{AD} $\left(\sim N\left(0,10^6\right)\right)$.

Part IV - Combination of estimates of intervention effect including covariates

$$\begin{aligned} \delta_{jbk} &\sim N\left(d_{Ak} + d_{Ab} + \beta_{bk}^{B} \cdot X_j, \sigma^2\right) \\ \beta_{bk}^{B} &\sim N\left(\beta_{Ak}^{B} - \beta_{Ab}^{B}, \sigma_{B^B}^2\right) \end{aligned} \tag{3.9.4}$$

k = 1 . . . , total number of treatments

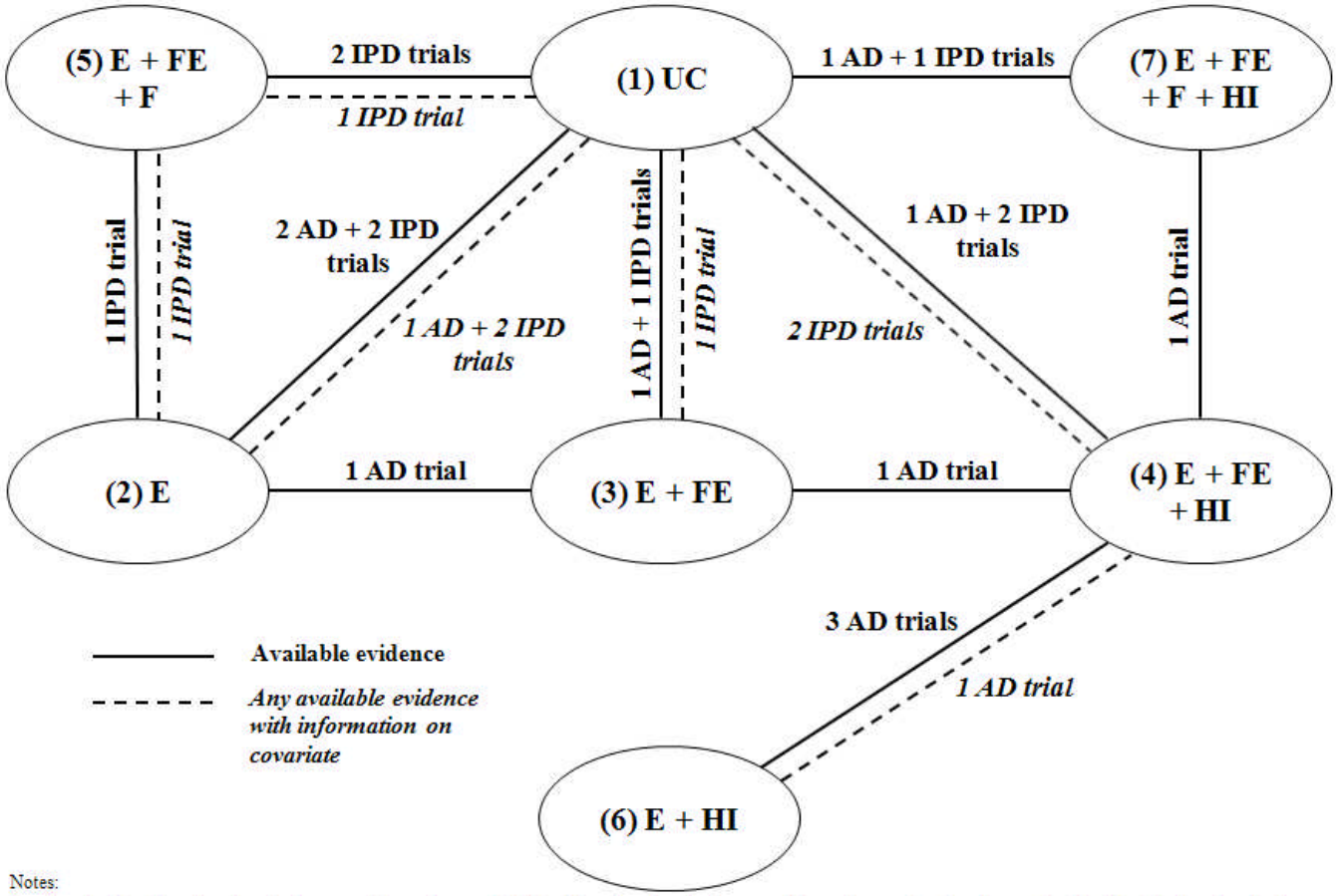
In equation (3.9.4), the study-specific individual level covariate regression terms, β_{0j}'s , are given vague prior distributions $\left(\sim N\left(0,10^6\right)\right)$. Exchangeability of β_{bk}^{B} is declared, with prior distributions needing to be specified for σ_{B^B} $\left(\sim Unif\left(0,2\right)\right)$. If independent treatment interactions are desired across the network, it would simply require defining β_{bk}^{W} and β_{bk}^{B} as following Normality with mean 0 and large variance. Extension to multiple regression coefficients is straightforward. Note: $d_{AA}, \beta_{AA}^{B}, \beta_{AA}^{W} = 0$.

3.4.5 Application

Evidence presented in Table 3.1 is structured in order to consider the different controls and treatments received by the various groups across studies. The resulting evidence network is presented in Figure 3.2. An MTC analysis of this data, ignoring the IPD, has been published elsewhere (Cooper, N. et al., 2012).

Notice that there are as many comparisons with (1) UC strategy as between non-UC interventions. Out of the 10 *direct* comparisons expressed by the network, 6 are informed by at least 1 study available in IPD format (continuous line). Figure 3.2 also indicates (dashed line) the network structure of evidence with information on the covariate of interest - household parent status (i.e. two or single parent household). Due to missing data issues, a smaller network is obtained (i.e. only 6 interventions are included in the comparison) when including studies which have information on this covariate, as there is no covariate information available for studies assessing intervention *(7) E + FE + F + HI* (i.e. the intervention of 'highest intensity'). The implications of these data restrictions/limitations are considered in section 3.5 and in the discussion section.

Figure 3.2 - Network diagram for the functioning smoke alarm outcome with information on the number and format of evidence available for each treatment comparison (continuous line) and on the number and format of evidence available for single parent status (dashed line).



Notes:

E - education; E + FE - education plus low cost / free equipment; E + FE + HI - education plus low cost / free equipment plus home inspection; E + FE + F - education plus low cost / free equipment plus fitting; E + HI - education plus home inspection; E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection; and

IPD - individual participant level data; AD - aggregate level data

As for the pair-wise models in section 3.3, the MTC models are fitted sequentially, 'building up' to the full complexity of model (3.9). Initially, the results of the MTC models for IPD, with AD plus IPD, including the IPD where available, ignoring covariates, are compared. The binary covariate 'single parent status' is then considered again, as a potential effect modifier. Firstly, a model is specified with exchangeable interaction coefficients separating the estimated *treatment x covariate* interaction effects from the within- and between-study variation. This is followed by a model with overall exchangeable interaction coefficients and comparisons between (overall) exchangeable and independent interaction effects. These latter models are slight modifications of model (3.9) in which within- and between-study associations are merged into an overall parameter.

Once again, unless stated otherwise, for all models presented in this section, the same model setting as before is used (i.e. 10,000 'burn-in' iterations, followed by 5,000 iterations, on which inferences are based). The WinBUGS code used for model (3.9) of the synthesis of AD plus IPD, including the binary covariate and separating within- and between-study associations, is provided in Appendix 3, with specific prior distributions used for this example.

3.4.5.1 Analysis of example without covariates

Table 3.7 shows parameter estimates obtained for the novel approaches, without covariates. The first column of results relates to the analysis which combines the nine IPD studies using model (3.7) – akin to doing an IPD synthesis excluding studies for which IPD could not be obtained. In the second results column, and through the use of model (3.8), all 20 studies are synthesised, using IPD where available (nine studies) and AD where not (11 studies).

When only IPD is considered, a smaller evidence network (i.e. with six rather than seven interventions) is evaluated with intervention *E + HI* being excluded as there are no IPD studies including this treatment strategy.

Table 3.7 - Parameter estimates from fitting different MTC synthesis models without including covariates to the functioning smoke alarm outcome data

Interpretation		Model (3.7) - 9 studies included		Model (3.8) - 20 studies included	
		Random effects MTC of IPD		**Random effects MTC of AD and IPD**	
		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter				
Log odds ratios for intervention effects (*vs* usual care) - *d's* [i]	***E***	-0.297	-2.045 to 1.425	-0.130	-1.116 to 0.82
	E + FE	2.036	-1.074 to 5.338	1.125	-0.05 to 2.427
	E + FE + HI	1.130	-0.86 to 3.391	0.956	0.033 to 2.177
	E + FE + F	0.930	-0.843 to 2.651	0.962	-0.235 to 2.171
	E + HI	--- [ii]	--- [ii]	1.169	-0.407 to 3.167
	E + FE + F + HI	1.165	-1.732 to 3.886	1.938	0.827 to 3.158
Between-study variance	τ^2	1.677	0.347 to 3.812	0.651	0.151 to 2.362
	Function of parameter				
Odds ratios for intervention effects (*vs* usual care) - e^{d}*'s* [i]	***E***	0.743	0.129 to 4.158	0.878	0.328 to 2.271
	E + FE	7.664	0.342 to 208.15	3.080	0.952 to 11.33
	E + FE + HI	3.096	0.423 to 29.69	2.601	1.033 to 8.823
	E + FE + F	2.534	0.431 to 14.17	2.618	0.791 to 8.768
	E + HI	--- [ii]	--- [ii]	3.220	0.666 to 23.743
	E + FE + F + HI	3.205	0.177 to 48.7	6.944	2.286 to 23.52
Deviance Information criteria [iii]	*DIC*	2937.67		3059.49	

Notes:

i) E - education; E + FE - education plus low cost / free equipment; E + FE + HI - education plus low cost / free equipment plus home inspection; E + FE + F - education plus low cost / free equipment plus fitting; E + HI - education plus home inspection; E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

ii) IPD evidence not available for this treatment comparison.

iii) Although presented models are not comparable, the DIC statistic is shown for completeness.

Carrying out the synthesis of only IPD through the use of model (3.7), results are found to be similar to the synthesis of the same studies using model (3.6), when all trial evidence is reduced to AD (results shown in Appendix 4 - Table A4). The same

is true for the results of synthesising all 20 studies at AD when compared to results when using synthesis model (3.8) (results not shown but published elsewhere (Cooper, N. et al., 2012)). This is to be expected, as there is no extra information in the IPD when overall mean effects are of interest (i.e. covariates are not considered). The most likely reasons why these two pairs of results do not agree exactly relates to how cluster adjustment is dealt with in the AD and IPD and potentially the slight influence of the prior distributions.

Larger differences are apparent between the use of all 20 studies (i.e. IPD + AD) and only the nine for which IPD is available. When synthesising the full set of evidence, the most 'intense' intervention – *E + FE + F + HI* – stands out, with an OR of 6.9 (95% CrI 2.29 – 23.52). When considering only evidence from the 9 IPD studies, the intervention *E +FE* is estimated as being the most effective compared to standard care, although this carries high uncertainty (OR 7.7 (95% CrI 0.34 – 208.15)).

3.4.5.2 Analysis of models including a binary covariate

Unfortunately, only 2 of the 11 studies for which only AD was available provided an estimate of the percentage of included subjects from 1Ps. Additionally, no parent status information was available for two of the nine studies for which IPD was available (see Table 3.1). The impact on the network diagram of the forced omission of 11 of the studies is indicated by the dashed lines in Figure 3.2. Compared to the network for all 20 studies (solid lines in Figure 3.2), this shows that even though six out of the initial seven interventions are still included, they have 'weaker links' in terms of the amount of evidence informing each of the comparisons.

The results of estimating treatment interactions terms separating the between- and the within-study associations are shown in Table 3.8. These results refer to the direct implementation of model (3.9) where IPD is included where possible (seven studies) and AD where not (two studies) and interactions are assumed to be exchangeable.

Table 3.8 - Parameter estimates from fitting a MTC model of AD and IPD with covariate 'single parent status' to the functioning smoke alarm outcome data, considering exchangeable treatment interactions and modelling separately within and between associations.

Model (3.9) - 9 studies included			
Random effects MTC of AD and IPD with covariate *			
Interpretation	**Parameter**	Median of MCMC posterior sample	95 per cent credible interval
Regression coefficients for within study association with single parent (vs two parent families)[i] - β^W's	***E***	-0.018	-0.779 to 0.704
	E + FE	0.363	-0.856 to 3.235
	E + FE + HI	0.366	-0.663 to 2.48
	E + FE + F	-0.023	-0.664 to 0.604
	E + HI	0.143	-1.855 to 2.934
Regression coefficients for between study association with single parent status (*vs* two parent families)[i] - β^B's	***E***	2.667	-1.67 to 6.059
	E + FE	2.603	-2.426 to 6.957
	E + FE + HI	2.688	-2.374 to 7.114
	E + FE + F	2.521	-2.387 to 6.97
	E + HI	2.761	-2.146 to 7.167
Difference between regression coefficients for within and between study associations with single parent (*vs* two parent families)[i] - β^{diff}'s	***E***	2.671	-1.692 to 6.186
	E + FE	1.997	-3.514 to 6.683
	E + FE + HI	2.151	-3.093 to 6.83
	E + FE + F	2.553	-2.423 to 7.002
	E + HI	2.462	-2.954 to 7.268
Deviance Information criteria	*DIC*	2452.65	

Notes:

* Model includes information on covariate related to household having one or two parents - treatment interactions are assumed exchangeable and are split in within- and between-study associations.

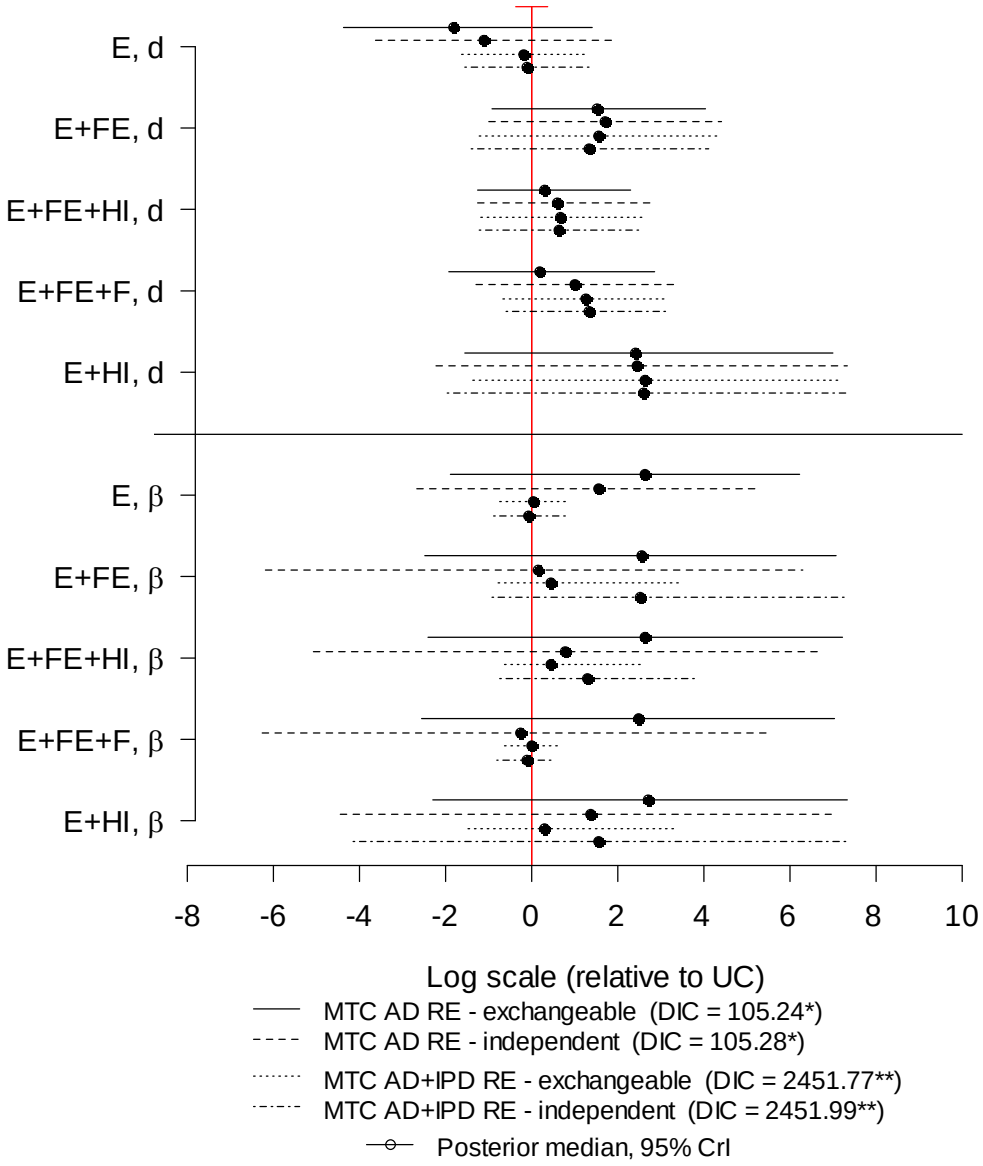
i) E - education; E + FE - education plus low cost / free equipment; E + FE + HI - education plus low cost / free equipment plus home inspection; E + FE + F - education plus low cost / free equipment plus fitting; E + HI - education plus home inspection; E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

To assess the existence of ecological bias, the difference between the association terms is estimated, β^{diff}'s. The CrIs of the difference between within- and between-study association estimates for household parent status include 0. Therefore, there is little evidence to suggest the presence of systematic differences between the covariate treatment interactions estimated by the within-study and between-study variation. In all subsequent analyses presented, these two sets of coefficients are replaced with a single one, combining the between- and within-study information.

The results of four different RE approaches to the analysis of the available evidence are presented in Figure 3.3. "MTC AD RE – exchangeable" refers to the analysis where all nine studies are fitted using AD (and all covariates are study-level) and the *treatment x covariate* interactions are assumed to be exchangeable; "MTC AD RE – independent" refers to a similar analysis, but with each *treatment x covariate* interaction assumed to be independent. "MTC AD + IPD RE – exchangeable" considers seven IPD and two AD studies. Finally, "MTC AD+IPD RE – independent" is similar to the previous one, only with independent interaction terms.

While the main treatment effects (the *d*'s) are reasonably consistent across all four models, the point estimates and the uncertainty in the interaction terms (the β*'s*) are considerably different. Uncertainty was much reduced when IPD information is available and used as such. Where it is not, estimates are shrunk towards the estimates where IPD is available in the model where interactions are assumed to be exchangeable. Interestingly, this resulted in all interaction effects being close to 0 for the exchangeable interactions model, including IPD where available. This contrasts with parameter estimates from the independent interaction model, using all evidence in the AD format, which are particularly large in magnitude and very uncertain.

Figure 3.3 - Parameter estimates from fitting the MTC of AD and the MTC of AD plus IPD models with information on covariate 'single parent status' to the functioning smoke alarm outcome data, considering exchangeable and independent treatment interactions, not separating between- and within-study interactions.



(Notes: * - comparable DIC statistics; ** - comparable DIC statistics)

On the basis of the DIC statistic (see Appendix 2 for details on this measure of model goodness of fit), all models including the covariate provide a 'better' fit to the data than the models where it is not included (e.g. "MTC AD+IPD RE – exchangeable":

DIC = 2451.77; “MTC AD+IPD RE – no covariate”: DIC = 2468.53). This indicates that a proportion of the existent heterogeneity/inconsistency in the synthesis is being explained when incorporating a relevant covariate (Cooper, Nicola J. et al., 2009). From all models considered in Figure 3.3, the ones with lowest DIC statistics are the ones which consider exchangeable *treatment x covariate* associations (i.e. “MTC AD RE – exchangeable” and “MTC AD+IPD RE – exchangeable” with DIC = 105.6 and DIC = 2451.9, respectively).

3.5 Challenges for the synthesis of (little) evidence of different formats: extending developed synthesis models

The novel methods described in section 3.4 extend the standard (pair-wise) meta-analysis by providing important information to decision makers on the optimal intervention strategy for a given purpose. The inclusion of covariates in the synthesis modelling provides a way of allowing for systematic variability between trials within pair-wise contrasts to be explained. This way, different participant subgroups may attain different intervention effects than would have been indicated by overall comparison with trials as a whole. To further evaluate the role of subgroup differences on the treatment effect, a baseline covariate effect assessment (i.e. the effect for a participant belonging to a particular subgroup – single or two parent families – under no intervention conditions) would facilitate a better understanding of how intervention relative benefits varied according to certain characteristics of participants. Policy decision makers could then consider the absolute magnitude of benefits while taking into account possible heterogeneity within the population of interest.

One additional complication is that information on the covariate(s) of interest can often be missing or incomplete in the dataset. In the case study there is missing covariate information at three levels, that is: *(i)* in the IPD where covariate information is available only for a percentage of subjects (i.e. not collected or

presented for some individuals in the dataset); and/or *(ii)* in the IPD where complete absence of information on the covariate is observed (i.e. not collected or presented for all individuals in the dataset) ; and/or *(iii)* in the AD where for some studies the mean (percentage) of individuals with the characteristic(s) of interest is missing.

The following section will present an extension of the MTC synthesis model (3.9), which takes into consideration both the incorporation of the baseline covariate effect and a multiple imputation procedure of the missing covariate values at AD and IPD through MCMC, and assuming these are 'missing at random'[21] – see further detail in Appendix 2.

3.5.1 Extensions to mixed treatment comparisons models including covariate(s): the synthesis of aggregate and individual level data

The following MTC model differs from model (3.9) in that: *(i)* it does not partition variability regarding interactions; *(ii)* it considers the issue of covariate missing values by imputing them using the posterior predictive distribution of the parameter to which it refers; and *(iii)* it estimates and incorporates the covariate baseline effect. Shorter versions of this model for AD and for IPD only can be easily derived and will not be described. A full specification of the synthesis model for AD and IPD used is:

(3.10) *Part I - Model for individually allocated IPD studies including covariates*

$$Y_{ijk} \sim Bernoulli\left(p_{ijk}\right)$$

(3.10.1)

[21] The missing-at-random assumption (sometimes called the ignorability assumption) considers that the probability that an observation is missing may depend on the observed values but not the missing values, as sufficient data has already been collected.

$$\text{logit}\left(p_{ijk}\right)=\begin{cases}\mu_{jb}^{IPD}+\beta_0\cdot x_{ij} & b=A,B,C,\ldots \quad \text{if } k=b\\ \mu_{jb}^{IPD}+\delta_{jbk}+\beta_0\cdot x_{ij}+\beta_{bk}\cdot x_{ij} & \text{if } k \text{ alphabetically after } b\end{cases}$$

$$x_{ij}\sim Bernoulli\left(p.x\right)$$

For i =1, 2, ... , number of participants in the individually allocated j^{th} IPD study; and j =1, 2, . . . , number of individually allocated IPD studies

Part II - Model for cluster allocated IPD studies including covariates

$$Y_{imjk}\sim Bernoulli\left(p_{imjk}\right)$$

$$\text{logit}\left(p_{imjk}\right)=\begin{cases}\mu_{mjb}^{c.IPD}+\beta_0\cdot x_{imj} & b=A,B,C,\ldots \quad \text{if } k=b\\ \mu_{mjb}^{c.IPD}+\delta_{jbk}+\beta_0\cdot x_{imj}+\beta_{bk}\cdot x_{imj} & \text{if } k \text{ alphabetically after } b\end{cases} \qquad (3.10.2)$$

$$\mu_{mjb}^{c.IPD}\sim N\left(\theta_j,\tau.c_j^2\right) \qquad x_{imj}\sim Bernoulli\left(p.x\right)$$

For i =1, 2, ... , number of participants in the m^{th} cluster of the j^{th} cluster allocated IPD study; m = 1, 2, . . . , number of clusters in the j^{th} study; and j = (number of individually allocated IPD studies + 1),…,(number of individually allocated IPD studies + number of cluster allocated IPD studies).

Compared to equations (3.9.1) and (3.9.2), the changes applied to (3.10.1) and (3.10.2) are the following: *(i)* the individual level covariate regression term, $\beta_0\cdot x_{ij}$ (or $\beta_0\cdot x_{imj}$ for cluster-allocated studies), becomes non-study specific and represents the baseline effect of the subgroups of interest, x_{ij} (x_{imj}); *(ii)* the split in within- and between-study variability is not considered, that is, these are merged into a single parameter, β_{bk}; and *(iii)* a distributional assumption is imposed on the covariate values, indicating that x_{ij} (x_{imj}) are *Bernoulli* distributed with event probability *p.x*, common across all IPD studies. By imposing this (prior) distributional assumption over the covariate values, these model sections consider existing covariate evidence

and uses this information to multiple impute the missing covariate information through the MCMC procedure.

Part III - Model for both cluster and individually allocated AD studies including covariates

$$r_{jk} \sim Bin\left(p_{jk}, n_{jk}\right)$$

$$\text{logit}\left(p_{jk}\right) = \begin{cases} \mu_{jb}^{AD} & b = A, B, C, \ldots \quad \text{if } k = b \\ \mu_{jb}^{AD} + \delta_{jbk} + \beta_{bk}^{B} \cdot X_{j} & \text{if } k \text{ alphabetic ally after } b \end{cases} \qquad (3.10.3)$$

$$X_{j} \sim Beta\left(pa.x, pb.x\right)$$

j = (number of individually allocated IPD studies + number of cluster allocated IPD studies + 1), . . . , (total number of studies)

Part IV - Combination of estimates of intervention effect including covariates

$$\delta_{jbk} \sim N\left(d_{Ak} + d_{Ab} + \beta_{bk} \cdot X_{j}, \sigma^{2}\right)$$

$$\beta_{bk} \sim N\left(\beta_{Ak} - \beta_{Ab}, \sigma_{B}^{2}\right) \qquad (3.10.4)$$

$$pa.x \sim Unif(0,1) \qquad pb.x \sim Unif(0,1) \qquad p.x \sim Beta(pa.x, pb.x)$$

k = 1 . . . , total number of treatments

As for model (3.9), the likelihood of the contribution of both cluster- and individually-allocated AD studies, r_{jk}, is described in the usual way in part three of the model. Subgroup baseline effect, represented by the β_0 term, is considered common to both IPD and AD and it requires a (vague) prior distribution to be specified $\left(\sim N\left(0,10^{6}\right)\right)$. The study-level covariate information, X_j, follows a Beta distribution with parameters *pa.x* and *pb.x*, both Uniformly distributed $\left(\sim Unif\ (0,1)\right)$. The same distributional assumption is imposed on *p.x* so that parameter sharing

exists throughout the different model parts, linking IPD and AD information and enabling the estimation of common predicted covariate distribution. Extension to multiple regression coefficients is straightforward. Note: $d_{AA}, \beta_{AA} = 0$.

3.5.2 Application

In the findings from the analysis reported in section 3.4.5.2, missing data are dealt with in a simple and straightforward way, that is, by discarding studies (with evidence at summary- and individual level) with the totality of the covariate information absent. Also, for the IPD for which covariate information is missing for some of the trial participants, these individuals are discarded, that is, a complete case analysis is performed. The impact on the evidence network of the 'forced' omissions of these studies is large, as not only did certain comparisons become "weaker" in terms of the amount of evidence informing them, but also the network included only 6 out of the 7 initial interventions. Many disadvantages are expected from using this type of approach to missing data, from which the one mainly highlighted here is the introduction of an element of ambiguity into the statistical analysis undertaken[22].

As reported in Table 3.1, information on the covariate relating to single parent status is absent in 55% of the trials[23]. This represents a significantly high proportion of missing information; nonetheless, 7 trials had information on this covariate at the individual level which means an imputation procedure over this covariate is expected to be robust.

[22] A complete case analysis is considered unbiased provided that the missingness mechanism is independent of the outcome. In particular, complete case analysis is still unbiased if missingness in baseline covariates is dependent on the baseline covariates themselves (i.e. missing not at random). The problem with complete case analysis here is therefore more that it is inefficient, rather than being biased (Rubin, Donald B., 2004, Rubin, D. B., 1976).

[23] That is from the 20 available studies, 2 out of the 11 AD and 7 out of the 9 available IPD trials have information on the households' single parent status.

Table 3.9 shows the results from fitting the 3 different synthesis models (i.e. of all evidence at summary level, of IPD alone and of both AD and IPD) including the covariate on single parent status. These results include the imputation of covariate values, increasing the available evidence base to a new level. Log-ORs (the $\hat{d}'s$) and ORs (the $\hat{e}^d's$) for each intervention (relative to the 'baseline' intervention) are provided for each model. This is complemented by the results of the single parent subgroup baseline effect, $\hat{\beta}_0$, and the *treatment x covariate* interaction terms, $\hat{\beta}'s$. Between-study heterogeneity is also illustrated through τ^2.

The imputation analysis performed over the covariate 'missings' allowed for each model evidence base to be comparable. This implies that the full set of "relevant" studies (i.e. 20) is now used for the AD and for the AD + IPD models (Table 3.9 – first and third column of results, respectively) and 9 IPD studies for the synthesis of IPD only analysis (Table 3.9 – second column of results).

With information on nine IPD studies, the synthesis model of the mixture of AD and IPD estimates a treatment effect on the log scale of 1.7 (95% CrI 0.33 – 3.13) for the most "intense" intervention (i.e. *(7) E + FE + F HI*). Similar conclusions in terms of the most effective intervention can be drawn from synthesizing all evidence at summary level, although in this case, CrIs include zero (log-OR 1.6, 95% CrI -0.19 – 3.19). The inclusion of IPD studies impacts upon the main treatment effect estimates compared to the AD model, upwardly inflating the between-study variance estimate and its uncertainty. Estimated baseline subgroup effect is shown to be very different and with opposite signs across models, once more flagging up the 'weaknesses' of the AD model estimates compared to the ones obtained by using the AD + IPD model.

With respect to interaction estimates, uncertainty is much reduced when IPD is available, resulting in some associations effects to be close to 0 (i.e. for *(2) E* and for *(5) E + FE + F*). Again, this shows that potential ecological bias is being accounted

for when IPD is included in the synthesis as it facilitates an appropriate estimation of treatment x covariate interactions.

Table 3.9 - Parameter estimates from fitting the MTC model of AD, IPD only and AD and IPD with covariate 'single parent status' to the functioning smoke alarm outcome data; considering exchangeable treatment interactions, estimating subgroup baseline effects and imputing missing covariate values.

		20 studies included		9 studies included		Model (3.10) - 20 studies included	
		Random effects MTC of AD with imputation		**Random effects MTC of IPD with imputation**		**Random effects MTC of AD and IPD with imputation**	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Parameter						
Log odds ratios for intervention effects (*vs* usual care) - d's [i]	E	-1.082	-3.004 to 0.884	-0.334	-2.141 to 1.41	-0.188	-1.218 to 0.819
	$E + FE$	0.875	-0.521 to 2.413	1.988	-1.155 to 5.349	1.028	-0.261 to 2.412
	$E + FE + HI$	0.515	-0.617 to 1.993	0.971	-1.044 to 3.277	0.681	-0.344 to 1.87
	$E + FE + F$	0.504	-0.981 to 2.169	0.932	-0.826 to 2.771	0.935	-0.263 to 2.13
	$E + HI$	0.835	-0.995 to 2.989	--- [iii]	--- [iii]	1.149	-0.649 to 3.298
	$E + FE + F + HI$	1.561	-0.189 to 3.188	1.019	-1.877 to 4.161	1.732	0.326 to 3.13
Regression coefficient [ii]	$\beta_{0\ single\ parent}$	1.063	-1.159 to 3.445	-0.346	-0.664 to -0.029	-0.313	-0.63 to 0.002
(Overall) Regression coefficients for intervention interactions (*vs* usual care) - β's [i]	E	1.597	-1.449 to 4.43	0.080	-0.534 to 0.659	0.089	-0.504 to 0.652
	$E + FE$	1.079	-2.46 to 4.546	0.509	-1.3 to 3.579	0.310	-1.223 to 2.332
	$E + FE + HI$	1.481	-2.065 to 5.039	1.121	-0.072 to 3.126	0.881	-0.159 to 2.592
	$E + FE + F$	1.071	-2.844 to 4.677	0.102	-0.519 to 0.716	0.071	-0.537 to 0.666
	$E + HI$	1.375	-2.22 to 5.116	--- [iii]	--- [iii]	0.401	-1.265 to 2.859
	$E + FE + F + HI$	1.231	-2.398 to 4.993	0.404	-1.993 to 3.429	0.394	-1.548 to 3.048
Between-study variance	τ^2	0.474	0.017 to 2.041	1.663	0.316 to 3.817	0.654	0.14 to 2.333
	Function of parameter						
Odds ratios for intervention effects (*vs* usual care) - e^d's [i]	E	0.339	0.05 to 2.421	0.716	0.118 to 4.094	0.828	0.296 to 2.268
	$E + FE$	2.399	0.594 to 11.171	7.298	0.315 to 210.4	2.797	0.77 to 11.156
	$E + FE + HI$	1.673	0.54 to 7.335	2.642	0.352 to 26.49	1.975	0.709 to 6.491
	$E + FE + F$	1.656	0.375 to 8.749	2.539	0.438 to 15.97	2.547	0.768 to 8.417
	$E + HI$	2.305	0.37 to 19.863	--- [iii]	--- [iii]	3.155	0.523 to 27.064
	$E + FE + F + HI$	4.765	0.828 to 24.246	2.772	0.153 to 64.11	5.651	1.385 to 22.869

Notes:

i) E - education; E + FE - education plus low cost / free equipment; E + FE + HI - education plus low cost / free equipment plus home inspection; E + FE + F - education plus low cost / free equipment plus fitting; E + HI - education plus home inspection; E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

ii) Covariate term for single parent status (baseline subgroup effect)

iii) IPD evidence not available for this treatment comparison.

3.6 Discussion

Findings summary

This chapter contributes to the synthesis of evidence methods literature by describing and applying a series of meta-analytic models, including novel MTC models that allow IPD and both AD and IPD to be included while considering a participant level covariate and making different assumptions about the covariate effects (Cooper, Nicola J. et al., 2009). Modelling of cluster allocation effects (Sutton, A. J. et al., 2008), distinct covariate effects based on between- and within-study variability (since the former is susceptible to ecological biases (Riley, Richard D. & Steyerberg, 2010)), and the multiple imputation of missing covariate information were also considered.

The motivating example showed that the use of evidence at the individual level, whether or not in the combination with summary evidence, provided more precise estimates and estimates of greater accuracy of the *treatment x covariate* interaction effects, when compared to those estimated through AD only. Additionally, different assumptions about the covariate interactions were tested. In this example, assuming interaction 'exchangeability' provided the 'best' fit to the data compared to assuming common or independent regression slopes.

The motivating example – assessing the effectiveness of interventions to increase the uptake of functioning smoke alarms in households – showed that more 'intense' interventions are more effective than those which are less so, with the one providing education plus low cost/free equipment plus fitting plus home inspection having the highest level of effectiveness from the set.

Evidence synthesis in Health Technology and Public Health programmes appraisal

Estimating the effectiveness of alternative health care interventions is at the heart of not only clinical but also economic evaluations. The NICE for England and Wales uses economic analyses to recommend health care technologies for use in the NHS. As stated in Chapter 2, the guide to methods for HTA published by the NICE (2008) acknowledges that the construction of '…an analytical framework to synthesise the available evidence in order to estimate clinical and cost effectiveness…' should be performed and recommends that '… all relevant evidence must be identified, quality assessed and pooled using explicit criteria and justifiable and reproducible methods' (page 27). Meta-analyses techniques are often used to summarize evidence on clinical effectiveness in the NICE technology appraisals, and to subsequently inform related economic analyses. The use of indirect and MTC methodologies in informing decision-making is also becoming more common.

The notion of what should be considered to be relevant evidence in HTA and PH is yet to be unequivocally determined. Consequently, issues surrounding the use of IPD compared to AD also remain unclear. Resistance from authors/researchers to release IPD data, the costs related to time and computational burdens compared to analysis of AD only, and the delay in producing the evidence for decision making all work against the routine collection and use of IPD. The benefits of obtaining IPD over and above the existing AD should be taken into account, since clear benefits, such as more accurate estimation of subgroup effects, as demonstrated here, make a strong case for using IPD in synthesis models whenever possible. However, IPD may not always be available for all studies; hence the methods developed here. It is believed that these are an improvement on existing alternatives identified by a recent review of the literature (Riley, R. D. et al., 2007).

Strengths and limitations

This chapter did not explore issues related to sub-optimum data quality, such as the use of non-randomised studies in the synthesis (although bias adjustment proposals have been published (Turner, R. M. et al., 2009, Welton et al., 2009, Spiegelhalter, David J. & Best, 2003)) or adjusting for arm imbalances in the covariate of interest (although approaches recently developed for pair-wise meta-analysis could be applied to the MTC models developed here (Turner, R. M. et al., 2009)).

The evidence base with respect to the covariate of interest on single parent status was enhanced through an imputation procedure. Using all the available evidence enabled establishing comparisons between results from models with and without covariate information. The multiple imputation procedure assumed a 'missingness' mechanism of 'missing at random'. Nonetheless, no sensitivity analysis was performed to verify the validity of obtained estimates, despite the advice in some of the literature (Carpenter et al., 2007).

For heterogeneity to be realistic, restrictions were imposed in the main analysis through the use of 'not so vague' priors, which may constrain the interpretation of results. Nonetheless, sensitivity analyses have been conducted over these prior distributions for between-study variances, which showed that results were not sensitive to the choice of distribution parameterization.

The MTC synthesis modelling could have explored other potential extensions. These could come from examining other differential effects by child and family factors (i.e. more subgroup analysis), from including and exploring the effect of study quality in the analysis, and from the assessment of possible network inconsistencies between direct and indirect evidence. A further worthwhile extension would be to develop a generalized linear modelling framework to extend the proposed approach to other types of outcome measures (e.g. categorical, continuous) building on recent work (Dias et al., 2011b).

Conclusions

Using IPD from all studies is desirable. Nonetheless, this will not be possible in the majority of instances and thus it is believed that the models presented have a valuable role in the evaluation of interventions, particularly where there is inconsistency in the network and/or the treatment subgroup effects are of interest. This chapter brings into question the often publicised view that IPD syntheses are the 'gold standard' if only a fraction of the available studies can be included; i.e. it is argued that it is better not to exclude any studies from the analysis, irrespective of the format in which they are available.

Models herein described were applied to a particular PH example. Nevertheless, they are potentially of use in other health care contexts, including HTA assessments of drugs and devices where IPD may be available for a particular product, but not for competitor products. In the following chapter, the cost effectiveness of the PH programmes evaluated is assessed. This is performed by using the outputs of the synthesis models described in this chapter as decision model inputs, with further discussion of the impact of using evidence at the individual level in comparison with using only AD.

CHAPTER 4

4. ECONOMIC EVALUATION OF SMOKE ALARM PROGRAMS FOR PREVENTING FIRE-RELATED INJURIES OF PRE-SCHOOL CHILDREN IN THE HOME

4.1 Defining the context: Economic Evaluation of Public Health interventions

Chapter 3 developed a series of novel evidence synthesis methods based on the format of evidence available (AD and IPD). The estimates derived from these new methods can be used to simultaneously evaluate the effectiveness of several interventions. The aim of the current chapter is to use derived synthesis estimates based on AD in order to explore the cost effectiveness of PH programmes (encompassing seven competing strategies) to increase the uptake of functioning fire alarm equipment and to reduce fire related injuries in children younger than five years of age. The next chapter (Chapter 5) will consider the use of IPD for cost effectiveness modelling. The current and following chapters are, therefore, inter-related in the sense that, in turn, they explore the impact on model based cost effectiveness estimates of two alternative methods to estimate the effectiveness

parameters in a decision model; either through the synthesis of AD or by combining AD and IPD. In doing so, both chapters build on synthesis of the evidence of effectiveness presented and discussed in Chapter 3.

As highlighted in Chapter 1, economic evaluation compares the costs and consequences of alternative courses of action. Resource allocation decisions based on this tool aim to maximise health gains from the available resources – resources which are finite within the public sector. While this instrument has been widely used in the past two decades to support decision making in the health care setting, it has only been in recent years that an increase has occurred in the awareness of its advantages, and, consequently, an increase in its use, for PH. Reasons for this relate to the fact that issues affecting the evaluation of the effectiveness of PH interventions also affect their economic evaluation. These issues include, for example, the general lack of data on the interventions and general poor quality where it does exist, lack of randomised data on interventions (or data from which the effect of interventions may be appropriately isolated), and issues of bias and confounding.

Given that the use of cost effectiveness in the assessment of PH policies is still in its infancy, many still consider effectiveness to be the only (or the principal) relevant factor in the implementation of a policy. To inform effectiveness, a review of the literature is commonly undertaken, the results of which may require aggregation (i.e. combination of quantitative evidence from multiple sources) in order to attain a legitimate understanding of the effects of these multifaceted programmes. To undertake this aggregation, the data (usually from RCTs) needs to be sufficiently homogenous, or heterogeneity needs explicit consideration. When data carries some degree of heterogeneity (for example, due to differences in population characteristics, settings, or differences in the design and conduct of trials), it is of interest to explore how the treatment effect changes across patient subgroups. However, it is not sufficient to examine the effectiveness of a PH programme (and the impact of heterogeneity) – the analysis of the effectiveness of public policies should also be cross-referenced with data on their costs (Drummond et al., 2005).

The Canadian National Centre for Healthy Public Policy has laid out the principles for assessing PH policies (Morestin et al., 2010) and recognises the need for evaluating costs alongside effects. It highlights many methodological challenges specific to PH, some of which are: *(i)* the attribution of effects (both intended and unintended) of the policy on the targeted population and problem; *(ii)* the costs and consequences which should be analysed, considering the feasibility of the programme; and *(iii)* the acceptability of the policy by the relevant stakeholders, which often involves subjective judgements, beliefs, values and interests of the actors concerned. In addition, a recent study by Weatherly *et al.* (2009) identifies further challenges regarding outcomes valuation and in considering equity (i.e. of obtaining an equilibrium between an efficient and an equitable allocation of resources). The authors recognise that empirical literature offers very limited insight on how to appropriately respond to all these factors.

In the UK, there has been increasing awareness on the need to assess PH programmes, and on the need to develop methods specific to this area. As an example, guidance on the methods for conducting systematic reviews in PH has recently been released from the Centre for Reviews and Dissemination (CRD, 2009) and from the Cochrane Collaboration (Armstrong et al., 2007)) . Also, the NHS Health Development Agency (Kelly et al., 2005) and the NICE (NICE, 2009) has released guidance for the economic appraisal of PH interventions. In these, a general framework is presented highlighting the mechanisms that should be applied in the economic evaluation of PH interventions. These include, for instance, guidance on identifying, reviewing, extracting, synthesising and presenting evidence, as well as guidance on possible modelling approaches, perspectives and the identification and selection of model inputs.

This chapter develops a cost effectiveness model of the PH strategies described and evaluated in Chapter 3. Section 4.2 briefly summarises existing economic literature on the prevention of fire related accidents in the home. Section 4.3 presents the methodology used to perform the proposed economic evaluation and includes, for example, the decision model structure and a description of the evidence used to

populate it. This is followed by a discussion of the cost effectiveness results at population and subgroup levels in section 4.4. Finally, section 4.5 discusses the findings, highlights possible policy implications and flags up scope for further work.

4.2 Review of the existing economic evidence in accident prevention

This section focuses on the economics of injury prevention programmes. In particular, it highlights relevant economic literature concerning interventions aimed at reducing fire-related injuries.

The purpose of this review is to illustrate: *(i)* the types of accident prevention programmes that have been economically assessed previously; *(ii)* how these programmes were evaluated; (iii) the economic model framework used; *(iii)* the data inputs informing them; *(iv)* the main modelling assumptions implemented; *(v)* the viewpoint of the analysis; *(vi)* and the sensitivity analysis put into practice. Studies were identified through searches on the Medline, Embase, Cinahl, Assia, Psychinfo and Web of Science databases, from the inception of the database until December 2009. The inclusion criteria included full economic evaluation (cost effectiveness) studies, examining interventions that could be included in injury prevention briefings or implemented by centres for the primary and secondary prevention of thermal injuries in children aged 0-19 years and their families – the participants of interest. The outcome measure assessed was the possession and use of home fire alarm safety equipment. From the more than 400 studies and/or abstracts identified, three were found to be relevant (Pitt et al., 2009, Ginnelly et al., 2005b, Haddix et al., 2001), revealing that little economic evidence is available in the literature on this subject. A discussion of these three studies follows.

Both Haddix *et al.* (2001) and Ginnelly *et al.* (2005b) carried out economic evaluations of smoke alarm give-away schemes using decision modelling. The first study evaluated a give-away scheme implemented in the USA, while the latter was a

UK scheme. These were not aimed specifically at children. Using a societal and/or health care system perspective, these studies gave some insights into the key trade-offs between resources and outcomes within such programmes. In the study by Ginnelly *et al.* (2005b) the scheme was not found to be cost effective, while in the Haddix *et al.* (2001) evaluation it was considered an *'…economically beneficial program for preventing fatal and non-fatal residential fire related injuries…'*. Probabilistic sensitivity analysis was implemented in the former, whereas in the latter univariate and multivariate deterministic analyses were performed.

Pitt *et al.* (2009) used the Ginnelly *et al.* (2005b) study as a starting point and implemented a cost utility analysis of the lifetime costs and effectiveness of relevant home safety interventions with particular emphasis on programmes that provided smoke alarms. The aim of this study was to evaluate interventions that reduced unintentional injuries in children under 15 years of age. Using a public sector perspective and a Markov-state transition model, the authors evaluated the provision and installation of free smoke alarms *vs.* 'no intervention'. The intervention was found to be cost effective and three model parameters were highlighted as result drivers: the intervention uptake level (i.e. intervention effectiveness); the prevalence of smoke alarms in households of the targeted population; and the functional decay rate of the equipment. This study failed to compare all relevant options by not exploring, for instance, alternative programmes such as: free supply of the home safety device *vs.* free supply and installation; free device supply *vs.* tailored device supply and advice; or different amounts of safety education and information alongside the safety device-based programme components. Pitt *et al.* (2009) was used as basis for the NICE PH guidance on the prevention of unintentional injuries among under-15s in the home (NICE, 2010).

The small number of existing studies, together with the issues highlighted above (e.g. interventions evaluated and target population), justifies the development of a new economic evaluation in this chapter.

4.3 Methods

The decision problem and the structure of the decision models used to evaluate the cost effectiveness of smoke alarm programmes aimed at reducing household fires involving children up to the age of five are presented next. The principles for model design set out by Philips *et al.* (2006) were used.

4.3.1 The decision problem

The study population includes UK households with pre-school children (unit of analysis). The aim of the interventions assessed here is to reduce injuries, whether minor, moderate or disabling, and deaths in children, as a consequence of a fire accident in the home. The strategies evaluated are those described in Chapter 3. Only households with a single child are considered, and it is assumed that the intervention is offered when they are born.

4.3.2 Decision model structure

In modelling the decision problem described above there was a general awareness that, in many aspects of the evaluation, data was going to be sparse or inexistent. As stated by Weatherly and colleagues (2009), this is usually the case when modelling PH type interventions compared to health technologies, like pharmaceuticals, for which greater and higher quality evidence may be available.

While the developed model is mainly focused on possible reductions of household fires and fire-related injuries in children under five, the lifetime costs and benefits of each of the strategies are considered. The model's time horizon was set to one hundred years, by which time the majority of individuals will have died. The model used a cycle length of one year, which was considered an appropriate reflection of potential transitions between model states.

The model structure is depicted in Figure 4.1. The model is split into three parts. Parts one and two model the process up to 5 years, while part three focuses on the evaluation of long term consequences and costs. Each part is described in turn in the next subsections.

Part 1 – Intervention model

Part one is labelled as the **intervention model** (Figure 4.1 - part 1) and it models the number of households accepting the intervention or not, the uptake of smoke alarms in a family that has accepted the intervention and the likelihood of having functioning equipment. The interventions modelled are expected to act on the rate of uptake of the safety equipment.

Part 2 – Five year Markov structure

Part two of the model uses a Markov structure to evaluate the occurrence of events until children are aged five, in households with and without functioning smoke alarm equipment (labelled as a **5 year Markov structure**). Households with functioning equipment can then see reduction in the risk of a household fire. The state transition diagram for part 2 is shown in Figure 4.1 - part 2. This part considers conditions labelled ‘well’, ‘disabled’ and ‘death’, entailing six model states, namely:

(S1) (Household with) functioning smoke alarm (and child ‘well’);

(S2) (Household with) no functioning smoke alarm (and child ‘well’);

(S3) (Household with) functioning smoke alarm and (child) disability;

(S4) (Household with) no functioning smoke alarm and (child) disability;

(S5) (Child) death due to fatal fire injury; and

(S6) (Child) death due to other causes.

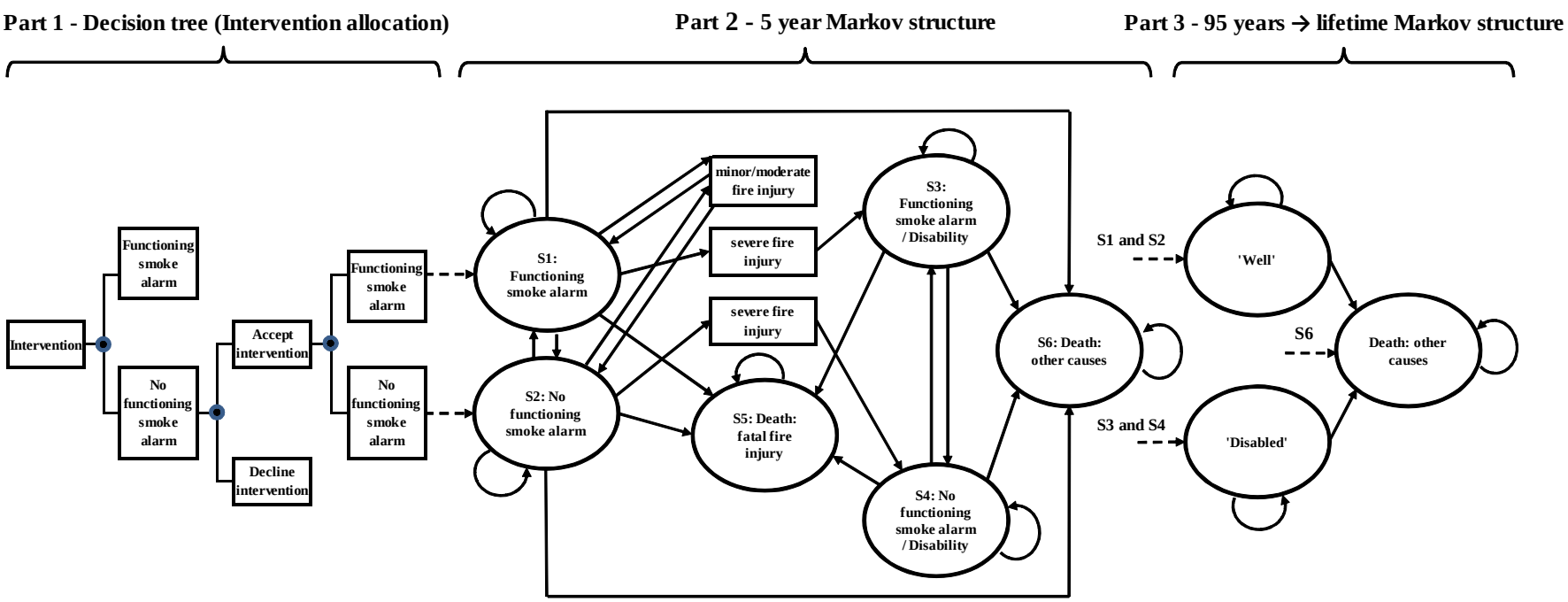
The progress between model states is conditional on the occurrence of fires in the household, and on the consequences for the child (fire-related injuries). The health

consequences include six event types: *(a)* no injury; *(b)* minor injury; *(c)* moderate injury; *(d)* severe injury; as well as fatal injuries and probability for all cause mortality, namely *(e)* fatal fire injury; and *(f)* all cause mortality.

From states S1 and S2 the child can move to any of the other model states. Once in the 'disabled' condition (S3 and S4), the child can either transit to one of the absorbing states or stay in one of the 'disabled' states, without the possibility of transiting to S1 or S2. The model also considers the possibility of the safety equipment ceasing to function and, in the case of it failing to function, that it is repaired. This is achieved with the introduction of a decay/repair factor, which establishes the transition rates, from 'functioning' to 'non-functioning' equipment and vice-versa (e.g. S1 to S2).

A severe fire-related injury was defined as one requiring an inpatient stay greater than five days of treatment in an intensive care unit. It was assumed that any child suffering a severe injury (particularly burns) would suffer some form of disability and would carry that impairment for the rest of its life. A child experiencing these events would therefore suffer a decrement in (health related) quality of life and would be subject to additional health costs for the rest of its lifetime. In the event of a minor fire-related injury, a child is assumed not to have any significant decrease in their quality of life or any additional ongoing health costs and in theory this type of injury might occur more than once. Like for minor injuries, moderate injuries also involve a utility decrement (higher than for a minor injury), however this is restricted to that event/cycle and does not imply a lifetime disability.

Figure 4.1 - Decision analytical model structure, part 1: model for households receiving interventions; part 2: Markov state transition model for pre-school children aged 1 to 5; and part 3: Markov state transition model for rest of life (5 years onwards).



Part 3 – Lifetime Markov structure

The final part of the decision model, that is, model part three, labelled as **lifetime Markov structure**, is a simple Markov process which is used to model the progression of the child for the next 95 years (i.e. the remainder of their lives) – Figure 4.1 - part 3. An individual who is 'well' or 'disabled' at the age of five stays in this state until death – enabling the incorporation of the lifetime impact of a disabling fire-related injury in the evaluation.

Explicit assumptions imposed by using this model structure are highlighted in the next section and discussed in section 4.5.

4.3.3 Decision analytic model assumptions/simplifications

As with any model, simplifications and assumptions are required. In this work it was assumed:

i) The probability of a household accepting an intervention was the same across interventions, owing to a lack of information on the uptake of the different programmes;

ii) Lack of evidence on the benefits and detriments of interventions where multiple children are involved, the decision model only considers households with a single child. This may be a conservative assumption, as on one hand multiple children could benefit from the same equipment, on the other it would increase the probability of at least one of them being injured in a home fire;

iii) The model ignores potential (positive or negative) spill over effects on sibling(s) and/or parent(s). For instance, parents could benefit from their own actions in installing and maintaining the safety equipment in the home;

iv) The probability of a future fire-related injury is assumed not to be dependent on past accidents, and remains constant throughout the relevant model time frame (i.e. 5 years for part 2 of the model). This assumption is common to most Markov models and implies that a household's awareness of the risk of accident remains the same, irrespective of whether a previous event occurred;

v) The model only allows for one fire-related accident and injury in any one year.

4.3.4 Model implementation

As described in Chapter 3, effectiveness evidence was synthesised within a Bayesian framework. The synthesis results were used in the decision model using 5,000 posterior samples (extracted from the Convergence Diagnostic and Output Analysis WinBUGS output (CODA)). Evidence to inform other model parameters was identified from the literature and is described in the next section. Measures of uncertainty related to each of the model input parameters were sought in the literature and used to define probability distributions to represent parameter related uncertainty in the decision model. The model is probabilistic and 5,000 samples were run in the probabilistic sensitivity analysis – the number of simulations performed in the decision model was conditional to the amount of MCMC simulations in the synthesis analysis. The decision model was implemented in software package R version 2.11.0 (Copyright © 2010 The R Foundation for Statistical Computing) and decision model R code is supplied in Appendix 5.

4.3.5 Identifying, combining and analysing existing relevant evidence

Public health methods guidance states that all 'relevant' evidence should be considered. Ideally, systematic reviews of the literature should be conducted, but this is not necessary for all types of information in economic modelling (NICE, 2009). Non-exhaustive reviews are commonly put into practice for model parameters other than relative effectiveness (Anderson, 2010).

A non-exhaustive literature search was performed in order to identify evidence to populate model parameters other than effectiveness. Although a variety of sources of evidence were identified and used, for a large number of parameters evidence was scarce. In the following subsections, a summary of the key literature sources are described, starting with the effectiveness evidence and followed by evidence to populate parameters for each model part.

When a measure of uncertainty was available, a distribution was defined and used probabilistically in the model. For a significant proportion of the parameters, however, no data existed and an informed estimate was made after consultation with the collaborating team (Denise Kendrick, Alex J. Sutton and Nicola J. Cooper).

Table A6 (in Appendix 6) presents a complete list of model parameters and data sources. This table also shows the parameter estimates and distributional assumptions used in the model.

4.3.5.1 Relative effectiveness

A description of how effectiveness evidence was identified and used to populate the cost effectiveness model follows.

Methods of synthesis with and without covariates

Evidence at aggregate level and methods of synthesis have been described in Chapter 3 (sections 3.3 and 3.4). This chapter uses evidence at the summary level (20 studies) to inform the effectiveness model parameter – a common approach in many analyses. It is important to appropriately provide the link between what is provided by the clinical effectiveness synthesis modelling and how the output of that synthesis is used to inform cost effectiveness model input parameters. This work considered: *(i)* the use of predictive treatment effect distributions; and *(ii)* the importance of modelling baseline effects and how these are incorporated in the synthesis outcome.

With respect to i), when *(a)* treatment effect heterogeneity results from unavoidable variation in implementation; or when *(b)* the intervention is aimed at individuals in groups of potentially heterogeneous effect sizes – as is often the case with PH interventions, estimated treatment effects will not reflect the efficacy of implementing the programmes in the future. In order to overcome this issue, Ades and colleagues (2005) recommend that modellers consider using the predictive distribution of a future treatment effect, or assume that the future implementation will result in a distribution of treatment effects. The posterior predictive distributions are used to model our knowledge of possible values the probability of the outcome of interest could take in the control group and in each of the treatment groups.

With respect to ii), in order to obtain an unbiased and statistically efficient treatment comparison, it is optimal to account for baseline factors that influence the outcome. There is, in fact, some credibility attached to demonstrating that effects adjusted for baseline do not alter the conclusion derived from the unadjusted analysis (Pocock et al., 2002). In using the output of the synthesis to inform the cost effectiveness model, it is beneficial to consider these baseline effects. Such effect adjustment may be estimated by adding the (log scale) treatment effects to the baseline to obtain the absolute efficacy of each of the treatments, on the assumption of a certain probability of 'success' (i.e. the uptake of safety equipment) of the common comparator (say, treatment A). To estimate the baseline for this common comparator, the usual approach is to conduct a separate synthesis of the evidence on this treatment alone.

To consider these components in the models developed in Chapter 3, it is necessary to extend them. Algebraic descriptions of the modelling extensions which consider these two issues are given below. For synthesis models which do not consider covariate effects (models 3.6, 3.7 and 3.8 described in Chapter 3):

$$\begin{aligned} \text{logit}(\text{p.treat}_{\text{A}}) &= mA \\ \text{logit}(\text{p.treat}_{\text{k}}) &= mA + d.pred_k \qquad \text{for k} \neq \text{A} \end{aligned} \tag{4.1}$$

$$mA \sim N(\mu_A, \sigma_A^2) \qquad d.pred_k \sim N(d_{Ak}, \sigma^2)$$

k = 1 . . . , total number of interventions

The absolute log odds of 'success' (i.e. uptake of 'functioning' smoke alarm equipment in the household) of intervention A, *mA*, is based on a separate model. This parameter is assumed to be Normally distributed, with mean μ_A and variance σ_A^2. The probability of 'success' for intervention *k* (where $k \neq A$) was calculated by adding the predictive distribution of each relative treatment effect, $d.pred_k$, to the predictive distribution of the baseline probability of 'success' (i.e. probability of safety equipment uptake), *mA*.

With respect to the synthesis modelling with covariate information (models 3.9 and 3.10 in Chapter 3):

$$\text{logit}(\text{p.treat}_{0A}) = \text{logit}(\text{p.treat}_{1A}) = mA$$

$$\begin{cases} \text{logit}(\text{p.treat}_{0k}) = mA + d.pred_k \\ \text{logit}(\text{p.treat}_{1k}) = mA + d.pred_k + \beta_0 + \beta.pred_k \end{cases} \quad \text{for k} \neq \text{A} \qquad (4.2)$$

$$mA \sim N(\mu_A, \sigma_A^2) \qquad d.pred_k \sim N(d_{Ak}, \sigma^2)$$

$$\beta_0 \sim N(0, 10^6) \qquad \beta.pred_k \sim N(\beta_{bk}, \sigma_B^2)$$

k = 1 . . . , total number of interventions

Compared to 4.1, in extension 4.2 the probabilities of 'success' for each strategy *k* (where $k \neq A$) are calculated as before, but now differs between subgroups. The probability of 'success' in subgroup = 0 (e.g. subgroup of 2Ps), $p.treat_{0k}$, is compounded by the 'common comparator' baseline effect and the treatment effect predictive distribution. The probability of 'success' in subgroup = 1 (e.g. subgroup of

1Ps), $p.treat_{1k}$, is formed by the two previous components and augmented by the subgroup baseline score, β_0, and the predicted distribution of the covariate effect, $\beta.pred_k$.

The following sections summarises the overall population results – using extension 4.1 – and the subgroup results with imputation of covariate information – using extension 4.2.

Results for the synthesis of aggregate data without covariates

The results presented here relate to the MTC model which synthesises AD. The network of evidence, its structure, and the results for these models have been described in Chapter 3. Here the focus is on those estimates from the MTC that represent key input parameters in the decision model, that is, the absolute probability of 'success' of each intervention. These results are shown in Table 4.1. This table of results presents medians and 95% CrIs of the MCMC posterior samples. Results show that, except for usual care and interventions supplying only education, most interventions have a rate of success in excess of 80% (median estimate).

Table 4.1 – ORs and absolute probabilities of success estimates for each intervention result of fitting MTC model for AD without including covariates to the functioning smoke alarm outcome data.

Interpretation		**Random effects MTC of AD - network of 20 studies**	
		Median of MCMC posterior sample	95 per cent credible interval
	Parameter		
Odds ratios for intervention effects (*vs* usual care)	***(2) E***	0.876	0.34 to 2.234
	(3) E + FE	3.130	0.916 to 11.28
	(4) E + FE + HI	2.598	1.023 to 8.618
	(5) E + FE + F	2.646	0.792 to 9.057
	(6) E + HI	3.232	0.674 to 24.15
	(7) E + FE + F + HI	6.934	2.255 to 23.93
	Probability		
Absolute probability of success of interventions	***(1) UC***	0.695	0.647 to 0.74
	(2) E	0.671	0.207 to 0.942
	(3) E + FE	0.876	0.459 to 0.986
	(4) E + FE + HI	0.852	0.448 to 0.983
	(5) E + FE + F	0.859	0.4 to 0.982
	(6) E + HI	0.880	0.413 to 0.991
	(7) E + FE + F + HI	0.941	0.651 to 0.993

Note: (1) UC - usual care; (2) E - education; (3) E + FE - education plus low cost / free equipment; (4) E + FE + HI - education plus low cost / free equipment plus home inspection; (5) E + FE + F - education plus low cost / free equipment plus fitting; (6) E + HI - education plus home inspection; (7) E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

Results for the synthesis of aggregate data with covariates

The reader is here reminded of the policy rationale behind performing subgroup CEA. As intervention effects may change across population subgroups, different decisions may be performed for each of these subgroups with respect to the cost effectiveness of particular interventions. The rationale is to enable the derivation of subgroup specific estimates – estimates that work as vehicles to obtain subgroup cost effectiveness outcomes to support decision making.

This subsection revisits analyses by fitting alternative synthesis models which use covariate information. The first model (model (a)) includes as a covariate the number of parents in the household (i.e. one parent household (1P) *vs.* two-parent household (2P)); the second (model (b)) considers only their employment status as covariate (i.e. employed parents (2U) *vs.* at least one parent unemployed (1U)); and the third (model (c)) includes both variables. For each model, medians and 95% CrIs of the posterior samples are shown.

Table 4.2 presents OR estimates for each intervention and the absolute probabilities of 'success' derived using models (a) and (b). As in the unadjusted analysis, when considering subgroups, results indicate that *(7) E+FE+F+HI* is the strategy with the highest probability of 'success' in the uptake of 'functioning' smoke alarms in the household.

Table 4.2 Relative intervention effect estimates and absolute probabilities of success estimates for each intervention, result of fitting MTC models for AD including: (a) a covariate relating to the number of parents (i.e. 1P *vs.* 2P) in the household; and (b) a covariate parents' employment status (i.e. 2U *vs.* 1U), to the functioning smoke alarm outcome data.

Random effects MTC of AD including covariate information - network of 20 studies with imputation							
a) Including covariate relating to number of parents in the household				**b) Including covariate relating to parents employment status in the household**			
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	**Interpretation**		Median of MCMC posterior sample	95 per cent credible interval
	Function of parameter				***Function of parameter***		
Odds ratios for intervention effects (*vs* usual care)	***(2) E***	0.339	0.05 to 2.421	Odds ratios for intervention effects (*vs* usual care)	***(2) E***	0.268	0.021 to 2.692
	(3) E + FE	2.399	0.594 to 11.17		***(3) E + FE***	1.114	0.082 to 16.77
	(4) E + FE + HI	1.673	0.54 to 7.335		***(4) E + FE + HI***	0.849	0.163 to 5.46
	(5) E + FE + F	1.656	0.375 to 8.749		***(5) E + FE + F***	1.054	0.055 to 22.84
	(6) E + HI	2.305	0.37 to 19.86		***(6) E + HI***	1.595	0.121 to 34.53
	(7) E + FE + F + HI	4.765	0.828 to 24.25		***(7) E + FE + F + HI***	2.652	0.235 to 36.72
	Probability				***Probability***		
Absolute probability of success of interventions for two parent families (2P)	***(1) UC***	0.714	0.565 to 0.813	Absolute probability of success of interventions for families with employed parents (2E)	***(1) UC***	0.803	0.421 to 0.956
	(2) E	0.446	0.066 to 0.926		***(2) E***	0.525	0.036 to 0.955
	(3) E + FE	0.857	0.402 to 0.983		***(3) E + FE***	0.814	0.149 to 0.993
	(4) E + FE + HI	0.799	0.378 to 0.978		***(4) E + FE + HI***	0.768	0.192 to 0.984
	(5) E + FE + F	0.798	0.295 to 0.977		***(5) E + FE + F***	0.813	0.105 to 0.993
	(6) E + HI	0.848	0.354 to 0.989		***(6) E + HI***	0.861	0.199 to 0.997
	(7) E + FE + F + HI	0.923	0.509 to 0.992		***(7) E + FE + F + HI***	0.915	0.307 to 0.995
Absolute probability of success of interventions for single parent families (1P)	***(1) UC***	0.714	0.565 to 0.813	Absolute probability of success of interventions for families with at least one parent unemployed (1E)	***(1) UC***	0.803	0.421 to 0.956
	(2) E	0.925	0.256 to 0.998		***(2) E***	0.833	0.046 to 0.999
	(3) E + FE	0.981	0.432 to 1		***(3) E + FE***	0.954	0.17 to 1
	(4) E + FE + HI	0.982	0.455 to 1		***(4) E + FE + HI***	0.966	0.233 to 1
	(5) E + FE + F	0.973	0.274 to 1		***(5) E + FE + F***	0.917	0.081 to 0.999
	(6) E + HI	0.986	0.475 to 1		***(6) E + HI***	0.968	0.2 to 1
	(7) E + FE + F + HI	0.992	0.615 to 1		***(7) E + FE + F + HI***	0.979	0.232 to 1

Model (c) uses information on both covariates, enabling the estimation of covariate effects for four population subgroups: two employed parents (2EP), employed single parent families (1EP), two parent families with at least one parent unemployed (2UP), and unemployed single parent families (1UP). Results are shown in Table 4.3. Again,

results show that, in each of these subgroups, the probability of 'success' is higher for intervention *(7) E + FE + F + HI*.

Table 4.3 – Absolute probabilities of success estimates for each intervention result of fitting MTC model for AD including covariates 'single parent status' and 'parents' employment status' to the functioning smoke alarm outcome data.

Random effects MTC of AD including information for both covariates - network of 20 studies with imputation							
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	**Interpretation**		Median of MCMC posterior sample	95 per cent credible interval
	Probability				***Probability***		
Absolute probability of success of interventions for two employed parent families (2EP)	***(1) UC***	0.600	0.097 to 0.967	Absolute probability of success of interventions for families with two parents and at least one parent unemployed (2UP)	***(1) UC***	0.600	0.097 to 0.967
	(2) E	0.125	0.002 to 0.939		***(2) E***	0.905	0.004 to 1
	(3) E + FE	0.615	0.025 to 0.993		***(3) E + FE***	0.992	0.074 to 1
	(4) E + FE + HI	0.528	0.034 to 0.98		***(4) E + FE + HI***	0.994	0.074 to 1
	(5) E + FE + F	0.474	0.01 to 0.988		***(5) E + FE + F***	0.978	0.024 to 1
	(6) E + HI	0.717	0.038 to 0.997		***(6) E + HI***	0.994	0.085 to 1
	(7) E + FE + F + HI	0.753	0.024 to 0.995		***(7) E + FE + F + HI***	0.995	0.074 to 1
Absolute probability of success of interventions for employed single parent families (1EP)	***(1) UC***	0.600	0.097 to 0.967	Absolute probability of success of interventions for unemployed single parent families (1UP)	***(1) UC***	0.600	0.097 to 0.967
	(2) E	0.957	0.002 to 1		***(2) E***	0.678	0 to 1
	(3) E + FE	0.992	0.01 to 1		***(3) E + FE***	0.938	0 to 1
	(4) E + FE + HI	0.991	0.01 to 1		***(4) E + FE + HI***	0.957	0 to 1
	(5) E + FE + F	0.987	0.007 to 1		***(5) E + FE + F***	0.830	0 to 1
	(6) E + HI	0.996	0.024 to 1		***(6) E + HI***	0.966	0.001 to 1
	(7) E + FE + F + HI	0.997	0.018 to 1		***(7) E + FE + F + HI***	0.967	0 to 1

4.3.5.2 Evidence used in model part 1 – the intervention model

Table 4.4 lists and describes the model parameters used in the intervention model, and provides references for the sources from which information was extracted. Table 4.4 is a subsection of Table A6 presented in Appendix 6, which shows the complete list of parameters used in the analysis.

The first parameters listed are the absolute probabilities of 'success' of each intervention in increasing the uptake of 'functioning' smoke alarms in a household. The sources of evidence for this model parameter are the results from the MTC models shown in the previous section (4.3.5.1). Lack of evidence forces an assumption to be made on families' level of acceptance of the interventions. This assumption was based on the expert knowledge of one (thesis) external advisor (Denise Kendrick) and assumed that a 90% acceptance level, constant across alternatives. The baseline probability of a household owning a 'functioning smoke alarm' and the incidence of fires in a household where there is a functioning smoke alarm were derived from official governmental statistics.

Table 4.4 - List of model input parameters used within part 1 (intervention) of the decision model for functioning smoke alarms. Parameter descriptions and sources of evidence used to inform the parameter are shown.

Model input parameter	Parameter description	Source(s) of evidence informing the parameter
Parameter type: Probabilities		
p_MTCfunc	Absolute probability of a functioning smoke alarm specific to each intervention	From MTC, as described in section 4.3.5.1
p_accept	Probability accept intervention (assumed same for all interventions)	Assumption
pop_fsa	Probability a household having a functioning smoke alarm	Survey of English Housing 2004/5 (Government, 2006) - Table 5.2
Parameter type: Resource cost		
Interventions Costs		
c_hsi	Cost of home safety inspection based on cost of LA home care worker for 40 minutes of their time including travel	PSSRU 2008 (Curtis, 2008)
c_alarmg	Cost of smoke alarm giveaway (with ten-year sealed battery)	Personal communication Jane Zdanowska
c_educ	Cost of providing education programme per household accepting intervention - based on cost of home care worker for 20 minutes of their time including travel	DiGuiseppi *et al.* (1999) – updated to 2009 prices
c_fixed	Fixed cost of an intervention scheme (e.g. set-up, administration, etc). Composite value derived from cost analysis of DiGuiseppi *et al.* (1999)	
c_acc	Additional cost incurred for each household that accept intervention (composite value)	
c_install	Cost of having the smoke alarm installed	

Parameter type: Utility parameters per cycle		
u_pop	General background mean utilities for non-injured population	UK Population Norms (Kind et al., 1999)

A fixed cost for each intervention of approximately £55,000 was obtained from the cost analysis implemented by DiGuiseppi and colleagues (1999). This fixed cost involved a variety of items such as intervention set-up (e.g. pilot test, distribution, staff training and reminders) and administration (e.g. programme coordination, brochures and photocopying). The study by DiGiuseppi *et al.* (1999) also provided estimates of the cost of installing a fire alarm, providing education to parents and the cost of parents accepting the intervention. The Personal Social Services Research Unit (PSSRU) 2008 report (Curtis, 2008) was used to obtain an estimate (based on assumptions) of the cost of performing a home inspection. Personal communication with field trialists was used to obtain an estimate of the cost of a smoke alarm within a giveaway programme. Where possible, cost data was inflated to 2009 prices (Curtis, 2009).

A vector of the general population mean utility values for the (non-injured) population were obtained from Kind *et al.* (1999). This study provided UK population norms and uncertainty estimates by age group, which were used across the entire decision model.

4.3.5.3 Evidence used in model part 2 – the 5 year Markov structure

The sources of evidence used to inform part two of the decision model are provided below, according to the type of model parameter being informed.

Event rates

The Survey of English Housing (Government, 2006) was used to obtain information on smoke alarm battery testing, important in determining the repair rate and the decay level of the equipment. Official reports also provided data on the likelihood of there being no injuries or fatalities following a fire (Government, 2007). The likelihood of

children incurring a fire-related injury, disaggregated by type of injury, was obtained from personal communication with an NHS Burns service specialist (Ken Dunn). This estimate was informed by a study developed by the Manchester burns unit in collaboration with the International Burn Injury Database (available at www.ibidb.org). Fire and rescue services' probabilities of attending a household fire were obtained from official sources (Government, 2006), as well as the likelihood of intensive treatment being involved. All-cause mortality estimates for the UK population were extracted from the Office for National Statistics (ONS) website – see Table 4.5 for detailed estimates used.

Table 4.5 - All cause mortality.

Year	All cause probability of death	Year	All cause probability of death	Year	All cause probability of death
0	0.004881	34	0.000837	68	0.015637
1	0.000371	35	0.000906	69	0.017008
2	0.0002155	36	0.0009105	70	0.018694
3	0.0001635	37	0.000981	71	0.0206325
4	0.000129	38	0.0010745	72	0.0230315
5	0.0001165	39	0.0011525	73	0.0256835
6	0.000102	40	0.001245	74	0.028487
7	0.000089	41	0.001345	75	0.0318565
8	0.0001075	42	0.0014625	76	0.0357665
9	0.000098	43	0.001584	77	0.03985
10	0.0000965	44	0.001693	78	0.044449
11	0.0001045	45	0.001894	79	0.050377
12	0.0001095	46	0.0020245	80	0.055856
13	0.0001385	47	0.0022465	81	0.0625195
14	0.0001465	48	0.0024795	82	0.0697915
15	0.000189	49	0.002651	83	0.0778325
16	0.0002535	50	0.003029	84	0.087146
17	0.0003615	51	0.003282	85	0.096125
18	0.000401	52	0.003505	86	0.106745
19	0.0004225	53	0.003948	87	0.1128245
20	0.0004475	54	0.00434	88	0.125178
21	0.000441	55	0.004743	89	0.135927
22	0.000455	56	0.0051635	90	0.1540795
23	0.0004525	57	0.005472	91	0.1751405
24	0.0004785	58	0.0059335	92	0.191869
25	0.000488	59	0.006512	93	0.2098855
26	0.0005445	60	0.007054	94	0.225663
27	0.0005335	61	0.008026	95	0.248501
28	0.000574	62	0.008787	96	0.2689845
29	0.000598	63	0.009764	97	0.2897915
30	0.000643	64	0.0107225	98	0.308781
31	0.000664	65	0.011695	99	0.324315
32	0.0007345	66	0.0127985	100	0.3522955
33	0.000794	67	0.0140685		

Costs

Health care costs

Emergency ambulance and paramedic unit costs were obtained from official national estimates (Curtis, 2008), charged *per minute* and multiplied by the expected number of minutes these units take to arrive at the accident scene (only applied to severe

events, i.e. severe fire-related child injuries). Average costs (and standard errors) of fire-related events were obtained from personal communication with a NHS Burn service specialist in Manchester (Ken Dunn), based on a non-published study carried out by this burns unit on a patient level costing system. Deaths resulting from fires incurred a cost of around £185 (updated to 2009 prices), which include both coroners and autopsy fees (Ginnelly et al., 2005b). The cost of a yearly precautionary check-up of the safety equipment was put at an average of £62 (*2008/9 NHS Reference Costs Guidance*, 2010). Mean incurred NHS costs of disabilities *per year* was reported in the Long Term Health and Healthcare outcomes of Accidental Injury study (HALO) (Nicholl et al., 2009).

Out of pocket / private costs

The total cost of property damage caused by a fire was derived from the British Crime Survey 2002/3 (Government, 2004b). The cost of a battery for a smoke alarm for one year was obtained from web-based safety equipment providers.

Law Enforcement and Rescue Services Costs

The costs of law enforcement and rescue services were also taken into consideration. Police presence at the fire scene where severe injuries occurred was assigned the same cost as that assumed by Ginnelly *et al*. (2005b) updated to 2009 prices. Fire rescue services were costed at approximately £3,000, with the value obtained from official governmental statistics (Government, 2004a). Both of these are assumed fixed.

Utilities

Utility values were assigned to all the model states. Utility data (decrements) were drawn from Sanchez and colleagues (2008) to inform each of the non-fatal fire-related events (i.e. minor, moderate and severe injuries). Along with many other elements, this study assessed EQ-5D information collected prospectively from burn victims, categorizing them in terms of the severity / degree of burns. Evidence extracted from

this source was deemed fixed in the decision model. A reduction in an individual's quality of life following an event leading to disability was obtained from the HALO report (Nicholl et al., 2009). This study investigated long term health effects and health-related quality of life of patients who had sustained serious injuries from fire-related accidents. A sustained yearly mean reduction of 0.1 in EQ-5D score for patients suffering a permanent injury was estimated, from a population norm of 0.8 for patients monitored for up to 11 years after an accident. Using the method of moments and through a Beta distribution, this information was probabilistically modelled.

The following table (Table 4.6) compiles the information described above for part two of the decision model.

Table 4.6 – List of model input parameters used within part 2 (5 year Markov structure) of the decision model for functioning smoke alarms. Parameter descriptions and sources of evidence used to inform the parameter are shown.

Model input parameter	Parameter description	Source(s) of evidence informing the parameter
Parameter type: Probabilities		
p_checkup	Probability have a precautionary checkup following a fire	Fire Statistics 2007 (Government, 2007) - Table 8
p_fire.func	Probability of a fire where functioning smoke alarms present	Fire Statistics 2007 (Government, 2007) - Table 2.4
p_fire.nonfunc	Probability of a fire where non-functioning smoke alarms present	
p_fire.noSA	Probability of a fire where no smoke alarms present or unspecified	
p_fatalSA	Probability of a fatality following a fire where functioning smoke alarm present	
p_fatalnSA	Probability of a fatality following a fire where non-functioning or no smoke alarm	
p.1yrbattery	Probability own a smoke alarm with battery life of 1 year	Survey of English Housing 2004/5 (Government, 2006) - Table 5.3
p.test1yr	Probability test smoke alarm at least once a year	Survey of English Housing 2004/5 (Government, 2006) - Fig 5.1
p.testless1yr	Probability test smoke alarm less than once a year	

p_noinjury	Probability of incurring 'no injuries' following a house fire (given functioning smoke alarm/ non-functioning or no smoke alarm)	Survey of English Housing 2004/5 (Government, 2006) – Table 3.7; Fire statistics 2007 (Government, 2007) - Table 2.4
p_FRSattend	Probability of inside household fire being attended by the Fire and Rescue Service	Survey of English Housing 2004/5 (Government, 2006) - Table 3.4
p_ITU	The additional proportion of burn unit costs incurred in ITU	Assumption based on analysis in Hemington-Gorse *et al.* (2009)
p_minor	Probability a child aged 0-4 incurs a minor injury following a house fire	Ken Dunn (personal communication)
p_moderate	Probability a child aged 0-4 incurs a moderate injury following a house fire	
p_severe	Probability a children aged 0-4 incurs a severe injury following a house fire	
p_allcause	Probability of all cause mortality for a UK citizen from 0 to 100 years old (for use in each decision model cycle)	(ONS, 2010b)

Parameter type: Resource cost

Health Care Costs		
c_minPU	Cost per minute of a Paramedic Unit	PSSRU 2008 (Curtis, 2008) - updated to 2009 prices
c_minEA	Cost per minute of a Emergency Ambulance	
mn.minor	Mean cost (and standard error) of a minor injury	Ken Dunn (Personal communication)
mn.moderate	Mean cost (and standard error) of a moderate injury	
mn.severe	Mean cost (and standard error) of a severe injury	
c_fatal	Cost of a fatality following a household fire (updated to 2008/9 prices) – includes coroners and autopsy costs	Ginnelly *et al*. (2005b) - updated to 2009 prices
c_dispyr	Mean incurred NHS costs of disability per year	HALO study (Nicholl et al., 2009) and personal communication with Jon Nicholl
c_checkup	Cost of precautionary check-up of safety equipment	NHS reference costs 2008/9 (*2008/9 NHS Reference Costs Guidance*, 2010) - code VB11Z
Out of Pocket / Private Costs		
c_battery	Cost of smoke alarm 1 year battery to individual	www.safelincs.co.uk – 2009 price
c_property	Total cost of damage caused by the fire	British Crime Survey: Fires in the Home 2002/3 (Government, 2004b) - updated to 2009 prices
Law Enforcement and Rescue Services Costs		
c_police	Cost of police attending – assumed only to attend where severe injuries	Ginnelly *et al*. (2005b) - updated to 2009 prices
c_FRSresponse	Cost of Fire and rescue Service attending a fire	Economic Cost of Fire 2004 (Government, 2004a) - Table 3.6 - updated to 2009 prices

Parameter type: Utility parameters per cycle

u_min	Deficit in utilities for minor injury (DRG 460 + 459)	Sanchez *et al.* (2008)
u_mod	Deficit in utilities for moderate injury (DRG 458 + 457)	
u_sev	Deficit in utilities for severe injury (DRG 472)	

u_deficit	Deficit in utilities following a disability	HALO study (Nicholl et al., 2009) and personal communication with Jon Nicholl
u_pop	As in Table 4.4	

4.3.5.4 Evidence used in model part 3 – the lifetime Markov structure

In the final part of the decision model, few parameterizations were needed to reflect the lifetime Markov structure. The sources of evidence used to inform these parameters are the same as the ones described for parts 1 and 2 of the decision model and are referred to in Table 4.7.

Table 4.7 – List of model input parameters used within the decision model for functioning smoke alarms. Sources of evidence used to inform the parameter and parametric assumption used to model parameter uncertainty is also shown.

Model input parameter	Parameter description	Source(s) of evidence informing the parameter
Parameter type: Probabilities		
p_allcause	As in Table 4.5	
Parameter type: Resource cost		
c_dispyr	As in Table 4.5	
Parameter type: Utility parameters per cycle		
u_deficit	As in Table 4.5	
u_pop	As in Table 4.4	

4.3.6 Base case and scenario analysis

4.3.6.1 Base case analysis

A key element of any economic evaluation is the definition of the 'perspective' from which the analysis is considered. Economic evaluations of health interventions commonly take a health service perspective. Nonetheless, given the inherent

complexity of PH interventions, their costs and effects will often be borne outside of the health care system, due to its impact on the welfare of the whole of society, not just on the individuals or organisations directly involved. A wider perspective, such as public sector or societal, allows for the capture of the impact of the interventions across sectors, such as health or education.

The base case analysis of the current study follows the NICE PH reference case recommendations (NICE, 2009), from the public sector perspective. This includes health care, law enforcement and rescue services related costs.

Table 4.8 presents a summary of the base case characteristics used in the current evaluation, adapted from Table 6.1 in "Methods for the development of the NICE PH guidance" (NICE, 2009).

Table 4.8 - Summary of the base case

Element of assessment	Base case
Type of economic evaluation	Cost-effectiveness analysis
Perspective on costs	Public sector, including the NHS and PSS
Perspective on outcomes	All health effects on individuals
Evidence on outcomes	Simultaneous synthesis of evidence of multiple interventions
Measure of health effects	QALYs
Main source of data for measurement of health-related quality of life (HRQL)	Reported directly by patients (HALO report (Nicholl et al., 2009))
Source of preference data for valuation of changes in HRQL	Representative sample of the public (UK Population Norms (Kind et al., 1999))
Discount rate	An annual rate of 3.5% was used on both costs and health effects
Equity weighting	An additional QALY has the same weight, regardless of the characteristics of the individuals who gain the health benefit
Size of the cohort simulated	100,000
Time horizon	100 years - until population all dead in order to account for all outcomes

4.3.6.2 Further analyses

Four alternative analyses were conducted, consisting of: a quasi-societal viewpoint including not only health care and public sector costs but also out of pocket costs – not including indirect or intangible costs (*scenario 2*); and NHS and Personal Social Services (PSS) viewpoint including only health care related costs (*scenario 3*). In order to explore the impact of fixed intervention costs, an extra scenario was considered in which these were excluded (*scenario 4*). This latter scenario is a simplification of the base case scenario. The impact of varying the time horizon was also evaluated by implementing a scenario that considered only the short term effects (i.e. the initial five years) of the programmes (*scenario 1*). For further details on the types of costs that were included in each of the scenarios, please consult Table 4.9 below.

Table 4.9 – List of scenarios considered in current analysis (each one assessed as a specific scenario, includes base case) showing how these are interconnected and their inherent hierarchical structure.

Item	Base case (Public sector)	Scenario 1: Public sector viewpoint with short-term effects	Scenario 2: Quasi-societal viewpoint	Scenario 3: NHS and PSS viewpoint	Scenario 4: NHS and PSS viewpoint with no implementation costs
Interventions costs	☑	☑	☑	☑	☒
Health care costs	☑	☑	☑	☑	☑
Out of pocket / private costs	☒	☒	☑	☒	☒
Law Enforcement and Rescue Services Costs	☑	☑	☑	☒	☒
Time horizon of 100 years	☑	☒	☑	☑	☑

4.4 Cost effectiveness results

The cost effectiveness results of the decision model described in the previous section are presented in this section. This section is split into two subsections: section 4.4.1 where overall population results (no subgroups) are shown, and section 4.4.2 where subgroup cost effectiveness results are discussed. The latter considers results by family type (2P and 1P) and by family employment status (i.e. 2U and 1U).

The methods guidance for health technology appraisal (NICE, 2008) states that the £20,000 to £30,000 threshold values should be considered when evaluating the cost effectiveness of health technologies. Although these values are occasionally used throughout to support the interpretation of the results shown next, it is important to note that a predefined threshold does not exist outside of the health sector. Section 4.5 comes back to this subject.

4.4.1 Overall population results (no covariates)

The base case results are shown in Table 4.10. The mean change in benefits and costs relative to intervention (1) UC and for a cohort size of 10,000 individuals, as well as ICERs for each of the programmes, are presented. The estimated incremental changes, in both QALYs and costs, relative to (1) UC are found to be considerably small when considering a cohort size of 10,000 households. The estimated ICER for the strategy involving the delivery of education and free (or sponsored) equipment, *(3) E+FE*, is approximately £33,000/QALY gained. Only two interventions are on the efficiency frontier (*(3) E + FE* and *(7) E + FE + F + HI*), while the remaining 4 interventions are either dominated or extendedly dominated (having higher costs or ICERs than more effective interventions, respectively). Figure 4.2a is a graphic representation of the incremental expected outcomes (in the *x-axis* effect gains and in the *y-axis* cost differences).

Table 4.10 – Results of the base case scenario for all interventions (for when effectiveness AD were synthesised and used to populate the decision model)

Intervention	Mean Δ (relative to (1) UC) for cohort size of 10,000		ICER (£s)
	QALYs	Costs (£s)	
(1) UC	----	----	---
(2) E	0.253	10.25	Extended dominated
(3) E + FE	0.558	18.44	33,045
(4) E + FE + HI	0.551	39.96	Dominated
(5) E + FE + F	0.548	37.55	Dominated
(6) E + HI	0.330	31.91	Dominated
(7) E + FE + F + HI	0.591	59.15	1,244,477

Figure 4.2 – (a) Cost effectiveness plane and (b) acceptability curves for the functioning smoke alarms decision model.

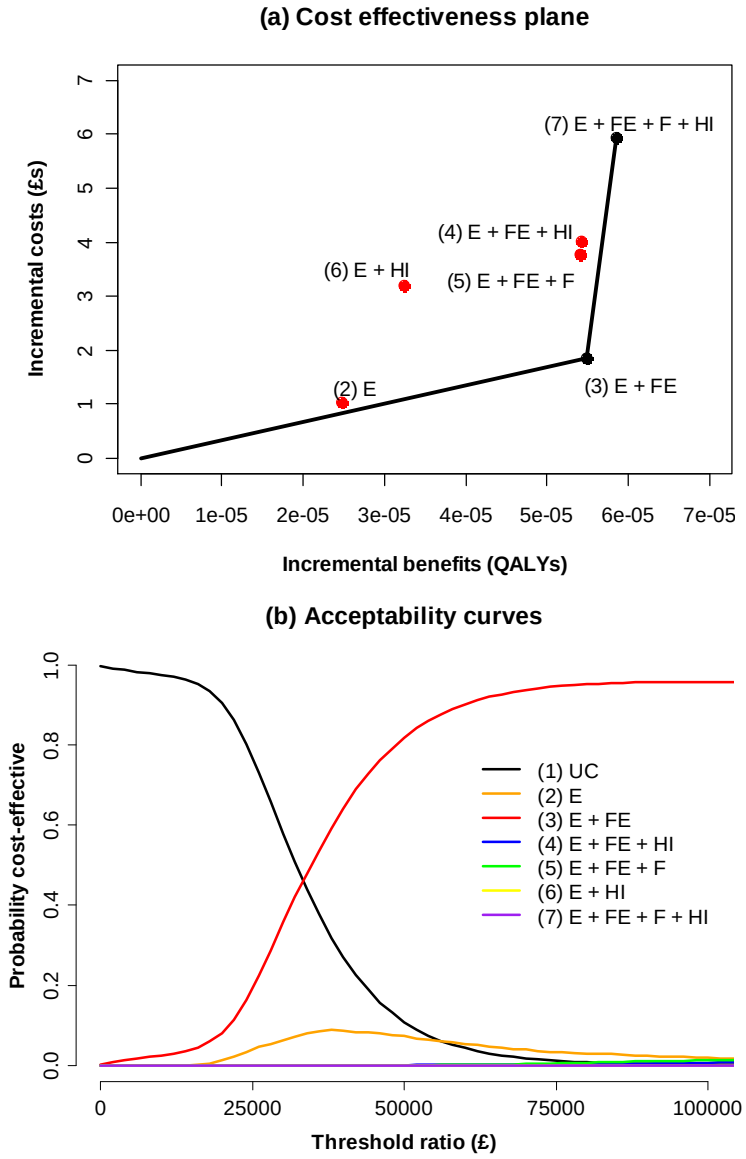


Figure 4.2b shows a graphical illustration of the probability of the alternative interventions being cost effective. It depicts the typical '*ogive*' shape of the Cost Effectiveness Acceptability Curves (CEACs), a characteristic observed when the joint density of costs and effects is contained in the *first quadrant* (positive incremental costs and effects). If threshold values of £20,000 and £30,000 are used, the usual care (*(1) UC*) intervention has the highest probability of being cost effective (0.906 and

0.579, respectively). Nonetheless, for threshold values of approximately £34,000 cost/QALY and above, programmes including education and free / low-cost equipment are considered most cost effective.

The intervention which yields the highest health benefits (most effective intervention) is the one with the "highest intensity", that is, intervention *(7)* which includes, education, free / low-cost equipment and its installation as well as home safety inspections. This result is consistent with what was found in Chapter 3 (section 3.5.2) and in section 4.3.5.1 of this chapter. Nevertheless, a decision maker would need to be willing to pay £1.2m per additional QALY (Table 4.10) to fund this programme.

As explained by Fenwick and colleagues (2001), it should be noted that an alternative with the highest probability of being cost effective for a particular threshold value may not be the one yielding the highest NBs. In Figure 4.3 the probabilities of being cost effective are shown for the interventions that attained the highest expected NB (Cost Effectiveness Acceptability Frontier (CEAF)).

Figure 4.3 – Cost effectiveness acceptability frontier for the functioning smoke alarms decision model.

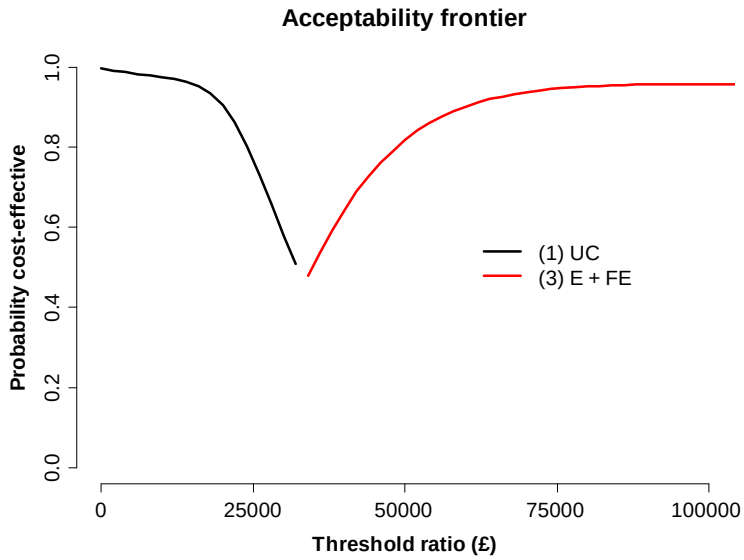


Figure 4.3 highlights that for threshold values between £0 and £100,000 threshold ratios, only two interventions should be considered, and that uncertainty is the highest (i.e. probability = 0.5) close to the ICER for intervention *(3) E + FE*. Beyond that, decision uncertainty falls with the increase in threshold values.

4.4.2 Subgroup cost effectiveness analysis results

It is important to assess the cost effectiveness of interventions by population subgroup, as an intervention may be cost effective for one subgroup of the population and not for another. Thus, there may be population health gains from stratifying decisions based on subgroup membership. The current subsection shows results from exploring: *(i)* the number of parents in the household (i.e. 1P or 2P) – section 4.4.2.1; *(ii)* parent(s)' employment status (i.e. 2U or 1U) – section 4.4.2.2; and *(iii)* both these factors – section 4.4.2.3.

4.4.2.1 Subgroups analyses: number of parents in the household (two vs. single)

For both subgroups, decisions to adopt or reject interventions are found to be identical to the ones made above for the overall population (results not shown). Interestingly, for 1Ps, the programme including only education (i.e. *(2) E*) achieves a considerable probability of being cost effective (i.e. of around 0.2) at approximately £30,000 per QALY gained. This implies higher decision uncertainty at around this threshold value.

4.4.2.2 Subgroups analyses: household employment status (employed vs. at least one parent unemployed)

When considering the subgroup of 2U, for a £30,000 threshold ratio, the probability of being cost effective is higher for *(1) UC* (of approximately 0.67). For the other subgroup, not only interventions *(3) E + FE* and *(7) E + FE + F + HI* are dominant over other alternatives, but also strategy *(4) E + FE + HI* is. Estimated ICERs for these two latter interventions are extremely large, in excess of £2m per QALY gained

(Table 4.11a). Again, although extended dominated by other alternatives, intervention *(2) E* has around 20% probability of being cost effective at approximately £32,000 threshold value – see Table 4.11b for further details.

Table 4.11 - ICERs and probabilistic sensitivity analysis results for each intervention for the base case scenario and for the subgroup of (a) 2Us and of (b) 1Us.

(a) Employed Parent' Households								
Interventions		**(1) UC**	**(2) E**	**(3) E + FE**	**(4) E + FE + HI**	**(5) E + FE + F**	**(6) E + HI**	**(7) E + FE + F + HI**
ICER (£s)		---	Extended dominated	35,965	Dominated	Dominated	Dominated	786,894
Probability \ Cost effectiveness threshold value (£)	20,000	0.914	0.011	0.075	0.000	0.000	0.000	0.000
	30,000	0.674	0.038	0.288	0.000	0.000	0.000	0.000
	50,000	0.185	0.054	0.758	0.001	0.002	0.000	0.000
(b) At Least One Unemployed Parent' Households								
Interventions		**(1) UC**	**(2) E**	**(3) E + FE**	**(4) E + FE + HI**	**(5) E + FE + F**	**(6) E + HI**	**(7) E + FE + F + HI**
ICER (£s)		---	Extended dominated	32,149	2,048,905	Dominated	Dominated	2,628,186
Probability \ Cost effectiveness threshold value (£)	20,000	0.861	0.048	0.091	0.000	0.000	0.000	0.000
	30,000	0.498	0.162	0.340	0.000	0.000	0.000	0.000
	50,000	0.104	0.108	0.786	0.001	0.000	0.000	0.000

4.4.2.3 Subgroups analyses: number of parents in the household and their employment status

Results shown in this subsection (Table 4.12) reveal that, at around £30,000 per QALY gained, *(1) UC* is the best strategy for all subgroups. If the threshold is £35,000 per QALY gained, then *(3) E + FE* becomes the best strategy for subgroups labelled in Table 4.12 as (b) and (c), and marginally (d). Usual care (*(1) UC*) remains

the best for subgroup labelled as (a). All 'active' interventions have a low probability of being cost effective at £30,000 per additional QALY.

Table 4.12 - Cost effectiveness results for the functioning smoke alarms decision model for four subgroups (a) two employed parent household (2EP); (b) employed single parent household (1EP); (c) two parent household with at least one unemployed (2UP); and (d) unemployed single parent household (1UP).

Interventions		(1) UC	(2) E	(3) E + FE	(4) E + FE + HI	(5) E + FE + F	(6) E + HI	(7) E + FE + F + HI
(a) Two Employed Parent' Households								
ICER (£s)		---	Extended dominated	40,721	Dominated	Dominated	Dominated	1,027,227
Probability \ Cost effectiveness threshold value	20,000	0.933	0.005	0.062	0.000	0.000	0.000	0.000
	30,000	0.753	0.019	0.228	0.000	0.000	0.000	0.000
	50,000	0.353	0.023	0.622	0.001	0.001	0.000	0.000
(b) Employed Single Parent Household								
ICER (£s)		---	Extended dominated	32,055	3,989,355	Dominated	Dominated	5,747,717
Probability \ Cost effectiveness threshold value	20,000	0.858	0.051	0.091	0.000	0.000	0.000	0.000
	30,000	0.505	0.151	0.344	0.000	0.000	0.000	0.000
	50,000	0.124	0.073	0.802	0.000	0.001	0.000	0.000
(c) Two Parent Household with at Least One Unemployed								
ICER (£s)		---	Extended dominated	33,961	Dominated	Dominated	Dominated	1,849,602
Probability \ Cost effectiveness threshold value	20,000	0.856	0.065	0.079	0.000	0.000	0.000	0.000
	30,000	0.526	0.191	0.283	0.000	0.000	0.000	0.000
	50,000	0.180	0.101	0.717	0.001	0.002	0.000	0.000
(d) Unemployed Single Parent Household								
ICER (£s)		---	Extended dominated	35,013	1,642,801	Dominated	Dominated	3,227,412
Probability \ Cost effectiveness threshold value	20,000	0.874	0.050	0.076	0.000	0.000	0.000	0.000
	30,000	0.554	0.154	0.292	0.000	0.000	0.000	0.000
	50,000	0.199	0.115	0.680	0.003	0.004	0.000	0.000

4.4.3 Results of analysis under alternative scenarios

A series of alternative scenarios were run to contrast with the base case results, as defined in section 4.3.6.2. Following a full incremental analysis, estimated ICERs for each of the scenarios are shown in Table 4.13. It can be observed that results for the scenarios considering NHS and PSS viewpoint are identical to the ones obtained for the reference case where no subgroups are considered, that is, estimated ICER for intervention *(3) E + FE* is approximately £33,000 per QALY gained. This result indicates that, if the NICE threshold acceptance range recommendations were to be used, and decisions were to be made by a health sector decision maker, none of the 'active' alternatives would be funded.

Table 4.13 – Cost effectiveness results of the 4 scenarios for all interventions and all participant households and for when AD on effectiveness was synthesised and used to populate the decision model, all the rest remaining constant.

Intervention	ICER (£s) - Full incremental analysis			
	Scenario 1: Public sector viewpoint considering short-term effects (5 years simulation process)	**Scenario 2: Quasi-societal viewpoint**	**Scenario 3: NHS and PSS viewpoint**	**Scenario 4: NHS and PSS viewpoint, excuding implementation costs**
(1) UC	---	---	---	Dominated
(2) E	Extended dominated	Dominated	Extended dominated	---
(3) E + FE	70,020	659,482	32,752	1,255
(4) E + FE + HI	Dominated	Dominated	Dominated	Extended dominated
(5) E + FE + F	Dominated	Dominated	Dominated	789
(6) E + HI	Dominated	Dominated	Dominated	Extended dominated
(7) E + FE + F + HI	2,380,430	1,037,609	1,254,939	1,480

Adopting the NHS and PSS perspective but excluding the upfront cost of implementing the interventions (i.e. assuming that these costs are assured by another sector or sectors of society), implies many changes in results when compared to other alternative scenarios. The set of dominated or extended dominated programmes is now completed by *(1) UC*, *(4) E + FE + HI* and *(6) E + HI*. The fact that *(1) UC* becomes

dominated implies that intervention *(2) E* is the comparison starting point, as this programme shows both lower costs and benefits across non-dominated alternatives. The magnitudes of the estimated ICERs are found to be low, not going over £2,000 per QALY gained. The programme with the lowest estimated ICER was the one which included education, supply of free or low cost equipment and its installation (i.e. *(5) E + FE + F*).

4.5 Discussion

Summary of findings

Assessing the effectiveness of alternative strategies is important in a health care system operating under fixed budget constraints, where decisions on the use of the technologies must be based on cost effectiveness. This study evaluated the cost effectiveness of alternative interventions to increase the household uptake of 'functional' smoke alarms and, consequently, reduce the number of home fire-related injuries in pre-school children.

The results of the analyses of the evidence of effectiveness presented in Chapter 3 indicate that more complex interventions (which include multiple components such as education, equipment and its fitting and inspection) have higher probability of increasing the possession of functioning smoke alarms than those less multifaceted. In this chapter it is shown that these are associated with higher costs and in order for them to be adopted, decision makers need to be 'willing to pay' or displace large amounts of funds. Strategies which provide education and free or sponsored equipment to families can be adopted at a lower willingness to pay value. Results are consistent with previous trial-based model results (Ginnelly et al., 2005b), and marginally similar to the decision model results obtained in the NICE PH guidance on the prevention of unintentional injuries among under-15s in the home (NICE, 2010).

Study strengths, limitations and further work

This study considers a number of methodological improvements over other modelling approaches previously undertaken in this area (Pitt et al., 2009, Ginnelly et al., 2005b, Haddix et al., 2001). For instance, in the current study: (i) multiple interventions were compared in both effect and cost dimensions – which, to the author's knowledge, is the first time that this has been done within a PH study; (ii) different scenarios with respect to the perspective of the analysis were implemented; and (iii) subgroup analysis, that, as discussed in the next paragraphs, considered two potential heterogeneous factors and explored the cost effectiveness for different population subsets.

Nevertheless, the framework presented carries some clear limitations and/or strong assumptions. These are briefly described in section 4.3.3. It is worth highlighting that, due to lack of evidence, the conclusions of the current study is limited to the population of UK homes with only one child under 5. The generalisation of the study findings to other populations, such as considering all UK homes (including the ones with multiple children), or even to all members of the household (i.e. to both parents and children), should not be performed without appropriately accounting for this in the bulk of the evidence used to populate the decision model.

Common to both effectiveness and economic assessments of injury prevention schemes is the lack of evidence regarding both their effects and costs. Generally, it is difficult to understand how best to design and deliver these interventions in order to efficiently increase home safety. Without this knowledge, policymakers are uninformed on how to reduce fire-related injuries and tackle potential inequalities in child injury rates (Dowswell & Towner, 2002).

Randomised controlled trials are usually highest in rank in the quality of evidence hierarchy, and are often considered the best design to evaluate the (not always straightforward) link between cause and effect. Nevertheless, these studies do carry certain limitations for the evaluation of PH interventions (Drummond et al., 2005). For

instance, RCTs are limited in their action, mainly owing to internal validity and ethical considerations. In fact, they tend to produce biased effect estimates (over-inflating) of the policy's efficacy (NICE, 2009). Other types of study designs, like observational studies, are also deemed to provide good quality evidence, conditional on how well potential sources of bias are taken into account.

In the assessment performed in this chapter, the evidence base on effectiveness was sparse, as many of the comparisons within the network contained only a small number of trials (Chapter 3). This led to high uncertainty in some estimates which, in turn, was propagated throughout the decision model, affecting the estimates of costs and QALYs. Additional studies may be required to augment the evidence base and reduce uncertainty over decisions on the cost effectiveness of alternative strategies in this area of PH. Moreover, there were numerous decision model input parameters for which limited data existed. This was addressed by ensuring that for all these parameters the source of evidence was reliable and preferably of official (governmental) origin.

The differences across interventions in terms of estimated mean costs and benefits were found to be considerably small. This fact may raise several issues, such as decision makers considering all strategies to be fairly equal, or even that the transaction costs of implementing more complex interventions not outweigh their benefits.

The complexity of PH interventions is also linked to the existence of heterogeneity within a targeted population. Variability in study populations, interventions and settings, and variations in study designs and outcomes, are all considered sources of heterogeneity. In the assessment of PH programmes, relevant sources of heterogeneity should be identified and explored explicitly (e.g. through subgroup analysis). Analysts and policy makers are often interested in examining variation among, for example, different social, ethnic, demographic and educational groups (Rychetnik et al., 2002). The current study has considered the analysis of subgroups and estimated subgroup specific cost effectiveness information. This evidence may allow decision makers to

consider the absolute magnitude of benefits and costs while taking into account possible heterogeneity within the population of interest.

The NICE PH methods guidance (2009) encourages the use of the cost per QALY as an economic outcome measure – as in Pitt *et al.* (2009) report, the current study used this framework. However, the guidance raises the point that the QALY measure may not be sufficient to capture the complex impact and context of some PH programmes, in particular programmes involving social support, education, and guidance to individuals. To address this other measures such as life years gained, cases averted or a more disease-specific outcome, are recommended in alternative..

The usual NICE ICERs acceptance region recommendations do not exist outside the health sector, making it difficult to judge whether the benefits accruing to the non-health sectors are cost effective (NICE, 2009). Therefore, decisions on whether to recommend interventions should not be based on cost effectiveness alone but also on equity. Generally, the relevant economic literature discussed in section 4.2 used a £0 – £50,000 or a £20,000 – £30,000 threshold range to evaluate programmes' cost effectiveness. In the analysis presented in this chapter, although the interpretation of results was supported by the same general ranges, these were made taking this fact into account.

Parents are usually the target population of the preventative upstream interventions assessed here. It is the parents' responsibility to supply a safe environment and safeguard the health of their children, in light of their vulnerability and lack of risk awareness in early years. Interventions aim to change parents' awareness, attitudes and behaviours across various safety issues. One structural limitation of the model developed in this study is that it ignores potential spill-over effects towards parents, who could also benefit from installing and maintaining 'functioning' safety equipment in their homes. Another limitation is the fact that other possible model states may exist in real life that are not considered by the current model. For example, the model could have examined varying degrees of disability or even a possible return to the 'well' state after a severe injury following a fire-related injury in the household.

Concluding remarks

In this chapter important findings were made about the cost effectiveness of interventions in promoting the uptake of functional smoke alarms and consequently, in reducing child injuries at home. However, there continues to be insufficient evidence to inform and support PH policy/decision making. This state of affairs can be changed, but it will require strong direction to ensure the priorities for economic evaluation evidence become organised and coordinated at local, regional and national levels.

Most of the evidence used to inform model parameters was available at summary level. It would be useful to understand the extent to which the availability of individual level evidence for at least some of the model parameters, would impact on final outcomes- particularly, when subgroup analysis is at stake. This issue will be explored in Chapter 5.

CHAPTER 5

5. USING AGGREGATE- AND INDIVIDUAL-PARTICIPANT LEVEL DATA FOR COST EFFECTIVENESS MODELLING

5.1 Background

The literature review carried out in Chapter 2 acknowledged that it is not uncommon to have access to IPD, possibly alongside study AD, to inform cost effectiveness decision analytic models. However, in this situation, methods of analyses were lacking. Motivated by the need to consider all 'relevant' evidence, Chapter 3 developed new synthesis methodology for when IPD, or when both AD and IPD, are available.

One of the recommendations made in Chapter 3 was to include, whenever possible, evidence at the individual level in the MTC analysis. Using a real life example, Chapter 3 illustrated that the inclusion of IPD enables the estimation of intervention effects at the subgroup level with greater accuracy and precision, compared to the use of summary evidence only. Ignoring available IPD may, therefore, compromise the validity of estimates, leading to erroneous interpretations of the underlying *treatment x covariate* associations and, consequently, of the true subgroup effects.

Using the same case study (presented in Chapter 1), this chapter aims to expand on Chapter 3 in evaluating the impact of using IPD, from relative effectiveness to cost effectiveness. The chapter begins by providing the framework of the analysis in Section 5.2, followed by a discussion in section 5.3, of the effectiveness and cost effectiveness results from using AD + IPD in the absence of subgroups. Section 5.4 extends this analysis and presents results when considering mutually exclusive subgroups. Finally, section 5.5 discusses the chapter's findings and how they fit within the current methods literature, highlighting limitations and scope for further work.

5.2 Framework of analyses

The CEA presented in Chapter 4 evaluated the use of alternative strategies that promoted the uptake of 'functioning' smoke alarm equipment in order to reduce child fire-related injuries in the home. Evidence on a multitude of parameters was collated in a decision model aiming at describing the short, medium and long term effects of preventing the consequences of fires in households. In the analyses presented in Chapter 4, effectiveness evidence for the alternative interventions was derived using AD from the existing evidence base. Here, the same evidence base on relative effectiveness is used but now considers IPD where available (given that IPD was made available for a proportion of all studies, the 'IPD model' contained a mixture of AD + IPD).

Use of effectiveness evidence generated using AD + IPD

The specific methods used in the synthesis (of AD and of IPD) were presented in detail in Chapter 3. The reader is reminded that relative effectiveness evidence was further transformed for inclusion in the decision model. In particular, extensions to the synthesis modelling were made to consider the use of predictive treatment effect distributions (Ades, A. E. et al., 2005) and the modelling of baseline effects (Pocock et al., 2002). Details on these extensions are provided in section 4.3.5.1 of Chapter 4,

where although they are presented within an AD context, these can be easily generalised to the use of IPD context. Methods will not be further reported here, though results are shown for each intervention, the probabilities of household uptake of a 'functioning' smoke alarm, estimated through the use of AD and AD + IPD. Information is provided through medians and 95% CrIs of the MCMC posterior samples.

Cost effectiveness

To facilitate comparisons between the effect of AD and AD + IPD in cost effectiveness, input data and model structure for parameters other than effectiveness remain unchanged (reported in Chapter 4). Any differences in cost effectiveness results are, therefore, a direct consequence of differences between the effectiveness input information. To evaluate the impact on cost effectiveness of using IPD rather than AD, ICERs and probabilities of being cost effective will be compared.

On the one hand, if effectiveness estimates are similar whether using AD or AD + IPD, when informing the decision model, no effect on cost effectiveness estimates is expected. If effectiveness estimates are not similar whether using AD or AD + IPD, it is difficult to predict the effects of those differences on cost effectiveness outcomes, which will depend mainly on the reasons behind those differences and their direction. A gain in precision[24] from using IPD may reduce decision uncertainty (e.g. when two interventions are at stake, one will have a higher probability of being cost effective and the other a lower probability). In contrast, it is not straightforward to infer the effects of gains in accuracy (i.e. bias[25] reduction) from IPD compared to AD on cost effectiveness.

---

[24] In general statistical terms, precision is defined to be the reciprocal of the variance, i.e. precision = 1 / variance.

[25] By bias it is meant that an inaccurate estimation of the association of interest is obtained. In the context above, bias = $\left|E\left[\hat{\theta}\right]-\theta\right| > 0$.

Subgroup effectiveness and cost effectiveness

Chapter 3 illustrated that when no covariates are taken into account, effectiveness results from using AD and AD + IPD should be equivalent, given the outcome assessed is binary[26]. However, AD and study level covariate information were shown to provide an incomplete and/or inaccurate understanding of the true nature of the association between the intervention effect and particular participant characteristics, generating imprecise and/or biased estimates. Ecological fallacy bias (Berlin et al., 2002, Lambert et al., 2002, Piantadosi et al., 1988) and/or confounding (Bland, 2000) are the main reasons for this inaccuracy, as discussed in Chapter 3.

The above mentioned issues mean that, when subgroup CEA is performed, the use of IPD, rather than AD, may affect cost effectiveness. The impact of using AD may be reflected, for instance, through obtaining imprecise and/or biased cost effectiveness estimates, resulting in wrong recommendations and/or decisions being made for all or specific groups of the population. By improving precision and/or reducing bias, IPD may facilitate approval/rejection decisions and, consequently, allow for efficiency increments in the maximisation of health gains.

As in previous chapters, the case study is used to explore the following alternative subgroup scenarios: *(i)* the number of parents in the household (i.e. 'single' or 'two parent' family); *(ii)* parent(s)' employment status (i.e. 2U or 1U); and *(iii)* both of these factors.

The following sections discuss the effectiveness and cost effectiveness results from using information for the mixture of AD + IPD, and compare these to using AD alone. Results are initially shown for the overall population, followed by results of the analyses considering population subgroups.

[26] As discussed in Chapter 2 and 3, summary data of binary outcomes are considered 'sufficient statistics', and there is no loss of information when using them (Fisher, 1922).

5.3 Results in the absence of subgroups

5.3.1 Effectiveness

Results in Table 5.1 show the probability of 'functioning' smoke alarm uptake for each intervention, when no covariates are considered. As expected, results are very similar between the AD and AD + IPD models. Any small differences may be a consequence of the MCMC iterative procedure (i.e. simulation error) or a consequence of different approaches being used with respect to cluster adjustment (i.e. ad-hoc in the AD case and built-in in the IPD case).

Intervention *(7) E + FE +F + HI* has the highest probability of uptake of ('functioning') smoke alarms. Except for interventions *(1) UC* and *(2) E*, both models estimate probabilities of 'functioning' smoke alarm uptake in excess of 80%.

Table 5.1 - Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) for each intervention result of fitting different MTC models (AD and AD plus IPD) without including covariates, to the functioning smoke alarm outcome data.

Interpretation		Random effects MTC of AD - network of 20 studies		Random effects MTC of AD and IPD - network of 20 studies*	
		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Probability				
Absolute probability of success of interventions **	***(1) UC***	0.695	0.647 to 0.74	0.686	0.412 to 0.869
	(2) E	0.671	0.207 to 0.942	0.655	0.155 to 0.951
	(3) E + FE	0.876	0.459 to 0.986	0.868	0.382 to 0.987
	(4) E + FE + HI	0.852	0.448 to 0.983	0.849	0.382 to 0.986
	(5) E + FE + F	0.859	0.4 to 0.982	0.853	0.327 to 0.985
	(6) E + HI	0.880	0.413 to 0.991	0.876	0.347 to 0.993
	(7) E + FE + F + HI	0.941	0.651 to 0.993	0.938	0.575 to 0.994

Notes:

* Nine of the 20 studies had individual level data available

** (1) UC - usual care; (2) E - education; (3) E + FE - education plus low cost / free equipment; (4) E + FE + HI - education plus low cost / free equipment plus home inspection; (5) E + FE + F - education plus low cost / free equipment plus fitting; (6) E + HI - education plus home inspection; (7) E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

5.3.2 Cost effectiveness

This section discusses the impact of using AD + IPD effectiveness evidence to populate the decision model, compared to using AD only. Table 5.2 depicts the results of the two alternative analyses, by reporting expected benefits and costs, ICERs and the probability of each treatment being cost effective. Given that relative effectiveness does not differ between these two analyses (section 5.3.1), cost effectiveness results are not affected. In both scenarios, intervention *(1) UC* has the highest probability of being cost effective with approximately 0.9 and 0.6 at £20,000 and £30,000 threshold values, respectively. At a threshold value of approximately £35,000 and above per QALY gained, programmes including education and free / low-cost equipment (*(3) E + FE*) are most cost effective for the population. Despite the high effectiveness of intervention *7*, its costs preclude recommendation at conventional threshold levels.

Table 5.2 – Cost effectiveness results for all interventions and for when AD and AD plus IPD on effectiveness was synthesised and used to populate the decision model. Both analyses use base case characteristics and, in both, evidence informing all other economic model parameters remained the same.

AD effectiveness evidence used				**Cost effectiveness threshold at**					
				£20,000 per QALY		**£30,000 per QALY**		**£50,000 per QALY**	
Intervention	**Expected Costs**	**Expected QALYs**	**ICER** (£/QALY)	**NMB** (£)	**Probability cost effective**	**NMB** (£)	**Probability cost effective**	**NMB** (£)	**Probability cost effective**
(1) UC	981,747	54.46134	---	107,480	0.906	652,093	0.579	1,741,320	0.108
(2) E	981,748	54.46137	Extended dominated	107,479	0.014	652,093	0.062	1,741,320	0.074
(3) E + FE	981,749	54.46140	33,045	107,479	0.080	652,093	0.359	1,741,321	0.818
(4) E + FE + HI	981,751	54.46140	Dominated	107,477	0.000	652,091	0.000	1,741,319	0.000
(5) E + FE + F	981,751	54.46140	Dominated	107,477	0.000	652,091	0.000	1,741,319	0.000
(6) E + HI	981,750	54.46138	Dominated	107,477	0.000	652,091	0.000	1,741,318	0.000
(7) E + FE + F + HI	981,753	54.46140	1,244,477	107,475	0.000	652,089	0.000	1,741,317	0.000
AD + IPD effectiveness evidence used									
(1) UC	972,862	54.46162	---	125,691	0.914	670,291	0.605	1,759,490	0.114
(2) E	972,863	54.46164	Extended dominated	125,691	0.011	670,290	0.060	1,759,490	0.074
(3) E + FE	972,864	54.46167	33,752	125,690	0.076	670,291	0.335	1,759,491	0.811
(4) E + FE + HI	972,866	54.46167	Dominated	125,688	0.000	670,288	0.000	1,759,489	0.000
(5) E + FE + F	972,866	54.46167	Dominated	125,689	0.000	670,289	0.000	1,759,489	0.001
(6) E + HI	972,865	54.46165	Dominated	125,689	0.000	670,289	0.000	1,759,488	0.000
(7) E + FE + F + HI	972,868	54.46167	1,107,554	125,686	0.000	670,287	0.000	1,759,487	0.000

5.4 Results in the presence of mutually exclusive subgroups

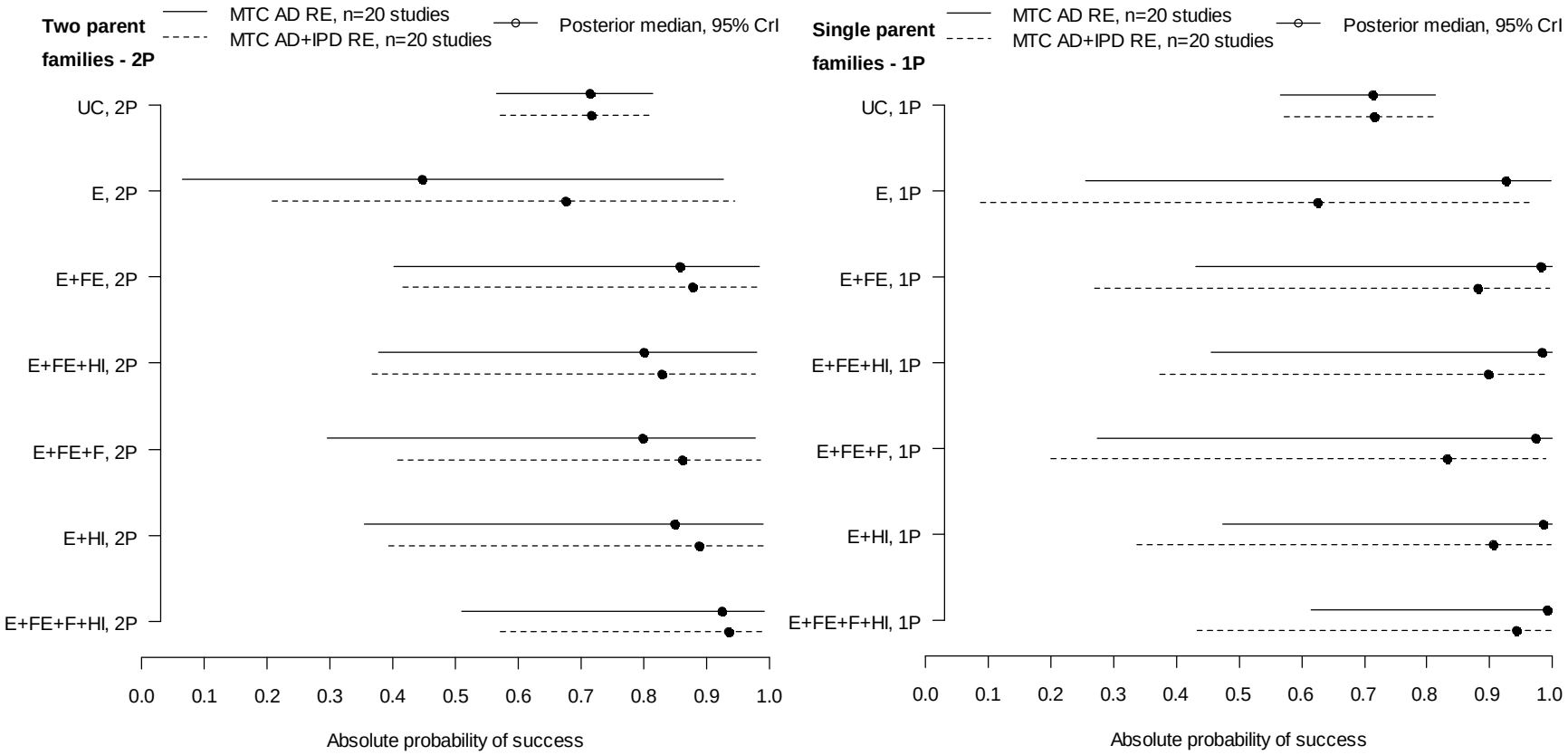
5.4.1 Subgroup analyses: number of parents in the household (two vs single)

5.4.1.1 Effectiveness

In this subsection the probabilities of smoke alarm uptake estimates for each intervention are presented for instances when a binary covariate on the number of parents in the household is included in the modelling. Results are depicted graphically in Figure 5.1.

In general, point estimates are lower when using AD within the 2P subgroup, compared to using AD + IPD. The reverse is observed in the 1P subgroup. The use of IPD is reflected by having narrower 95% CrI in the 2P subgroup, when compared to the use of AD only. In the 1P subgroup the gain in accuracy when using IPD is more evident, with large shifts in the point estimates when compared to using AD only. The largest point estimate difference between these two models is found in the 1P subgroup and for intervention *(2) E*: when using AD, the 'functioning' smoke alarm uptake probability is estimated to be approximately 0.93; when using AD + IPD it is estimated to be 0.63.

Figure 5.1 – Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) for two parent families (2P) and single parent families (1P) from fitting the MTC model for AD (i.e. MTC AD RE) and for AD plus IPD (i.e. MTC AD+IPD RE).



Note: (1) UC - usual care; (2) E - education; (3) E + FE - education plus low cost / free equipment; (4) E + FE + HI - education plus low cost / free equipment plus home inspection; (5) E + FE + F - education plus low cost / free equipment plus fitting; (6) E + HI - education plus home inspection; (7) E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

5.4.1.2 Cost effectiveness

Results in Table 5.3 relate to the subgroups analysis of two and single parent families. For the 2P subgroup, decisions to adopt or reject interventions are similar using AD and AD + IPD, which, in turn, are similar to those discussed above for the overall population. With respect to 1Ps, results are slightly different when effectiveness AD and AD + IPD are used. If AD is used, the intervention including only education (i.e. *(2) E*) achieves approximately 20% probability of being cost effective at approximately £30,000 per additional QALY, while combining AD + IPD, this probability is no more than 8% – the reader is reminded that the largest difference between effectiveness point estimates (i.e. smoke alarm uptake probabilities) from AD and AD + IPD was found for this intervention (see Figure 5.1). Despite this, in both situations, interventions *(2)*, *(4)*, *(5)* and *(6)* are dominated or extended dominated by other programmes and approval decisions are not altered between AD and AD + IPD if a £30,000 threshold value is used. However, when using AD, approval decisions are altered if the decision maker is willing to 'pay' (or displace) approximately £32,000 per additional QALY (i.e. intervention *(3) E + FE* is cost effective). This is indicative of higher decision uncertainty around this particular threshold value, which is not observed when IPD is used.

Table 5.3 - Cost effectiveness results for all interventions, for the 2P and 1P subgroups and for when AD and AD plus IPD effectiveness evidence was synthesised and used to populate the decision model.

Two Parent Households								
Interventions		**(1) UC**	**(2) E**	**(3) E + FE**	**(4) E + FE + HI**	**(5) E + FE + F**	**(6) E + HI**	**(7) E + FE + F + HI**
AD effectiveness evidence used								
ICER (£s)		---	Extended dominated	33,806	Dominated	Dominated	Dominated	1,443,376
Probability being CE / λ*	20,000	0.920	0.004	0.076	0.000	0.000	0.000	0.000
	30,000	0.629	0.031	0.339	0.000	0.000	0.000	0.000
	50,000	0.140	0.034	0.825	0.000	0.000	0.000	0.000
AD + IPD effectiveness evidence used								
ICER (£s)		---	Extended dominated	33,078	Dominated	Dominated	Dominated	1,414,877
Probability being CE / λ*	20,000	0.897	0.011	0.092	0.000	0.000	0.000	0.000
	30,000	0.592	0.064	0.344	0.000	0.000	0.000	0.000
	50,000	0.110	0.075	0.813	0.000	0.001	0.000	0.000
Single Parent Households								
Interventions		**(1) UC**	**(2) E**	**(3) E + FE**	**(4) E + FE + HI**	**(5) E + FE + F**	**(6) E + HI**	**(7) E + FE + F + HI**
AD effectiveness evidence used								
ICER (£s)		---	Extended dominated	30,877	Extended dominated	Dominated	Dominated	2,507,302
Probability being CE / λ*	20,000	0.843	0.066	0.091	0.000	0.000	0.000	0.000
	30,000	0.455	0.193	0.352	0.000	0.000	0.000	0.000
	50,000	0.066	0.116	0.817	0.001	0.000	0.000	0.000
AD + IPD effectiveness evidence used								
ICER (£s)		---	Extended dominated	33,583	Extended dominated	Dominated	Dominated	1,272,082
Probability being CE / λ*	20,000	0.903	0.013	0.084	0.000	0.000	0.000	0.000
	30,000	0.591	0.077	0.332	0.000	0.000	0.000	0.000
	50,000	0.129	0.089	0.779	0.001	0.002	0.000	0.000

Notes:

* where λ is the threshold ratio

5.4.2 Subgroup analyses: household employment status (employed vs at least one parent unemployed)

5.4.2.1 Effectiveness

Absolute probabilities of 'functioning' smoke alarms uptake for each subgroup relating to household employment status are shown in Table 5.4.

Table 5.4 – Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) from fitting the MTC model for AD and for AD plus IPD including covariate 'single parent status' to the functioning smoke alarm outcome data.

		Random effects MTC of AD - network of 20 studies		**Random effects MTC of AD and IPD - network of 20 studies***	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Probability				
Absolute probability of success of interventions for families with employed parents (2E) **	***(1) UC***	0.803	0.421 to 0.956	0.802	0.405 to 0.945
	(2) E	0.525	0.036 to 0.955	0.759	0.193 to 0.977
	(3) E + FE	0.814	0.149 to 0.993	0.910	0.405 to 0.993
	(4) E + FE + HI	0.768	0.192 to 0.984	0.890	0.355 to 0.991
	(5) E + FE + F	0.813	0.105 to 0.993	0.902	0.402 to 0.993
	(6) E + HI	0.861	0.199 to 0.997	0.926	0.429 to 0.996
	(7) E + FE + F + HI	0.915	0.307 to 0.995	0.961	0.627 to 0.997
Absolute probability of success of interventions for families with at least one parent unemployed (1E) **	***(1) UC***	0.803	0.421 to 0.956	0.802	0.405 to 0.945
	(2) E	0.833	0.046 to 0.999	0.718	0.139 to 0.977
	(3) E + FE	0.954	0.17 to 1	0.895	0.34 to 0.994
	(4) E + FE + HI	0.966	0.233 to 1	0.873	0.265 to 0.992
	(5) E + FE + F	0.917	0.081 to 0.999	0.874	0.275 to 0.992
	(6) E + HI	0.968	0.2 to 1	0.910	0.329 to 0.996
	(7) E + FE + F + HI	0.979	0.232 to 1	0.952	0.505 to 0.997

Notes:

* Nine of the 20 studies had individual level data available

** (1) UC - usual care; (2) E - education; (3) E + FE - education plus low cost / free equipment; (4) E + FE + HI - education plus low cost / free equipment plus home inspection; (5) E + FE + F - education plus low cost / free equipment plus fitting; (6) E + HI - education plus home inspection; (7) E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

The use of IPD in the MTC modelling implies a gain in precision, translated as having a narrower 95% CrI of the uptake probability posterior samples, when compared to the use of AD. This is verified for both subgroups. The largest point estimate difference from using AD and AD + IPD is found for intervention *(2) E* – this is true for both subgroups. Generally, differences between estimates of AD and AD + IPD are larger in 2Us, than in the 1Us.

5.4.2.2 Cost effectiveness

Table 5.5 shows the subgroup cost effectiveness results for ‘employed’ and ‘at least one unemployed parent’ subgroups. The use of AD leads to more marked difference between subgroups, whereas the use of IPD brings some consistency to the results in the sense that expected ICERs and the probabilities of being cost effective are similar for both subgroups for each intervention.

Table 5.5 – Cost effectiveness results for all interventions, for the 2U and 1P subgroups and for when AD and AD plus IPD effectiveness evidence was used.

Employed Parent' Households								
Interventions		**(1) UC**	**(2) E**	**(3) E + FE**	**(4) E + FE + HI**	**(5) E + FE + F**	**(6) E + HI**	**(7) E + FE + F + HI**
AD effectiveness evidence used								
ICER (£s)		---	Extended dominated	35,965	Dominated	Dominated	Dominated	786,894
Probability being CE / λ*	20,000	0.914	0.011	0.075	0.000	0.000	0.000	0.000
	30,000	0.674	0.038	0.288	0.000	0.000	0.000	0.000
	50,000	0.185	0.054	0.758	0.001	0.002	0.000	0.000
AD + IPD effectiveness evidence used								
ICER (£s)		---	Extended dominated	32,480	Dominated	Dominated	Dominated	1,308,633
Probability being CE / λ*	20,000	0.904	0.014	0.082	0.000	0.000	0.000	0.000
	30,000	0.555	0.096	0.349	0.000	0.000	0.000	0.000
	50,000	0.108	0.080	0.812	0.000	0.000	0.000	0.000
At Least One Unemployed Parent' Households								
Interventions		**(1) UC**	**(2) E**	**(3) E + FE**	**(4) E + FE + HI**	**(5) E + FE + F**	**(6) E + HI**	**(7) E + FE + F + HI**
AD effectiveness evidence used								
ICER (£s)		---	Extended dominated	32,149	2,048,905	Dominated	Dominated	2,628,186
Probability being CE / λ*	20,000	0.861	0.048	0.091	0.000	0.000	0.000	0.000
	30,000	0.498	0.162	0.340	0.000	0.000	0.000	0.000
	50,000	0.104	0.108	0.786	0.001	0.000	0.000	0.000
AD + IPD effectiveness evidence used								
ICER (£s)		---	Extended dominated	33,194	Dominated	Dominated	Dominated	1,272,219
Probability being CE / λ*	20,000	0.892	0.025	0.083	0.000	0.000	0.000	0.000
	30,000	0.574	0.086	0.340	0.000	0.000	0.000	0.000
	50,000	0.121	0.084	0.794	0.000	0.000	0.000	0.000

Notes:

* where λ is the threshold ratio

5.4.3 Subgroup analyses: number of parents in the household and their employment status

5.4.3.1 Effectiveness

Table 5.6 shows the probability of 'success' of the interventions for four subgroups. For subgroup 'two employed parents family' (2EP), the 'functioning' smoke alarm uptake for all interventions estimated when using AD is lower than the one estimated from AD + IPD. In general, it is also the case for this subgroup that differences between probability point estimates obtained from using AD and AD + IPD are larger across interventions. These differences are generally smaller in the subgroup of unemployed single parents (1UP). Again, the intervention for which the largest difference in point estimates between using AD and AD + IPD is found for is the one which provides only education, *(2) E*.

Table 5.6 - Absolute probabilities of 'success' estimates (i.e. uptake of 'functioning' safety equipment in the household) from fitting the MTC model for AD and for AD plus IPD including covariates 'single parent status' and 'parents' employment status' to the functioning smoke alarm outcome data.

Interpretation		Random effects MTC of AD - network of 20 studies		Random effects MTC of AD and IPD - network of 20 studies *	
		Median of MCMC posterior sample	95 per cent credible interval	Median of MCMC posterior sample	95 per cent credible interval
	Probability				
Absolute probability of success of interventions for two employed parent families (2EP) **	***(1) UC***	0.600	0.097 to 0.967	0.602	0.162 to 0.948
	(2) E	0.125	0.002 to 0.939	0.606	0.065 to 0.976
	(3) E + FE	0.615	0.025 to 0.993	0.832	0.155 to 0.993
	(4) E + FE + HI	0.528	0.034 to 0.98	0.764	0.129 to 0.989
	(5) E + FE + F	0.474	0.01 to 0.988	0.822	0.165 to 0.992
	(6) E + HI	0.717	0.038 to 0.997	0.868	0.157 to 0.996
	(7) E + FE + F + HI	0.753	0.024 to 0.995	0.917	0.281 to 0.997
Absolute probability of success of interventions for employed single parent families (1EP) **	***(1) UC***	0.600	0.097 to 0.967	0.602	0.162 to 0.948
	(2) E	0.957	0.002 to 1	0.551	0.03 to 0.981
	(3) E + FE	0.992	0.01 to 1	0.839	0.106 to 0.997
	(4) E + FE + HI	0.991	0.01 to 1	0.870	0.139 to 0.998
	(5) E + FE + F	0.987	0.007 to 1	0.788	0.085 to 0.994
	(6) E + HI	0.996	0.024 to 1	0.892	0.13 to 0.998
	(7) E + FE + F + HI	0.997	0.018 to 1	0.927	0.201 to 0.999
Absolute probability of success of interventions for families with two parents and at least one parent unemployed (2UP) **	***(1) UC***	0.600	0.097 to 0.967	0.602	0.162 to 0.948
	(2) E	0.905	0.004 to 1	0.541	0.039 to 0.973
	(3) E + FE	0.992	0.074 to 1	0.807	0.122 to 0.993
	(4) E + FE + HI	0.994	0.074 to 1	0.731	0.097 to 0.989
	(5) E + FE + F	0.978	0.024 to 1	0.780	0.109 to 0.991
	(6) E + HI	0.994	0.085 to 1	0.844	0.115 to 0.996
	(7) E + FE + F + HI	0.995	0.074 to 1	0.899	0.205 to 0.997
Absolute probability of success of interventions for unemployed single parent families (1UP) **	***(1) UC***	0.600	0.097 to 0.967	0.602	0.162 to 0.948
	(2) E	0.678	0 to 1	0.504	0.023 to 0.98
	(3) E + FE	0.938	0 to 1	0.831	0.092 to 0.997
	(4) E + FE + HI	0.957	0 to 1	0.858	0.108 to 0.998
	(5) E + FE + F	0.830	0 to 1	0.754	0.061 to 0.994
	(6) E + HI	0.966	0.001 to 1	0.884	0.1 to 0.998
	(7) E + FE + F + HI	0.967	0 to 1	0.919	0.173 to 0.999

Notes:

* For nine of the 20 studies individual level data was available

** (1) UC - usual care; (2) E - education; (3) E + FE - education plus low cost / free equipment; (4) E + FE + HI - education plus low cost / free equipment plus home inspection; (5) E + FE + F - education plus low cost / free equipment plus fitting; (6) E + HI - education plus home inspection; (7) E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

5.4.3.2 Cost effectiveness

Cost effectiveness results for the four subgroups are shown in Table 5.7. The use of IPD rather than just AD, contributes to the consistency of cost effectiveness results across subgroups. When considering the subgroup of 2EPs, a shift in the expected ICER estimate for intervention *(3) E +FE* is observed when using AD – that is, an estimated mean ICER of approximately £41,000 compared to £35,500 per QALY gained when IPD is used. This shift implies that, for instance, using a £30,000 threshold value, the estimated probability of intervention *(1) UC* being cost effective is 0.75 when using AD, but 0.64 when using AD + IPD.

If a threshold value of £30,000 is used, approval / rejection decisions are the same for each subset of the population – *(1) UC* is the intervention recommended. This is valid when using AD + IPD. When using AD, and for a threshold value of £33,000 per QALY gained, intervention *(3) E + FE* would be recommended for use in the 1EP subgroup. If AD + IPD is used, intervention *(3) E + FE* is not considered cost effective at that particular threshold value, indicating that cost effectiveness decisions may alter depending on the level of disaggregation of the evidence informing model inputs and the ability to appropriately capture the true underlying effects and associations.

Table 5.7 – Cost effectiveness results for all interventions, for 4 subgroups of families: (a) 2 employed (2EP); (b) employed single (1EP); (c) 2 with at least one unemployed (2UP); and (d) unemployed single parents (1UP). Results are shown for when using AD and AD plus IPD to populate the decision model.

AD effectiveness evidence used

Interventions		(1) UC	(2) E	(3) E + FE	(4) E + FE + HI	(5) E + FE + F	(6) E + HI	(7) E + FE + F +
(a) Two Employed Parent' Households (2EP)								
ICER (£s)		---	Extended dominated	40,721	Dominated	Dominated	Dominated	1,027,227
Probability being CE / λ*	20,000	0.933	0.005	0.062	0.000	0.000	0.000	0.000
	30,000	0.753	0.019	0.228	0.000	0.000	0.000	0.000
	50,000	0.353	0.023	0.622	0.001	0.001	0.000	0.000
(b) Employed Single Parent Household (1EP)								
ICER (£s)		---	Extended dominated	32,055	3,989,355	Dominated	Dominated	5,747,717
Probability being CE / λ*	20,000	0.858	0.051	0.091	0.000	0.000	0.000	0.000
	30,000	0.505	0.151	0.344	0.000	0.000	0.000	0.000
	50,000	0.124	0.073	0.802	0.000	0.001	0.000	0.000
(c) Two Parent Household with at Least One Unemployed (2UP)								
ICER (£s)		---	Extended dominated	33,961	Dominated	Dominated	Dominated	1,849,602
Probability being CE / λ*	20,000	0.856	0.065	0.079	0.000	0.000	0.000	0.000
	30,000	0.526	0.191	0.283	0.000	0.000	0.000	0.000
	50,000	0.180	0.101	0.717	0.001	0.002	0.000	0.000
(d) Unemployed Single Parent Household (1UP)								
ICER (£s)		---	Extended dominated	35,013	1,642,801	Dominated	Dominated	3,227,412
Probability being CE / λ*	20,000	0.874	0.050	0.076	0.000	0.000	0.000	0.000
	30,000	0.554	0.154	0.292	0.000	0.000	0.000	0.000
	50,000	0.199	0.115	0.680	0.003	0.004	0.000	0.000

AD + IPD effectiveness evidence used

Interventions		(1) UC	(2) E	(3) E + FE	(4) E + FE + HI	(5) E + FE + F	(6) E + HI	(7) E + FE + F +
(a) Two Employed Parent' Households (2EP)								
ICER (£s)		---	Extended dominated	35,538	Dominated	Dominated	Dominated	1,058,776
Probability being CE / λ*	20,000	0.909	0.018	0.072	0.000	0.000	0.000	0.000
	30,000	0.636	0.076	0.288	0.000	0.000	0.000	0.000
	50,000	0.197	0.074	0.726	0.000	0.002	0.000	0.000
(b) Employed Single Parent Household (1EP)								
ICER (£s)		---	Extended dominated	35,937	Dominated	Dominated	Dominated	1,021,409
Probability being CE / λ*	20,000	0.912	0.017	0.072	0.000	0.000	0.000	0.000
	30,000	0.644	0.072	0.284	0.000	0.000	0.000	0.000
	50,000	0.203	0.074	0.720	0.001	0.002	0.000	0.000
(c) Two Parent Household with at Least One Unemployed (2UP)								
ICER (£s)		---	Extended dominated	35,839	Extended dominated	Dominated	Dominated	989,129
Probability being CE / λ*	20,000	0.912	0.019	0.069	0.000	0.000	0.000	0.000
	30,000	0.640	0.077	0.283	0.000	0.000	0.000	0.000
	50,000	0.198	0.084	0.712	0.002	0.003	0.000	0.000
(d) Unemployed Single Parent Household (1UP)								
ICER (£s)		---	Extended dominated	36,186	Extended dominated	Dominated	Dominated	946,311
Probability being CE / λ*	20,000	0.905	0.018	0.076	0.000	0.000	0.000	0.000
	30,000	0.645	0.070	0.285	0.000	0.000	0.000	0.000
	50,000	0.214	0.089	0.691	0.003	0.003	0.000	0.000

5.5 Discussion

Summary of findings

In a binary outcome setting, the use of IPD is particularly useful in guiding decision making for particular population subgroups. In comparison with the use of evidence in summary format, the ability to appropriately estimate *treatment x covariate* associations means an accurate and/or more precise judgement of an intervention's cost effectiveness for a particular subgroup of the population. Thus, access and use of IPD is important in providing an appropriate answer to one of the key questions of any appraisal; that is, from the available evidence, for which subset of the targeted population is the programme considered cost effective?

The case study results revealed that all active interventions were not cost effective at population level using a £30,000 threshold value. This result was consistent using AD and AD + IPD. Results supported the idea that, if only effectiveness AD was used, uncertainty about the intervention to approve for particular subgroups was higher. Similar conclusions to the population average were obtained in the subgroup analyses. Nonetheless, the fact that the case study conclusions pointed to the same decision for each population subgroup does not imply that in other case studies, or in other circumstances, the decision would be the same for all subgroups[27].

Study strengths, limitations and further work

The main contribution of the current chapter was to illustrate and discuss the expected benefits from integrating the results of a novel MTC model for the simultaneous

[27] As emphasised in Chapter 4, the reader is reminded of the inexistence of a predefined threshold outside the health sector. As the case study results shown in this chapter were obtained using a public sector viewpoint, it implies that approval decisions are not straightforward.

synthesis of IPD and AD within a decision model. The main tools for the analysis developed here were the novel synthesis models discussed in Chapter 3 and the decision analytic model described in Chapter 4.

As discussed in Chapter 3, the novel evidence synthesis models enhance the use of evidence to the individual level, in a binary outcome setting. By comparison with the use of AD only, advantages from the IPD approach were discussed. Nonetheless, these synthesis models could be extended to consider other types of outcome measures. For instance, if the MTC outcome was of continuous or time to event nature, it would be interesting to assess the impact on cost effectiveness results when using AD and IPD due to, for example, the existence of nonlinearities, among other potential issues.

Another issue that may be considered a limitation, relates to the fact that, in the comparisons between 1 (overall population), 2 or 4 subgroups, it was assumed that the analyst had, in turn, access to: *(i)* no information on any covariate of interest; *(ii)* information on one covariate; or *(iii)* information on two covariates simultaneously. This implies that different synthesis models, with different specifications, were used to obtain relative effectiveness estimates.

The findings from this chapter show the importance of using all relevant evidence when informing a specific decision problem, in particular the role of accessing and using evidence at the individual level. Nonetheless, analysts / modellers are aware that the quality of the evidence used is a major factor in enabling decision makers to reach appropriate judgements on the cost effectiveness of alternative interventions, even if available at the individual level. The quality of the evidence and its adequacy for the problem at hand, are also determinants in supplying support for the decision making process.

What is the value of acquiring individual level data?

Individual participant data are regarded as the optimal vehicle for an appropriate estimation of cost effectiveness outcomes and quantification of uncertainty through

decision modelling. Formal modelling and assessment of baseline risks of particular events, of intervention effects and of *treatment x covariate* associations, with attendant uncertainty, are some of the fundamental advantages of using this data format. It is argued that, when wanting to reduce decision uncertainty in a specific decision problem, funding bodies should consider exploration of all existing evidence, in particular IPD, before directing additional funds to sponsor new trials aimed at resolving (some of) the uncertainty.

Nevertheless, there is a series of obstacles in obtaining and using evidence at the individual level. These obstacles stem mainly from the difficulty in acquiring evidence at this level from the various possible sources (e.g. pharmaceutical companies, trial funding bodies), but also from the (time and computational) burden of exploring and analysing it. The question for which there is no clear or direct answer is whether these IPD 'barriers' outweigh its advantages, some of which are highlighted in the current exercise.

Individual level data are unquestionably valuable. Quantifying the expected costs of uncertainty when using AD or IPD is an important issue as it may inform decision makers on the added value of this supplementary level of evidence. It is here flagged up the need to explore this issue further.

Concluding remarks

This chapter showed how IPD and AD may affect cost effectiveness and allocation decisions. Although the use of IPD may be challenging, higher accuracy and/or precision are achieved when analysing subgroup effectiveness. Consequently, cost effectiveness estimates derived using IPD will lead, in principle, to better decisions.

The quantification of decision uncertainty may inform decision makers on the additional benefits provided by IPD. It is expected that it will improve the estimation of the upper bound value of conducting further research. This topic will be subject to extensive discussion in the following chapter.

CHAPTER 6

6. THE VALUE OF FURTHER RESEARCH: THE ADDED VALUE OF INDIVIDUAL-PARTICIPANT LEVEL DATA

6.1 Introduction

6.1.1 Background

In health care, decisions are inevitably made under uncertainty. In the presence of uncertainty, a decision maker should not only consider *(i)* the suitability of the provision of health care (given the available information), but also *(ii)* whether it is worthwhile to fund supplementary research (to decrease existing decision uncertainty).

Chapter 5 argued the possible advantages of using evidence at the individual level (rather than at the aggregate level) when assessing cost effectiveness, thus directly informing question i) and laying the foundation for informing question ii) above. Using an example from the PH field, Chapter 5 showed that the use of effectiveness evidence at the aggregate level provided only partial and/or incorrect insight into the true relationship between the relative effects of interventions and particular patient

characteristics[28]. Subgroup cost effectiveness estimates obtained when including evidence as IPD were more accurate (i.e. less biased) and/or more precise.

In estimating the value of conducting further research [item ii) listed above], characterising and quantifying decision uncertainty and its consequences are key components. One of the main reasons for the existence of decision uncertainty is that decisions are based on sampled data, so that true model parameter values cannot be known with certainty[29]. Whilst a particular health technology may be (on average) cost effective, our confidence in this assessment may be low. Moreover, it may also be important to explore heterogeneity between patients (e.g. through defining subgroups of the targeted population). In fact, population average based judgements may disguise sources of heterogeneity that should be reflected in decision making. When conducting subgroup analyses, quantifying uncertainty for each population strata can and should be performed (NICE, 2008, Sculpher, M., 2008). Not only is the possibility of making different decisions for different subsets of the population important, but also determining the value of performing further research in each subset is. As highlighted previously, for the class of models used in this thesis, it is in situations like this that the role of (and the benefits from) using IPD, compared with AD, becomes more evident, i.e. quantifying uncertainty in subgroups is generally most accurately attained when this type of evidence is available.

There has been an increase in awareness of the need to deal with subgroups and heterogeneity in health care decision making. Defining appropriate and meaningful subgroups is complex in itself. Sculpher (2008) highlights several potential forms of subgroups and heterogeneity, pointing the reader in the direction of correctly identifying subgroups and the advantages of reflecting these in decision making. This paper was the basis for the NICE methods guidance on subgroup analysis (NICE,

[28] It is important to remind the reader that, in this example, binary outcome measures were synthesised across studies using a MTC.

[29] As highlighted in the introductory chapter, structural uncertainty is also another key element for decision uncertainty.

2008). Coyle *et al.* (2003) outline a more quantitative approach, defining population NBs in the presence of subgroups. Taking into account equity considerations, the authors allow for different decisions between subgroups and also identify optimal criteria in access to health care technologies, leading to the maximisation of efficiency gains, though decision uncertainty was not explored to inform the need for further research.

In the limit, the existence of heterogeneity implies that decisions may be made at the individual level. This was discussed by Basu (2009), who put forward the idea of performing individual level effectiveness assessments (individualized comparative effectiveness research, *i*-CER). The *i*-CER replaces the usual average treatment effect estimation for the entire target population (or subgroup) by assessing the effect of the treatment at the individual level. Basu and Meltzer (2007) outlined a method to quantify the potential gains of providing decision making at the individual level. Their framework proposes the estimation of the *expected value of individualised care* (EVIC) to show what society is willing to pay in order that individually efficient decisions may be made (Basu & Meltzer, 2010).

More recently, Basu (2011) demonstrated how individual preferences can be estimated to facilitate the estimation of the EVIC. It was shown that market failures such as the presence of asymmetric, imperfect and/or incomplete information [between the person (patient) and the insurer] may generate inefficiencies in the society owed to moral hazard effects (Pauly & Blavin, 2008). In this paper Basu presented a new modelling approach, which allowed for both within- and between-treatment heterogeneity in outcomes. This novel econometric approach facilitates the estimation of marginal benefit curves, derived from both observed and unobserved patient characteristics, aimed at representing individual treatment effects heterogeneity. This methodological development contributed to the understanding of the value of considering heterogeneity in CEA, although it has not formally addressed the issue of uncertainty associated with decisions based on other than the average.

Espinoza *et al.* [unpublished, (2011)] advocated that whilst a higher number of subgroups may generate more population net health, the transaction costs associated may not compensate these at the margin. The framework developed, combined the work of Coyle (2003) and Basu and Meltzer (2010, 2007), generalising this framework and defining in practice how to conduct CEA in the presence of subgroups. Espinoza *et al.* (2011) extended Coyle's (2003) work by addressing the potential benefits of characterizing heterogeneity not only in terms of increased individual benefits based on current information, but also in estimating the value of conducting further research. The authors generalised the concept of *expected value of perfect information* (EVPI) adopted by Claxton and others (1996) in order to consider mutually exclusive population subgroups. They extended Basu and Meltzer's (2010, 2007) work by acknowledging that a continuum exists from decisions made at the mean-, at the subgroup- and at the individual level, the latter with all existing heterogeneity explained and in absence of (decision) uncertainty. It also provided support for the understanding of EVIC by highlighting the concept of the *value of heterogeneity* (VoH), separating it into *static* and *dynamic* values – concepts that will be discussed later in this chapter.

6.1.2 Aims and objectives

This chapter uses Basu and Meltzer's (2010, 2007) and Espinoza and colleagues' (2011) framework to explore issues around the potential benefits of accessing data at the individual level. While Espinoza and colleagues' paper conceptually highlighted the potential sources of heterogeneity and the choices to effect an appropriate selection of subgroup specifications for the analysis, it provided few insights on the particular type of data needed to perform such tasks[30] (i.e AD and/or IPD). Therefore, this chapter seeks to assess the added value of having access to IPD, compared to using AD only, in appropriately performing subgroup value of information analysis. This chapter begins by examining the situation of having access to relevant evidence at

[30] In fact, Espinoza and colleagues' (2011) paper demonstrated their conceptual framework by implementing it in an already existing trial base economic model.

summary level and/or at individual level without considering subgroups. The objective is not to prospectively identify optimal research designs using these two evidence types (i.e. with optimal sample size and patient allocation procedure, correct follow-up periods and endpoints (Briggs, A. H. et al., 2006)), but to provide sufficient evidence to support the use of particular intervention(s).

The framework developed in this chapter aims to be as generic as possible in understanding the implications of considering IPD. These implications are illustrated through the use of a case study, as explored in previous chapters. The effectiveness of different programmes, intended to increase the uptake of functioning fire alarm equipment in households (binary outcome, 1 – the presence of functioning smoke alarm equipment, 0 – no functioning equipment), was explored in Chapter 3. It was extended in Chapter 4 with a cost effectiveness assessment of the different relevant alternatives. Assessments were made with the purpose of verifying the advantages of having access and using all relevant evidence, in particular, of using IPD (Chapter 5). The current chapter assesses the added value of acquiring IPD in providing information for further research on fire alarm program awareness.

In the case study used in this chapter – as in Chapters 3, 4 and 5 – the use of IPD only (i.e. ignoring the existence of AD) implies a smaller evidence network and, consequently, a smaller number of alternative interventions to be evaluated. As a result, comparisons here are only made between results from using the evidence at AD and AD + IPD, with IPD only analysis excluded. Therefore, in this chapter and in the context of the motivating example, the use of the expressions ‘IPD’ and ‘AD + IPD’ may be considered exchangeable.

This chapter also seeks to provide an assessment of the value in considering heterogeneity in decisions made on the basis of existing and additional information. In this respect, the case study makes use of two binary covariates: number of parents in the family (0 = 2P; 1 = 1P) and their employment status (0 = 2U; 1 = 1U). With the

use of these two covariates, two (and two different specifications[31]) or four subgroups are defined. To fully understand the value of considering IPD in the presence of subgroups, four scenarios are considered: *(i)* no subgroups (AD) *vs.* no subgroups (IPD) ; *(ii)* no subgroups (AD) *vs.* all subgroups (IPD); *(iii)* some subgroups (AD) *vs.* all subgroups (IPD); and *(iv)* all subgroups (AD) *vs.* all subgroups (IPD). Scenario (i) proposes providing an answer to the expected benefits from considering IPD compared to AD when making population average decisions. Scenarios (ii), (iii) and (iv) propose to assess the gains in disentangling existing heterogeneity at different levels of population strata. Thus, the maximum possible NBs gained from considering IPD (i.e. IPD with all subgroups) are compared with different AD disaggregation levels (i.e. AD with no, some and all considered population subgroups).

As mentioned in Chapters 3, 4 and 5, different synthesis models are used to analyse different subgroup specifications (i.e. in previous chapters, different synthesis models were considered where: no covariate information is available; is available for just one binary covariate; or is available for two covariates). The main reason for this approach was to observe the consequences of a gradual increase in the availability of evidence (i.e. from AD to IPD) and of information on covariates (i.e. from total absence of information on covariates to availability of information on two covariates). This approach continues to be used throughout the current chapter[32].

This chapter starts by describing the main concepts of value of information analysis – section 6.2, comprising an assessment of how available evidence may shape the

[31] A specification is a subgroup definition based on an available particular level of data disaggregation. For instance, a certain population of interest may be split in 2 mutually exclusive subgroups on the basis of whether or not patients are diabetic (*specification 1*), on whether or not they are hypertense (*specification 2*) or any other disaggregation based on patients characteristics. As explained by Espinoza and colleagues (2011), the difficulty lies in obtaining information on subgroups and, more importantly, in how to select between alternative specifications.

[32] Nonetheless, a single synthesis model including the 2 covariates could be used to perform all the analyses in this chapter. That is, the results of a 2 binary covariates synthesis model (i.e. providing results for 4 population subgroups) could be used as a starting point to obtain results for when wanting to assess the scenarios of 'no subgroups' or '2 subgroups'. This exercise, named 'backward estimation', is explored in the Appendix 7.

quantification of the value of additional research. This is followed by an application that considers evidence at the summary-level and also the mixture of summary and IPD. Section 6.3 extends the methodology to the subgroup setting, exemplifying, through the case study, situations where two and four subgroups are defined. The final section summarizes the main chapter findings and discusses several issues including further research topics.

6.2 Value of additional research in the absence of subgroups

Given the presence of uncertainty surrounding the expected NBs associated with the use of alternative technologies, decision recommendations based on current information may change if these uncertainties are resolved. The joint probability that a decision based on existing information will be wrong and the consequences of making a wrong decision can be assessed and quantified (Claxton, K. et al., 2002). The methodologies used to enable this quantification are shown next. Definitions and descriptions of the methods are introduced first (section 6.2.1), followed by issues relating to available evidence and the value of further research (section 6.2.2). The results obtained through the motivating example will be given (section 6.2.3).

6.2.1 Definitions and methods

In a decision model with a vector of unknown parameters θ, with a choice to be made between a vector of mutually exclusive interventions $y = \{1,2,...,Y\}$, the best possible decision is the one that yields the highest expected NB[33], that is, under *current information* (Ades, A. et al., 2004, Claxton, K., 1999, Stinnett & Mullahy, 1998) the optimal strategy is the one that:

[33] The cost effectiveness of an alternative y can be expressed in terms of NMB ($NMB_y = Q_y \bullet \lambda - C_y$, where C_y represents the costs and Q_y the consequences for alternative y).

$$\max_y E_\theta NB(y,\theta) \tag{6.1}$$

Because true values of θ are unknown, under *perfect information* the NB attained is the average of the maximum NBs over the joint distribution of θ:

$$E_\theta \max_y NB(y,\theta) \tag{6.2}$$

The EVPI is the difference between the estimated payoffs of having perfect information from those obtained under current information, and represents the expected opportunity loss of uncertainty surrounding decisions (Claxton, K., 1999), that is:

$$\text{EVPI} = E_\theta \max_y NB(y,\theta) - \max_y E_\theta NB(y,\theta) \tag{6.3}$$

The above expression refers to decisions made for one individual. At the patient population level the EVPI reflects potential benefits of current and future patients from additional information. This requires an assessment of the period (T) over which information that may be acquired in the near future would be useful for the current decision problem and the incidence over this time period (I_t). The population EVPI (PEVPI), discounted at rate α, may be estimated as (Philips et al., 2008):

$$\text{PEVPI} = \text{EVPI} \cdot \sum_{\text{t}}^{T} \frac{I_t}{(1+\alpha)^t} \tag{6.4}$$

6.2.2 Available evidence and the value of further research

Although in many instances, access to IPD is preferred, decision analysis does not consider evidence as unsatisfactory if data is available only at summary level. Nonetheless, the type of evidence available may influence one's certainty about the adoption decision. The current subsection focuses on the important question of the

impact the type of evidence may have on the estimation of the need for further research[34].

Assuming the general case of a statistical synthesis model parameterized by θ and evaluating the probability distributions of observed data (i.e. *evidence* or *likelihood*), $p(x|\theta)$, and pooled statistic $\hat{\theta}$ which serves as an estimator of θ, based on any observed data x_i ($i=1, \ldots, m$ studies). It is assumed that the data follows some unknown distribution and, by constructing an estimator which reflects it, it is closer to the underlying true value, θ (whose posterior predictive distribution is then used to populate a decision model input parameter). The analyst can have access to these *m* studies at a summary level, an AD evidence base. Alternatively, it is possible to envisage a situation in which one has access to IPD from each of these *m* studies, an IPD evidence base. With access to a mixture of AD and IPD (only a proportion of studies are available at the individual level), the aim may be to enumerate possible benefits the synthesis of IPD may bring to the correct estimation of decision model input parameters, considering expected consequences over the EVPI estimation when compared to the synthesis of AD only.

In the absence of subgroups and with binary outcomes, estimated EVPI is not expected to differ when using AD or IPD to describe the evidence base, since AD is a sufficient statistic[35] (as is the case in this chapter's motivating example[36]). In contrast,

---

[34] A further relevant question (not addressed within the current work) may be to explore design issues regarding the additional evidence. This relates to a prospective assessment of the sampling and could be implemented using of the *expected value of sampling information* (EVSI) framework. The idea here is that resolving uncertainty completely can only be achieved by using infinitely large samples. Thus, to explore plausible designs of further research, it is sensible to calculate the EVSI, setting it alongside the costs of obtaining the sample. As defined by Claxton and Posnett (1996), the EVSI is "*…the difference between the reduction in the expected loss due to sample information and the costs of obtaining the sample…and represents the societal return to proposed research*". In practical terms, EVSI often informs the optimal sample size for a future study.

[35] Summary data of binary outcomes are considered 'sufficient statistics' as there is no loss of information in comparison to IPD (as discussed in Chapter 2 – section 2.2).

[36] In the case study, if no subgroups are being considered, EVPI estimates should be equivalent for when using AD and for when using the same evidence set partially at IPD format. Marginal differences

differences in EVPI are to be expected when synthesising continuous outcomes. Interpretations of such differences should be carried out with caution. These do not reflect the value of acquiring further evidence or the value of acquiring IPD rather than AD. In this case, AD and IPD represent the same evidence base, where the IPD is one realisation of all possible datasets described by the AD[37]. Hence, using IPD generates more efficient estimates[38] of EVPI than using AD.

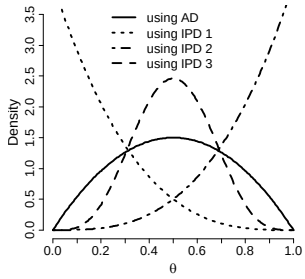
Three ways can be envisaged in which access to IPD, rather than AD, can influence results (in the absence of subgroups). The following scenarios are intended to be as generic as possible.

---

between these may however occur, caused by external factors (e.g. due to the cluster adjustment procedure in the synthesis that is done ad-hoc in the AD case and built-in in the IPD case, or due to simulation error – although we acknowledge that the latter can be resolved (Oakley et al., 2010)).

[37] This can be easily illustrated with a brief example: let's consider that the researcher has access to the following AD – four individuals are sampled from the targeted population ($n = 4$) in which two have had the event of interest ($r = 2$) and two of them have a relevant characteristic represented by the covariate (*cov*). Despite the researcher not having access to the IPD which generated the AD information, the IPD can only be one of the following: a) all individuals that had the event do not have the characteristic of interest (IPD 1 in the table below); b) all individuals that had the event have the characteristic of interest (IPD 2 in the table below); and c) half the individuals that had the event also have the characteristic of interest (IPD 3 in the table below).

cov	IPD 1	IPD 2	IPD 3
	r	*r*	*r*
0	1	0	0
0	1	0	1
1	0	1	1
1	0	1	0

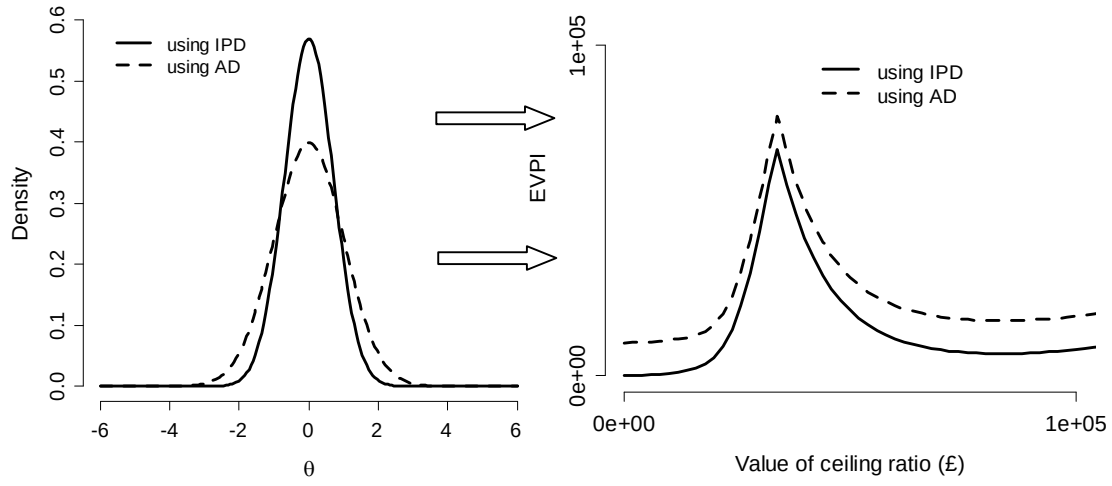


The IPD that was reduced to generate the AD can only be one of the 3 cases described above; that is, it is one realisation from a range of possible datasets from the AD. The figure above represents the evidence generated on a hypothetical parameter of interest (in this case the probability of observing an event in the subgroup of patients with the covariate), generated from each possible IPD dataset and from the AD. Estimates from AD are expected to consider the possibility of evidence being from any of the IPD sets. Thus, AD estimates are shown as less precise.

[38] With IPD we expect to obtain an unbiased statistic with sufficiently smaller variance (efficient statistic) when compared to using AD. As defined in chapter 5, by bias it is meant that an inaccurate estimation of the association of interest is obtained.

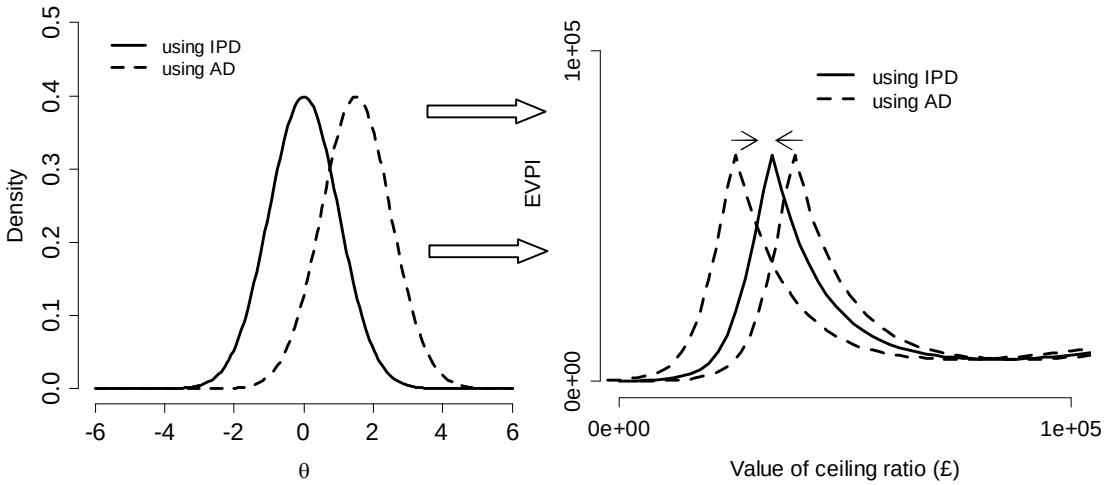
Scenario 1.1: If, by comparison with the use of AD evidence base, the use of IPD involves an increase in parameter(s) precision (efficiency), obtained distribution of relative treatment effects will have lower or equal variance, i.e. $Var\left[\hat{\theta}^{AD}\right] \geq Var\left[\hat{\theta}^{IPD}\right]$ – Figure 6.1a. Therefore, across a range of possible threshold values, EVPI values are expected to decrease due to the precision gained on this particular decision model parameter, all other model parameters constant – Figure 6.1b.

Figure 6.1 - Representation of (a) an increase in estimates precision when IPD is considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 1.1).



Scenario 1.2: If, by comparison with the use of an AD evidence base, the use of IPD may involve a reduction or elimination of bias in estimates, which may imply a shift in the distribution of effects (for simplicity it is here assumed an impact on the scale parameter, and not on the shape and/or dispersion parameters) – Figure 6.2a depicts a hypothetical situation where the estimates derived from AD are biased upwards, i.e. $E\left[\hat{\theta}^{AD}\right] \geq E\left[\hat{\theta}^{IPD}\right]$. This may be translated in shifts in the mean ICER and consequently shifts in EVPI curves as depicted in Figure 6.2b – assuming constant information for all other model parameters.

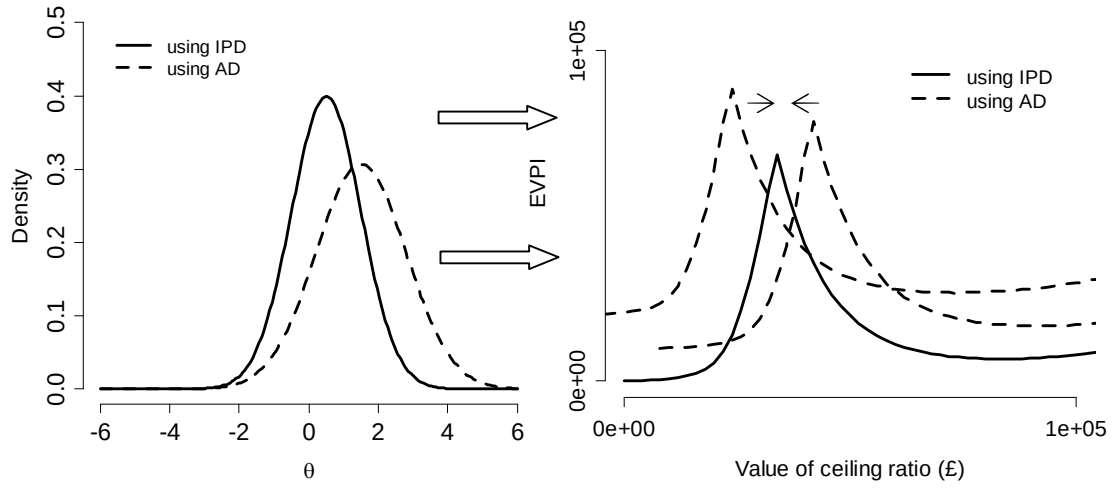
Figure 6.2 - Representation of (a) a decrease (removal) of bias in obtained estimates when IPD is considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 1.2).



Whilst in scenario 1.1 it can be predicted that EVPI is lower when using IPD, in the current scenario it cannot.

Scenario 1.3: This situation entails both an increase in precision and a reduction in bias from using IPD relatively to AD. The following graphical representation illustrates this scenario and interpretations can be inferred from the two previous examples – Figure 6.3. In this scenario it cannot be predicted how the availability of IPD will affect the EVPI estimates.

Figure 6.3 - Representation of (a) an increase in precision and a decrease (removal) of bias in obtained estimates when IPD is considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 1.3).



In the following subsection comparisons will be made between using AD and AD + IPD to inform the case study decision model.

6.2.3 Application: value of additional research in the context of having access to individual participant level data

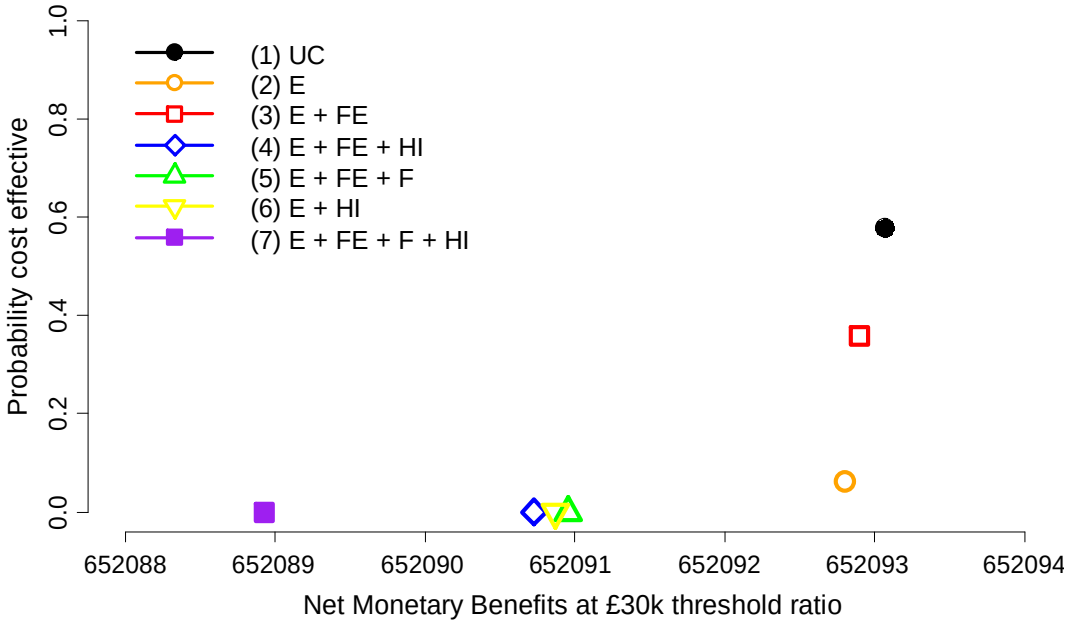
The analyses carried out in this subsection are made under the PH perspective. A 10 years expected lifetime of the programmes[39] and an annual effective population (i.e. expected number of new households with dependent/s under 5 years old per year in the UK) of 31,000 households (ONS, 2010a) were considered adequate for the value of information analysis calculations.

Figure 6.4 summarizes the main cost effectiveness results when AD is used to inform the smoke alarms decision model. In this analysis, no subgroups are taken into account (i.e. it is assumed that no covariate information is available) and relate to population

[39] The choice of the programme lifetime is linked to the fact that most alarms installed today have a life span of about 8-10 years. After this time the entire unit should be replaced, as recommended by the National Fire Protection Association (NFPA).

average decisions – see Chapter 4, section 4.4 for additional details. Usual care, identified as *(1)* in the figure, has the highest expected NBs and is also associated with the highest probability of being cost effective at a £30,000 threshold ratio, though this might not always be the case (Fenwick et al., 2004).

Figure 6.4 – Cost effectiveness results for the smoke alarms decision model: NMBs estimates at £30,000 threshold ratio versus the probability of the intervention(s) being cost effective. Results shown are for when AD only was used to inform the effectiveness decision model parameters and when no subgroups are considered.



Note: (1) UC – usual care; (2) E - education; (3) E + FE - education plus low cost / free equipment; (4) E + FE + HI - education plus low cost / free equipment plus home inspection; (5) E + FE + F - education plus low cost / free equipment plus fitting; (6) E + HI - education plus home inspection; (7) E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

As highlighted in Chapter 5, since the outcome variable analysed in the MTC model is binary, population average (cost) effectiveness results obtained using AD and AD +

IPD are expected to be similar[40] in the absence of covariates (no subgroups). These similarities were expected. In Table 6.1 a more detailed view of the cost effectiveness results when AD + IPD effectiveness evidence is used to inform the decision model, all other evidence informing the model remaining constant.

Table 6.1 – Expected cost effectiveness of functioning smoke alarms interventions per participant for when using AD plus IPD effectiveness evidence to inform the economic model.

		Cost effectiveness threshold at			
		£20,000 per QALY		**£30,000 per QALY**	
Intervention	**ICER** (£/QALY)	**NMB** (£)	**Probability cost effective**	**NMB** (£)	**Probability cost effective**
(1) UC	---	125,691.2	0.914	670,290.8	0.605
(2) E	Extended dominated	125,690.7	0.011	670,290.5	0.060
(3) E + FE	33,752	125,690.5	0.076	670,290.6	0.335
(4) E + FE + HI	Dominated	125,688.3	0.000	670,288.4	0.000
(5) E + FE + F	Dominated	125,688.5	0.000	670,288.6	0.000
(6) E + HI	Dominated	125,688.7	0.000	670,288.5	0.000
(7) E + FE + F + HI	1,107,554	125,686.4	0.000	670,286.6	0.000

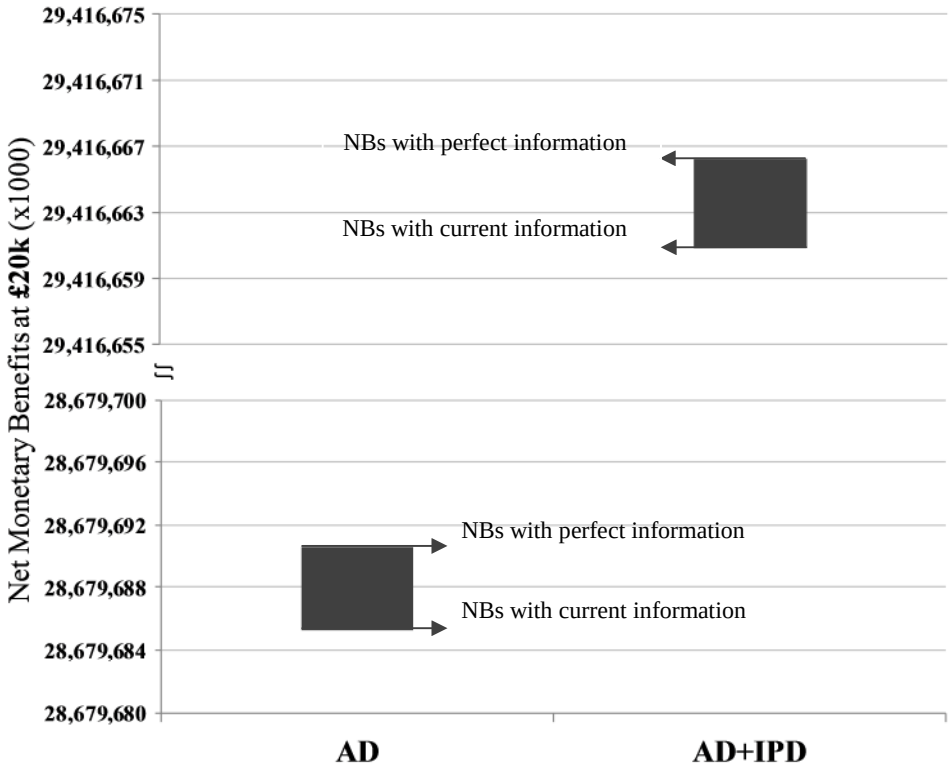
The figures presented in the above table are at the participant level. Estimates of the NMBs obtained are of approximately £28.7 and £29.4 billions, when using AD and AD + IPD, respectively[41]. In Figure 6.5 estimates obtained with current information are represented at the lower end of the bar. The upper end represents the population NBs attained with perfect information. The population EVPI is the difference between the NBs with perfect information and with current information. As expected, the population EVPI estimates are almost equivalent – approximately £5,420 and £5,450

[40] Although estimates from both scenarios should be equivalent, small differences were found and are discussed above and in Chapter 5.

[41] Again, population EVPI estimates from using AD and AD + IPD should be equivalent. Nonetheless, marginal differences exist. Reasons for these differences are discussed above and in Chapter 5.

when using AD and when using AD + IPD, respectively. For a threshold value of £30,000, estimated population EVPI is approximately £45,000 – results are not shown. Additionally, and for both situations (AD and AD + IPD), population EVPI estimate reaches its maximum value at approximately £33,000, when the threshold equals the ICER of the intervention which involves providing education and free or sponsored smoke alarm equipment (i.e. *(3) E + FE*) – see Chapter 4 for additional details.

Figure 6.5 – Population expected NMBs at a threshold value of £20,000 per QALY (2009 values) derived from the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD effectiveness evidence.



The next section explores the implications of performing similar analysis in the context of considering population stratifications.

6.3 Value of additional research in the presence of mutually exclusive subgroups

This section restates and rephrases the two important questions (referred to in the introduction of this chapter) that are part of the health technologies appraisal process, but now are made with subgroup considerations. Namely: *i*)* with existing information, is the technology considered cost effective for all population strata – approval of the health care intervention for which population?; and *ii*)* is there justification for requesting further information for all population strata – and if not, is further research warranted for making decisions regarding the use of the intervention in specific subgroups of the population?

This section extends the previous one to accommodate considerations regarding the presence of subgroups, considering also the concept of *value of heterogeneity*. Again, definitions and descriptions of the methods are introduced first (section 6.3.1), followed by issues relating to available evidence in considering subgroups (section 6.3.2). Finally, the results of these analyses are presented considering the example (section 6.3.3).

6.3.1 Definitions and methods

Espinoza *et al.* (2011) show that when the decision maker is interested in subgroup specific results, the maximum expected NBs for each subgroup *k* (out of a total of *K* mutually exclusive subgroups), under current information can be calculated as (analogous to equation (6.1)):

$$\max_y E_{\theta_k} NB_k\left(y,\theta_k\right) \tag{6.5}$$

With respect to the situation with perfect information, the expected value of the decision for subgroup *k* is:

$$E_{\theta_k} \max_y NB_k(y,\theta_k) \tag{6.6}$$

The expected opportunity cost of uncertainty for subgroup *k* can be expressed as:

$$\text{EVPI}_\text{k} = E_{\theta_k} \max_y NB_k(y,\theta_k) - \max_y E_{\theta_k} NB_k(y,\theta_k) \tag{6.7}$$

The EVPI considering subgroups (EVPI_K) is simply the weighted average across subgroups considering the proportion of patients in each subgroup (w_k). If, for instance, two subgroups are considered (i.e. $k = \{1,2\}$), with w_1 and w_2 representing the proportion of each subgroup in the population (where $w_k \in (0,1)$ and $\sum_k^K w_k = 1$), EVPI_K is:

$$\begin{aligned}\text{EVPI}_\text{K} = & \sum_\text{k}^\text{K} \text{EVPI}_k \cdot w_k \overset{\text{under K = 2}}{=} \\ & = E_{\theta_k}\left[\left(\max_y NB_1(y,\theta_1)\right)\cdot w_1 + \left(\max_y NB_2(y,\theta_2)\right)\cdot w_2\right] - \\ & \left[\left(\max_y E_{\theta_1} NB_1(y,\theta_1)\right)\cdot w_1 + \left(\max_y E_{\theta_2} NB_2(y,\theta_2)\right)\cdot w_2\right]\end{aligned} \tag{6.8}$$

The population EVPI (PEVPI_K) considering subgroups is given by:

$$\text{PEVPI}_\text{K} = \text{EVPI}_\text{K} \cdot \sum_k^\text{K} \sum_\text{t}^{T_k} \frac{I_{k,t}}{(1+\alpha)^t} \tag{6.9}$$

where the time period for which the information on interventions is pertinent to the k^{th} subgroup is represented by T_k and where $I_{k,t}$ represents the incidence over period *t*.

Espinoza *et al.* (2011) coined the concept of *value of heterogeneity*, which indicates the merit of resolving existing heterogeneity in a subgroup setting. That is, the value of making different decisions in different subgroups with current information and the value of resolving parameter uncertainty conditional on a particular level of heterogeneity. A brief description of the issues surrounding this concept is provided here, with the aid of a schematic representation. Figure 6.6 shows the value of heterogeneity using hypothetical examples – it compares the maximum NBs obtained

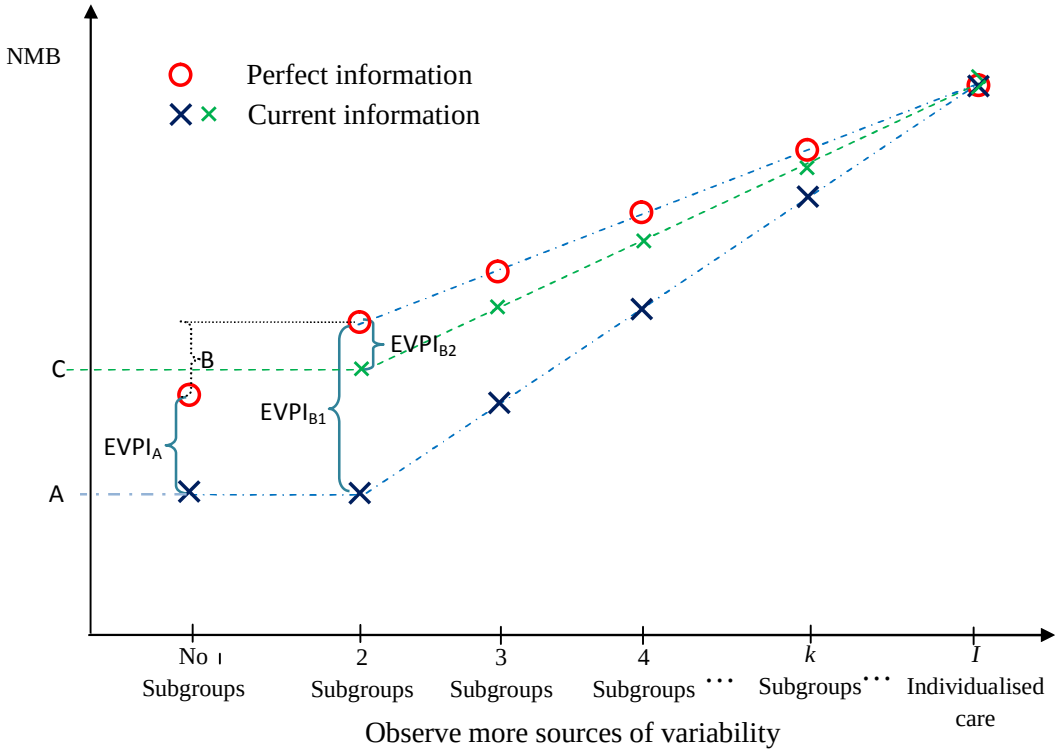
with current and perfect information when the population is analysed as a whole and when it is split into two, three, ..., *k* subgroups to the limit of explaining all existing heterogeneity and maximizing potential benefits by performing decisions at the individual level[42] (right hand side of the figure). This diagram reflects the idea of the existence of a continuum amongst decisions for each of these population stratifications. Additionally, it provides grounds for the two dimensions of the VoH: the *static* and the *dynamic* values[43].

If equal gains are observed with current information when subgroups are considered or not (gains A in Figure 6.6), the *static* VoH is equal to zero. If, when considering two subgroups, the NMBs obtained with current information are higher than the ones obtained for the average population (i.e. C > A), the static VoH is positive. Again when considering two subgroups, if the estimated NMB with perfect information is greater than the gains obtained for the population average (i.e. the quantity represented by B), the *dynamic* VoH is positive. If the specification used to define the subgroups facilitates the split of the targeted population and results in NMB distributions that do not overlap, little decision uncertainty is expected and it may imply a situation like the one illustrated with $EVPI_A > EVPI_{B2}$.

[42] In this conceptual diagram it is assumed that, when performing decisions at the individual level (*I* - individualised care), the total characterization of the heterogeneity is accomplished. This implies the absence of decision uncertainty and that the decision maker has all information possible to efficiently allocate resources given the complete knowledge of the individual characteristics that determine heterogeneity (including the counterfactual). Additionally, it is assumed that transaction costs associated with obtaining/exploring more granular subgroup specifications exist, but are neglectable. That is, the costs of analysing/exploring available data in order to obtain an optimal level of disaggregation is considered small and outweighed by the benefits of considering those subgroups specifications. See section 6.4 for further discussion on these issues.

[43] As described by Espinoza *et al.* (2011), when the decision maker wants to consider the value of exploring further the available data (i.e. current information) with a view to explore issues of heterogeneity, this is called the *static value of heterogeneity*, since this activity does not involve further data collection. On the other hand, the *dynamic value of heterogeneity* represents the absolute value of collecting further information when considering a particular level of subgroup disaggregation.

Figure 6.6 – Two forms of VoH: (i) the *static* value (represented by gain A and representing the NMB obtained with existing information for the average population, equivalent to the gains obtained when considering 2 subgroups); and (ii) the dynamic value (represented by the vertical distance B, where $EVPI_{B1} > EVPI_{A}$). The *x-axis* reflects the number of subgroups and the *y-axis* the NMBs. Reproduced from Espinoza *et al.* (2011) and Claxton (2011), with permission from the authors.



The following section will discuss issues relating to available evidence when quantifying the expected cost of uncertainty in a subgroup framework.

6.3.2 Available evidence and the value of further research in the presence of subgroups

Returning to the general scenarios discussed in section 6.2.2, and taking into account the need to consider the presence of subgroups, the use of IPD, compared to the use of AD only, may correct for bias and/or increase precision in relation to population average estimates. This section will analyse the impact of these effects now in the presence of subgroup effects. Scenarios 2.1 to 2.3 (listed below) will consider

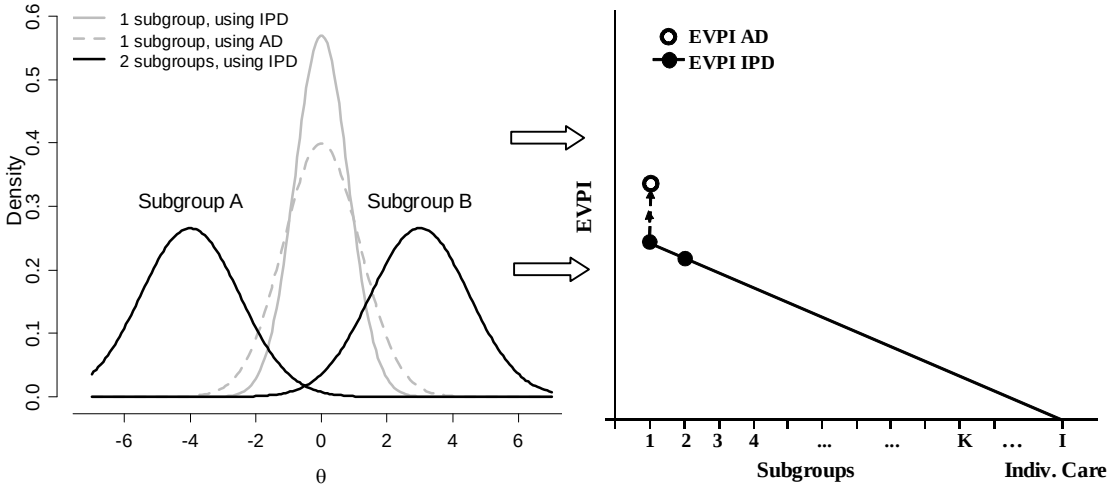
potential gains in precision. Given the multiplicity of directions in which the means of the treatment effect distributions (relating to each subgroup) may shift, and the unpredictable effects on the total (weighted average) EVPI estimation, the impact of bias is not presented here as it becomes essentially an empirical question. Following the rationale described in section 6.1.2, three situations may be considered in any type of clinical outcome setting (i.e. discrete or continuous): no subgroups (AD) *vs.* all subgroups (IPD) – scenario 2.1; some subgroups (AD) *vs.* all subgroups (IPD) – scenario 2.2; and all subgroups (AD) *vs.* all subgroups (IPD) – scenario 2.3. These are subject to discussion below.

Scenario 2.1: The use of IPD may facilitate the use of formal modelling of *treatment x covariate* associations[44], which, in particular circumstances, may not be attainable when using AD. In this scenario, the case is considered where EVPI for subgroups can be only estimated when IPD is available, allowing the appropriate quantification of uncertainty in subgroup related effect estimates. A possible graphical representation may be found in Figure 6.7a, where a distribution of effects for the overall population was obtained from both evidence formats. In this case, as with scenario 1.1, the EVPI is expected to be higher or equal with AD than with IPD (empty circle on or above full circle in Figure 6.7b – a reminder that IPD is being considered only to increase the precision of the estimates). Given that IPD has the ability to disentangle existing heterogeneity when stratifying, an estimate of EVPI for two subgroups is only obtained when using IPD. With two subgroups, different adoption decisions may be made for each subgroup, which implies distinct NMB distributions for each. Consequently, the EVPI is here expected to be lower when compared to the overall population (i.e. $EVPI^{IPD}_{k=1} > EVPI^{IPD}_{k=2}$, with *k* number of subgroups – as represented in Figure 6.7b). If the two subgroup specification does not facilitate the characterization of uncertainty and, consequently, NMB distributions overlap, higher decision

[44] While this is the most frequently encountered form of subgroup analysis (i.e. on treatment effect) it is not the only one, and indeed not the most important one in CEA. Modelling heterogeneity in baseline risk is by far the most used in cost effectiveness modelling.

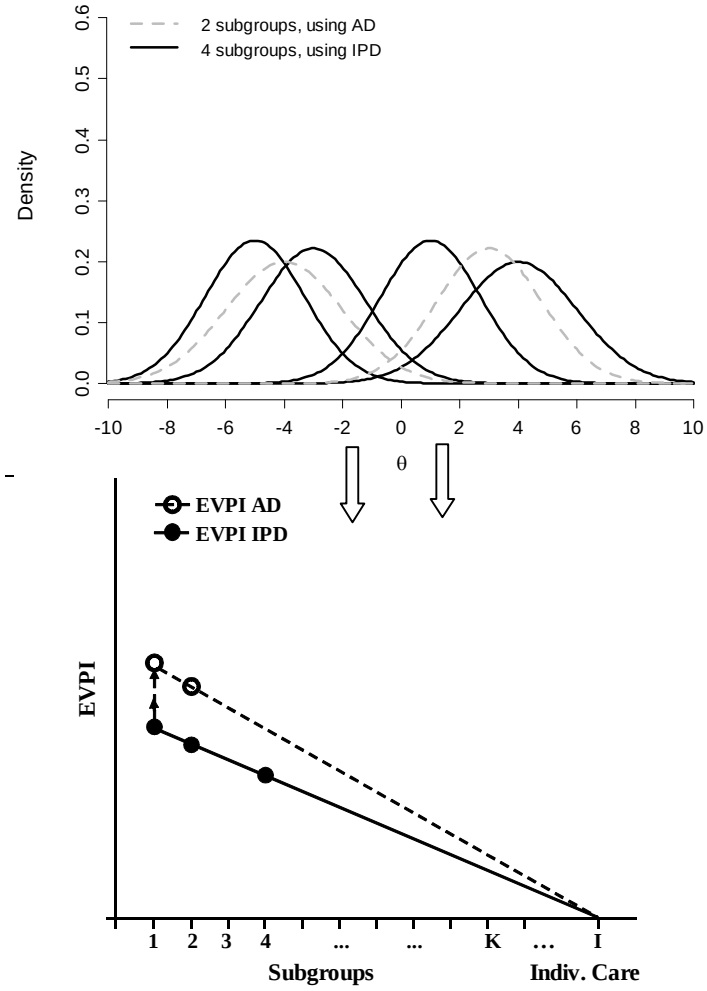
uncertainty may be expected, implying $EVPI_{k=1}^{IPD} < EVPI_{k=2}^{IPD}$ – a situation which is not represented graphically. As the use of IPD allows the characterization of heterogeneity, in extending this from two to *K* subgroups, to (*I*) individualised care, the situation illustrated in Figure 6.7b may be obtained.

Figure 6.7 – Representation of (a) considering existing heterogeneity in the identification of 2 subgroups and possible impact in the distribution of subgroup effects when IPD is considered and the inability of performing this analysis when using AD; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 2.1).



Scenario 2.2: If AD allow the estimation of *q* subgroup effects (through, for instance, meta-regression analysis), and the use of IPD facilitates resolving existing heterogeneity for *q* and *l* subgroups (where $q < l < k$, with *k* representing all possible subgroups), a situation as represented in Figure 6.8a may be obtained.

Figure 6.8 - Representation of (a) considering existing heterogeneity in the identification of 4 subgroups and possible impact in the distribution of subgroup effects when IPD is considered and the inability of performing this analysis when using AD, although possible for 2 subgroups; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 2.2).

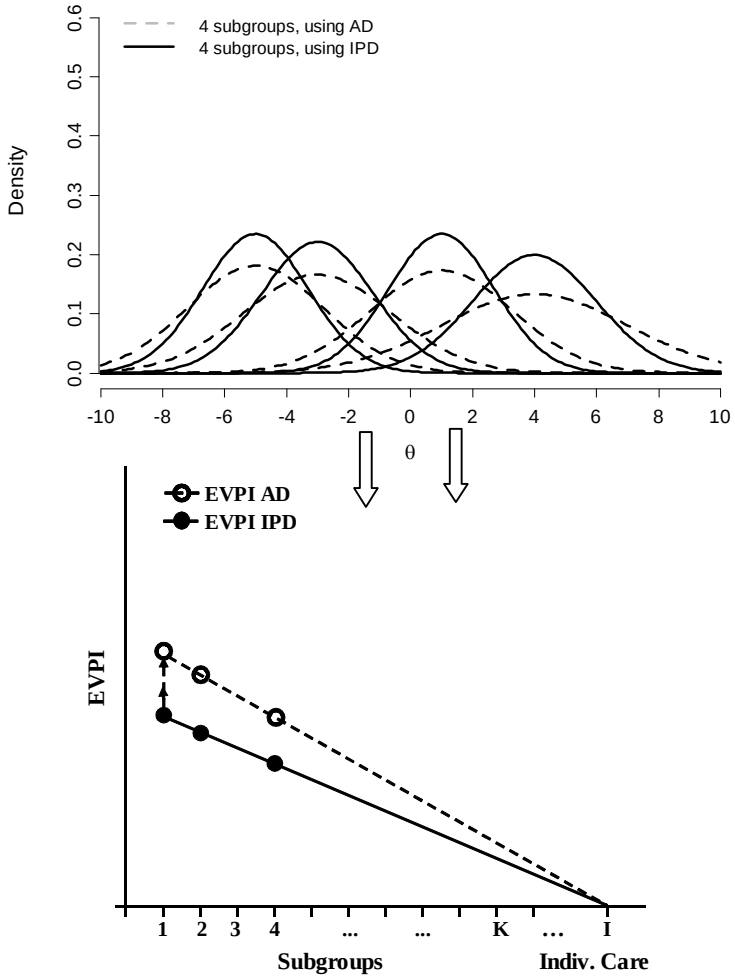


This diagram reinforces the idea that, if the interest is in performing analysis on *l* number of subgroups, this may be restricted by the type of evidence available (e.g. AD). If IPD is available, it may meet the necessary conditions to perform the analysis, with the additional potential benefit of obtaining estimates with greater precision (for *q* and *l* subgroup effect estimates) in comparison with the distribution of effects derived from AD. Replicating this scenario in real world decisions, where these are made on a

continuum, EVPI is expected to be lower or equal when IPD are available and used in the model compared to estimated EVPI when AD is used – Figure 6.8b.

Scenario 2.3: Assuming that AD also enables estimation of all required subgroup effects, a situation as represented in Figure 6.9a may exist. Less precision is expected in interaction estimates derived from AD when compared to IPD. This fact is translated in a higher capacity of IPD to explore heterogeneity (and reduce the level of uncertainty within each subgroup) – Figure 6.9b.

Figure 6.9 – Representation of (a) considering existing heterogeneity in the identification of 4 subgroups and possible impact in the distribution of subgroup effects when AD and IPD are considered; and (b) the possible consequences over the EVPI when using AD and IPD in these circumstances (scenario 2.3).



In the next subsection the motivating example results of using both AD and IPD (i.e. AD + IPD) to estimate the expected cost of uncertainty are shown.

6.3.3 Application: value of additional research in the presence of subgroups and in the context of having access to individual participant level data

The following analyses consider three alternative subgroup specifications separately, i.e. when considering the number of parents in the household and/or the parents' employment status. By using the formulations expressed in equations (6.7) to (6.9) (and Appendix 7), NB and EVPI estimates are obtained when using AD only and when using the mixture of AD and IPD to inform key effectiveness parameters. The alternative specifications analysed comprise:

a) the average population (no subgroups) – addressed in section 6.2.3;
b) two subgroups (k = 2) specification 1, defined by the number of parents in the family. The proportion of patients in each of the two subgroups is estimated from the available trial evidence (results across trials are averaged) and will be used further as the weights, w_k – in regards to the number of parents in the household, 71.0% of which are 2Ps and 29.0% 1Ps.
c) two subgroups (k = 2) specification 2, defined by the parents' employment status. In the samples of the trials used to evaluate effectiveness, 45.7% are 2Us and 54.3% have at least one parent unemployed in the family (defining w_k);
d) when the analyst/modeller has simultaneous access to information on the two binary covariates (i.e. number of parents in the household and their employment status), these can be used jointly to define four subgroups (k = 4). For each of the four subgroups the following weights (w_k) are used: two employed parents (2EP), 38.0%; two parents, at least one of them unemployed (2UP), 33.0%; single parent employed (1EP), 7.7%; and single unemployed parent (1UP), 21.3%.

Permutations of these four alternatives represent situations such as the ones described by scenarios 2.1 to 2.3 in section 6.3.2. For instance, comparisons of AD in a) with AD + IPD in b) or c) are those considered in scenario 2.1.

6.3.3.1 Further research: for which population?

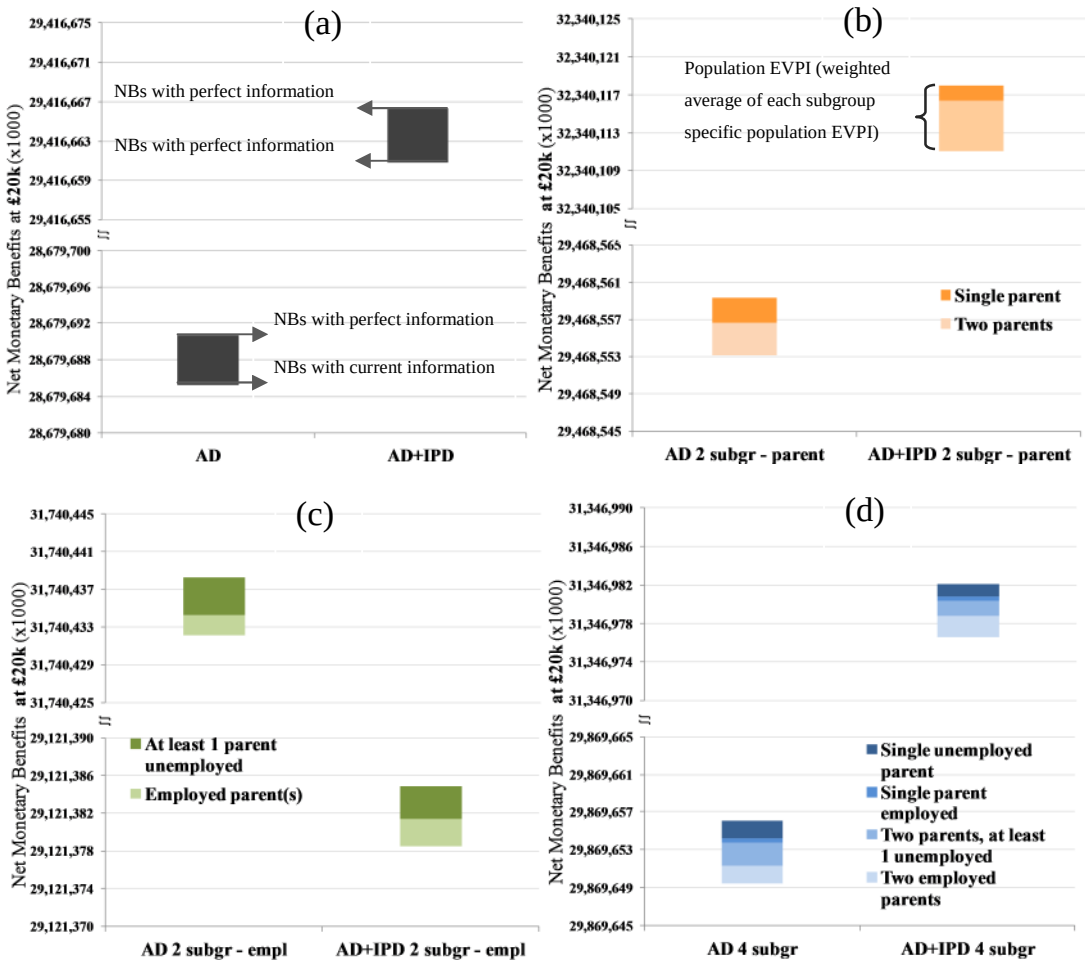
Figure 6.10 shows the estimated population NBs for a £20,000 threshold for each of the specifications a) to d) above. In this figure the situation in which covariate information is not considered is reiterated for completeness (i.e. Figure 6.10a repeats the graphical representation of Figure 6.5). In each graphical representation of Figure 6.10, the lower ends of the bars represent the NMB estimates obtained with current information. The upper ends of the bars represent the NMBs attained with perfect information. The population EVPI is the difference between the upper ends and the lower ends of the bars. The graphs considering subgroups (i.e. *b)* to *d)* above), have different shading for each population subgroup and each bar section corresponds to the contribution of a particular subgroup towards the total weighted population EVPI.

At a £20,000 threshold value, the same decision is being made when the population is split (i.e. rejection of any of the 'active' alternatives – see Chapter 4 and Chapter 5 for additional detail). In Figure 6.10 it is observed that the estimated (weighted) population EVPI is marginally higher when it is split into two subgroups than when it is considered as a whole – no subgroups ($PEVPI_{k=2} > PEVPI_{k=1}$, with *k* representing the number of subgroups – e.g. $PEVPI_{k=2}$ = £6,938 > £5,456 = $PEVPI_{k=1}$ for single parent status when using AD + IPD) [45]. This is true when using AD and AD + IPD to inform the decision model. However, and as expected, as the number of subgroups increases, population EVPI decreases ($PEVPI_{k=2} > PEVPI_{k=4}$, e.g. $PEVPI_{k=2}$ = £6,938 > £5,417 = $PEVPI_{k=4}$ for four subgroups when using AD + IPD). This is true when using AD + IPD to inform the decision model, but not verified when AD is used. With current information, higher benefits are obtained when considering two subgroups compared with considering the average population, implying a positive *static* VoH.

[45] Nonetheless, as different synthesis models are being used across the different subgroup scenarios, it is believed that care must be taken in interpreting these differences – see Appendix 7 and 8 for further details.

A potentially more appropriate analysis would employ both covariates simultaneously. When four subgroups are being considered and AD + IPD is being used, population EVPI estimates are observed to be marginally lower than when using AD estimates (i.e. $\text{PEVPI}_{k=4}^{\text{AD}} \geq \text{PEVPI}_{k=4}^{\text{AD+IPD}}$). This is consistent with the fact that using IPD resolves some of the existing decision uncertainty, lowering the EVPI.

Figure 6.10 - Population expected NMBs at a threshold value of £20,000 per QALY (2009 values) derived from the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD for when decisions are made for (a) the average population (repeated from above for completeness); (b) 2 subgroups, specification 1 – single parent status specification; (c) 2 subgroups, specification 2 – parent's employment status specification; and (d) 4 subgroups.



In graphs (b) to (d) of Figure 6.10, it can be observed that the shading areas within each of the bars, which represent the contribution of each of the subgroups, are of different sizes. This implies that different subgroups contribute differently to the (total weighted) population EVPI. Moreover, for the same subgroup, this estimated contribution is different if IPD is being considered or if it is not. To understand how this EVPI was estimated, it is of interest to disentangle this value in the individual subgroup expected values – Table 6.2. In Figure 6.10b (and corresponding 4th and 5th columns in Table 6.2), the effects of considering IPD are observed, with the contribution of each individual subgroup to the total EVPI estimate for the "single parent" subgroup specification differing when using AD and when using AD + IPD. While the use of AD indicates a small imbalance between the contributions of these subgroups (i.e. £2,668 and £3,576 for single and two parent families, respectively), the use of AD + IPD points to a stronger imbalance between these two groups of participants, with the 2Ps subgroup having an estimated value of further research of £5,337.

The EVPI estimates obtained when considering four subgroups – represented in Figure 6.10d and the last four columns of Table 6.2 – are of £6,591 for AD and £5,417 for AD + IPD. For AD + IPD, for instance, the subgroup for which the expected cost of uncertainty is higher is the 2EP subgroup, contributing £2,135 to the total of £5,417. The same is not true if AD estimates are used. In this case the 2UP subgroup is the one that would require a larger amount of investment (£2,402 to the total of £6,591). At the other extreme, the subgroup for which the estimated value of additional research is lower is 1EP with £427 and £480 for AD + IPD and AD, respectively.

Table 6.2 – Population EVPI at a threshold value of £20,000 per QALY (2009 values) derived from the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD for when decisions are made for: 1 subgroup (the entire targeted population); 2 subgroups – single parent status specification; 2 subgroups – parent's employment status specification; and 4 subgroups.

			Subgroups							
			2 subgroups		2 subgroups		4 subgroups			
	Effectiveness evidence (from MTC)	1 subgroup (population)	single parent	two parents	employed parent(s)	at least 1 unemployed parent	single unemployed parent	single parent employed	two parents, at least 1 unemployed	two employed parents
Population EVPI (£20,000 threshold value)	**using AD**	£5,414	£2,668	£3,576	£2,135	£4,003	£1,841	£480	£2,402	£1,868
	using AD+IPD	£5,456	£1,601	£5,337	£2,935	£3,469	£1,254	£427	£1,601	£2,135

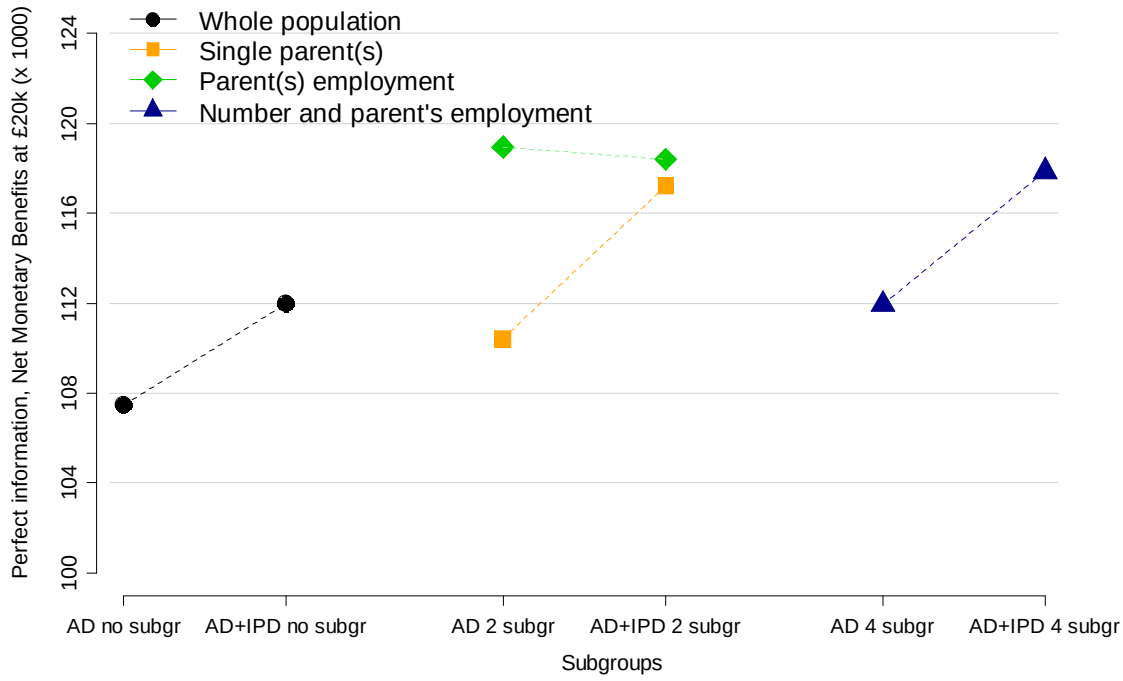
6.3.3.2 Further research: optimal number of subgroups?

The following exercise is intended to understand the optimal number of subgroups to be defined for intervention(s) approval and subsequent further research development. Underlying this problem are two main issues that need addressing: *(a)* if **no** further research is undertaken, how many subgroups should be defined? (akin to asking which specification maximizes the *static* VoH); and *(b)* if further research **is** undertaken, how many subgroups should be defined? (akin to having *dynamic* VoH).

In the case study, the NB estimates with current and with perfect information are fairly similar. Given this, only the perfect information case is represented graphically (Figure 6.11). Thus, a direct answer to question (b) above will be provided, whereas an answer to issue (a) will be provided indirectly. Figure 6.11 depicts the (individual level) NMBs obtained with perfect information for different subgroup specifications (i.e. situations a) to d) described above) and for when AD and AD + IPD estimates are used to inform the modelling. The reader is reminded that in the current assessment, the choice of the number of subgroups is limited to the definition of a maximum of four subgroups.

If four subgroups are considered, similar NMB gains are obtained, as in the two subgroup specifications (i.e. 'number of parents in the family' or 'parents' employment status'). With perfect information, it can be said that no apparent substantial additional gains are obtained if more than two subgroups are taken into account – answering question (b) above. If no further research is undertaken, that is, judgements are made with existing current information, the same conclusion can be derived – answering question (a) above[46].

Figure 6.11 – Expected (individual level) NMBs with perfect information (for £20,000 threshold values and 2009 values) for the cost effectiveness decision model for functioning smoke alarms. Results are shown for the use of AD and AD plus IPD effectiveness evidence to inform the decision model when heterogeneity is considered. That is, no subgroups, weighted average for 2 subgroups (2 specifications – number of parents in the family and their employment status) and the weighted average for the 4 subgroups are considered.



[46] Interpretations are assuming neglectable transactions costs or that these do not outweigh the above mentioned gains. If transactions costs are found to be relevant, decisions on performing further research and on the selection of the optimal number of subgroups must be made by taking into account both estimated gains and costs.

It can be observed that under perfect information, higher NMBs are attained if two subgroups are considered and employment specification used (approximately £119k). This is obtained when AD is used and, in fact, in the four situations reflected in the figure, there is the only one where NMB estimates for AD are found to be higher than for AD + IPD. Reasons underlying this phenomenon are related to: *(i)* the fact that different synthesis models are analysing different subgroup specifications, as described in section 6.1.2 and in Appendices 7 and 8; *(ii)* confounding factors and/or ecological bias affecting the AD estimates; and *(iii)* due to simulation error.

6.4 Discussion

This chapter examines the advantages of having access to evidence at the individual participant level when quantifying decision uncertainty with the purpose of estimating an upper boundary for the value of additional research. The rationale behind this is that parameter uncertainty is translated into decision uncertainty. However, different formats of the same evidence set (i.e. AD and IPD) may provide different distributional 'scenarios' about the same set of parameters. These will provide different estimations of existing decision uncertainty, which can be quantified in expected costs. Additionally, and making use of the expected NB approach, this chapter considers this framework in the context of subgroup analysis, reshaping the question of the value of acquiring additional research to: *for which population stratum is it valuable to conduct this additional research?*

One of the most important recent contributions in the field of heterogeneity and subgroup analysis in health care has been made by Basu and Meltzer (2007). These authors formalised the concept of EVIC, attaching it to the gain from improved decision making from understanding heterogeneity. Espinoza *et al.* (2011) extended the understanding of heterogeneity by considering its static and dynamic value – a

crucial distinction when questions about performing further research are at stake – and by considering the idea of the existence of a continuum amongst decisions when stratifying the targeted population. This chapter adds to these existing frameworks by assessing the advantages and disadvantages of using AD and IPD in appropriately disentangling existing heterogeneity. Starting by exploring the impact of these evidence types on the estimation of the need for further research, this chapter extends previous work developed in this area by highlighting the capabilities of available evidence in quantifying the value of further research in the presence of mutually exclusive subgroups. Additionally, and considering the existence of population subgroups, the framework discussed in this chapter allows the investigation of whether requesting further information for particular population strata is justified, and, moreover, how those judgements depend on the type of evidence (i.e. AD and/or IPD) being used in the analysis. It is argued that, conceptually, the use of an evidence set containing IPD is translated in a more appropriate characterisation of decision uncertainty compared to using AD only. It is pointed out that this is true when making decisions for the whole population and when splitting the population into subgroups.

Chapters 2, 3 and 5 highlighted the fact that exploring heterogeneity is problematic when AD only is available: analyses may be underpowered, the control for confounding may be limited and biased interpretations of the *treatment x covariate* associations may result. In addition, subgroup analyses can only be performed for those studies that publish equivalent subgroup information. For the majority of the studies, covariates are defined or stratified differently. The issues with using AD for this effect can be tackled when IPD is available for synthesis. This chapter highlights the advantages of using IPD to explore heterogeneity in CEA, compared to the use of AD only. Two main features of using IPD rather than AD were explored: gains in precision; and reduction/elimination of bias in decision model input parameter estimates. The effects of these over the estimation of the expected cost of uncertainty are discussed at length and compared and contrasted against the effects of using AD only.

In this chapter a motivating example is analysed where data stratification is conditioned by data availability on covariates. Like in previous chapters, the rationale behind each subgroup scenario analysed assumes that the covariate information available is the one used in the analysis (e.g. in the population average scenario, no subgroups are considered because it is assumed that information on covariates was non-existent). Moreover, it is not the purpose of this chapter to seek to determine suitable grounds for the stratification applied to the case study. In particular, no equity considerations are placed on the subgroup formation – as demonstrated by Coyle *et al.* (2003), imposing equity constraints regarding subgroup identification could provide the opportunity to assess if the costs incurred in inflicting these constraints would be worthwhile. These issues will be considered for future research.

In the motivating example used in this chapter, comparisons made across subgroup scenarios and between NMB estimates with current and perfect information obtained using AD and IPD (i.e. AD + IPD) have been performed with caution. This is because different synthesis models are used to analyse different subgroup specifications. The approach taken may have limited the comparisons' interpretability and is highlighted as the main factor contributing to some unpredictable results obtained for particular subgroup specifications (i.e. for the analysis considering the subgroup of 2U and of 1U). Nonetheless, an alternative approach is proposed which, as a starting point, considers the synthesis model results of two binary covariates (i.e. providing results for four population subgroups). This approach is described in Appendix 7 and is explored with a brief illustrative example. The application to the chapters' case study is presented in Appendix 8.

Throughout this chapter, EVPI estimates obtained by using AD and IPD have been compared and contrasted depending on a number of factors, the most important being the number of subgroups considered. However, it is highlighted that care must be taken in interpreting the differences between EVPI estimates coming from these two different data structures. To view this EVPI difference in terms of the value of acquiring additional evidence or of acquiring IPD rather than AD to resolve existing

decision uncertainty is incorrect, and the temptation to do so may be considered a limitation of the framework and of the analysis presented.

Important elements that need to be considered in the framework discussed in this chapter are the transaction costs related to exploring available evidence, understanding information about existing heterogeneity and also the costs attached to implementing treatment decisions for each subgroup. These, as highlighted by Espinoza *et al.* (2011), are expected to increase with the number of population subgroups being considered. In the framework discussed throughout this chapter, and in the case study used, transaction costs are assumed negligible or not outweighing the gains.

The availability of IPD may add an extra layer of complexity to the transaction costs equation, as the task of obtaining and exploring the IPD is in itself a burden. In addition to the costs of obtaining the IPD, to fully understand heterogeneity, new research may be needed in order to attempt to resolve it. Nonetheless, a single IPD (or even multiple) is not an 'unlimited' source of evidence. As more and more stratifications of the dataset(s) are performed (and hence, expected population health gains potentially increased), subgroup sample size decreases (smaller amounts of evidence available) and subgroup-specific uncertainty is expected to increase. Thus, the saturation of a particular dataset may be achieved at different stratification levels. All these (transaction) costs should be factored in when assessing the impact of using IPD when quantifying subgroup level decision uncertainty. These are not under consideration in the framework presented in this chapter.

The current work could be extended by evaluating the *expected value of perfect information for parameters*[47] (EVPPI). Such analysis focuses on prioritising the parameters for which further research would be most valuable (Briggs, A. H. et al., 2006). It would be interesting to understand how the EVPPI estimates would vary for the key effectiveness parameters, depending on the evidence format used (i.e. AD and

[47] Similar to EVPI, the EVPPI represents the value of reducing uncertainty surrounding a particular parameter or group of parameters in the decision model (Briggs, A. H. et al., 2006).

IPD), and comparing these with the EVPI estimate obtained in the analysis performed in this chapter (due to model input parameter correlation, it is expected that EVPI estimates obtained and the sum of EVPPI estimates will be similar but not equivalent). This issue will be considered in future research.

The framework developed in this chapter clearly forces the position of the standard economic analysis of health technologies currently being performed to a non-optimal choice for a number of reasons. In the first instance, obtaining the usual mean cost effectiveness estimates may not take into account underlying population heterogeneous factors. Regarding subgroup CEA may offer better decisions with respect to the optimisation of (limited) available health resources and maximize population health. Secondly, considering subgroups in the analysis also enables the understanding of which subgroup(s) justify investment of additional funds in order to increase certainty about approval recommendations. Finally, these two issues may only be attainable if the type and characteristics of the evidence used in the analysis are adequate to perform such tasks. The use of IPD rather than AD may influence not only the extent to which an appropriate understanding of heterogeneity is attained, but, more importantly, it may shape approval decisions for particular population subgroups and judgements in future research.

Overall, this chapter describes and explores some key issues relating to heterogeneity and subgroup analysis within the value of information setting and in the context of using IPD. The discussion of the results from a motivating example supplies insights over the added value of having access to IPD, in particular in the synthesis of evidence context.

CHAPTER 7

7. DISCUSSION

The overall purpose of this thesis was to contribute to the methodological debate on the use of evidence, particularly, the use of IPD and AD, for cost effectiveness decision analytic modelling in health care economic evaluation. In response to the objectives and research targets initially presented in Chapter 1, this chapter summarizes the contents, main findings, and contributions to knowledge of this thesis, as well as identifying future research possibilities.

7.1 Summary of the main thesis findings

Chapter 2 summarised the methodological and analytical issues in the use and synthesis of evidence for cost effectiveness modelling. The current state of the art in this field was presented, alongside a discussion of the challenges that analysts/modellers may face when dealing with evidence from a variety of sources to inform a range of model inputs. Through the development of a taxonomy, this chapter offered guidance on the appropriate synthesis methodologies to use for a given model

parameter. It also identified areas where further methodological and applied contributions are needed.

This review indicated that most evidence synthesis methods development in recent years has been carried out with respect to optimizing the use and synthesis of clinical evidence, often used to inform effectiveness decision model parameters. While for other, not less important, model parameters (such as the ones explicitly considered in this chapter: disease natural progression; cost/resource use; and health utility data), methodological advances have been fairly limited. This unbalanced rate of methods development across parameter types is intrinsically related to their characteristics, and the characteristics of the data that often informs them.

Chapter 2 was devoted to pursuing the first research target of this thesis, that is, to review the literature with respect to methods and applications on the use of evidence for the economic modelling of health care programmes, setting the scene for the four subsequent chapters.

When populating a decision model and, in particular, when populating effectiveness model input parameters, it is often the case that IPD may only be available for a proportion of the relevant studies which constitute the evidence base. Chapter 3 explored the methods for the synthesis of binary effectiveness evidence. This exploration included recent modelling frameworks which allowed both IPD and AD to be jointly modelled in the estimation of effect(s) of intervention(s). Given the apparent non-existence of synthesis methodology for multiple interventions' comparison considering IPD, Chapter 3 described and discussed developed novel modelling approaches to not only model IPD on its own, but also in conjunction with summary-level data. The development of novel synthesis models tackled some of the methodological literature gaps identified in Chapter 2.

The methods developed in Chapter 3 also considered extensions to allow covariate information at both participant- and study-level to be included, raising important methodological issues with respect to the estimation of covariate effects. Through the

use of IPD, whether or not in combination with summary evidence, estimates of the covariate effects obtained were more precise and of greater accuracy, when compared to those estimated through AD only. In addition, assumptions about the covariate effect were tested. The MTC model, assuming different but exchangeable *treatment x covariate* interactions, was shown to provide the 'best' fit in the situation where scarcity of data is evident, compared to modelling without considering exchangeability of regression slopes.

Furthermore, the synthesis of evidence at the individual level allowed estimating distinct covariate effects based on between- and within-study variability. Extending the developed synthesis methods to consider these interactions provided a formal approach for assessing the presence and impact of aggregation/ecological fallacy bias. The issue of having limited availability of covariate information at both participant- and study-level was explored and was dealt with by making use of the multiple imputation features available through the MCMC simulation in order to impute the missing covariate information. By not discarding evidence for which no covariate information was available, comparisons between models with and without covariate effect estimation were enabled.

Chapter 3 methodological findings were supported by the use of a motivating example within the PH accident prevention scheme. The assessment of the effectiveness of interventions aiming to increase the uptake of functioning smoke alarms in households with pre-school children showed that more 'intense' interventions were identified as more effective than less 'intensive' ones, with the programme providing education, low cost or sponsored equipment, equipment fitting and inspection having the highest level of effectiveness from the set.

Chapter 3 provided an answer to one of the stipulated research targets which aimed to use all relevant evidence by exploring and enhancing the use of IPD to best (simultaneously) evaluate the effectiveness of multiple intervention options.

Chapter 4 showed how the issues identified in Chapter 2 and addressed in Chapter 3 were also relevant for CEA. This is an applied addition to the thesis, which showed the impact on the cost effectiveness outcomes of using AD only. The chapter provided the decision modelling grounds for the remainder of the thesis, as the question of whether interventions to promote the uptake of functioning smoke alarms are cost effective is raised, and if proved to be so, are worthy of public investment. Economic evaluation is introduced as a tool to inform decision-making in this regard and a brief review of existing economic evidence on this subject is provided. To the author's knowledge, this was one of the first times a MTC framework was used to evaluate the effectiveness of PH programmes. It was also one of the first times a simultaneous assessment over the cost effectiveness of seven programmes was made towards the final outcome of reducing the occurrence of household fires and related child injuries.

Despite the fact that the decision model developed in Chapter 4 was informed by summary-level effectiveness evidence only, important findings were made about the cost effectiveness of the interventions at the population and at the subgroup-level. In the base case scenario (i.e. using the public sector viewpoint), it was shown that, if a £30,000 threshold ratio were to be used, the same decision would be provided whether for the population average or for subgroups, that is, rejection of all 'active' interventions. Additionally, scenarios were incorporated in this chapter, which evaluated different (analysis) viewpoints and timeframes. These scenarios showed that taking an NHS and PSS perspective led to similar mean ICER estimates for dominant interventions and, consequently, came to the same decision as for the base case (i.e. public sector viewpoint).

Chapter 5 extended the case study cost effectiveness assessment performed in Chapter 4 to the evaluation of the impact of using different formats of evidence. Chapter 5 looked at one of the goals from the second research objective set out in Chapter 1, that is, the assessment of the added value of having access and using IPD for cost effectiveness decision making, by exploring approval decisions for the average population and population subgroups. The simultaneous use of IPD and AD was found to be particularly useful in clarifying decision making for subgroups, when

compared to the use of AD only. This followed from the ability of the synthesis models used to model IPD in appropriately estimating *treatment x covariate* associations which enhanced the possibility of performing more accurate judgements over interventions' cost effectiveness. In general, when subgroup evaluations were made, decision uncertainty was found to be higher for a particular threshold range, when using aggregate effectiveness evidence only, compared to using IPD. Subgroup cost effectiveness estimates derived from using IPD, lead, in principle, to better decisions, and thus, to better use of limited health resources, when considering population heterogeneity.

As in the previous chapters, Chapter 6 again highlights the advantages of having access to evidence at the individual level, compared to having access and using AD only. This chapter discusses the advantages of using IPD in the estimation and quantification of decision uncertainty. This is done with the purpose of estimating an upper boundary for the value of additional research, providing an answer to the objective set out in Chapter 1 of wanting to perform judgements over the feasibility of funding additional research at the population level and at different population strata.

A series of conceptual scenarios were presented, where the use of IPD, compared to AD, were assumed to lead to an increase in precision, reduction/elimination of bias, or both. The expected effects of these factors on the estimation of the expected loss of the expected opportunity loss were discussed. The conceptual scenarios presented considered the absence and the presence of population subgroups. The important issue that different decisions may be recommended for different subsets of the population, and that these may well maximize population health in a more efficient way than population average judgements, was discussed. Additionally, issues surrounding the estimation of the value of further research in each of these population subsets were considered. The framework proposed in this chapter showed that the use of an evidence set containing IPD may be translated in a more accurate estimation of decision uncertainty for subgroups of the population compared to using AD only. Generally, IPD generates more robust estimates of the (total weighted) population EVPI than the use of AD only, particularly when exploring population heterogeneity.

Overall, this chapter described and explored some key issues relating to heterogeneity and subgroup analysis within the value of information setting and in the context of having access and using IPD.

7.2 What could have been done differently?

Taxonomy characteristics

The taxonomy considers three dimensions – number and format of data sources and number of parameters to inform; though these may not capture all possible scenarios, potentially limiting its guidance role in some circumstances. The taxonomy could have been developed to include other (not less important) dimensions or sub-dimensions, such as taking into account covariate information within each taxonomy section.

Evidence informing cost effectiveness decision modelling

A systematic review process is at the basis of the effectiveness evidence base used to inform the cost effectiveness modelling. The same kind of methodical process was not put into practice to obtain evidence to populate other decision model parameters. This procedure is not considered necessary for all types of information in economic modelling, as highlighted in the NICE PH methods guidance (2009). Nonetheless, engaging in a more exhaustive search process would potentially provide additional short- and long-term evidence, particularly with respect to costs and consequences for cost effectiveness modelling.

Subgroup analysis and exploring heterogeneity

The systematic reviews (Wynn et al., 2010, Kendrick et al., 2007) which identified the studies of the motivating example used in this thesis explored the association of the

risk of injury with six explanatory factors (i.e. child age, gender, ethnic group, single parent status, residing in rented accommodation and employment status). However, only two of these factors were chosen for analysis in this thesis (i.e. single parent family status and employment status), because these two covariates were the ones for which more information was available – that is, the choice was data driven. No aims were set to determine what were and were not suitable grounds for the covariates included in the modelling and, consequently, for the stratifications applied to the case study. It would have been interesting to explore this issue further, irrespective of the data limitations identified.

7.3 Recommendations for accessing and analysing individual-level evidence

One theme running through this thesis has been the advantages of having access and analysing IPD. In fact, in cost effectiveness analyses multiple sources of evidence are often used for which, in most cases, only the AD evidence is available. However, the challenge imposed is to understand when it is worth obtaining individual patient data – i.e. to IPD or not to IPD? Should it be when the 'relevant' information (e.g. on adverse effects, population groups) is not reported in related published papers? Or, regardless, when purely wanting to increase precision and/or reduce bias in the exploration of individual level characteristics and how these affect approval decisions? Should a threshold be set for the amount of evidence needed (and its disaggregation level) to appropriately address a decision problem at hand?

Analysts / modellers should consider which methodological factors are expected to influence results in their particular problem setting. Additionally, in deciding whether to obtain the IPD, one should equate the burden in terms of time and resources needed in the analysis, as well as the constraints in terms of expertise required to conduct these analyses.

This thesis claims that, in most circumstances, there are advantages in having access and analysing IPD. It showed that IPD has clear value over summary data for both synthesis and decision modelling aspects of the analysis. A key advantage is that individual level data provides a wealth of data that facilitates the specification and evaluation of cost effectiveness for particular population subgroups. I believe these benefits are important in achieving consensus in approval decisions by multidisciplinary decision makers.

7.4 Recommendations for future research

The taxonomy developed in Chapter 2 focused mainly on clinical effectiveness evidence, discussing also the application of the taxonomy to other key input parameters in any economic model (i.e. disease natural history, resource use/costs, and preferences). It is encouraged *(i)* a fuller application of the taxonomy to all model parameters in future modelling; *(ii)* the continued exploration of the adequacy of the taxonomy dimensions to the individual characteristics of other decision model parameters; and *(iii)* the maintenance of the characteristics, contents and dimensions of the taxonomy, to continue viewing it as a tool designed to help health economics modellers to adequately address the research problems faced.

As highlighted throughout this thesis, there is a set of advantages in considering evidence at the individual level, when compared to the use of AD only. These advantages are derived mainly from the additional flexibility provided by this evidence format to explore and answer clinical and economic research questions. Nonetheless, there is a series of obstacles relating to IPD, from the difficulty in having access to this sort of data to the burden of having to explore and model large datasets. However, and before engaging in trying to access or explore IPD, it would be interesting to investigate the idea of a mechanism or framework that would allow the quantification of the benefits of considering IPD beforehand. For instance, quantifying the gains that would be obtained if IPD enabled the appropriate selection of relevant

covariates and subgroup formation (adequately explore existing heterogeneity) and the extent to which it would correct existing covariate effect biases.

A series of meta-analytic models, including novel MTC models, were developed and implemented in Chapter 3. These allowed the synthesis modelling of IPD and of AD and IPD simultaneously in a binary outcome setting. The use of a motivating example within the PH setting allowed the application of the novel synthesis models to real world data – and to a great extent, these developments were driven by the challenges presented to me by this particular example. Although the novel methods are generalisable to other situations (of the same characteristics as the motivating example), it would be interesting to apply these to other case studies or to simulated data – although it is acknowledged that many considerations are needed when designing and generating simulated data (Burton et al., 2006).

It would be interesting to extend the developed modelling work to explicitly appreciate the type and quality of the study information being synthesised – following up on bias adjustment proposals already existing in the current literature (Turner, R. M. et al., 2009, Welton et al., 2009, Spiegelhalter, David J. & Best, 2003). This thesis acknowledges that the inclusion of study-level covariates in the synthesis modelling may explain some of the between-study heterogeneity and reduce inconsistency in the network – reinforcing some of the recent work published in this area (Cooper, Nicola J. et al., 2009). Nonetheless, and following up on the recent work by Dias and colleagues (Dias et al., 2010b), it would be interesting to extend this work to explicitly evaluate consistency across the evidence base, considering that a proportion of this evidence may be available at the individual level. Last but not least, a natural extension of the synthesis modelling developed could be to consider a generalized linear modelling framework, broadening the approach to other types of outcome measures (e.g. categorical, continuous, time to event).

A keystone of any economic evaluation is the perspective employed by the analysis. This dictates the range of costs and consequences included in the evaluation. Different perspectives foster different outcomes, which may result in different resource

allocation decisions. This thesis follows the NICE recommendations (2009) for the evaluation of PH interventions by using a public sector perspective as base case. Using this broader perspective, it enables the identification and inclusion of the majority of the consequences. This analysis allows the exploration of the impact of taking a narrower perspective, such as the usual health services viewpoint, which may be required by a decision maker. The development of consistent methods for the measurement and valuation of the broader inter-sectoral consequences of PH interventions is a topic for further research.

Finally, alternative approaches to subgroup analysis have been proposed in the recent literature which could have been followed in this thesis. For instance, the framework introduced by Basu and Meltzer (Basu & Meltzer, 2007), cited in Chapter 6, could have been applied to analyse the IPD with a view to quantifying residual unexplained heterogeneity and the EVIC. This would have allowed the estimation of the value that society is willing to pay so that individually efficient decisions are made, optimizing the resource allocation process.

7.5 Conclusions

This thesis explored the use of evidence, and in particular, the use of IPD in all the different stages of an economic evaluation: the synthesis of the evidence, cost effectiveness and analysis of subgroups, and value of further research. Although the author would like to see the use of IPD to be made compulsory (or, at least strongly recommended) in certain HTA assessments, it may be too early to make such recommendations. It is hoped that the research conducted throughout this thesis provides guidance for a better use of available evidence, bridges some of the existing gaps in the methods literature and stimulates the broader use of IPD in the conduct of more and better economic evaluations of health care and/or PH interventions.

APPENDICES

Appendix 1

Relevant issues in decision analytic modelling for health care

Defining the decision problem

A clear identification of the question to be addressed in the analysis must be made. This includes defining the target population or population subgroup(s) and the relevant options subject to comparison. The location and setting in which the alternative interventions are being delivered should also be detailed.

Structuring the decision model

After specifying the problem at hand the structure of the model can be outlined, conditional to predefined boundaries. These boundaries may be based in the availability of data, or translated in limiting the modelling to possible consequences of the options under evaluation – parsimony rules apply. The nature of the interventions under assessment and their potential consequences will be key elements in shaping the model.

Perspectives, time horizon and discounting

The viewpoint assumed in an analysis is an important factor as economic outcomes may vary largely depending on what perspective is specified. The possible set of perspectives includes broader views, as the societal ones, to narrower ones, such as the patient. As set out by the NICE methods guidance (NICE, 2008), technology appraisals in the UK should consider only costs incurred by the NHS and PSS budgets

in their base case analysis. Nonetheless, evaluations of PH interventions may consider effects on sector budgets other than health (NICE, 2009).

Technologies, programmes or services comparative assessment is usually made at a specific point in time. The timing of costs and consequences related to these programmes must be considered even if they occur in the future. Thus, these should be long term enough in order to consider all of the main health and economic consequences (Drummond et al., 2005).

It is instinctive to think that future costs and benefits should be reduced or 'discounted' to reflect the fact that they should not have the same weight in decision-making as those incurred in the present. The time preference concept underlies this idea; that is, there is a preference for having the resources now rather than in the future, with the advantage of benefiting from them in the meantime. It is important to highlight that, in recent years, much has been debated and written about considering equal or differential discount rates for health outcomes and costs (Claxton, K. et al., 2011, Gravelle et al., 2007, Claxton, K. et al., 2006, Gravelle & Smith, 2001). The current discount rate recommendation from the NICE is that both costs and effects should be discounted at the rate of 3.5% (NICE, 2008). The NICE official guidance on discounting is being used in this thesis.

Identifying and synthesising evidence

When an adequate structure of the model has been established and the discounting mechanism, perspective and time horizon of the analysis decided on, the task of identifying available relevant evidence to populate it starts. This procedure should make use of the general principles of evidence-based medicine set out by Sackett *et al.* (1996). Guidelines exist for systematic identification and quality assessment of effectiveness evidence (CRD, 2009). With respect to other decision model parameters (e.g. resource use and utilities), modest contributions to the literature have been made (these are highlighted and discussed throughout the thesis). When a set of sources are available to inform a specific decision model parameter, evidence synthesis plays an

important role and is currently seen as a requirement for decision modelling. (Sutton, A. J. & Abrams, 2001, Sutton, A. J. et al., 2000a).

Evaluating variabilities, uncertainties and heterogeneities

As uncertainty surrounds model estimates, any decision based on cost effectiveness will also be uncertain (NICE, 2008). Models and uncertainty go hand in hand: uncertainty looms not only in the input evidence that informs them but also in their selection and quality assurance. When quantifying uncertainty, it is important to distinguish between patient variation within a cohort and uncertainty in knowledge. Variability or "first-order" uncertainty refers to between-individual diversity in the population to which the policy is applicable, reflecting "chance" (Stinnett & Paltiel, 1997). Not being a source of uncertainty, heterogeneity relates to the differences between patients that can be explained, e.g. differences between sub-groups of patients (Briggs, A. H. et al., 2006). Uncertainty can arise from parameters (Spiegelhalter, David J. & Best, 2003, Briggs, A. H., 2000). Parameter uncertainty, or "second-order" uncertainty, usually refers to the uncertainty surrounding an expected value, i.e. the fact that there is never certainty as to the expected costs and effects when the treatment is provided for a particular population (Briggs, A. H. et al., 2006). Consequently, it is reducible with more evidence. Uncertainty may also arise from the structure used in the decision model. Structural uncertainty reflects the assumptions imposed by the modelling framework and has received relatively little attention in the literature (Bojke et al., 2009, Philips et al., 2006).

In cost effectiveness, uncertainty in its outcomes is an opportunity for a technology (or technologies) to be funded such that population health is not maximised. This raises an opportunity cost to uncertainty in terms of health forgone. The evaluation of uncertainty in decision models is often called probabilistic sensitivity analysis (PSA) or probabilistic modelling. The ground of probabilistic modelling is to reveal the uncertainty in the input parameters and illustrate its consequences on the outputs of interest (Briggs, A. H. et al., 2006). However, the mathematical relationship between inputs is often too complex to return the exact distribution of the estimator for the cost

effectiveness measure. In that case, Monte Carlo simulation procedures are often used to quantify uncertainty over the expected outcome measures. Using the joint distribution of costs and effects and through the use of Bayesian decision theory it is possible to address questions on whether current available evidence is considered sufficient to support the choice of a specific intervention and whether additional research is valuable.

Assessing the value of additional research

Decision uncertainty may address the question of whether additional research is required or should be undertaken. Probabilistic modelling outputs can be used as a vehicle to quantify the cost of making a wrong decision or delaying it. This entails patients' forgone health gains and the waste of available resources. The difference between obtained net gains and those that could have been attained by reimbursing the optimal technology (or set of technologies) represents the opportunity cost of uncertainty, the EVPI (Claxton, K. P. & Sculpher, 2006, Ginnelly et al., 2005a). Carrying out further research is meaningful if this estimated quantity surpasses the costs of conducting such research. Where estimated EVPI signifies that undertaking additional research is potentially worthwhile, further questions may be of interest with respect to the type of research that is more valuable. Decisions are also required as to the optimal design of the additional research. Value of information methods may be extended to account for patient allocation, trial quality, and trial design issues (Claxton, K. & Thompson, 2001).

Appendix 2

Bayesian methods

The Bayesian approach to statistics (Lee, 1997) is well established in the literature and is used in a number of areas of health care research, including data synthesis (Sutton, A. J. & Abrams, 2001, Sutton, A. J. et al., 2000a) and the economic evaluation of heath care technologies (Ades, A. E. et al., 2006b, Spiegelhalter, David J. & Best, 2003, O'Hagan et al., 2001, Spiegelhalter, D. J. et al., 2000, 1999). Within the evidence synthesis framework, Bayesian methodology has made solid progress, being used extensively over the last decade (Sutton, A. J. & Abrams, 2001, Sutton, A. J. et al., 2000a). In fact, economic evaluation research has benefited from these advances, with the implementation of Bayesian decision analysis as an explicit working platform and, particularly, with the increased use and requirement for probabilistic analysis (NICE, 2008, Claxton, K. et al., 2005).

Bayesian methods can be considered an alternative to the classical (frequentist) approach to statistical inference. It is a more appealing and intuitive framework, since both the data and model parameters are considered as random quantities. The key feature is the likelihood function, that defines how reasonable are the data given values of those model parameters. The increased use of Bayesian methods is also linked to advances in computational methods, allowing for a more efficient computation in combining non-conjugate probability distributions. A key difference between the two approaches is that Bayesian methods allow the model to incorporate external information alongside available data – prior distributions. When very little or no information is accessible, subjective, non-informative or 'vague' beliefs are set as priors.

In Bayesian theory, the joint prior probability density function for all model parameters, $p(\theta)$, is combined with the study data, *y*, in the form of a likelihood function, $L(\theta|y)$, to obtain the joint posterior probability density function, $p(\theta|y)$.

This is entitled Bayes' Theorem (Lee, 1997), that is, $p(\theta|y) \propto L(\theta|y) \cdot p(\theta)$. The following subsections will describe Bayesian simulation techniques and their characteristics.

MCMC simulation and WinBUGS software

Simulation based methods, like MCMC, have been broadly used (Carlin & Louis, 2009, Gelman, 2004, Carlin & Louis, 2000, Gilks et al., 1998). One of these methods, Gibbs sampling, has received significant attention in the last few years, with the algorithm being used to generate a sequence of samples from the joint probability distribution of two or more random variables.

WinBUGS, an interactive Windows version of the BUGS (Bayesian inference Using Gibbs Sampling) software, performs Bayesian analysis of complex statistical models using Gibbs sampling. Using simulation, inferences need to consider the initial transient behaviour of the chain(s) and the consequent need for visual or numerical assessment of convergence. The initial set of iterations must be discarded, as these are not independent (i.e. 'burn-in' period). For each inferential procedure, a series of random samples are obtained from parameters' joint posterior distributions. To make sure that these come from a stationary distribution, this thesis generally considers an initial set of 10,000 iterations of the MCMC sampler as 'burn-in' and a subsequent set of 5,000 iterations for inferential purposes.

MCMC simulation and multiple imputation

In Bayesian inference, MCMC is one of the primary methods for generating multiple imputations in nontrivial problems (Schafer, 1997). When imputing missing information, MCMC generates independent draws of the missing data from its predictive distribution. Multiple imputation through MCMC techniques is attractive for exploratory or multi-purpose analyses involving a large number of estimands. Rather than using a small number of imputed data sets (as frequentist approaches do), MCMC makes it possible to take an independent draw from the imputation dataset

within each iteration. One advantage of this is that the posterior standard errors of regression coefficients already summarize the uncertainty about the process for the missing data and also the uncertainty about the coefficients themselves. The disadvantage is that the imputations will be much slower and care in checking convergence is needed. Thus, although WinBUGS modelling is attractive owing to its practicality, careful checking of convergence and sensitivity to prior distributions and initial values is needed when using its imputation features.

Model selection and criticism

As with any inferential technique used to explore the complexities of real world data, comparison of alternative model formulations is important in identifying those which appear to adequately portray data information. In the classical modelling framework, model comparison is usually performed through the use of two quantities, a measure of model fit and of its complexity. As an increase in model complexity (i.e. an increase in the number of free model parameters) is attended by an improvement in model fit (in general defined by a deviance statistic), a trade-off situation arises, as described in the early work of Akaike (1973).

The DIC (Deviance information criterion (Spiegelhalter, D. J. et al., 2002)) is a generalization of the AIC (Akaike information criterion (Akaike, 1973)) and the BIC (Bayesian information criterion (Schwarz, G., 1978)) and is considered a useful criteria for Bayesian model comparison. The DIC can be considered a Bayesian measure of fit or adequacy, penalized by additional model complexity. The DIC statistic, $DIC = \overline{D} + p_D$, comprises a classical estimate of model fit, posterior expectation $\overline{D} = E_\theta[D(\theta)]$, where $D(\theta)$ is the deviance $\left[-2\log L(\theta|y)\right]$, plus the effective number of parameters, $p_D = \overline{D} - D(\hat{\theta})$, where $D(\hat{\theta})$ is the likelihood at the posterior expectation of a stochastic node, θ. The larger is $\overline{D}$, the worse the fit; the larger is p_D, the better the model will fit the data.

Advantages and disadvantages of the Bayesian approach to meta-analysis

Some authors have assessed the adoption of a Bayesian approach to meta-analytic modelling (Sutton, A. J. & Abrams, 2001, Sutton, A. J. et al., 2000a, Su & Po, 1996). The majority of the advantages of a Bayesian approach highlighted next constitute the reasons for opting for this methodological approach and these may be depicted throughout this thesis.

Adopting a Bayesian approach: (a) enables evidence from a variety of sources, regarding a specific problem, to be taken into account within a coherent modelling framework; (b) may consider prior beliefs to be included in the modelling in the form of prior distributions – not considered in this thesis as all priors are intended to be vague; (c) considers parameter uncertainty jointly by automatically accounting for it in the analysis; (d) allows probability statements to be made directly regarding quantities of interest – this is particularly useful if wanting to rank a series of interventions with respect to the probability of being best; and (e) leads naturally into a decision theory framework which may also consider costs and utilities regarding health care / public health decisions.

Nevertheless, as for any modelling or simulation exercise, there are disadvantages. Some of these are: (a) using non-informative priors may place impartiality and objectivity further away; (b) different (vague) prior distributions may be considered in the same modelling framework which may generate varying results (this means that, when using a Bayesian framework, sensitivity to the specification of the prior distributions is important); and (c) the modelling may be computationally complex and intense to implement, and therefore time consuming to perform.

Appendix 3

WinBUGS codes used to combine the two data formats, AD and IPD, including a binary patient covariate.

This code relates to model (3.9) described above and is designed to be as generic as possible and easy for the user to modify to adapt to specific applications. For example, if no data exist in one or more of the sections, the corresponding section of code can simply be deleted from the model.

Six datasets/ data-files are required to fit the complete model: two containing constants, one which indexes study treatments and specifies study baseline treatments, two for IPD and one for the AD. It should be noted again that for the clustered allocated AD adjustments should be done prior defining WinBUGS data model. All data should be loaded before the model is compiled.

Due to size and agreements of use, the original data sets are not included in their entirety, but a couple of lines of data are supplied for each study/data combination for illustration purposes.

```
model {

### Part 1: Model for non-clustered IPD trial data ###

for(i in 1:n.non.cluster.subjects) {
# Likelihood for non-clustered IPD data
outcome[i] ~ dbern(p[i])
# Model for non-clustered IPD data
logit(p[i]) <- mu[study[i]] + delta[index[i]] * (1 - equals(treat[i], baseline[i])) +
                    beta_cov[study[i]] * cov[i] + beta.w[index[i]] * (cov[i] - meancov[i]) *
                    (1-equals(treat[i],baseline[i]))) +
                    beta.b[index[i]] * (1-equals(treat[i], baseline[i])) * meancov[i]
          }

# Vague priors for non-clustered IPD
for(j in 1:n.ipd.non.cluster.trials) {
```

```
        beta_cov[j]~dnorm(0, 1.0E-6)
        mu[j]~dnorm(0, 1.0E-6)
        }


### Part 2: Model for IPD cluster trial data  ###

for(i in 1:n.cluster.subjects) {
# Likelihood for cluster IPD data
c.outcome[i] ~ dbern(c.p[i])
# Model for cluster IPD data
logit(c.p[i]) <- c.mu[c.study[i], c.cluster[i]] +
                delta[c.index[i] + n.non.cluster.arms] * (1 - equals(c.treat[i], c.baseline[i])) +
                c.beta_cov[c.study[i]] * c.cov[i] +
                beta.w[c.index[i] + n.non.cluster.arms] * (c.cov[i] - c.meancov[i]) *
                (1-equals(c.treat[i], c.baseline[i])) +
                beta.b[c.index[i] + n.non.cluster.arms] * (1 - equals(c.treat[i], c.baseline[i])) *
                c.meancov[i]
        }

# Vague priors for cluster IPD
for(i in 1:n.ipd.cluster.trials) {
        c.beta_cov[i]~dnorm(0, 1.0E-6)
        # Random-effects for clusters in IPD
        for(j in 1:n.cluster.max) {
                c.mu[i, j] ~ dnorm(mu.mean[i], inv.tau.sq.mu[i])
}
        mu.mean[i] ~ dnorm(0, 1.0E-6)
        inv.tau.sq.mu[i] <- 1 / (sigma.mu[i] * sigma.mu[i])
        sigma.mu[i] ~ dunif(0,2)
        tau.sq.mu[i] <- sigma.mu[i] * sigma.mu[i]
        }

for(i in 1:(n.non.cluster.arms + n.cluster.arms)) {
        md[i] <- d[treat1[i]] - d[baseline1[i]]
# Random-effects IPD trial-specific LORs
delta[i] ~ dnorm(md[i], prec)
        beta.w[i] <- bw[treat1[i]] - bw[baseline1[i]]
        beta.b[i] <- bb[treat1[i]] - bb[baseline1[i]]
}

# Part 3: Model for non-cluster and cluster aggregate data #

for(i in 1:n.agg.arms) {
# Binomial likelihood for AD data
        outcome.ad[i] ~ dbin(pa[i], n[i])
# Model for AD data
logit(pa[i]) <- mu.ad[a.study[i]] + delta[i + n.non.cluster.arms + n.cluster.arms] *
                (1 - equals(a.treat[i], a.base[i]))

delta[i + n.non.cluster.arms + n.cluster.arms] ~ dnorm(md.ad[i], prec)
```

```
        md.ad[i] <- d[a.treat[i]] - d[a.base[i]] + (bb[a.treat[i]] - bb[a.base[i]]) * a.cov[i]
        }

# Vague priors for AD trial baselines
for(j in 1:n.agg.trials) {
        mu.ad[j] ~ dnorm(0, 1.0E-6)
        }

### Part 4: Model for combining all treatment effect estimates ###

# Vague priors for basic parameters
bw[1] <- 0
bb[1] <- 0
d[1] <- 0
for (k in 2:max.treat) {
        bw[k] ~ dnorm(m.betaw,prec.betaw)
        bb[k] ~ dnorm(m.betab, prec.betab)
        d[k] ~ dnorm(0, 1.0E-6)
        }

# Vague priors for random-effects
m.betaw ~ dnorm(0, 1.0E-6)
tau.betaw ~ dunif(0,2)
tau.sq.betaw <- (tau.betaw * tau.betaw)
prec.betaw <- 1 / (tau.sq.betaw)

m.betab ~ dnorm(0, 1.0E-6)
tau.betab ~ dunif(0,2)
tau.sq.betab <- (tau.betab * tau.betab)
prec.betab <- 1 / (tau.sq.betab)

tau ~ dunif(0,2)
tau.sq <- tau * tau
prec <- 1 / (tau.sq)



### Dataset 1: Constants to define for IPD###

# Number of participants in all IPD individually allocated studies #
list(n.non.cluster.subjects = 2702,
# Number of IPD individually allocated studies #
n.ipd.non.cluster.trials = 6,
# Number of IPD individually allocated study arms #
n.non.cluster.arms = 12,
# Number of participants in all IPD cluster allocated studies #
n.cluster.subjects = 537,
# Number of IPD cluster allocated studies #
n.ipd.cluster.trials = 1,
# Number of IPD cluster allocated study arms #
n.cluster.arms = 2,
```

# Maximum number of clusters in any of the IPD cluster allocated studies #
n.cluster.max = 37,
# Number of interventions being assessed #
max.treat = 6)

Dataset 2: Constants to define for AD###

Number of AD studies
list(n.agg.trials = 2,
# Number of AD study arms #
n.agg.arms = 4)

Dataset 3: Indexing study treatments and specifying baseline treatments

treat[]	baseline1[]
1	1
3	1
1	1
4	1
2	2
5	2
...	...
...	...
...	...

END
treat1 = treatment group codification (coded 1 to n number of treatments – each line of
this dataset contains information for a treatment arm, therefore every 2 lines contains the
treatment codification for a 2-arm study), baseline1 = baseline treatment codification

Dataset 4: Individually allocated IPD studies

study[]	treat[]	outcome[]	baseline[]	cov[]	meancov[]	index[]
1	1	0	1	0	0 – 1	1
1	1	0	1	1	0 – 1	1
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...

END
study = study number, treat = treatment arm (coded 0,1),
outcome = outcome (coded 0,1), baseline = baseline treatment code,
cov = binary covariate of interest (coded 0,1),
meancov = binary covariate study average (proportion of individuals with characteristic,
0 to 1), index = treatment arm code.

Dataset 5: Cluster allocated IPD studies

c.study[]	c.treat[]	c.outcome[]	c.baseline[]	c.cov[]	c.meancov[]	c.index[]	c.cluster[]
1	0	0	1	1	0 – 1	1	1
1	0	0	1	0	0 – 1	1	1

...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

END

c.study = study number, c.treat = treatment arm (coded 0,1),
c.outcome = outcome (coded 0,1), c.baseline = baseline treatment code,
c.cov = binary covariate of interest (coded 0,1),
c.meancov = binary covariate study average (proportion of individuals with characteristic,
0 to 1), c.index = treatment arm code, c.cluster = cluster number

Dataset 6: Individually and cluster allocated AD studies

a.study[]	a.treat[]	outcome.ad[ ]	n[]	a.base[]	a.cov[]
1	1	50	100	1	0.99
1	2	100	200	1	0.98
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...

END

a.study = study number, a.treat = treatment arm code (coded 1 to n number of
treatments), outcome.ad = number of events, n = number of participants,
a.base = baseline treatment code, a.cov = treatment arm aggregate value expressed as a decimal

Initial values, either need specifying or generating for the below scalars, vectors and matrices

list(delta = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0),
d = c(NA,0,0,0,0, 0),
bw = c(NA,0,0,0,0, 0),
bb = c(NA,0,0,0,0, 0),
mu = c(0,0,0,0,0, 0),
beta_cov = c(0,0,0,0,0, 0),
mu.mean = c(0),
sigma.mu = c(1),
c.beta_cov = c(0),
tau = 1,
mu.ad = c(0,0),
m.betaw = 0,
tau.betaw = 1,
m.betab = 0,
tau.betab = 1,
mA = 0,
c.mu = structure(.Data=c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0), .Dim = c(1,37))
# This matrix will have dimension n.ipd.cluster.trials *x* n.cluster.max
)

Appendix 4

Table A4 - Parameter estimates from fitting the MTC synthesis model to IPD reduced to AD to the functioning smoke alarm outcome data without including covariates.

		Model (3.6) - 9 studies included	
		Random effects MTC of IPD reduced to AD	
Interpretation		Median of MCMC posterior sample	95 per cent credible interval
	Parameter		
Log odds ratios for intervention effects (*vs* usual care) - *d's* [i]	***E***	-0.302	-2.064 to 1.357
	E + FE	2.037	-1.081 to 5.351
	E + FE + HI	1.147	-0.795 to 3.432
	E + FE + F	0.925	-0.846 to 2.675
	E + HI	--- [ii]	--- [ii]
	E + FE + F + HI	1.150	-1.734 to 3.9
Between-study variance	τ^2	1.657	0.341 to 3.824
	Function of parameter		
Odds ratios for intervention effects (*vs* usual care) - e^d*'s* [i]	***E***	0.739	0.127 to 3.884
	E + FE	7.666	0.339 to 210.72
	E + FE + HI	3.148	0.452 to 30.94
	E + FE + F	2.522	0.429 to 14.52
	E + HI	--- [ii]	--- [ii]
	E + FE + F + HI	3.159	0.177 to 49.39
Deviance Information criteria	*DIC*	116.07	

Notes:

i) E - education; E + FE - education plus low cost / free equipment; E + FE + HI - education plus low cost / free equipment plus home inspection; E + FE + F - education plus low cost / free equipment plus fitting; E + HI - education plus home inspection; E + FE + F + HI - education plus low cost / free equipment plus fitting + home inspection.

ii) IPD evidence not available for this treatment comparison.

Appendix 5

Decision model R code for the economic evaluation of functioning smoke alarm programs.

The following code relates to modelling for the average population and under the public sector perspective, and intends to illustrate how the modelling was performed.

Two datasets / data-files are required to run the economic model: one relates to all cause mortality (as showed in Table 4.5) and the other one, the CODA, with the effectiveness estimates of the different interventions.

```
#P=Preventative Strategy
#1=Usual Care (includes usual safety education)
#2=Education
#3=Education + low cost/free equipment (10yr sealed battery unit)
#4=Education + low cost/free equipment (10yr sealed battery unit) + home inspection
#5=Education + low cost/free equipment (10yr sealed battery unit) + fitting
#6=Education + home inspection
#7=Education + low cost/free equipment (10yr sealed battery unit) + fitting + home inspection

#S=Health state
#1=functioning fire alarm
#2=non-functioning fire alarm
#3=functioning fire alarm /disability
#4=non-functioning fire alarm /disability
#5=death fatal injury
#6=death other causes

#N=Number of households
#C=Cycle
#T=Total number of years

#pi_[C, S, P]
#lambda[C, S(t), S(t-1), P]

#-----------------------------------------------------------------#
##### DATA 1: constants and vectors definition ######
```

```
#------------------------------------------------------------------#

## PARAMETER TYPE: OTHER

# Size of the cohort simulation going through the model
N <- 100000
# Number of prevention strategies being evaluated
P <- 7
# Number of cycles in part 2 of the model
C <-5
# Number of cycles in the whole decision model (model timeframe)
T <- 100
# Number of health states in the model
S <- 6
# Number of ceiling ratio values (from 0 to £200,000)
K <- 101
# Total number of households in the UK
n_hh <- 22539000
# Discount rate for utilities
disc_u <- 0.035
# Discount rate for costs
disc_c <- 0.035
# Number of minutes of Paramedic Unit - assumed only attend where severe injuries (SE assumption)
mnPU <- 49.5
precPU <- 0.038
sdPU <- 1/sqrt(precPU)
# Number of minutes of Emergency Ambulance - assumed only attend where moderate injuries (SE
assumption)
mnEA <- 38.6
precEA <- 0.038
sdEA <- 1/sqrt(precEA)

## PARAMETER TYPE: PROBABILITIES

# Probability have precautionary check-up following a fire
r_checkup <- 5658
n_checkup <- 12935
# Probability of a fire where functioning smoke alarms present
r_fire_func <- 22771
# Probability of a fire where non-functioning smoke alarms present
r_fire_nonfunc <- 7052
# Probability of a fire where no smoke alarms present or unspecified
r_fire_noSA <- 22883
r_nonfunc_noSA <- r_fire_nonfunc + r_fire_noSA
r_total_fires <- r_fire_func + r_fire_nonfunc + r_fire_noSA
r_fire_nfunc<-r_fire_nonfunc + r_fire_noSA
# Probability of a fatality following a fire where functioning smoke alarm present
r_fatalSA <- 109
# Probability of a fatality following a fire where non-functioning or no smoke alarm
r_fatalnSA <- 232
# Probability of owning a smoke alarm with battery life of 1 year
r_1yrbattery <- 11888
n_1yrbattery <- 15850
# Probability test smoke alarm at least once a year
r_1yrtest <- 15616
n_test <- 18372
```

```
# Probability test smoke alarm less than once a year
r_testless1yr <- 367
# We assume no decay if part of giveaway scheme as given 10 year sealed unit smoke alarm
# Probability of incurring 'no injuries' following a house fire (given functioning smoke alarm/ non-functioning or no smoke alarm)
r_noinjury <- c(47967, 47967)
#Calculated from r_total.fires * 0.91 = prob of no injury
# Probability of inside household fire being attended by the Fire and Rescue Service
r_FRSattend <- 9
n_fires <- 272
# The additional proportion of burn unit costs incurred in ITU
#p_ITU <- 0.4
p_ITUa <- 9.2
p_ITUb <- 13.8
# Probability a child aged 0-4 incurs a moderate injury following a house fire
r_moderate <- 3
# Probability a child aged 0-4 incurs a minor injury following a house fire
r_minor <- 7
# Total number of children aged 0-4 which incured in an injury following a house fire
n_burnsinjury <- 19
# Probability accept intervention (assumed same for all interventions)
#p_accept <- 0.9
p_accepta <- 0.7
p_acceptb <- 0.077778
# Probability a household having a functioning smoke alarm
r_fsa <- 14709
n_sa <- 18386
# Probability of having a standard 1 year smoke alarm battery
p_battery <- 0.75
p_batterya <- 3.25
p_batteryb <- 1.083333
# Probability of having a 10 year long life lithium smoke alarm battery
#p_battery10y <- 0.06
#p_battery10ya <- 14.98
#p_battery10yb <- 234.686667

##
## PARAMETER TYPE: RESOURCE COST
##

#INTERVENTIONS COSTS

# Cost of home safety inspection based on cost of LA home care worker for 40 minutes of their time including travel
c_hsi <- 12
# Cost of providing education programme per household accepting intervention - based on cost of home care worker
# for 20 minutes of their time including travel
c_educ <- 6
# Fixed cost of an intervention scheme (e.g. set-up, administration, etc).
c_fixed <- 54977
# Additional cost incurred for each household that accept intervention (composite value)
c_acc <- 0.49
# Cost of smoke alarm giveaway (with ten-year sealed battery)
c_alarmg <- 4.41
# Cost of having the smoke alarm installed
```

```
c_install <- 10.66

#HEALTH CARE COSTS

# Cost per minute of a Paramedic Unit
c_minPU <- 7.21
# Cost per minute of a Emergency Ambulance
c_minEA <- 7.11
# Mean cost (and standard error) of a minor injury
mn_minor <- 1087
se_minor <- 209
var_minor <- se_minor * se_minor
mn_minorsq<- mn_minor * mn_minor
sigmasq_minor<- log(1+ var_minor / mn_minorsq)
mu_minor<- log(mn_minor) - 0.5 * sigmasq_minor
# Mean cost (and standard error) of a moderate injury
mn_moderate <- 2573
se_moderate <- 1415
var_mod<- se_moderate * se_moderate
mn_modsq<- mn_moderate * mn_moderate
sigmasq_mod<- log(1 + var_mod / mn_modsq)
mu_mod<- log(mn_moderate) - 0.5 * sigmasq_mod
# Mean cost (and standard error) of a severe injury
mn_severe <- 58519
se_severe <- 32019
var_sev<- se_severe * se_severe
mn_sevsq<- mn_severe * mn_severe
sigmasq_sev<- log(1 + var_sev / mn_sevsq)
mu_sev<-log(mn_severe) - 0.5 * sigmasq_sev
# Cost of a fatality following a household fire (updated to 2008/9 prices) - includes coroners and
autopsy costs
c_fatal <- 185.16
# Cost of a disability per year
#c_dispyr <- 342
c_dispyra <- 16
c_dispyrb <- 0.046784
# cost of precautionary check-up NHS reference costs 2008/9 code VB11Z
mu_c_checkup <-62
#Calculate cost of precautionary check-up standard error from interquartile range
qrt1_c_checkup <- 45
qrt3_c_checkup <- 74
se_c_checkup<-((qrt3_c_checkup - qrt1_c_checkup) / (2 * 0.6745))

##
##OTHER COSTS
##
#OUT OF POCKET / PRIVATE COSTS

#Cost to individual - assume would buy standard unit
#c_alarmind <- c(0,4.41,0,0,0,4.41,0)
#Smoke alarm cost
c_sa <- 0
# http://www.safelincs.co.uk/Ionisation-Smoke-Alarms/ # Cost of smoke alarm if falls on individual
# Smoke alarm battery cost
c_battery <- 0      # 1 year battery
#c_battery10y <- 8.89       # 10 year long life lithium batteries
```

```
#c_battery10ya <- 16
#c_battery10yb <- 1.799775
#c_sa <- 0
#c_battery <- 0
#c_battery.std <- 0                          #Use when including private costs
#c_property <- 980                           # Use when including private costs
#c_propertya <- 16
#c_propertyb <- 0.016327

##
#LAW ENFORCEMENT AND RESCUE SERVICES COSTS

# Cost of police attending - assumed only to attend where severs injuries
c_police <- 156.67
# Cost of Fire and rescue Service attending a fire
c_FRSresponse <- 3051

## PARAMETER TYPE: UTILITY PARAMATERS PER CYCLE

# Deficit in utilities for minor injury  (DRG 460 + 459)
u_min <- 0.0487
# Deficit in utilities for moderate injury  (DRG 458 + 457)
u_mod <- 0.069
# Deficit in utilities for severe injury  (DRG 472)
u_sev <- 0.107
# Deficit in utilities following a disability
#u_deficit <- 0.1
u_deficita <- 14.3
u_deficitb <- 128.7

#-----------------------------------------------------------------#
##### DATA 2: Mortality and population utilities ######
#-----------------------------------------------------------------#

mortutil_data<-read.table('…', header=T, sep="\t", quote="\"", dec=".",
fill=T,na.strings=c(""),as.is=1:3)

#-----------------------------------------------------------------#
##### DATA 3: Probability intervention is effective ####
#-----------------------------------------------------------------#

p_MTCfunc_pred_temp<-read.table('…', header=F, sep="\t", quote="\"", dec=".",
fill=T,na.strings=c(""),as.is=1:1)

sims <- seq(1,5000,1)
ProbMatrix<- matrix(data = NA, nrow = 5000, ncol = 7, dimnames = list(sims,
c("preff1","preff2","preff3","preff4","preff5","preff6","preff7")))

for (i in 1:5000) {
  for (k in 1:7) {
    ProbMatrix[i,1] <- p_MTCfunc_pred_temp[i, 2]
    ProbMatrix[i,2] <- p_MTCfunc_pred_temp[i + 5000, 2]
    ProbMatrix[i,3] <- p_MTCfunc_pred_temp[i + 10000, 2]
    ProbMatrix[i,4] <- p_MTCfunc_pred_temp[i + 15000, 2]
    ProbMatrix[i,5] <- p_MTCfunc_pred_temp[i + 20000, 2]
    ProbMatrix[i,6] <- p_MTCfunc_pred_temp[i + 25000, 2]
```

```
   ProbMatrix[i,7] <- p_MTCfunc_pred_temp[i + 30000, 2]
   }
  }

#-----------------------------------------------------------------#
#-----------------------------------------------------------------#
###### DECISION MODEL FOR SMOKE ALARMS ##
#-----------------------------------------------------------------#
#-----------------------------------------------------------------#

Nsim <- 5000

u_pop = array(NA, dim=c(Nsim,T), dimnames=list(seq(1,Nsim,1),seq(1,T,1)))
pi_ = array(NA, dim=c(T,S,P),
dimnames=list(seq(1,T,1),c("fsa","nfsa","fsad","nfsad","dfi","doc"),seq(1,P,1)))
CHECK = array(NA, dim=c(T,P), dimnames=list(seq(1,T,1),seq(1,7,1)))
ct_ = array(NA, dim=c(T,P), dimnames=list(seq(1,T,1),seq(1,7,1)))
ut_ = array(NA, dim=c(T,P), dimnames=list(seq(1,T,1),seq(1,7,1)))
decay = array(NA, dim=c(P), dimnames=list(seq(1,P,1)))
p_fsa = array(NA, dim=c(4,P), dimnames=list(seq(1,4,1),seq(1,P,1)))
o1 = array(NA, dim=c(C,4,P),
dimnames=list(c("cycle1","cycle2","cycle3","cycle4","cycle5"),seq(1,4,1),seq(1,P,1)))
o2 <- o1; o3 <- o1; o4 <- o1; o5 <- o1; o6 <- o1; o7 <- o1; o8 <- o1; o9 <- o1; o10 <- o1; o11 <- o1; o12
<- o1; o13 <- o1; o14 <- o1; o15 <- o1; o16 <- o1
TOT = array(NA, dim=c(C,4,P),
dimnames=list(c("cycle1","cycle2","cycle3","cycle4","cycle5"),seq(1,4,1),seq(1,P,1)))
lambda = array(NA, dim=c(T,S,S,P),
dimnames=list(seq(1,T,1),c("fsa","nfsa","fsad","nfsad","dfi","doc"),c("fsa","nfsa","fsad","nfsad","dfi","
doc"),seq(1,P,1)))
TOTAL = array(NA, dim=c(T,S,P),
dimnames=list(seq(1,T,1),c("fsa","nfsa","fsad","nfsad","dfi","doc"),seq(1,P,1)))
c_equip = array(NA, dim=c(P,4), dimnames=list(seq(1,P,1),seq(1,4,1)))
c_fire = array(NA, dim=c(4), dimnames=list(seq(1,4,1)))
c_o1 = array(NA, dim=c(C,P,4),
dimnames=list(c("cycle1","cycle2","cycle3","cycle4","cycle5"),seq(1,P,1),seq(1,4,1)))
c_o2 <- c_o1; c_o3 <- c_o1; c_o4 <- c_o1; c_o5 <- c_o1; c_o6 <- c_o1; c_o7 <- c_o1; c_o8 <- c_o1;
c_o9 <- c_o1; c_o10 <- c_o1;
c_o11 <- c_o1; c_o12 <- c_o1; c_o13 <- c_o1; c_o14 <- c_o1; c_o15 <- c_o1; c_o16 <- c_o1
cost = array(NA, dim=c(T,S,P),
dimnames=list(seq(1,T,1),c("fsa","nfsa","fsad","nfsad","dfi","doc"),seq(1,P,1)))
TotC = array(NA, dim=c(P), dimnames=list(seq(1,P,1)))
mean_C = array(NA, dim=c(P), dimnames=list(seq(1,P,1)))
TotU = array(NA, dim=c(P), dimnames=list(seq(1,P,1)))
mean_U = array(NA, dim=c(P), dimnames=list(seq(1,P,1)))
u = array(NA, dim=c(T,S,P),
dimnames=list(seq(1,T,1),c("fsa","nfsa","fsad","nfsad","dfi","doc"),seq(1,P,1)))
u_o1 = array(NA, dim=c(C,P,4),
dimnames=list(c("cycle1","cycle2","cycle3","cycle4","cycle5"),seq(1,P,1),seq(1,4,1)))
u_o2 <- u_o1; u_o3 <- u_o1; u_o4 <- u_o1; u_o5 <- u_o1; u_o6 <- u_o1; u_o7 <- u_o1; u_o8 <- u_o1;
u_o9 <- u_o1; u_o10 <- u_o1;
u_o11 <- u_o1; u_o12 <- u_o1; u_o13 <- u_o1; u_o14 <- u_o1; u_o15 <- u_o1; u_o16 <- u_o1
Rc = array(NA, dim=c(K), dimnames=list(seq(1,K,1)))

ProbMatrix <- cbind(ProbMatrix, "p_accept"=rbeta(Nsim, p_accepta, p_acceptb))
ProbMatrix <- cbind(ProbMatrix, "pop_fsa1"=rbeta(Nsim, r_fsa+1, n_sa-r_fsa+1))
ProbMatrix <- cbind(ProbMatrix, "p_fire1"=rbeta(Nsim, r_fire_func+1, n_hh-r_fire_func+1))
```

```
ProbMatrix <- cbind(ProbMatrix, "p_fire2"=rbeta(Nsim, r_nonfunc_noSA+1, n_hh-
r_nonfunc_noSA+1))
ProbMatrix <- cbind(ProbMatrix, "p_fire3"=rbeta(Nsim, r_fire_func+1, n_hh-r_fire_func+1))
ProbMatrix <- cbind(ProbMatrix, "p_fire4"=rbeta(Nsim, r_nonfunc_noSA+1, n_hh-
r_nonfunc_noSA+1))
ProbMatrix <- cbind(ProbMatrix, "p_1yrtest"=rbeta(Nsim, r_1yrtest+1, n_test-r_1yrtest+1))
ProbMatrix <- cbind(ProbMatrix, "p_testless1yr"=rbeta(Nsim, r_testless1yr+1, n_test-r_testless1yr+1))
ProbMatrix <- cbind(ProbMatrix, "p_1yrbattery"=rbeta(Nsim, r_1yrbattery+1, n_1yrbattery-
r_1yrbattery+1))
ProbMatrix <- cbind(ProbMatrix, "p_noinjury"=rbeta(Nsim, r_noinjury+1, r_total_fires-r_noinjury+1))
#Assuming probability of injury the same regardless of functioning or non-functioning SA
ProbMatrix <- cbind(ProbMatrix, "p_injury1"=1-ProbMatrix[,"p_noinjury"])
ProbMatrix <- cbind(ProbMatrix, "p_injury2"=1-ProbMatrix[,"p_noinjury"])
ProbMatrix <- cbind(ProbMatrix, "p_injury3"=1-ProbMatrix[,"p_noinjury"])
ProbMatrix <- cbind(ProbMatrix, "p_injury4"=1-ProbMatrix[,"p_noinjury"])
ProbMatrix <- cbind(ProbMatrix, "p_fatal1"=rbeta(Nsim, r_fatalSA+1, r_fire_func-r_fatalSA+1))
ProbMatrix <- cbind(ProbMatrix, "p_fatal3"=ProbMatrix[,"p_fatal1"])
ProbMatrix <- cbind(ProbMatrix, "p_fatal2"=rbeta(Nsim, r_fatalnSA+1, r_fire_nfunc-r_fatalnSA+1))
ProbMatrix <- cbind(ProbMatrix, "p_fatal4"=ProbMatrix[,"p_fatal2"])
ProbMatrix <- cbind(ProbMatrix, "p_minor"=rbeta(Nsim, r_minor+1, n_burnsinjury-r_minor+2))
ProbMatrix <- cbind(ProbMatrix, "p_moderate"=rbeta(Nsim, r_moderate + (1-
ProbMatrix[,"p_minor"]), n_burnsinjury - r_moderate + (1-ProbMatrix[,"p_minor"])))
ProbMatrix <- cbind(ProbMatrix, "p_severe"= 1 - ProbMatrix[,"p_minor"] -
ProbMatrix[,"p_moderate"])
ProbMatrix <- cbind(ProbMatrix, "p_ITU"= rbeta(Nsim, p_ITUa, p_ITUb))
ProbMatrix <- cbind(ProbMatrix, "p_FRSattend"= rbeta(Nsim, r_FRSattend+1, n_fires-
r_FRSattend+1))
ProbMatrix <- cbind(ProbMatrix, "p_checkup"= rbeta(Nsim, r_checkup+1, n_checkup-r_checkup+1))
ProbMatrix <- cbind(ProbMatrix, "p_battery"= rbeta(Nsim, p_batterya, p_batteryb))
#ProbMatrix <- cbind(ProbMatrix, "p_battery10y"= rbeta(Nsim, p_battery10ya, p_battery10yb))
ProbMatrix <- cbind(ProbMatrix, "c_educ"=rep(c_educ,5000))
ProbMatrix <- cbind(ProbMatrix, "c_fixed"=rep(c_fixed,5000))
ProbMatrix <- cbind(ProbMatrix, "c_alarmg"=rep(c_alarmg,5000))
#ProbMatrix <- cbind(ProbMatrix, "c_alarmind"=ProbMatrix[,"c_alarmg"])
ProbMatrix <- cbind(ProbMatrix, "c_sa"=rep(c_sa,5000))
ProbMatrix <- cbind(ProbMatrix, "c_hsi"=rep(c_hsi,5000))
ProbMatrix <- cbind(ProbMatrix, "c_install"=rep(c_install,5000))
ProbMatrix <- cbind(ProbMatrix, "c_acc"=rep(c_acc,5000))
ProbMatrix <- cbind(ProbMatrix, "c_battery"=rep(c_battery,5000))
#ProbMatrix <- cbind(ProbMatrix, "c_battery10y"=rgamma(Nsim, c_battery10ya, c_battery10yb))
ProbMatrix <- cbind(ProbMatrix, "c_interv1"=rep(0,5000))
ProbMatrix <- cbind(ProbMatrix, "c_interv2"=ProbMatrix[,"c_educ"])
ProbMatrix <- cbind(ProbMatrix, "c_interv3"=ProbMatrix[,"c_educ"] + ProbMatrix[,"c_alarmg"])
ProbMatrix <- cbind(ProbMatrix, "c_interv4"=ProbMatrix[,"c_educ"] + ProbMatrix[,"c_alarmg"] +
ProbMatrix[,"c_hsi"])
ProbMatrix <- cbind(ProbMatrix, "c_interv5"=ProbMatrix[,"c_educ"] + ProbMatrix[,"c_alarmg"] +
ProbMatrix[,"c_install"])
ProbMatrix <- cbind(ProbMatrix, "c_interv6"=ProbMatrix[,"c_educ"] + ProbMatrix[,"c_hsi"])
ProbMatrix <- cbind(ProbMatrix, "c_interv7"=ProbMatrix[,"c_educ"] + ProbMatrix[,"c_alarmg"] +
ProbMatrix[,"c_hsi"] + ProbMatrix[,"c_install"])
ProbMatrix <- cbind(ProbMatrix, "PU_mins"=rnorm(Nsim, mnPU, sdPU))
ProbMatrix <- cbind(ProbMatrix, "EA_mins"=rnorm(Nsim, mnEA, sdEA))
ProbMatrix <- cbind(ProbMatrix, "c_min"=rlnorm(Nsim, mu_minor, sqrt(sigmasq_minor)))
ProbMatrix <- cbind(ProbMatrix, "c_minor"=ProbMatrix[,"c_min"] + ProbMatrix[,"p_ITU"] *
ProbMatrix[,"c_min"])
ProbMatrix <- cbind(ProbMatrix, "c_mod"=rlnorm(Nsim, mu_mod, sqrt(sigmasq_mod)))
```

```
ProbMatrix <- cbind(ProbMatrix, "c_moderate"=ProbMatrix[,"c_mod"] + ProbMatrix[,"p_ITU"] *
ProbMatrix[,"c_mod"] + c_minEA * ProbMatrix[,"EA_mins"])
ProbMatrix <- cbind(ProbMatrix, "c_sev"=rlnorm(Nsim, mu_sev, sqrt(sigmasq_sev)))
ProbMatrix <- cbind(ProbMatrix, "c_severe"=ProbMatrix[,"c_sev"] + ProbMatrix[,"p_ITU"] *
ProbMatrix[,"c_sev"] + c_minPU * ProbMatrix[,"PU_mins"] + c_police)
ProbMatrix <- cbind(ProbMatrix, "c_dispyr"=rgamma(Nsim, c_dispyra, c_dispyrb))
ProbMatrix <- cbind(ProbMatrix, "c_checkup"=rnorm(1,mu_c_checkup,se_c_checkup))
ProbMatrix <- cbind(ProbMatrix, "c_noinjury"=ProbMatrix[,"p_checkup"] *
ProbMatrix[,"c_checkup"])
ProbMatrix <- cbind(ProbMatrix, "u_deficit"=rbeta(Nsim, u_deficita, u_deficitb))

ProbMatrix <- cbind(ProbMatrix, "c_property"=rep(0,5000))

for(l in 1:T) {
        u_pop[,l] <-rnorm(Nsim,mortutil_data[l,"mnu_pop"], mortutil_data[l,"u_pop.se"])
        ProbMatrix <- cbind(ProbMatrix, u_pop[,l])
 colnames(ProbMatrix)[ncol(ProbMatrix)] <- paste("u_pop",l,sep="")
        }


#PART 1 - Intervention model
#(separating the cohort into the different Markov health states)

CEmodelSA <- function(i,CEmodeldata= ProbMatrix) {
# i=1 ; CEmodeldata= ProbMatrix
input <- CEmodeldata[i,]
#input <- CEmodeldata

n1 <- NULL
n2 <- NULL
n3 <- NULL
n4 <- NULL
c_n1 <- NULL
c_n2 <- NULL
c_n3 <- NULL
c_n4 <- NULL

n1[1] <- N * input["pop_fsa1"]                          #Standard care
n2[1] <- 0
n3[1] <- 0
n4[1] <- N * (1 - input["pop_fsa1"])
#To choose between absolute effects (T[p]=p_MTCfunc[p]) or predictive distribution of the absolute
effects #(T_pred[p]=p_MTCfunc_pred[p])

for (p in 2:P) {
        n1[p]<- N * input["pop_fsa1"]
        n2[p]<- N * (1 - input["pop_fsa1"]) * input["p_accept"] * input[paste("preff",p, sep="")]
        n3[p]<- N * (1 - input["pop_fsa1"]) * input["p_accept"] * (1 - input[paste("preff",p, sep="")])
        n4[p]<- N * (1 - input["pop_fsa1"]) * (1 - input["p_accept"])
        }

for (p in 1:P) {
        pi_[1,1,p]<- n1[p] + n2[p]
        pi_[1,2,p]<- n3[p] + n4[p]
        pi_[1,3,p]<- 0
        pi_[1,4,p]<- 0
```

```
            pi_[1,5,p]<- 0
            pi_[1,6,p]<- 0

            CHECK[1,p]<- pi_[1,1,p] + pi_[1,2,p] + pi_[1,3,p] + pi_[1,4,p] + pi_[1,5,p] + pi_[1,6,p]

            c_n1[p]<- n1[p] * 0
            c_n2[p]<- n2[p] * (input[paste("c_interv",p, sep="")] + input["c_acc"])
            c_n3[p]<- n3[p] * (input[paste("c_interv",p, sep="")] + input["c_acc"])
            c_n4[p]<- n4[p]

ct_[1,p]<- c_n1[p] + c_n2[p] + c_n3[p] + c_n4[p] + input["c_fixed"]

            ut_[1,p]<- input["u_pop1"] * N
            }

#PART TWO - MARKOV MODEL (5 year period)

decay[1] <- (1 - input["p_1yrtest"] - 0.5 * input["p_testless1yr"]) * input["p_1yrbattery"]
decay[2] <- decay[1]
decay[3] <- 0
decay[4] <- 0
decay[5] <- 0
decay[6] <- decay[1]
decay[7] <- 0

for(p in 1:P) {
   #This part of the code allows to set different decays for different interventions
   #e.g. wired, 1 yr battery, 10 yr battery, etc.
   p_fsa[1,p]<- input["pop_fsa1"] * (1 - decay[p])
             #'Decay' is proportion of smoke alarms no longer functional after 12 months
             p_fsa[2,p]<- 0
             #Probability go from not functioning to functioning included in the decay variable
             p_fsa[3,p]<- p_fsa[1,p]
             p_fsa[4,p]<- 0
            }

for(p in 1:P) {
for(c in 2:C) {
for(k in 1:4) {
#1=functional, 2=non-functional, 3=functional / disability, 4=non-functional / disability
            o1[c,k,p]<- (1 - input[paste("p_fire",k, sep="")]) * mortutil_data[c,"p_allcause"]
            o2[c,k,p]<- (1 - input[paste("p_fire",k, sep="")]) * (1 - mortutil_data[c,"p_allcause"]) *
                              p_fsa[k,p]
            o3[c,k,p]<- (1 - input[paste("p_fire",k, sep="")]) * (1 - mortutil_data[c,"p_allcause"]) * (1 –
                              p_fsa[k,p])
            o4[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] *
                              input[paste("p_fatal",k, sep="")]
            o5[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
                              input[paste("p_fatal",k, sep="")]) * input["p_minor"] * (1 –
                              mortutil_data[c,"p_allcause"]) * p_fsa[k,p]
            o6[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
                              input[paste("p_fatal",k, sep="")]) * input["p_minor"] * (1 –
                              mortutil_data[c,"p_allcause"]) * (1 - p_fsa[k,p])
            o7[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
                              input[paste("p_fatal",k, sep="")]) * input["p_minor"] *
                              mortutil_data[c,"p_allcause"]
```

```
		o8[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
				input[paste("p_fatal",k, sep="")]) * input["p_moderate"] * (1 –
				mortutil_data[c,"p_allcause"]) * p_fsa[k,p]
		o9[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
				input[paste("p_fatal",k, sep="")]) * input["p_moderate"] * (1 –
				mortutil_data[c,"p_allcause"] ) * (1 - p_fsa[k,p])
		o10[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
				input[paste("p_fatal",k, sep="")]) * input["p_moderate"] *
				mortutil_data[c,"p_allcause"]
		o11[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
				input[paste("p_fatal",k, sep="")]) * input["p_severe"] * (1 –
				mortutil_data[c,"p_allcause"]) * p_fsa[k,p]
		o12[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
				input[paste("p_fatal",k, sep="")]) * input["p_severe"] * (1 –
				mortutil_data[c,"p_allcause"]) * (1 - p_fsa[k,p])
		o13[c,k,p]<- input[paste("p_fire",k, sep="")] * input[paste("p_injury",k, sep="")] * (1 –
				input[paste("p_fatal",k, sep="")]) * input["p_severe"] *
				mortutil_data[c,"p_allcause"]
		o14[c,k,p]<- input[paste("p_fire",k, sep="")] * (1 - input[paste("p_injury",k, sep="")]) *
				mortutil_data[c,"p_allcause"]
		o15[c,k,p]<- input[paste("p_fire",k, sep="")] * (1 - input[paste("p_injury",k, sep="")]) * (1 –
				mortutil_data[c,"p_allcause"]) * p_fsa[k,p]
		o16[c,k,p]<- input[paste("p_fire",k, sep="")] * (1 - input[paste("p_injury",k, sep="")]) *
				(1 - mortutil_data[c,"p_allcause"]) * (1 - p_fsa[k,p])

		TOT[c,k,p]<-o1[c,k,p]+o2[c,k,p]+o3[c,k,p]+o4[c,k,p]+o5[c,k,p]+o6[c,k,p]+o7[c,k,p]+
o8[c,k,p]+o9[c,k,p]+o10[c,k,p]+o11[c,k,p]+o12[c,k,p]+o13[c,k,p]+o14[c,k,p]+o15[c,k,p]+o16[c,k,p]
		}
		}
}


#Transition probabilities

for (p in 1:P){
		for(c in 2:C){
				#From 'functioning' state
						lambda[c,1,1,p]<-o2[c,1,p]+o5[c,1,p]+o8[c,1,p]+o15[c,1,p]
						lambda[c,1,2,p]<-o3[c,1,p]+o6[c,1,p]+o9[c,1,p]+o16[c,1,p]
						lambda[c,1,3,p]<-o11[c,1,p]
						lambda[c,1,4,p]<-o12[c,1,p]
						lambda[c,1,5,p]<-o4[c,1,p]
						lambda[c,1,6,p]<-o1[c,1,p]+o7[c,1,p]+o10[c,1,p]+o13[c,1,p]+o14[c,1,p]
						TOTAL[c,1,p]<-
lambda[c,1,1,p]+lambda[c,1,2,p]+lambda[c,1,3,p]+lambda[c,1,4,p]+lambda[c,1,5,p]+lambda[c,1,6,p]
				#From 'non-functioning' state
						lambda[c,2,1,p]<-o2[c,2,p]+o5[c,2,p]+o8[c,2,p]+o15[c,2,p]
						lambda[c,2,2,p]<-o3[c,2,p]+o6[c,2,p]+o9[c,2,p]+o16[c,2,p]
						lambda[c,2,3,p]<-o11[c,2,p]
						lambda[c,2,4,p]<-o12[c,2,p]
						lambda[c,2,5,p]<-o4[c,2,p]
						lambda[c,2,6,p]<-o1[c,2,p]+o7[c,2,p]+o10[c,2,p]+o13[c,2,p]+o14[c,2,p]
						TOTAL[c,2,p]<-
lambda[c,2,1,p]+lambda[c,2,2,p]+lambda[c,2,3,p]+lambda[c,2,4,p]+lambda[c,2,5,p]+lambda[c,2,6,p]
				#From 'functioning / disability' state
						lambda[c,3,1,p]<-0
```

```
                        lambda[c,3,2,p]<-0
                        lambda[c,3,3,p]<-o2[c,3,p]+o5[c,3,p]+o8[c,3,p]+o15[c,3,p] + o11[c,3,p]
                        lambda[c,3,4,p]<-o3[c,3,p]+o6[c,3,p]+o9[c,3,p]+o16[c,3,p] + o12[c,3,p]
                        lambda[c,3,5,p]<-o4[c,3,p]
                        lambda[c,3,6,p]<-o1[c,3,p]+o7[c,3,p]+o10[c,3,p]+o13[c,3,p]+o14[c,3,p]
                        TOTAL[c,3,p]<-
lambda[c,3,1,p]+lambda[c,3,2,p]+lambda[c,3,3,p]+lambda[c,3,4,p]+lambda[c,3,5,p]+lambda[c,3,6,p]
                #From 'non-functioning' state
                        lambda[c,4,1,p]<-0
                        lambda[c,4,2,p]<-0
                        lambda[c,4,3,p]<- o2[c,4,p]+o5[c,4,p]+o8[c,4,p]+o15[c,4,p] + o11[c,4,p]
                        lambda[c,4,4,p]<-o3[c,4,p]+o6[c,4,p]+o9[c,4,p]+o16[c,4,p] + o12[c,4,p]
                        lambda[c,4,5,p]<-o4[c,4,p]
                        lambda[c,4,6,p]<-o1[c,4,p]+o7[c,4,p]+o10[c,4,p]+o13[c,4,p]+o14[c,4,p]
                        TOTAL[c,4,p]<-
lambda[c,4,1,p]+lambda[c,4,2,p]+lambda[c,4,3,p]+lambda[c,4,4,p]+lambda[c,4,5,p]+lambda[c,4,6,p]
                #From 'fatal' state
                        lambda[c,5,1,p]<-0
                        lambda[c,5,2,p]<-0
                        lambda[c,5,3,p]<-0
                        lambda[c,5,4,p]<-0
                        lambda[c,5,5,p]<-1
                        lambda[c,5,6,p]<-0
                        TOTAL[c,5,p]<-
lambda[c,5,1,p]+lambda[c,5,2,p]+lambda[c,5,3,p]+lambda[c,5,4,p]+lambda[c,5,5,p]+lambda[c,5,6,p]
                #From 'all cause' state
                        lambda[c,6,1,p]<-0
                        lambda[c,6,2,p]<-0
                        lambda[c,6,3,p]<-0
                        lambda[c,6,4,p]<-0
                        lambda[c,6,5,p]<-0
                        lambda[c,6,6,p]<-1
                        TOTAL[c,6,p]<-
lambda[c,6,1,p]+lambda[c,6,2,p]+lambda[c,6,3,p]+lambda[c,6,4,p]+lambda[c,6,5,p]+lambda[c,6,6,p]
                }
        }

#Number of individuals (households) in each state at time t>1
for (p in 1:P) {
        for(c in 2:C) {
        for (s in 1:S) {
                pi_[c,s,p]<- pi_[(c - 1),,p] %*% lambda[c,,s,p]
        }
          CHECK[c,p]<-pi_[c,1,p] + pi_[c,2,p] + pi_[c,3,p] + pi_[c,4,p] + pi_[c,5,p] + pi_[c,6,p]
        }
        }

#
#PART THREE - MARKOV MODEL (T-5 year period)
#

for (p in 1:P) {
        for(c in (C+1):T) {
                        lambda[c,1,1,p]<-1 - mortutil_data[c,"p_allcause"]
                        lambda[c,1,2,p]<-0
                        lambda[c,1,3,p]<-0
```

```
                    lambda[c,1,4,p]<-0
                    lambda[c,1,5,p]<-0
                    lambda[c,1,6,p]<-mortutil_data[c,"p_allcause"]
                    TOTAL[c,1,p]<-
lambda[c,1,1,p]+lambda[c,1,2,p]+lambda[c,1,3,p]+lambda[c,1,4,p]+lambda[c,1,5,p]+lambda[c,1,6,p]

                    lambda[c,2,1,p]<-0
                    lambda[c,2,2,p]<-1 - mortutil_data[c,"p_allcause"]
                    lambda[c,2,3,p]<-0
                    lambda[c,2,4,p]<-0
                    lambda[c,2,5,p]<-0
                    lambda[c,2,6,p]<-mortutil_data[c,"p_allcause"]
                    TOTAL[c,2,p]<-
lambda[c,2,1,p]+lambda[c,2,2,p]+lambda[c,2,3,p]+lambda[c,2,4,p]+lambda[c,2,5,p]+lambda[c,2,6,p]

                    lambda[c,3,1,p]<-0
                    lambda[c,3,2,p]<-0
                    lambda[c,3,3,p]<-1 - mortutil_data[c,"p_allcause"]
                    lambda[c,3,4,p]<-0
                    lambda[c,3,5,p]<-0
                    lambda[c,3,6,p]<-mortutil_data[c,"p_allcause"]
                    TOTAL[c,3,p]<-
lambda[c,3,1,p]+lambda[c,3,2,p]+lambda[c,3,3,p]+lambda[c,3,4,p]+lambda[c,3,5,p]+lambda[c,3,6,p]

                    lambda[c,4,1,p]<-0
                    lambda[c,4,2,p]<-0
                    lambda[c,4,3,p]<-0
                    lambda[c,4,4,p]<-1 - mortutil_data[c,"p_allcause"]
                    lambda[c,4,5,p]<-0
                    lambda[c,4,6,p]<-mortutil_data[c,"p_allcause"]
                    TOTAL[c,4,p]<-
lambda[c,4,1,p]+lambda[c,4,2,p]+lambda[c,4,3,p]+lambda[c,4,4,p]+lambda[c,4,5,p]+lambda[c,4,6,p]

                    lambda[c,5,1,p]<-0
                    lambda[c,5,2,p]<-0
                    lambda[c,5,3,p]<-0
                    lambda[c,5,4,p]<-0
                    lambda[c,5,5,p]<-1
                    lambda[c,5,6,p]<-0
                    TOTAL[c,5,p]<-
lambda[c,5,1,p]+lambda[c,5,2,p]+lambda[c,5,3,p]+lambda[c,5,4,p]+lambda[c,5,5,p]+lambda[c,5,6,p]

                    lambda[c,6,1,p]<-0
                    lambda[c,6,2,p]<-0
                    lambda[c,6,3,p]<-0
                    lambda[c,6,4,p]<-0
                    lambda[c,6,5,p]<-0
                    lambda[c,6,6,p]<-1
                    TOTAL[c,6,p]<-
lambda[c,6,1,p]+lambda[c,6,2,p]+lambda[c,6,3,p]+lambda[c,6,4,p]+lambda[c,6,5,p]+lambda[c,6,6,p]
                }
        }

#Number of individuals (households) in each state at time >C
for (p in 1:P) {
        for(c in (C+1):T) {
```

```
        for (s in 1:S) {
                        pi_[c,s,p]<- pi_[(c - 1),,p] %*% lambda[c,,s,p]
                        }
        CHECK[c,p]<- pi_[c,1,p] + pi_[c,2,p] + pi_[c,3,p] + pi_[c,4,p] + pi_[c,5,p] + pi_[c,6,p]
        }
}


##########################################
#-------- COSTS IN EACH STATE ----------#
##########################################

# Equipment costs of having functioning smoke alarm at end of each cycle

#Usual Care, intervention 2 and 6
for(k in 1:4) {
        c_equip[1,k]<- input["c_battery"] * input["p_battery"] # + input["c_battery10y"] *
                        input["p_battery10y"]
        c_equip[2,k]<- input["c_sa"] + input["c_battery"] * input["p_battery"]
                        #+ input["c_battery10y"] * input["p_battery10y"]
        c_equip[6,k]<- input["c_sa"] + input["c_battery"] * input["p_battery"]
                        #+ input["c_battery10y"] * input["p_battery10y"]
 }

# All other intervention groups
for (p in 3:5) {
        c_equip[p, 1]<- input["c_battery"] * input["p_battery"]
                        #+ input["c_battery10y"] * input["p_battery10y"]
        c_equip[p, 2]<- input["c_sa"] + input["c_battery"] * input["p_battery"]
                        #+ input["c_battery10y"] * input["p_battery10y"]
        c_equip[p, 3]<- c_equip[p,1]
        c_equip[p, 4]<- c_equip[p,2]
        }

c_equip[7, 1]<- input["c_battery"] * input["p_battery"]
                #+ input["c_battery10y"] * input["p_battery10y"]
c_equip[7, 2]<- input["c_sa"] + input["c_battery"] * input["p_battery"]
                #+ input["c_battery10y"] * input["p_battery10y"]
c_equip[7, 3]<- c_equip[p,1]
c_equip[7, 4]<- c_equip[p,2]


# Cost of having a fire at the household

for (k in 1:4) {
 c_fire[k]<- input["p_FRSattend"] * c_FRSresponse + input["c_property"]
 }


#Cost of taking each pathway through the model for functional, non-functional  functional/disability
and non-functional/disability

for (p in 1:P){
        for (c in 2:C){
                for(k in 1:2){      #1=functional, 2=non-functional
                        c_o1[c,p,k]<- 0
```

```
                c_o2[c,p,k]<- c_equip[p,k]
                #If already have smoke alarm this is cost of maintenance (ie.g. battery)
                # If not, cost of equipment (i.e. smoke alarm)
                c_o3[c,p,k]<- 0
                c_o4[c,p,k]<- (c_fire[k] + c_fatal)
                c_o5[c,p,k]<- (c_fire[k] + input["c_minor"] + c_equip[p,k])
                c_o6[c,p,k]<- (c_fire[k] + input["c_minor"])
                c_o7[c,p,k]<- (c_fire[k] + input["c_minor"])
                c_o8[c,p,k]<- (c_fire[k] + input["c_moderate"] + c_equip[p,k])
                c_o9[c,p,k]<- (c_fire[k] + input["c_moderate"])
                c_o10[c,p,k]<- (c_fire[k] + input["c_moderate"])
                c_o11[c,p,k]<- (c_fire[k] + input["c_severe"] + c_equip[p,k])
                c_o12[c,p,k]<- (c_fire[k] + input["c_severe"])
                c_o13[c,p,k]<- (c_fire[k] + input["c_severe"])
                c_o14[c,p,k]<- (c_fire[k] + input["c_noinjury"])
                c_o15[c,p,k]<- (c_fire[k] + input["c_noinjury"] + c_equip[p,k])
                c_o16[c,p,k]<- (c_fire[k] + input["c_noinjury"])
                }
        for(k in 3:4) {      #3=functional / disability, 4=non-functional / disability
                c_o1[c,p,k]<- 0
                c_o2[c,p,k]<- (c_equip[p,k] + input["c_dispyr"])
                #If already have smoke alarm  this is cost of maintenance (ie.g. battery)
                # If not, cost of equipment (i.e. smoke alarm)
                #c_disability is annual cost associated with being disabled by fire
                c_o3[c,p,k]<- (0 + input["c_dispyr"])
                c_o4[c,p,k]<- (c_fire[k] + c_fatal)
                c_o5[c,p,k]<- (c_fire[k] + input["c_minor"] + c_equip[p,k] +
                                input["c_dispyr"])
                c_o6[c,p,k]<- (c_fire[k] + input["c_minor"] + input["c_dispyr"] )
                c_o7[c,p,k]<- (c_fire[k] + input["c_minor"])
                c_o8[c,p,k]<- (c_fire[k] + input["c_moderate"] + c_equip[p,k] +
                                input["c_dispyr"])
                c_o9[c,p,k]<- (c_fire[k] + input["c_moderate"] + input["c_dispyr"])
                c_o10[c,p,k]<- (c_fire[k] + input["c_moderate"])
                c_o11[c,p,k]<- (c_fire[k] + input["c_severe"] + c_equip[p,k] +
                                input["c_dispyr"])
                c_o12[c,p,k]<- (c_fire[k] + input["c_severe"] + input["c_dispyr"])
                c_o13[c,p,k]<- (c_fire[k] + input["c_severe"])
                c_o14[c,p,k]<- (c_fire[k] + input["c_noinjury"])
                c_o15[c,p,k]<- (c_fire[k] + input["c_noinjury"] + c_equip[p,k] +
                                input["c_dispyr"])
                c_o16[c,p,k]<- (c_fire[k] + input["c_noinjury"] + input["c_dispyr"])
        }

        # From 'functioning fire alarm state
        cost[c,1,p]<-
c_o1[c,p,1]+c_o2[c,p,1]+c_o3[c,p,1]+c_o4[c,p,1]+c_o5[c,p,1]+c_o6[c,p,1]+c_o7[c,p,1]+c_o8[c,p,1]+c
_o9[c,p,1]+c_o10[c,p,1]+c_o11[c,p,1]+c_o12[c,p,1]+c_o13[c,p,1]+c_o14[c,p,1]+c_o15[c,p,1]+c_o16[c
,p,1]

#From 'non-functioning fire alarm' state
        cost[c,2,p]<-
c_o1[c,p,2]+c_o2[c,p,2]+c_o3[c,p,2]+c_o4[c,p,2]+c_o5[c,p,2]+c_o6[c,p,2]+c_o7[c,p,2]+c_o8[c,p,2]+c
_o9[c,p,2]+c_o10[c,p,2]+c_o11[c,p,2]+c_o12[c,p,2]+c_o13[c,p,2]+c_o14[c,p,2]+c_o15[c,p,2]+c_o16[c
,p,2]
```

```
		#From 'functioning fire alarm / disability' state
		cost[c,3,p]<-
c_o1[c,p,3]+c_o2[c,p,3]+c_o3[c,p,3]+c_o4[c,p,3]+c_o5[c,p,3]+c_o6[c,p,3]+c_o7[c,p,3]+c_o8[c,p,3]+c
_o9[c,p,3]+c_o10[c,p,3]+c_o11[c,p,3]+c_o12[c,p,3]+c_o13[c,p,3]+c_o14[c,p,3]+c_o15[c,p,3]+c_o16[c
,p,3]

		#From 'non-functioning fire alarm / disability' state
		cost[c,4,p]<-
c_o1[c,p,4]+c_o2[c,p,4]+c_o3[c,p,4]+c_o4[c,p,4]+c_o5[c,p,4]+c_o6[c,p,4]+c_o7[c,p,4]+c_o8[c,p,4]+c
_o9[c,p,4]+c_o10[c,p,4]+c_o11[c,p,4]+c_o12[c,p,4]+c_o13[c,p,4]+c_o14[c,p,4]+c_o15[c,p,4]+c_o16[c
,p,4]

		#From 'death fatal injury' state
		cost[c,5,p]<-0
		#From 'death all cause' state
		cost[c,6,p]<-0
	}
	for (c in (C+1):T) { #Cost of disability per year for the disability health states
		cost[c,1,p]<- 0
		cost[c,2,p]<- 0
		cost[c,3,p]<- input["c_dispyr"]
		cost[c,4,p]<- input["c_dispyr"]
		cost[c,5,p]<- 0
		cost[c,6,p]<- 0
		}
	}

for (p in 1:P) {
	u[1,1,p]<- input["u_pop1"]
	u[1,2,p]<- input["u_pop1"]
	u[1,3,p]<- input["u_pop1"]
	u[1,4,p]<- input["u_pop1"]
	u[1,5,p]<- 0
	u[1,6,p]<- 0

#Utilities of taking each pathway through the model for functional, non-functional,
#functional/disability and non-functional/disability

	for (c in 2:C) {
		for(k in 1:2) {	#1=functional, 2=non-functional
			u_o1[c,p,k]<- 0
			u_o2[c,p,k]<- input[paste("u_pop",c, sep="")]
			u_o3[c,p,k]<- input[paste("u_pop",c, sep="")]
			u_o4[c,p,k]<- 0
			u_o5[c,p,k]<- (input[paste("u_pop",c, sep="")] - u_min)
			u_o6[c,p,k]<- (input[paste("u_pop",c, sep="")] - u_min)
			u_o7[c,p,k]<- 0
			u_o8[c,p,k]<- (input[paste("u_pop",c, sep="")] - u_mod)
			u_o9[c,p,k]<- (input[paste("u_pop",c, sep="")] - u_mod)
			u_o10[c,p,k]<- 0
			u_o11[c,p,k]<- (input[paste("u_pop",c, sep="")] - u_sev)
			u_o12[c,p,k]<- (input[paste("u_pop",c, sep="")] - u_sev)
			u_o13[c,p,k]<- 0
			u_o14[c,p,k]<- 0
			u_o15[c,p,k]<- input[paste("u_pop",c, sep="")]
			u_o16[c,p,k]<- input[paste("u_pop",c, sep="")]
```

```
                }

        for(k in 3:4) {      #3=functional / disability, 4=non-functional / disability
                u_o1[c,p,k]<- 0
                u_o2[c,p,k]<- (input[paste("u_pop",c, sep="")] - input["u_deficit"])

                u_o3[c,p,k]<- (input[paste("u_pop",c, sep="")] - input["u_deficit"])
                u_o4[c,p,k]<- 0
                u_o5[c,p,k]<- ((input[paste("u_pop",c, sep="")] - input["u_deficit"]) –
                        u_min)
                u_o6[c,p,k]<- ((input[paste("u_pop",c, sep="")] - input["u_deficit"]) –
                        u_min)
                u_o7[c,p,k]<- 0
                u_o8[c,p,k]<- ((input[paste("u_pop",c, sep="")] - input["u_deficit"]) –
                        u_mod)
                u_o9[c,p,k]<- ((input[paste("u_pop",c, sep="")] - input["u_deficit"]) –
                        u_mod)
                u_o10[c,p,k]<- 0
                u_o11[c,p,k]<- ((input[paste("u_pop",c, sep="")] - input["u_deficit"]) –
                        u_sev)
                u_o12[c,p,k]<- ((input[paste("u_pop",c, sep="")] - input["u_deficit"]) –
                        u_sev)
                u_o13[c,p,k]<- 0
                u_o14[c,p,k]<- 0
                u_o15[c,p,k]<- (input[paste("u_pop",c, sep="")] - input["u_deficit"])
                u_o16[c,p,k]<- (input[paste("u_pop",c, sep="")] - input["u_deficit"])
                }

        # From 'functioning fire alarm' state
        u[c,1,p]<-
u_o1[c,p,1]+u_o2[c,p,1]+u_o3[c,p,1]+u_o4[c,p,1]+u_o5[c,p,1]+u_o6[c,p,1]+u_o7[c,p,1]+u_o8[c,p,1]+
u_o9[c,p,1]+u_o10[c,p,1]+u_o11[c,p,1]+u_o12[c,p,1]+u_o13[c,p,1]+u_o14[c,p,1]+u_o15[c,p,1]+u_o16
[c,p,1]

        #From 'non-functioning fire alarm' state
        u[c,2,p]<-
u_o1[c,p,2]+u_o2[c,p,2]+u_o3[c,p,2]+u_o4[c,p,2]+u_o5[c,p,2]+u_o6[c,p,2]+u_o7[c,p,2]+u_o8[c,p,2]+
u_o9[c,p,2]+u_o10[c,p,2]+u_o11[c,p,2]+u_o12[c,p,2]+u_o13[c,p,2]+u_o14[c,p,2]+u_o15[c,p,2]+u_o16
[c,p,2]

        #From 'functioning fire alarm / disability' state
        u[c,3,p]<-
u_o1[c,p,3]+u_o2[c,p,3]+u_o3[c,p,3]+u_o4[c,p,3]+u_o5[c,p,3]+u_o6[c,p,3]+u_o7[c,p,3]+u_o8[c,p,3]+
u_o9[c,p,3]+u_o10[c,p,3]+u_o11[c,p,3]+u_o12[c,p,3]+u_o13[c,p,3]+u_o14[c,p,3]+u_o15[c,p,3]+u_o16
[c,p,3]

        #From 'non-functioning fire alarm / disability' state
        u[c,4,p]<-
u_o1[c,p,4]+u_o2[c,p,4]+u_o3[c,p,4]+u_o4[c,p,4]+u_o5[c,p,4]+u_o6[c,p,4]+u_o7[c,p,4]+u_o8[c,p,4]+
u_o9[c,p,4]+u_o10[c,p,4]+u_o11[c,p,4]+u_o12[c,p,4]+u_o13[c,p,4]+u_o14[c,p,4]+u_o15[c,p,4]+u_o16
[c,p,4]

        #From 'death fatal injury' state
        u[c,5,p]<-0

        #From 'death all cause' state
```

```
                u[c,6,p]<-0
                }

        for (c in (C+1):T) {
                u[c,1,p]<- input[paste("u_pop",c, sep="")]
                u[c,2,p]<- input[paste("u_pop",c, sep="")]
                u[c,3,p]<- input[paste("u_pop",c, sep="")] - input["u_deficit"]
                u[c,4,p]<- input[paste("u_pop",c, sep="")] - input["u_deficit"]
                u[c,5,p]<- 0
                u[c,6,p]<- 0
                }
        }

#Costs in each cycle of model
for (p in 1:P) {
        for(c in 2:T) {
                ct_[c,p] <- pi_[c,,p] %*% cost[c,,p] / ((1 + disc_c)^(c - 1))
                }
        TotC[p] <- sum(ct_[,p])
        mean_C[p] <- TotC[p] / N
        }

#Utlities in each cycle of model
for (p in 1:P){
        for(c in 2:T) {
                ut_[c,p] <- pi_[c,,p] %*% u[c,,p] / ((1 + disc_u)^(c - 1))
                }
        TotU[p] <- sum(ut_[,p])
        mean_U[p] <- TotU[p] / N
        }

return(c(mean_C, mean_U))

}



Nsim1<-5000

system.time(CE_out <- sapply(1:Nsim1,CEmodelSA))


#Cost effectiveness
NB = array(NA, dim=c(P,K, Nsim1), dimnames=list(seq(1,P,1),seq(1,K,1), seq(1,Nsim1,1)))

######## CALCULATING NMBs (CEILING RATIOS FROM £0 TO £200,000 ############

for(i in 1:Nsim1) {
  for(k in 1:K) {
        Rc[k]<- (k - 1) * 2000
        for (p in 1:P) {
                NB[p,k,i]<- Rc[k] * CE_out[7+p,i] - CE_out[p,i]
                }
        }
}
```

```
### CALCULATING THE PROBABILITY OF INTERVENTION IS COST EFFECTIVE ###
### CALCULATING THE INTERVENTION WITH HIGHEST NMB ##########
### CALCULATING THE INDIVIDUAL (HOUSEHOLD) EXPECTED VALUE OF PERFECT
INFORMATION ##########

aux = array(NA, dim=c(Nsim1,K), dimnames=list(seq(1,Nsim1,1),seq(1,K,1)))
pCE = array(NA, dim=c(P,K), dimnames=list(seq(1,P,1),seq(1,K,1)))
meanNB = array(NA, dim=c(K,P), dimnames=list(seq(1,K,1),seq(1,P,1)))
CI = array(NA, dim=c(K), dimnames=list(seq(1,K,1)))
PI = array(NA, dim=c(K), dimnames=list(seq(1,K,1)))
EVPI = array(NA, dim=c(K), dimnames=list(seq(1,K,1)))

for(k in 1:K) {
  aux[,k] <- apply(NB[,k,],2, max, na.rm=T)
  for (p in 1:P) {
            pCE[p,k]<- sum(aux == NB[p,k,], na.rm=T)/Nsim1
            meanNB[k,p]<- mean(NB[p,k,])
    }
  CI[] <- apply(meanNB[,],1,max)
  PI[k] <- mean(aux[,k])
  EVPI[k] <- PI[k] - CI[k]
}

########## CALCULATING POPULATION EVPI ##########

discEffpop<- matrix(data = NA, nrow = 10, ncol = 1)
popEVPI<- matrix(data = NA, nrow = 101, ncol = 1)
Effpop<- 31000

for(k in 1:10) {
  discEffpop[k]<- Effpop /(1 + disc_c)^(k-1)
  }
sumdiscEffpop<-sum(discEffpop)

for(k in 1:K) {
  popEVPI[k]<- sumdiscEffpop * EVPI[k]
  }

########## CALCULATING THE PCE FOR FRONTIER ###########

pCE_CEAF = array(NA, dim=c(K,P), dimnames=list(seq(1,K,1),seq(1,P,1)))

for(k in 1:K) {
   if (CI[k] == meanNB[k,1]) pCE_CEAF[k,1]<- pCE[1,k] else pCE_CEAF[k,1]<- NA
   if (CI[k] == meanNB[k,2]) pCE_CEAF[k,2]<- pCE[2,k] else pCE_CEAF[k,2]<- NA
   if (CI[k] == meanNB[k,3]) pCE_CEAF[k,3]<- pCE[3,k] else pCE_CEAF[k,3]<- NA
   if (CI[k] == meanNB[k,4]) pCE_CEAF[k,4]<- pCE[4,k] else pCE_CEAF[k,4]<- NA
   if (CI[k] == meanNB[k,5]) pCE_CEAF[k,5]<- pCE[5,k] else pCE_CEAF[k,5]<- NA
   if (CI[k] == meanNB[k,6]) pCE_CEAF[k,6]<- pCE[6,k] else pCE_CEAF[k,6]<- NA
   if (CI[k] == meanNB[k,7]) pCE_CEAF[k,7]<- pCE[7,k] else pCE_CEAF[k,7]<- NA
   }
```

Appendix 6

Table A6 – List of model input parameters used within the decision model for smoke alarms. Sources of evidence used to inform the parameter and parametric assumption used to model parameter uncertainty is also shown.

Model input parameter	Parameter description	Source(s) of evidence informing the parameter	Distributional assumption	Estimates
Parameter type: Other				
N	Size of the cohort simulation going through the model	---	Fixed	100,000
P	Number of prevention strategies being evaluated	Defined by the MTC analysis	Fixed	7
C	Number of cycles in part 2 of the model		Fixed	5
T	Number of cycles in the whole decision model (model timeframe)		Fixed	100
S	Number of health states in the model		Fixed	6
K	Number of ceiling ratio values λ (from 0 to £200,000)		Fixed	101
n.hh	Total number of households in the UK	2001 UK Census (Statistics, 2001)	Fixed	22,539,000
disc.u	Discount rate for utilities	Treasury (NICE, 2008)	Fixed	3.5%
disc.c	Discount rate for costs		Fixed	3.5%
PU_mins	Number of minutes of Paramedic Unit – assumed only attend where severe injuries (SE assumption)	PSSRU 2008 (Curtis, 2008) (updated to 2009 prices)	Normal	E[PU] = 49.5 Var[PU] = 26.32 (assumption)
EA_mins	Number of minutes of Emergency Ambulance – assumed only attend where moderate injuries (SE assumption)		Normal	E[EA] = 38.6 precEA = 26.32 (assumption)
Parameter type: Probabilities				
p_MTCfunc	Probability of a functioning smoke alarm specific to each intervention	From MTC analysis		
p_accept	Probability accept intervention (assumed same for all interventions)	Assumption	Beta	p_accepta = 0.7 p_acceptb = 0.077778 based on E[p_accept] = 0.9 and assuming se[p_accept] = 0.225 **
pop_fsa	Probability a household having a functioning smoke alarm	Survey of English Housing 2004/5 (Government, 2006) -Table 5.2	Binomial	r_fsa=14,709 n_sa=18,386

p_checkup	Probability have a precautionary checkup following a fire	Fire Statistics 2007 (Government, 2007) - Table 8	Binomial	r_checkup = 5,658 n_checkup = 12,935
p_fire.func	Probability of a fire where functioning smoke alarms present	Fire Statistics 2007 (Government, 2007) - Table 2.4	Binomial	r_fire.func = 22,771 n_hh = 22,539,000
p_fire.nonfunc	Probability of a fire where non-functioning smoke alarms present		Binomial	r_fire.nonfunc = 7,052 n_hh = 22,539,000
p_fire.noSA	Probability of a fire where no smoke alarms present or unspecified		Binomial	r_fire.noSA = 22,883 n_hh = 22,539,000
p_fatalSA	Probability of a fatality following a fire where functioning smoke alarm present		Binomial	r_fatalSA = 109 r_fire.func = 22,771
p_fatalnSA	Probability of a fatality following a fire where non-functioning or no smoke alarm		Binomial	r_fatalnSA = 232 r_fire.nonfunc = 22,771
p.1yrbattery	Probability own a smoke alarm with battery life of 1 year	Survey of English Housing 2004/5 (Government, 2006) - Table 5.3	Binomial	r.1yrbattery = 11,888 n.1yrbattery=15,850
p.test1yr	Probability test smoke alarm at least once a year	Survey of English Housing 2004/5 (Government, 2006) - Fig 5.1	Binomial	r.1yrtest = 15,616 n.test = 18,372
p.testless1yr	Probability test smoke alarm less than once a year		Binomial	r.testless1yr = 367 n.test = 18,372
p_noinjury	Probability of incurring 'no injuries' following a house fire (given functioning smoke alarm/ non-functioning or no smoke alarm)	Survey of English Housing 2004/5 (Government, 2006) – Table 3.7; Fire statistics 2007 (Government, 2007) - Table 2.4	Binomial	r_noinjury = 47,967 n_total.fires= r_fire.func + r_fire.nonfunc + r_fire.noSA
p_FRSattend	Probability of inside household fire being attended by the Fire and Rescue Service	Survey of English Housing 2004/5 (Government, 2006) - Table 3.4	Binomial	r_FRSattend = 9 n_fires = 272
p_ITU	The additional proportion of burn unit costs incurred in ITU	Assumption based on analysis in Hemington-Gorse *et al.* (2009)	Binomial	P_ITUa = 9.2 P_ITUb = 13.8 based on E[p_ITU] = 0.4 and assuming se[p_ITU] = 0.1 **
p_minor	Probability a child aged 0-4 incurs in a minor injury following a house fire	Ken Dunn (personal communication)	Multinomial	r_minor = 7 n_burnsinjury = 19
p_moderate	Probability a child aged 0-4 incurs in a moderate injury following a house fire		Multinomial	r_moderate = 3 n_burnsinjur = 19
p_severe	Probability a children aged 0-4 incurs in a severe injury following a house fire		Multinomial	r_severe = 1 – r_moderate – r_minor = 9; n_burnsinjury = 19
p_allcause	Probability of all cause mortality for a UK citizen from 0 to 100 years	(ONS, 2010b)		Please see Table 4.5

	old (for use in each decision model cycle)			

Parameter type: Resource cost

Interventions Costs				£
c_hsi	Cost of home safety inspection based on cost of LA home care worker for 40 minutes of their time including travel	PSSRU 2008 (Curtis, 2008)	Fixed	12
c_alarmg	Cost of smoke alarm giveaway (with ten-year sealed battery)	Personal communication Jane Zdanowska	Fixed	4.41
c_educ	Cost of providing education programme per household accepting intervention - based on cost of home care worker for 20 minutes of their time including travel	DiGuiseppi *et al.* (1999) – updated to 2009 prices	Fixed	6
c_fixed	Fixed cost of an intervention scheme (e.g. set-up, administration, etc). Composite value derived from cost analysis of DiGuiseppi *et al.* (1999)		Fixed	54,997
c_acc	Additional cost incurred for each household that accept intervention (composite value)		Fixed	0.49
c_install	Cost of having the smoke alarm installed		Fixed	10.66
Health Care Costs				£
c_minPU	Cost per minute of a Paramedic Unit	PSSRU 2008 (Curtis, 2008) - updated to 2009 prices	Fixed	7.21
c_minEA	Cost per minute of a Emergency Ambulance		Fixed	7.11
mn.minor se_minor	Mean cost (and standard error) of a minor injury	Ken Dunn (Personal communication)	Lognormal	E[minor] = 1,087 se[minor] = 209
mn.moderate se_moderate	Mean cost (and standard error) of a moderate injury		Lognormal	E[moderate] = 2,573 se[moderate] = 1,415
mn.severe se_severe	Mean cost (and standard error) of a severe injury		Lognormal	E[severe] = 58,519 se[severe] = 32,019
c_fatal	Cost of a fatality following a household fire (updated to 2008/9 prices) – includes coroners and autopsy costs	Ginnelly *et al*. (2005b) - updated to 2009 prices	Fixed	185.16
c_dispyr	Mean incurred NHS costs of disability per year	HALO study (Nicholl et al., 2009) and personal communication with Jon Nicholl	Gamma	c_dispyra = 16 c_dispyrb = 0.046784 based on E[c_dispyr] = 342 and assuming se[c_dispyr] = 85.5 **
c_checkup	Cost of precautionary check-up of safety equipment	NHS reference costs 2008/9 (*2008/9 NHS Reference Costs Guidance*, 2010) - code VB11Z	Normal	E[c_checkup] = 62 se[c_checkup] = 21.5 based on interquantile range
Out of Pocket / Private Costs				£

c_battery	Cost of smoke alarm 1 year battery to individual	www.safelincs.co.uk – 2009 price	Fixed	1.39
c_property	Total cost of damage caused by the fire	British Crime Survey: Fires in the Home 2002/3 (Government, 2004b) - updated to 2009 prices	Gamma	c_propertya = 16 c_propertyb = 0.016327 based on E[c_property] = 980 and assuming se[c_property] = 245 **
Law Enforcement and Rescue Services Costs				£
c_police	Cost of police attending – assumed only to attend where severe injuries	Ginnelly *et al*. (2005b) - updated to 2009 prices	Fixed	156.67
c_FRSresponse	Cost of Fire and rescue Service attending a fire	Economic Cost of Fire 2004 (Government, 2004a) - Table 3.6 - updated to 2009 prices	Fixed	3,051

Parameter type: Utility parameters per cycle

u_min	Deficit in utilities for minor injury (DRG 460 + 459)	Sanchez et al. (2008)	Fixed	0.0487
u_mod	Deficit in utilities for moderate injury (DRG 458 + 457)		Fixed	0.069
u_sev	Deficit in utilities for severe injury (DRG 472)		Fixed	0.107
u_deficit	Deficit in utilities following a disability	HALO study (Nicholl et al., 2009) and personal communication with Jon Nicholl	Beta	u_deficita = 14.3 u_deficitb = 128.7 based on E[u_deficit] = 0.1 and assuming se[u_deficit] = 0.025 **
u_pop	General background mean utilities for non-injured population	UK Population Norms (Kind et al., 1999)	Normal	Under 25yrs, mean=0.94 (se=0.0021) 25-34yrs, mean=0.93 (se=0.0026) 35-44yrs, mean=0.91 (se=0.0027) 45-54yrs, mean= 0.85 (se=0.0043) 55-64yrs, mean=0.80 (se=0.0045) 65-74yrs, mean=0.78 (se=0.0045) >75yrs, mean=0.73 (se=0.0045)

** using the method of moments

Appendix 7

Chapter 6 formalises the concept of value of additional research in the presence of mutually exclusive subgroups and of IPD. In this context, NMB estimates under current and perfect information may increase (or decrease) from 1 to *k* subgroups (*k* representing the maximum number of possible subgroups, considering observable and non-observable individual characteristics). The decision made in each subgroup is dependent on determining whether the same decision is made across subgroups or a different decision is made in at least one of the subgroups considered. However, under current information, if the same decision is made for all subgroups, estimated total expected NMBs are the same, irrespective of the number of subgroups considered.

In practice, if considering *q* subgroups from an existing set of *l* subgroups (i.e. $q < l < k$) an available set of NB estimates can be used for *l* to perform EVPI estimations over *q* subgroups, from a total of *k* existing subgroups – what it is termed here 'backward estimation'. For instance, in the case of modelling for 2 (*l*) subgroups and wanting to derive the EVPI for the whole population (1 (*q*) subgroup), with $NB_l^{n \cdot y}$ representing a matrix of *n* samples, from probabilistic modelling, for *y* strategies, and w_l the proportion of each subgroup in the population, the following formula should be applied:

$$NB_{q,\theta}^{n \cdot y} = \sum NB_l^{n \cdot y} \cdot w_l \overset{\text{under } l=2}{=}$$

$$\begin{bmatrix} NB_{q,\theta}^{1,1} & \dots & NB_{q,\theta}^{1,y} \\ \dots & \dots & \dots \\ NB_{q,\theta}^{1,1} & \dots & NB_{q,\theta}^{n,y} \end{bmatrix} = \begin{bmatrix} NB_1^{1,1} & \dots & NB_1^{1,y} \\ \dots & \dots & \dots \\ NB_1^{1,1} & \dots & NB_1^{n,y} \end{bmatrix} \cdot w_1 + \begin{bmatrix} NB_2^{1,1} & \dots & NB_2^{1,y} \\ \dots & \dots & \dots \\ NB_2^{1,1} & \dots & NB_2^{n,y} \end{bmatrix} \cdot w_2 \quad \text{(A7.1)}$$

The weights are applied directly on the $n \cdot y$ samples of the NB. The result is a matrix of weighted average NBs of same dimensions. On this matrix, the usual functions of the maximum and arithmetic mean are applied to derive the perfect and current information as in equation (6.3). The difference between this framework and the one set out in equations (6.7) and (6.8) is that the weights are applied to the NB estimates before applying the maximum and mean functions, rather than after they are applied .

If, for instance, NBs estimates for 4 (*l*) subgroups are available and one wants to obtain estimates for a particular 2 (*q*) subgroups specification, initially equation (6.10) could be applied to obtain the weighted averages for each of the 2 *(q)* subgroups ($NB_1^{n \cdot y}$ and $NB_2^{n \cdot y}$), using weights w_l^* , before proceeding with the formula described in equation (6.3).

Illustrative example

To illustrate the backward estimation procedure, an illustrative numerical example is provided in Table A7.1. In this example a simple hypothetical decision problem compares three interventions (i.e. *Int 1*, *Int 2*, and *Int 3*) and four (mutually exclusive) subgroups representing different proportions of the total targeted patient population (i.e. diabetic and hypertensive – 10%; not diabetic and not hypertensive – 45%; diabetic and not hypertensive – 15%; and not diabetic and hypertensive – 30%). For simplicity, only five iterations of a probabilistic sensitivity analysis (PSA) are reported for each subgroup and intervention. The expected maximum NMBs are calculated for each alternative (averages at the bottom of the table) and the maximum expected NMBs are calculated at each iteration and for each subgroup. Under current information, *Int 3* is the optimal choice for all subgroups (i.e. the same decision is made for each of the subgroups being considered). The total weighted average NMB with current information is of £12.48 [(7.2 x 0.1) + (13.8 x 0.45) + (14.6 x 0.15) + (11.2 x 0.3) = 12.48]. The total weighted average NMB with perfect information is of £15.4 [(11.8 x 0.1) + (15.8 x 0.45) + (16.6 x 0.15) + (15.4 x 0.3) =

15.4]. The EVPI considering the four subgroups is estimated to be £2.92 [15.4 – 12.48 = 2.92].

However, comparisons across each subgroup specification may be compromised if estimates used in each of these subgroup analyses resulted from different synthesis models. This approach was used in the chapter's case study, making use of equations 6.5 to 6.7 as described in section 6.3.1.

Table A7.1 – Value of Heterogeneity: the four subgroup illustrative example.

Iteration	Subgroup 1 - 10% DH				Subgroup 2 - 45% NDNH				Subgroup 3 - 15% DNH				Subgroup 4 - 30% NDH			
	Int 1	*Int 2*	*Int 3*	*Max*	*Int 1*	*Int 2*	*Int 3*	*Max*	*Int 1*	*Int 2*	*Int 3*	*Max*	*Int 1*	*Int 2*	*Int 3*	*Max*
1	13	2	1	13	10	6	10	10	7	12	17	17	18	15	15	18
2	5	10	1	10	6	19	19	19	12	4	12	12	1	13	4	13
3	9	9	7	9	8	14	4	14	18	9	12	18	11	1	3	11
4	0	4	18	18	3	12	17	17	0	14	17	17	5	12	19	19
5	4	2	9	9	7	2	19	19	19	9	15	19	16	9	15	16
Average	6.20	5.40	**7.20**	11.80	6.80	10.60	**13.80**	15.80	11.20	9.60	**14.60**	16.60	10.20	10.00	**11.20**	15.40

NMB with current information = 7.2 x 0.1 + 13.8 x 0.45 + 14.6 x 0.15 + 11.2 x 0.3 = 12.48

NMB with perfect information = 11.8 x 0.1 + 15.8 x 0.45 + 16.6 x 0.15 + 15.4 x 0.3 = 15.4

$EVPI_K$ = 15.4 - 12.48 = 2.92

Legend: DH - diabetic and hipertense; NDNH - not diabetic and not hipertense; DNH - diabetic and not hipertense; NDH - not diabetic and hipertense

The four subgroup specifications can be used to estimate the NMB estimates for the population average, and for each of the two subgroup specifications. Table A7.2 follows on from Table A7.1, illustrating how the backwards estimating procedure can be performed to estimate population average estimates from four subgroups. Following equation (A7.1), each NMB estimate (of each PSA iteration, each subgroup and each interaction from Table A7.1) is weighted by the subgroup proportion of the total population. For instance, for subgroup 1 (i.e. the DH subgroup

which represents 10% of the population), iteration 1 and intervention 1, a weighted NMB estimate of £1.3 is obtained (13 x 0.1 = 1.3). Once all the subgroup weighted NMB estimates are obtained the weighted average population NMBs are simply the sum of these. For instance, the population NMB for intervention 2 and iteration 3 is calculated to be £8.85 (9 x 0.1 + 14 x 0.45 + 9 x 0.15 + 1 x 0.3 = 0.9 + 6.3 + 1.35 + 0.3 = 8.85). With current information, the optimal choice is *Int 3* with an estimated NMB of £12.48. With perfect information, an expected maximum NMB of £14.14 is estimated. Thus, the EVPI estimate using the backward estimation procedure from 4 to 1 subgroup is £1.66 (14.14 – 12.48 = 1.66).

Table A7.2 – Value of Heterogeneity: backward estimation from four subgroups to population average estimates.

Iteration	**Subgroup 1 - 10% DH**			**Subgroup 2 - 45% NDNH**			**Subgroup 3 - 15% DNH**			**Subgroup 4 - 30% NDH**			**Population**			
	Int 1	*Int 2*	*Int 3*	*Int 1*	*Int 2*	*Int 3*	*Int 1*	*Int 2*	*Int 3*	*Int 1*	*Int 2*	*Int 3*	*Int 1*	*Int 2*	*Int 3*	*Max*
1	1.3	0.2	0.1	4.5	2.7	4.5	1.05	1.8	2.55	5.4	4.5	4.5	12.25	9.2	11.65	12.25
2	0.5	1	0.1	2.7	8.55	8.55	1.8	0.6	1.8	0.3	3.9	1.2	5.3	14.05	11.65	14.05
3	0.9	0.9	0.7	3.6	6.3	1.8	2.7	1.35	1.8	3.3	0.3	0.9	10.5	8.85	5.2	10.5
4	0	0.4	1.8	1.35	5.4	7.65	0	2.1	2.55	1.5	3.6	5.7	2.85	11.5	17.7	17.7
5	0.4	0.2	0.9	3.15	0.9	8.55	2.85	1.35	2.25	4.8	2.7	4.5	11.2	5.15	16.2	16.2
Average	0.62	0.54	**0.72**	3.06	4.77	**6.21**	1.68	1.44	**2.19**	3.06	3.00	**3.36**	8.42	9.75	**12.48**	14.14

e.g. Subgroup 1 - Intervention 1 - Iteration 1 = 13 x 0.1 = 1.3 ; Subgroup 4 - Intervention 3 - Iteration 5 = 15 x 0.3 = 4.5

e.g. Population - Intervention 2 - Iteration 3 = 9 x 0.1 + 14 x 0.45 + 9 x 0.15 + 1 x 0.3 = 8.85

NMB with current information = 12.48 (same as before)

NMB with perfect information = 14.14

EVPI using backward estimation (4 to 1 subgroups) = 14.14 - 12.48 = 1.66

As the same choice of intervention is being made for each population subgroup, the same NMBs with current information are expected to be obtained whether the population average or four subgroups are considered. As expected, the NMB estimate with current information considering subgroups (from Table A7.1) is

equivalent to the weighted average population NMB estimate with current information from using the backward estimation procedure (Table A7.2), that is, £12.48. With respect to the NMBs obtained with perfect information, these are estimated to be higher when considering the four subgroups (i.e. £15.4 – Table A7.1) compared to the (weighted average) population (i.e. £14.14 – Table A7.2).

Table A7.3 and A7.4 illustrate the backward estimation approach from four subgroups to two subgroups (for specification 1 – diabetic and not diabetic populations, and for specification 2 – hypertensive and not hypertensive populations, respectively) and from two subgroups to the population average NMB estimates. The first step is to calculate the proportions for each (nested) population (e.g. the proportion of hypertensive individuals in the diabetic population is of 0.4 [i.e. 0.1 / (0.1 + 0.15) = 0.4].

Again following equation (A7.1), each NMB estimate (of each PSA iteration, each subgroup and each interaction from Table A7.1) is weighted by the subgroup proportion of the population. For instance, for the subgroup Diab, iteration 1 and intervention 1, a weighted NMB estimate of £9.4 is obtained (13 x 0.4 + 7 x 0.6 = 9.4). Once all the NMB estimates are obtained for each subgroup, the NMBs with current and perfect information can be obtained (e.g. the NMBs with current and perfect information for the Diab subgroup are £ 11.64 and £12.92, respectively). With these values, equations 6.5 to 6.7 may be applied to obtain weighted population average estimates. For the Diab / Not Diab specification, the NMBs obtained with perfect information are £14.36 [(12.92 x 0.25) + (14.84 x 0.75) = 14.36] and the EVPI is estimated to be £1.88 (14.36 – 12.48 = 1.88). Similar calculations can be made for the Hyper / Not Hyper subgroup specification. For that subgroup, estimated EVPI is £2.47 (14.95 – 12.48 = 2.47).

Table A7.3 – Value of Heterogeneity: backward estimation from four to two subgroups (specification 1) and from two subgroups to population estimates.

Iteration	Subgroup Diab				Subgroup Not Diab			
	Int 1	*Int 2*	*Int 3*	*Max*	*Int 1*	*Int 2*	*Int 3*	*Max*
1	9.4	8	10.6	10.6	13.2	9.6	12	13.2
2	9.2	6.4	7.6	9.2	4	16.6	13	16.6
3	14.4	9	10	14.4	9.2	8.8	3.6	9.2
4	0	10	17.4	17.4	3.8	12	17.8	17.8
5	13	6.2	12.6	13	10.6	4.8	17.4	17.4
Average	9.20	7.92	**11.64**	12.92	8.16	10.36	**12.76**	14.84

Proportion diabetic = 0.1 + 0.15 = 0.25

Proportion not diabetic = 0.45 + 0.3 = 0.75

Proportion of hipertense in diabetic population = 0.1 / 0.25 = 0.4

Proportion of not hipertense in diabetic population = 0.15 / 0.25 = 0.6

Proportion of hipertense in not diabetic population = 0.3 / 0.75 = 0.4

Proportion of not hipertense in not diabetic population = 0.45 / 0.75 = 0.6

e.g. Subgroup Diab - Intervention 1 - Iteration 1 = 13 x 0.4 + 7 x 0.6 = 9.4 ;

Subgroup Not Diab - Intervention 3 - Iteration 5 = 19 x 0.4 + 15 x 0.6 = 17.4

Subgroup Diab:

NMB with current information = 11.64

NMB with perfect information = 12.92

$EVPI_K$ using backward estimation (4 to 2 subgroups) = 12.92 - 11.64 = 1.28

Subgroup Not Diab:

NMB with current information = 12.76

NMB with perfect information = 14.84

$EVPI_K$ using backward estimation (4 to 2 subgroups) = 14.84 - 12.76 = 2.08

Population:

NMB with current information = 11.64 x 0.25 + 12.76 x 0.75 = 12.48 (same as before)

NMB with perfect information = 12.92 x 0.25 + 14.84 x 0.75 = 14.36

$EVPI_K$ using backward estimation (2 to 1 subgroup) = 14.36 - 12.48 = 1.88

Table A7.4 – Value of Heterogeneity: backward estimation from four to two subgroups (specification 2) and from two subgroups to population estimates.

Iteration	Subgroup Hiper				Subgroup Not Hiper				
	Int 1	*Int 2*	*Int 3*	*Max*	*Int 1*	*Int 2*	*Int 3*	*Max*	Proportion hipertense = 0.1 + 0.3 = 0.4
1	16.8	11.8	11.5	16.8	9.25	7.5	11.8	11.8	Proportion not hipertense = 0.15 + 0.45 = 0.6
2	2	12.3	3.25	12.3	7.5	15.3	17.3	17.3	Proportion of diabetic in hipertense population = 0.1 / 0.4 = 0.25
3	10.5	3	4	10.5	10.5	12.8	6	12.8	Proportion of not diabetic in hipertense population = 0.3 / 0.4 = 0.75
4	3.75	10	18.8	18.8	2.25	12.5	17	17	Proportion of diabetic in not hipertense population = 0.15 / 0.6 = 0.25
5	13	7.25	13.5	13.5	10	3.75	18	18	Proportion of not diabetic in not hipertense population = 0.45 / 0.6 = 0.75
Average	9.20	8.85	**10.20**	14.35	7.90	10.35	**14.00**	15.35	

e.g. Subgroup Hiper - Intervention 1 - Iteration 1 = 13 x 0.25 + 18 x 0.75 = 16.75 ;

Subgroup Not Hiper - Intervention 3 - Iteration 5 = 19 x 0.25 + 15 x 0.75 = 18

Subgroup Hiper

NMB with current information = 10.20

NMB with perfect information = 14.35

$EVPI_K$ using backward estimation (4 to 2 subgroups) = 14.35 - 10.20 = 4.15

Subgroup Not Hiper

NMB with current information = 14.00

NMB with perfect information = 15.35

$EVPI_K$ using backward estimation (4 to 2 subgroups) = 15.35 - 14.00 = 1.35

Population:

NMB with current information = 10.20 x 0.4 + 14.00 x 0.6 = 12.48 (same as before)

NMB with perfect information = 14.35 x 0.4 + 15.35 x 0.6 = 14.95

$EVPI_K$ using backward estimation (2 to 1 subgroup) = 14.95 - 12.48 = 2.47

Appendix 8

Case study results using the 'backward estimation'

Throughout the chapter, results are characterized by using different synthesis models, depending on whether no covariate information was available, whether this was available for just one binary covariate or whether it was available for both covariates. Using the NBs estimates for 4 subgroups (i.e. considering the 2 covariates), equation (A7.1) in addition to (6.7) to (6.9) may be applied to derive EVPI estimates for the population (1 subgroup) and for the 2 subgroup specifications. As the same decision is made for all four subgroups, estimated NBs under current information should be the same when considering for 4, 2 or 1 subgroup. However, differences are expected in the NBs estimated under perfect information.

Table A8.1 and A8.2 show the NBs results when using the output of synthesis model with AD and with AD + IPD, respectively. It can be observed that in either case estimated NBs obtained under current information are the same irrespective of the number of subgroups or specifications considered. Under current information, the NBs obtained from using AD + IPD are higher than those obtained from using AD only, indicating that, using only the latter evidence format, the population net gains may be underestimated. With respect to the population EVPI results, the expected cost of uncertainty increases with the number of subgroups. This phenomenon is observed for both threshold ratios used and for both scenarios with and without the use of IPD. Moreover, except for the 1 subgroup estimates, population EVPI estimates are higher when considering AD only compared to when considering AD + IPD. These results are consistent with the conceptual idea reflected throughout the chapter, that using IPD translates into a reduction of the expected cost of uncertainty.

Table A8.1 – Expected (individual level) NMBs under current information and for population EVPI (2009 values) of the cost effectiveness decision model for functioning smoke alarms, for a threshold value of £20,000 and of £30,000. Results are shown for the use of AD only and when heterogeneity is considered, that is, no subgroups, weighted average for 2 subgroups (2 specifications – number of parents in the family and their employment status) and the weighted average for the 4 subgroups are considered.

			Using AD effectiveness data			
		Threshold value	**4 subgroups**	**2 subgroups - number parents**	**2 subgroups - employment status**	**1 subgroup**
Net monetary benefits (£)	**under Current Information**	**£20,000**	£111,939	£111,939	£111,939	£111,939
		£30,000	£656,550	£656,550	£656,550	£656,550
	Population EVPI	**£20,000**	£6,478	£1,262	£1,790	£144
		£30,000	£46,168	£24,407	£29,865	£13,255

The fact that estimates obtained with AD are lower than with AD + IPD for one subgroup is contra-intuitive. As discussed in section 6.2.2 of Chapter 6, one reason for this phenomenon may be the fact that in the synthesis procedure, cluster adjustment is performed within the synthesis model when modelling IPD, while for AD these adjustments are made outside of the model. This implies that, for IPD modelling, there is an additional layer of uncertainty relating to the design effect that is not being considered when modelling AD only. See Chapter 3 for further details on this issue. Another potential reason, albeit less probable, is that of simulation error.

Table A8.2 – Expected (individual level) NMBs under current information and for population EVPI (2009 values) of the cost effectiveness decision model for functioning smoke alarms, for a threshold value of £20,000 and of £30,000. Results are shown for the use of AD + IPD and when heterogeneity is considered, that is, no subgroups, weighted average for 2 subgroups (2 specifications – number of parents in the family and their employment status) and the weighted average for the 4 subgroups are considered.

			Using AD+IPD effectiveness data			
		Threshold value	**4 subgroups**	**2 subgroups - number parents**	**2 subgroups - employment status**	**1 subgroup**
Net monetary benefits (£)	**under Current Information**	**£20,000**	£118,447	£118,447	£118,447	£118,447
		£30,000	£663,061	£663,061	£663,061	£663,061
	Population EVPI	**£20,000**	£5,499	£957	£1,699	£155
		£30,000	£42,235	£27,188	£29,429	£17,723

LIST OF ABBREVIATIONS (IN ALPHABETIC ORDER)

1P	Single parent or one parent household/family
1U	At least one unemployed parent in household/family
1EP	Employed single parent household/family
1UP	Unemployed single parent household/family
2P	Two parent household/family
2U	Employed parent in household/family
2EP	Two employed parents in household/family
2UP	Two parent household/family with at least one parent unemployed
AD	Aggregate level data
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
BUGS	Bayesian inference Using Gibbs Sampling
CA	Cluster Allocation of trial participants
CADTH	Canadian Agency for Drugs and Technologies in Health
CBA	Controlled Before and After trial
CBG-MEB	Medicines Evaluation Board - Netherlands
CEA	Cost Effectiveness Analysis
CEAC	Cost Effectiveness Acceptability Curve
CEAF	Cost Effectiveness Acceptability Frontier

CODA	Convergence Diagnostic and Output Analysis
CrI	Credibility Interval
DIC	Deviance Information Criteria
DH	Diabetic and Hypertensive patients
DNH	Diabetic and Non-Hypertensive patients
DoHA - HTA	Department of Health and Ageing's Health Technology Assessment
DSU	Decision Support Unit
E	Education
E + FE	Education plus Free Equipment
E + FE + HI	Education plus Free Equipment plus Home Inspection
E + FE + F	Education plus Free Equipment plus Fitting of equipment
E + HI	Education plus Home Inspection
E + FE + F + HI	Education plus Free Equipment plus Fitting of equipment plus Home Inspection
EQ-5D	EuroQol Five-Dimensional
EVIC	Expected Value of Individualised Care
EVPI	Expected Value of Perfect Information
EVPPI	Expected Value of Perfect Information for parameters or Partial Expected Value of Perfect Information
EVSI	Expected Value of Sampling Information
FE	Fixed-effect
FINOHTA	Finnish Office for Health Technology Assessment
HALO	The Long Term Health and Healthcare outcomes of Accidental Injury study

HIV/AIDS	Human Immunodeficiency Virus / Acquired ImmunoDeficiency Syndrome
HMO	Health Maintenance Organisation
HTA	Health Technology Assessment
HUI	Health Utility Index
IA	Individual Allocation of trial participants
ICC	Intra Class Correlation Coefficient
ICD	International Classification of Diseases
***i*-CER**	Individualized-Comparative Effectiveness Research
ICER	Incremental Cost Effectiveness Ratio
INFARMED	National Authority of Medicines and Health Products
IPD	Individual participant level data
MCMC	Markov Chain Monte Carlo
MPA	Medical Products Agency - Sweden
MPES	Multi Parameter Evidence Synthesis
MTC	Mixed Treatment Comparison
NB	Net Benefit
NDH	Non-Diabetic and Hypertensive
NDNH	Non-Diabetic and Non-Hypertensive
NHB	Net Health Benefit
NHS	National Health System
NMB	Net Monetary Benefit
NICE	National Institute for Health and Clinical Excellence

NRT	Non-Randomised Trial
ONS	Office for National Statistics
OR	Odds Ratio
PEVPI	Population Expected Value of Perfect Information
PH	Public Health
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality Adjusted Life Year
QWB	Quality of Well-Being
RCT	Randomised Controlled Clinical Trials
RE	Random-effect
RITA 3	Third Randomised Intervention Trial of unstable Angina
SBU	Swedish Council on Health Technology Assessment
SF-6D	Short Form Six-Dimensional
UC	Usual Care
UK	United Kingdom
USA	United States of America
VoH	Value of Heterogeneity

REFERENCES

2008/9 NHS Reference Costs Guidance (2010). London.

Abrams, K., Gillies, C. and Lambert, P. (2005). Meta-analysis of heterogeneously reported trials assessing change from baseline. *Statistics in Medicine,* 24(24), 3823-3844.

Ades, A., Lu, G. and Claxton, K. (2004). Expected Value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making,* 24(2), 207-227.

Ades, A. E. (2003). A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med,* 22(19), 2995-3016.

Ades, A. E., Claxton, K. and Sculpher, M. (2006a). Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Econ,* 15(4), 373-381.

Ades, A. E. and Lu, G. (2003). Correlations between parameters in risk models: estimation and propagation of uncertainty by Markov Chain Monte Carlo. *Risk Anal,* 23(6), 1165-1172.

Ades, A. E., Lu, G. and Higgins, J. P. (2005). The interpretation of random-effects meta-analysis in decision models. *Med Decis Making,* 25(6), 646-654.

Ades, A. E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N. and Lu, G. B. (2006b). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics,* 24(1), 1-19.

Ades, A. E. and Sutton, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society Series a-Statistics in Society,* 169, 5-35.

Akaike, H. (1973). Information theory and the maximum likelihood principle. *2nd International Symposium in Information Theory*. Springer-Verlag, pp. 267-281.

Anderson, R. (2010). Systematic reviews of economic evaluations: utility or futility? *Health Economics,* 19(3), 350-364.

Appleby, J., Devlin, N. and Parkin, D. (2007). NICE's cost effectiveness threshold. *BMJ,* 335(7616), 358-359.

Ara, R. and Brazier, J. E. (2010). Populating an economic model with health state utility values: moving toward better practice. *Value Health,* 13(5), 509-518.

Armstrong, R., Waters, E., Jackson, N., Oliver, S., Popay, J., Shepherd, J., Petticrew, M., Anderson, L., Bailie, R., Brunton, G., Hawe, P., Kristjansson, E., Naccarella, L.,

Norris, S., Pienaar, E., Roberts, H., Rogers, W., Sowden, A. and Thomas, H. (2007). *Guidelines for Systematic reviews of health promotion and public health interventions. Version 2*. Melbourne University: Australia.

Augustovski, F., Iglesias, C., Manca, A., Drummond, M., Rubinstein, A. and Marti, S. G. (2009). Barriers to generalizability of health economic evaluations in Latin America and the Caribbean region. *Pharmacoeconomics,* 27(11), 919-929.

Baker, S. G. (2006). A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. *Biostatistics,* 7(1), 58-70.

Barone, V. J. (1988). *An analysis of well-child parenting classes: the extent of parenting compliance with health-care recommendations to decrease potential injury of their toddlers.* Kansas, USA: University of Kansas.

Basu, A. (2009). Individualization at the heart of comparative effectiveness research: The time for i-CER has come. *Medical Decision Making,* 29(6), N9-N11.

Basu, A. (2011). Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *Journal of Health Economics,* 30(3), 549-559.

Basu, A. and Meltzer, D. (2007). Value of information on preference heterogeneity and individualized care. *Med Decis Making,* 27(2), 112-127.

Basu, A. and Meltzer, D. (2010). Quantitative Methods for Valuing Comparative Effectiveness Information. *Biopharmaceutical Report,* 17(2), 2-10.

Berlin, J. A., Santanna, J., Schmid, C. H., Szczech, L. A. and Feldman, H. I. (2002). Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med,* 21(3), 371-387.

Birch, S. and Gafni, A. (2003). Economics and the evaluation of health care programmes: generalisability of methods and implications for generalisability of results. *Health Policy,* 64(2), 207-219.

Birch, S. and Gafni, A. (2006). The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold. *J Health Serv Res Policy,* 11(1), 46-51.

Bland, M. (2000). *An introduction to medical statistics*. Oxford medical publications. 3rd ed. Oxford ; New York: Oxford University Press.

Bojke, L., Claxton, K., Sculpher, M. and Palmer, S. (2009). Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. *Value Health*.

Bower, P., Byford, S., Barber, J., Beecham, J., Simpson, S., Friedli, K., Corney, R., King, M. and Harvey, I. (2003). Meta-analysis of data on costs from trials of counselling in primary care: using individual patient data to overcome sample size limitations in economic analyses. *BMJ,* 326(7401), 1247-1250.

Brazier, J., Roberts, J. and Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *J Health Econ,* 21(2), 271-292.

Bremner, K. E., Chong, C. A., Tomlinson, G., Alibhai, S. M. and Krahn, M. D. (2007). A review and meta-analysis of prostate cancer utilities. *Med Decis Making,* 27(3), 288-298.

Brennan, A., Bansback, N., Nixon, R., Madan, J., Harrison, M., Watson, K. and Symmons, D. (2007). Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. *Rheumatology (Oxford),* 46(8), 1345-1354.

Briggs, A. (2000). Economic evaluation and clinical trials: size matters. *BMJ,* 321(7273), 1362-1363.

Briggs, A., Mihaylova, B., Sculpher, M., Hall, A., Wolstenholme, J., Simoons, M., Deckers, J., Ferrari, R., Remme, W. J., Bertrand, M. and Fox, K. (2007). Cost effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study. *Heart,* 93(9), 1081-1086.

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

Briggs, A. H., Claxton, K. and Sculpher, M. J. (2006). *Decision modelling for health economic evaluation*. Oxford handbooks in health economic evaluation. Oxford: Oxford University Press.

Brooks, R. (1996). EuroQol: the current state of play. *Health Policy,* 37(1), 53-72.

Bulzacchelli, M. T., Gielen, A. C., Shields, W. C., McDonald, E. M. and Frattaroli, S. (2009). Parental safety-related knowledge and practices associated with visiting a mobile safety center in a low-income urban population. *Family & Community Health,* 32(2), 147-158.

Burns, F. (1998). *An Information Strategy for the Modern NHS 1998-2005*. NHS Executive.

Burton, A., Altman, D., Royston, P. and Holder, R. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine,* 25(24), 4279-4292.

Caldwell, D. M., Ades, A. E. and Higgins, J. P. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ,* 331(7521), 897-900.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes methods for data analysis*. Chapman & Hall/CRC texts in statistical science series. 2nd ed. Boca Raton: Chapman & Hall/CRC.

Carlin, B. P. and Louis, T. A. (2009). *Bayesian methods for data analysis*. Chapman & Hall/CRC texts in statistical science series. 3rd ed. Boca Raton: CRC Press.

Carpenter, J., Kenward, M. and White, I. (2007). Sensitivity analysis after multiple imputation under missing at random: A weighting approach. *Statistical Methods in Medical Research,* 16(3), 259-275.

Chao, T. K. and Chen, T. H. (2009). Predictive model for progression of hearing loss: meta-analysis of multi-state outcome. *Journal of Evaluation in Clinical Practice,* 15(1), 32-40.

Cheng, A. K. and Niparko, J. K. (1999). Cost-utility of the cochlear implant in adults: a meta-analysis. *Arch Otolaryngol Head Neck Surg,* 125(11), 1214-1218.

Clamp, M. and Kendrick, D. (1998). A randomised controlled trial of general practitioner safety advice for families with children under 5 years. *BMJ,* 316(7144), 1576-1579.

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

Claxton, K. (2011). Heterogeneity in the Cost-effectiveness of Medical Interventions: The Challenge of Matching Patients to Appropriate Care. *ISPOR,* Madrid.

Claxton, K., Paulden, M., Gravelle, H., Brouwer, W. and Culyer, A. (2011). Discounting and decision making in the economic evaluation of health-care technologies. *Health Economics,* 20(1), 2-15.

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

Claxton, K., Sculpher, M., Culyer, A., McCabe, C., Briggs, A., Akehurst, R., Buxton, M. and Brazier, J. (2006). Discounting and cost-effectiveness in NICE - stepping back to sort out a confusion. *Health Econ,* 15(1), 1-4.

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

Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., Brazier, J. and O'Hagan, T. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ,* 14(4), 339-347.

Claxton, K., Sculpher, M. J. and Culyer, A. J. (2007). *Mark versus Luke? Appropriate methods for the evaluation of public health interventions.* Centre for Health Economics, Research Paper 31. University of York.

Claxton, K. and Thompson, K. M. (2001). A dynamic programming approach to the efficient design of clinical trials. *J Health Econ,* 20(5), 797-822.

Claxton, K. P. and Sculpher, M. J. (2006). Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics,* 24(11), 1055-1068.

Collett, D. (2003). *Modelling survival data in medical research.* Chapman & Hall/CRC texts in statistical science series. 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC.

Conner-Spady, B. and Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care,* 41(7), 791-801.

Connock, M., Frew, E., Evans, B. W., Bryan, S., Cummins, C., Fry-Smith, A., Li Wan Po, A. and Sandercock, J. (2006). The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. A systematic review. *Health Technol Assess,* 10(7), iii, ix-118.

Cookson, R., Flood, C., Koo, B., Mahon, D. and Rhodes, M. (2005). Short-term cost effectiveness and long-term cost analysis comparing laparoscopic Nissen fundoplication with proton-pump inhibitor maintenance for gastro-oesophageal reflux disease. *Br J Surg,* 92(6), 700-706.

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

Cooper, N., Kendrick, D., Achana, F., Dhiman, P., He, Z., Wynn, P., Le Cozannet, E., Saramago, P. and Sutton, A. (2012). Network meta-analysis to evaluate the effectiveness of interventions to increase the uptake of smoke alarms. *Epidemiologic Reviews,* 34(1), 32-45.

Cooper, N. J., Sutton, A. J., Abrams, K. R., Turner, D. and Wailoo, A. (2004). Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Econ,* 13(3), 203-226.

Cooper, N. J., Sutton, A. J., Ades, A. E., Paisley, S. and Jones, D. R. (2007). Use of evidence in economic decision models: practical issues and methodological challenges. *Health Econ,* 16(12), 1277-1286.

Cooper, N. J., Sutton, A. J., Morris, D., Ades, A. E. and Welton, N. J. (2009). Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine,* 28(14), 1861-1881.

Coyle, D., Buxton, M. J. and O'Brien, B. J. (2003). Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Econ,* 12(5), 421-427.

CRD (2009). *Systematic Reviews - CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination, University of York.

Culyer, A., McCabe, C., Briggs, A., Claxton, K., Buxton, M., Akehurst, R., Sculpher, M. and Brazier, J. (2007). Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *J Health Serv Res Policy,* 12(1), 56-58.

Curtis, L. (2008). *Unit Costs of Health & Social Care*. Kent: University of Kent.

Curtis, L. (2009). *Unit Costs of Health & Social Care*. Kent: University of Kent.

Dias, S., Sutton, A. J., Welton, N. J. and Ades, A. E. (2011a). NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment.

Dias, S., Welton, N. J. and Ades, A. E. (2010a). Study designs to detect sponsorship and other biases in systematic reviews. *Journal of Clinical Epidemiology,* 63(6), 587-588.

Dias, S., Welton, N. J., Caldwell, D. M. and Ades, A. E. (2010b). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine,* 29(7-8), 932-944.

Dias, S., Welton, N. J., Sutton, A. J. and Ades, A. E. (2011b). *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials*. NICE DSU Technical Support Document Series.

Dias, S., Welton, N. J., Sutton, A. J. and Ades, A. E. (2011c). *NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model*.

DiGuiseppi, C. and Higgins, J. P. (2001). Interventions for promoting smoke alarm ownership and function. *Cochrane Database of Systematic Reviews,* (2), CD002246.

DiGuiseppi, C., Roberts, I., Wade, A., Sculpher, M., Edwards, P., Godward, C., Pan, H. and Slater, S. (2002). Incidence of fires and related injuries after giving out free smoke alarms: cluster randomised controlled trial. *BMJ,* 325(7371), 995.

DiGuiseppi, C. and Roberts, I. G. (2000). Individual-level injury prevention strategies in the clinical setting. *Future of Children,* 10(1), 53-82.

DiGuiseppi, C., Slater, S., Roberts, I., Adams, L., Sculpher, M., Wade, A. and McCarthy, M. (1999). The "Let's Get Alarmed!" initiative: a smoke alarm giveaway programme. *Inj Prev,* 5(3), 177-182.

Dijkers, M. (1997). Quality of life after spinal cord injury: a meta analysis of the effects of disablement components. *Spinal Cord,* 35(12), 829-840.

Donnan, P. T., McLernon, D., Dillon, J. F., Ryder, S., Roderick, P., Sullivan, F. and Rosenberg, W. (2009). Development of a decision support tool for primary care management of patients with abnormal liver function tests without clinically apparent liver disease: a record-linkage population cohort study and decision analysis (ALFIE). *Health Technology Assessment (Winchester, England),* 13(25), iii-iv.

Donner, A. and Klar, N. (2002). Issues in the meta-analysis of cluster randomized trials. *Statistics in Medicine,* 21(19), 2971-2980.

Dowswell, T. and Towner, E. (2002). Social deprivation and the prevention of unintentional injury in childhood: a systematic review. *Health Education Research,* 17(2), 221-237.

Drummond, M. F. and McGuire, A. (2001). *Economic evaluation in health care : merging theory with practice*. Oxford ; New York: Oxford University Press.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. j. and Stoddart, G. L. (2005). *Methods for the economic evaluation of health care programmes*. Oxford medical publications. 3rd ed. Oxford ; New York: Oxford University Press.

Edwards, P., Roberts, I., Green, J. and Lutchmun, S. (2006). Deaths from injury in children and employment status in family: analysis of trends in class specific death rates. *BMJ,* 333(7559), 119.

Elkan, R., Kendrick, D., Hewitt, M., Robinson, J. J., Tolley, K., Blair, M., Dewey, M., Williams, D. and Brummell, K. (2000). The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature. *Health Technol Assess,* 4(13), i-v, 1-339.

Epstein, D. and Sutton, A. (2011). Modelling correlated clinical outcomes in health technology appraisal. *Value in Health,* 14(6), 793-799.

Epstein, D. M., Sculpher, M. J., Manca, A., Michaels, J., Thompson, S. G., Brown, L. C., Powell, J. T., Buxton, M. J. and Greenhalgh, R. M. (2008). Modelling the

long-term cost-effectiveness of endovascular or open repair for abdominal aortic aneurysm. *Br J Surg,* 95(2), 183-190.

Espinoza, M. A., Manca, A., Claxton, K. and Sculpher, M. J. (2011). *The Value of Heterogeneity: A Framework for cost-effectiveness subgroup analysis*. Paper presented at Health Economics Study Group Winter 2011 Meeting, York: University of York.

Farlow, M. R., Small, G. W., Quarg, P. and Krause, A. (2005). Efficacy of rivastigmine in Alzheimer's disease patients with rapid disease progression: results of a meta-analysis. *Dement Geriatr Cogn Disord,* 20(2-3), 192-197.

Feeny, D., Wu, L. and Eng, K. (2004). Comparing short form 6D, standard gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: results from total hip arthroplasty patients. *Quality of Life Research,* 13(10), 1659-1670.

Fenwick, E., Claxton, K. and Sculpher, M. (2001). Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ,* 10(8), 779-787.

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

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London,* Series A(222), 309–368.

Furukawa, T., Barbui, C., Cipriani, A., Brambilla, P. and Watanabe, N. (2006). Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology,* 59(1), 7-10.

Gandjour, A. (2010). Theoretical foundation of patient v. population preferences in calculating QALYs. *Medical Decision Making,* 30(4), E57-63.

Garber, A. M. and Phelps, C. E. (1997). Economic foundations of cost-effectiveness analysis. *J Health Econ,* 16(1), 1-31.

Gelman, A. (2004). *Bayesian data analysis*. Texts in statistical science. 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC.

Gielen, A. C., McDonald, E. M., Wilson, M. E., Hwang, W. T., Serwint, J. R., Andrews, J. S. and Wang, M. C. (2002). Effects of improved access to safety counseling, products, and home visits on parents' safety practices: results of a randomized trial. *Arch Pediatr Adolesc Med,* 156(1), 33-40.

Gielen, A. C., McKenzie, L. B., McDonald, E. M., Shields, W. C., Wang, M. C., Cheng, Y. J., Weaver, N. L. and Walker, A. R. (2007). Using a computer kiosk to promote child safety: results of a randomized, controlled trial in an urban pediatric emergency department. *Pediatrics,* 120(2), 330-339.

Gielen, A. C., Wilson, M. E., McDonald, E. M., Serwint, J. R., Andrews, J. S., Hwang, W. T. and Wang, M. C. (2001). Randomized trial of enhanced anticipatory guidance for injury prevention. *Archives of Pediatrics & Adolescent Medicine,* 155(1), 42-49.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1998). *Markov chain Monte Carlo in practice*. Boca Raton, Fla.: Chapman & Hall.

Ginnelly, L., Claxton, K., Sculpher, M. J. and Golder, S. (2005a). Using value of information analysis to inform publicly funded research priorities. *Appl Health Econ Health Policy,* 4(1), 37-46.

Ginnelly, L., Sculpher, M., Bojke, C., Roberts, I., Wade, A. and Diguiseppi, C. (2005b). Determining the cost effectiveness of a smoke alarm give-away program using data from a randomized controlled trial. *Eur J Public Health,* 15(5), 448-453.

Glick, H. (2007). *Economic evaluation in clinical trials*. Handbooks in health economic evaluation series. Oxford ; New York: Oxford University Press.

Gold, M. R. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Goldstein, H., Yang, M., Omar, R., Turner, R. and Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society Series C-Applied Statistics,* 49, 399-412.

Government, U. (2004a). *The Economic Cost of Fire: Estimates for 2004*. London: Office of the Deputy Prime Minister.

Government, U. (2004b). *Fires in the Home: findings from the 2002/3 British Crime Survey*. London: Office of the Deputy Prime Minister.

Government, U. (2005). *The National Institute for Clinical Excellence (Establishment and Constitution) Amendment Order 2005*. (ed. Government, U.).

Government, U. (2006). *Fires in the Home: findings from the 2004/05 Survey of English Housing*. London: Office of the Deputy Prime Minister.

Government, U. (2007). *Fire Statistics, United Kingdom*. Great Britain, London, 2009: Department for Communities and Local Government: London.

Government, U. (2009). *Fire statistics, United Kingdom*. Great Britain, London, 2010: The Office. Home Office of the Deputy Prime Minister.

Gravelle, H., Brouwer, W., Niessen, L., Postma, M. and Rutten, F. (2007). Discounting in economic evaluations: Stepping forward towards optimal decision rules. *Health Economics,* 16(3), 307-317.

Gravelle, H. and Smith, D. (2001). Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Econ,* 10(7), 587-599.

Haddix, A. C., Mallonee, S., Waxweiler, R. and Douglas, M. R. (2001). Cost effectiveness analysis of a smoke alarm giveaway program in Oklahoma City, Oklahoma. *Inj Prev,* 7(4), 276-281.

Harvey, P. A., Aitken, M., Ryan, G. W., Demeter, L. A., Givens, J., Sundararaman, R. and Goulette, S. (2004). Strategies to increase smoke alarm use in high-risk households. *Journal of Community Health,* 29(5), 375-385.

Hasselblad, V. and McCrory, D. C. (1995). Meta-analytic tools for medical decision making: a practical guide. *Med Decis Making,* 15(1), 81-96.

Hawkins, N. and Scott, D. A. (2010). Cost-effectiveness analysis: discount the placebo at your peril. *Medical Decision Making,* 30(5), 536-543.

Haynes, R., Reading, R. and Gale, S. (2003). Household and neighbourhood risks for injury to 5-14 year old children. *Social Science & Medicine,* 57(4), 625-636.

Hemington-Gorse, S. J., Potokar, T. S., Drew, P. J. and Dickson, W. A. (2009). Burn care costing: the Welsh experience. *Burns : journal of the International Society for Burn Injuries,* 35(3), 378-382.

Hendrickson, S. G. (2005). Reaching an underserved population with a randomly assigned home safety intervention. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention,* 11(5), 313-317.

Henriksson, M., Epstein, D. M., Palmer, S. J., Sculpher, M. J., Clayton, T. C., Pocock, S. J., Henderson, R. A., Buxton, M. J. and Fox, K. A. (2008). The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. *Heart,* 94(6), 717-723.

Higgins, J. P. and Green, S. (2008). Cochrane Handbook for Systematic Reviews of Interventions. [Online]. Available at: www.cochrane-handbook.org.

Higgins, J. P., Thompson, S. G. and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc,* 172(1), 137-159.

Higgins, J. P. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Stat Med,* 15(24), 2733-2749.

Higgins, J. P., Whitehead, A., Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001). Meta-analysis of continuous outcome data from individual patients. *Stat Med,* 20(15), 2219-2241.

Isaman, D. J., Herman, W. H. and Brown, M. B. (2006). A discrete-state discrete-time model using indirect observation. *Statistics in Medicine,* 25(6), 1035-1049.

Isaman, D. J. M., Barhak, J. and Ye, W. (2009). Indirect estimation of a discrete-state discrete-time model using secondary data analysis of regression data. *Statistics in Medicine,* 28(16), 2095-2115.

Jackson, C., Best, N. and Richardson, S. (2006). Improving ecological inference using individual-level data. *Statistics in Medicine,* 25(12), 2136-2159.

Jackson, C., Best, N. and Richardson, S. (2008). Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *Journal of the Royal Statistical Society Series a-Statistics in Society,* 171, 159-178.

Jansen, J. P. (2006). Self-monitoring of glucose in type 2 diabetes mellitus: a Bayesian meta-analysis of direct and indirect comparisons. *Current medical research and opinion,* 22(4), 671-681.

Jeng, G. T., Scott, J. R. and Burmeister, L. F. (1995). A comparison of meta-analytic results using literature vs individual patient data. Paternal cell immunization for recurrent miscarriage. *JAMA,* 274(10), 830-836.

Johnston, B. D., Britt, J., D'Ambrosio, L., Mueller, B. A. and Rivara, F. P. (2000). A preschool program for safety and injury prevention delivered by home visitors. *Injury Prevention,* 6(4), 305-309.

Jones, A. P., Riley, R. D., Williamson, P. R. and Whitehead, A. (2009). Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials,* 6(1), 16-27.

Kaplan, R. M., Ganiats, T. G., Sieber, W. J. and Anderson, J. P. (1998). The Quality of Well-Being Scale: critical similarities and differences with SF-36. *Int J Qual Health Care,* 10(6), 509-520.

Kelly, M. P., McDaid, D., Ludbrook, A. and Powell, J. (2005). *Economic appraisal of public health interventions*. NHS Health Development Agency UK.

Kendrick, D., Coupland, C., Mulvaney, C., Simpson, J., Smith, S. J., Sutton, A., Watson, M. and Woods, A. (2007). Home safety education and provision of safety equipment for injury prevention. *Cochrane Database Syst Rev,* (1).

Kendrick, D., Marsh, P., Fielding, K. and Miller, P. (1999). Preventing injuries in children: cluster randomised controlled trial in primary care. *BMJ,* 318(7189), 980-983.

Kendrick, D., Wynn, P., Pillet, E. and He, Z. (2010). Overview of reviews of home safety interventions for preventing fire-related injuries within the home in childhood. *(not published).*

Kind, P., Hardman, G. and Macran, S. (1999). UK Population Norms for EQ-5D. *Centre for Health Economics Discussion Paper Series* [Online], (Discussion paper 172).

Kind, P., Lafata, J. E., Matuszewski, K. and Raisch, D. (2009). The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health,* 12 Suppl 1, S27-30.

King, W. J., Klassen, T. P., LeBlanc, J., Bernard-Bonnin, A. C., Robitaille, Y., Pham, B., Coyle, D., Tenenbein, M. and Pless, I. B. (2001). The effectiveness of a home visit to prevent childhood injury. *Pediatrics,* 108(2), 382-388.

Kinney, M. R., Burfitt, S. N., Stullenbarger, E., Rees, B. and DeBolt, M. R. (1996). Quality of life in cardiac patient research: a meta-analysis. *Nursing research,* 45(3), 173-180.

Lambert, P. C., Sutton, A. J., Abrams, K. R. and Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol,* 55(1), 86-94.

Lee, P. M. (1997). *Bayesian statistics : an introduction*. 2nd ed. London: Wiley.

Lu, G. and Ades, A. (2009). Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics,* 10(4), 792-805.

Lu, G. and Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine,* 23(20), 3105-3124.

Lu, G., Ades, A. E., Sutton, A. J., Cooper, N. J., Briggs, A. H. and Caldwell, D. M. (2007). Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med,* 26(20), 3681-3699.

Lu, G. B. and Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association,* 101(474), 447-459.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine,* 21(16), 2313-2324.

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(11), 1101-1119.

Marshall, G. and Jones, R. H. (1995). Multi-state models and diabetic retinopathy. *Statistics in Medicine,* 14(18), 1975-1983.

Matthews, J. R. (1988). *An analysis of dangerous behaviour in toddlers*. Kansas, USA: University of Arkansas.

McCabe, C., Claxton, K. and Culyer, A. J. (2008). The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics,* 26(9), 733-744.

McGhan, W. F., Al, M., Doshi, J. A., Kamae, I., Marx, S. E. and Rindress, D. (2009). The ISPOR Good Practices for Quality Improvement of Cost-Effectiveness Research Task Force Report. *Value Health*.

McKenna, C., Burch, J., Suekarran, S., Walker, S., Bakhai, A., Witte, K., Harden, M., Wright, K., Woolacott, N., Lorgelly, P., Fenwick, L. and Palmer, S. (2010). A systematic review and economic evaluation of the clinical effectiveness and cost-effectiveness of aldosterone antagonists for postmyocardial infarction heart failure. *Health Technol Assess,* 14(24), 1-162.

McLernon, D. J., Dillon, J. and Donnan, P. T. (2008). Health-state utilities in liver disease: a systematic review. *Med Decis Making,* 28(4), 582-592.

Mihaylova, B., Briggs, A., Armitage, J., Parish, S., Gray, A. and Collins, R. (2005). Cost-effectiveness of simvastatin in people at different levels of vascular disease risk: economic analysis of a randomised trial in 20,536 individuals. *Lancet,* 365(9473), 1779-1785.

Mihaylova, B., Briggs, A., Armitage, J., Parish, S., Gray, A. and Collins, R. (2006). Lifetime cost effectiveness of simvastatin in a range of risk groups and age groups derived from a randomised trial of 20,536 people. *BMJ,* 333(7579), 1145.

Miller, R. E., Reisinger, K. S., Blatter, M. M. and Wucher, F. (1982). Pediatric counseling and subsequent use of smoke detectors. *American Journal of Public Health,* 72(4), 392-393.

Mock, C., Peden, M., McGee, K. S. and Krug, E. (2003). *A WHO plan for burn prevention and care. Injury: A Leading Cause of the Global Burden of Disease.* Geneva, Switzerland.

Morestin, F., Gauvin, F., Hogue, M. and Benoit, F. (2010). *Method for synthesizing knowledge about public policies*. Quebec, Canada: National Collaborating Centre for Healthy Public Policy - Canada.

NFPA, N. F. P. A. [Online]. Available at: www.nfpa.org [Accessed 01/10/2012].

NICE (2008). *Guide to the Methods of Thechnology Appraisal*. London: National Institute for Health and Clinical Excellence, Institute's Decision Support Unit.

NICE (2009). *Methods for the development of NICE public health guidance*. London: National Institute for Health and Clinical Excellence, UK.

NICE (2010). *NICE public health guidance 30: Preventing unintentional injuries among under-15s in the home*. London, UK: National Institute for Health and Clinical Excellence.

Nicholl, J., Turner, J., Young, T., Freeman, J., Chanakira, E., Dixon, S., Santarelli, M., Ohn, T., Cross, S. and Mason, S. (2009). *The Long Term Health and Healtcare Outcomes of Accidental Injury (The HALO Study).* Medical Care Research Unit and School of Health and Related Research, University of Sheffield and Sheffield Teaching Hospitals Foundation Trust. Dept of Health Policy. Research Programme. Project No 0010017.

Nixon, J., Khan, K. and Kleijnen, J. (2001). Summarising economic evaluations in systematic reviews: A new approach. *British Medical Journal,* 322(7302), 1596-1598.

Nixon, R. M., Bansback, N. and Brennan, A. (2007). Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statistics in Medicine,* 26(6), 1237-1254.

O'Connor, T. G., Davies, L., Dunn, J. and Golding, J. (2000). Distribution of accidents, injuries, and illnesses by family type. ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. *Pediatrics,* 106(5), E68.

O'Hagan, A., Stevens, J. W. and Montmartin, J. (2001). Bayesian cost-effectiveness analysis from clinical trial data. *Stat Med,* 20(5), 733-753.

Oakley, J., Brennan, A., Tappenden, P. and Chilcott, J. (2010). Simulation sample sizes for Monte Carlo partial EVPI calculations. *Journal of Health Economics,* 29(3), 468-477.

ONS (2010a). *Office for National Statistics, Household Projections, 2008 to 2033, England*. Housing Statistical Release. Office for National Statistics - Communities and Local Governement.

ONS (2010b). *Office for National Statistics. Mortality Statistics. Death registered in 2009.*

Papaioannou, D., Brazier, J. E. and Paisley, S. (2011). *NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature.*

Pauly, M. V. and Blavin, F. E. (2008). Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance. *Journal of Health Economics,* 27(6), 1407-1417.

Peasgood, T., Herrmann, K., Kanis, J. A. and Brazier, J. E. (2009). An updated systematic review of Health State Utility Values for osteoporosis related conditions. *Osteoporos Int,* 20(6), 853-868.

Peasgood, T., Ward, S. and Brazier, J. (2010). Health-state utility values in breast cancer. *Expert Review of Pharmacoeconomics and Outcomes Research,* 10(5), 553-566.

Phelan, K. J., Khoury, J., Xu, Y. and Lanphear, B. (2009). Validation of a HOME Injury Survey. *Injury Prevention,* 15(5), 300-306.

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

Philips, Z., Claxton, K. and Palmer, S. (2008). The half-life of truth: What are appropriate time horizons for research decisions? *Medical Decision Making,* 28(3), 287-299.

Piantadosi, S., Byar, D. P. and Green, S. B. (1988). The ecological fallacy. *American Journal of Epidemiology,* 127(5), 893-904.

Pignone, M., Saha, S., Hoerger, T., Lohr, K. N., Teutsch, S. and Mandelblatt, J. (2005). Challenges in systematic reviews of economic analyses. *Ann Intern Med,* 142(12 Pt 2), 1073-1079.

Pitt, M., Anderson, R. and Moxham, T. (2009). *Preventing unintentional injuries among under - 15s in the home. Report 3: cost-effectiveness modelling of home based interventions aimed at reducing unintentional injuries in children*. Exeter & Plymouth: Peninsula Technology Assessment Group (PenTAG), Peninsula Medical School, Universities of Exeter and Plymouth.

Pocock, S., Assmann, S., Enos, L. and Kasten, L. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. *Statistics in Medicine,* 21(19), 2917-2930.

Post, P. N., Stiggelbout, A. M. and Wakker, P. P. (2001). The utility of health states after stroke: a systematic review of the literature. *Stroke,* 32(6), 1425-1429.

Prevost, T. C., Abrams, K. R. and Jones, D. R. (2000). Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Stat Med,* 19(24), 3359-3376.

Raftery, J. (2009). Should NICE's threshold range for cost per QALY be raised? No. *BMJ,* 338, b185.

Reading, R., Langford, I. H., Haynes, R. and Lovett, A. (1999). Accidents to preschool children: comparing family and neighbourhood risk factors. *Social Science & Medicine,* 48(3), 321-330.

Reimers, A. and Laflamme, L. (2005). Neighbourhood social and socio-economic composition and injury risks. *Acta Paediatrica,* 94(10), 1488-1494.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M. and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol,* 58(10), 982-990.

Riley, R. D., Dodd, S. R., Craig, J. V., Thompson, J. R. and Williamson, P. R. (2008). Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Statistics in Medicine,* 27(29), 6111-6136.

Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. and Boutitie, F. (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med,* 27(11), 1870-1893.

Riley, R. D., Simmonds, M. C. and Look, M. P. (2007). Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *J Clin Epidemiol,* 60(5), 431-439.

Riley, R. D. and Steyerberg, E. W. (2010). Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods,* 1(1), 2-19.

Rodgers, M., Epstein, D., Bojke, L., Yang, H., Craig, D., Fonseca, T., Myers, L., Bruce, I., Chalmers, R., Bujkiewicz, S., Lai, M., Cooper, N., Abrams, K., Spiegelhalter, D., Sutton, A., Sculpher, M. and Woolacott, N. (2011). Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health technology assessment (Winchester, England),* 15(10), i-xxi, 1-329.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika,* 63(3), 581-590.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. Wiley classics library. Hoboken, N.J. ;: Wiley-Interscience.

Rychetnik, L., Frommer, M., Hawe, P. and Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology & Community Health,* 56(2), 119-127.

Sackett, D., Rosenberg, W., Gray, J., Haynes, R. and Richardson, W. (1996). Evidence based medicine: What it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. *British Medical Journal,* 312(7023), 71-72.

Salanti, G., Dias, S., Welton, N. J., Ades, A. E., Golfinopoulos, V., Kyrgiou, M., Mauri, D. and Ioannidis, J. P. (2010). Evaluating novel agent effects in multiple-treatments meta-regression. *Statistics in Medicine,* 29(23), 2369-2383.

Salanti, G., Higgins, J. P. T., Ades, A. E. and Ioannidis, J. P. A. (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research,* 17(3), 279-301.

Salanti, G., Marinho, V. and Higgins, J. P. (2009). A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol,* 62(8), 857-864.

Sanchez, J. L., Bastida, J. L., Martinez, M. M., Moreno, J. M. and Chamorro, J. J. (2008). Socio-economic cost and health-related quality of life of burn victims in Spain. *Burns,* 34(7), 975-981.

Sangvai, S., Cipriani, L., Colborn, D. K. and Wald, E. R. (2007). Studying injury prevention: practices, problems, and pitfalls in implementation. *Clinical Pediatrics,* 46(3), 228-235.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall / CRC Press.

Scholer, S. J., Hickson, G. B. and Ray, W. A. (1999). Sociodemographic factors identify US infants at high risk of injury mortality. *Pediatrics,* 103(6 Pt 1), 1183-1188.

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 Econ,* 7(7), 629-638.

Schwarz, D. F., Grisso, J. A., Miles, C., Holmes, J. H. and Sutton, R. L. (1993). An injury prevention program in an urban African-American community. *American Journal of Public Health,* 83(5), 675-680.

Schwarz, G. (1978). Estimating Dimension of a Model. *Annals of Statistics,* 6(2), 461-464.

Sculpher, M. (2008). Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics,* 26(9), 799-806.

Sculpher, M. and Claxton, K. (2005). Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty--when is there sufficient evidence? *Value Health,* 8(4), 433-446.

Sculpher, M. J., Claxton, K., Drummond, M. and McCabe, C. (2006). Whither trial-based economic evaluation for health care decision making? *Health Econ,* 15(7), 677-687.

Sculpher, M. J. and Drummond, M. F. (2006). Analysis sans frontieres: can we ever make economic evaluations generalisable across jurisdictions? *Pharmacoeconomics,* 24(11), 1087-1099.

Sculpher, M. J., Pang, F. S., Manca, A., Drummond, M. F., Golder, S., Urdahl, H., Davies, L. M. and Eastwood, A. (2004). Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess,* 8(49), iii-iv, 1-192.

Simmonds, M. C. and Higgins, J. P. (2007). Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med,* 26(15), 2982-2999.

Simmonds, M. C., Higgins, J. P., Stewart, L. A., Tierney, J. F., Clarke, M. J. and Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials,* 2(3), 209-217.

Smith, C. T., Williamson, P. R. and Marson, A. G. (2005). An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. *Journal of Evaluation in Clinical Practice,* 11(5), 468-478.

Smith, T. C., Spiegelhalter, D. J. and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine,* 14(24), 2685-2699.

Song, F., Loke, Y. K., Walsh, T., Glenny, A. M., Eastwood, A. J. and Altman, D. G. (2009). Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ,* 338, b1147.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS User Manual, version 1.4, January 2003. In: Unit, M. B. ed. [Online].

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. R. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology,* 64, 583-616.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R. and Abrams, K. R. (1999). Methods in health service research. An introduction to bayesian methods in health technology assessment. *BMJ (Clinical research ed),* 319(7208), 508-512.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R. and Abrams, K. R. (2000). Bayesian methods in health technology assessment: a review. *Health technology assessment (Winchester, England),* 4(38), 1-130.

Statistics, O. N. (2001). *UK Census 2001*. [Online]. Available at: http://www.statistics.gov.uk/census2001/census2001.asp.

Stewart, L. A. and Clarke, M. J. (1995). Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat Med,* 14(19), 2057-2079.

Stewart, L. A. and Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet,* 341(8842), 418-422.

Stewart, L. A. and Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof,* 25(1), 76-97.

Stinnett, A. A. and Mullahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making,* 18(2 Suppl), S68-80.

Stinnett, A. A. and Paltiel, A. D. (1997). Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. *Med Decis Making,* 17(4), 483-489.

Sturza, J. (2010). A review and meta-analysis of utility values for lung cancer. *Medical Decision Making,* 30(6), 685-693.

Su, X. and Po, A. (1996). Combining event rates from clinical trials: Comparison of Bayesian and classical methods. *Annals of Pharmacotherapy,* 30(5), 460-465.

Sullivan, P. W. and Ghushchyan, V. (2006). Preference-Based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making,* 26(4), 410-420.

Sullivan, P. W., Sculpher, M. J., Ghushchyan, V. H. and Slejko, J. F. (2009). Catalogue of Eq-5d Scores for the Uk. *Value in Health,* 12(7), A398-A398.

Sutton, A., Ades, A. E., Cooper, N. and Abrams, K. (2008). Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics,* 26(9), 753-767.

Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res,* 10(4), 277-303.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. and Song, F. (2000a). *Methods for meta-analysis in medical research*. Wiley Series in Probability and Statistics. First edition ed. London.

Sutton, A. J., Cooper, N. J. and Jones, D. R. (2009). Evidence synthesis as the key to more coherent and efficient research. *BMC medical research methodology,* 9, 29.

Sutton, A. J. and Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine,* 27(5), 625-650.

Sutton, A. J., Kendrick, D. and Coupland, C. A. (2008). Meta-analysis of individual- and aggregate-level data. *Stat Med,* 27(5), 651-669.

Sutton, A. J., Song, F., Gilbody, S. M. and Abrams, K. R. (2000b). Modelling publication bias in meta-analysis: a review. *Stat Methods Med Res,* 9(5), 421-445.

Sznajder, M., Leduc, S., Janvrin, M. P., Bonnin, M. H., Aegerter, P., Baudier, F. and Chevallier, B. (2003). Home delivery of an injury prevention kit for children in four French cities: a controlled randomized trial. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention,* 9(3), 261-265; discussion 265.

Tengs, T. O. and Lin, T. H. (2002). A meta-analysis of utility estimates for HIV/AIDS. *Med Decis Making,* 22(6), 475-481.

Tengs, T. O. and Lin, T. H. (2003). A meta-analysis of quality-of-life estimates for stroke. *Pharmacoeconomics,* 21(3), 191-200.

Tengs, T. O. and Wallace, A. (2000). One thousand health-related quality-of-life estimates. *Med Care,* 38(6), 583-637.

Thompson, S. G. and Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine,* 21(11), 1559-1573.

Towse, A. (2009). Should NICE's threshold range for cost per QALY be raised? Yes. *BMJ,* 338, b181.

Tudur, C., Williamson, P. R., Khan, S. and Best, L. Y. (2001). The value of the aggregate data approach in meta-analysis with time-to-event outcomes. *Journal of the Royal Statistical Society Series a-Statistics in Society,* 164, 357-370.

Tumeh, J. W., Moore, S. G., Shapiro, R. and Flowers, C. R. (2005). Practical approach for using Medicare data to estimate costs for cost-effectiveness analysis. *Expert Rev Pharmacoecon Outcomes Res,* 5(2), 153-162.

Turner, D., Wailoo, A., Nicholson, K., Cooper, N., Sutton, A. and Abrams, K. (2003). Systematic review and economic decision modelling for the prevention and treatment of influenza A and B. *Health Technol Assess,* 7(35), iii-iv, xi-xiii, 1-170.

Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. and Thompson, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med,* 19(24), 3417-3432.

Turner, R. M., Spiegelhalter, D. J., Smith, G. C. and Thompson, S. G. (2009). Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc,* 172(1), 21-47.

Unicef (2001). *A league table of child deaths by injury in rich nations.* Florence: Innocenti report card. Innocenti ResearchCentre.

Urdahl, H., Manca, A. and Sculpher, M. J. (2006). Assessing generalisability in model-based economic evaluation studies: a structured review in osteoporosis. *Pharmacoeconomics,* 24(12), 1181-1197.

Wailoo, A. J., Bansback, N., Brennan, A., Michaud, K., Nixon, R. M. and Wolfe, F. (2008). Biologic drugs for rheumatoid arthritis in the Medicare program: a cost-effectiveness analysis. *Arthritis Rheum,* 58(4), 939-946.

Wakefield, J. (2008). Ecologic studies revisited. *Annual Review of Public Health,* 29, 75-90.

Watson, M., Kendrick, D., Coupland, C., Woods, A., Futers, D. and Robinson, J. (2005). Providing child safety equipment to prevent injuries: randomised controlled trial. *BMJ (Clinical research ed),* 330(7484), 178.

Weatherly, H., Drummond, M., Claxton, K., Cookson, R., Ferguson, B., Godfrey, C., Rice, N., Sculpher, M. and Sowden, A. (2009). Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. *Health Policy,* 93(2-3), 85-92.

Wee, H. L., Machin, D., Loke, W. C., Li, S. C., Cheung, Y. B., Luo, N., Feeny, D., Fong, K. Y. and Thumboo, J. (2007). Assessing differences in utility scores: a comparison of four widely used preference-based instruments. *Value Health,* 10(4), 256-265.

Weinstein, M. C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C. and Luce, B. R. (2003). Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value Health,* 6(1), 9-17.

Weinstein, M. C., Torrance, G. and McGuire, A. (2009). QALYs: the basics. *Value Health,* 12 Suppl 1, S5-9.

Welte, R., Feenstra, T., Jager, H. and Leidl, R. (2004). A decision chart for assessing and improving the transferability of economic evaluation results between countries. *Pharmacoeconomics,* 22(13), 857-876.

Welton, N. J. and Ades, A. E. (2005). Estimation of markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making,* 25(6), 633-645.

Welton, N. J., Ades, A. E., Carlin, J. B., Altman, D. G. and Sterne, J. A. C. (2009). Models for potentially biased evidence in meta-analysis using empirically based priors. *Journal of the Royal Statistical Society Series a-Statistics in Society,* 172, 119-136.

Welton, N. J., Cooper, N. J., Ades, A. E., Lu, G. and Sutton, A. J. (2008). Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Stat Med,* 27(27), 5620-5639.

Whitehead, A. (2002). *Meta-analysis of Controlled Clinical Trials*. Statistics in Practice. Chichester: John Wiley & Sons.

Whitehead, A., Omar, R. Z., Higgins, J. P., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001). Meta-analysis of ordinal outcomes using individual patient data. *Stat Med,* 20(15), 2243-2260.

Wiebe, N., Vandermeer, B., Platt, R., Klassen, T., Moher, D. and Barrowman, N. (2006). A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of Clinical Epidemiology,* 59(4), 342-353.

Williams, A. (1985). Economics of coronary artery bypass grafting. *British medical journal (Clinical research ed),* 291(6491), 326-329.

Woolacott, N., Hawkins, N., Mason, A., Kainth, A., Khadjesari, Z., Vergel, Y. B., Misso, K., Light, K., Chalmers, R., Sculpher, M. and Riemsma, R. (2006). Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health Technol Assess,* 10(46), 1-233, i-iv.

Wu, O., Robertson, L., Twaddle, S., Lowe, G. D., Clark, P., Greaves, M., Walker, I. D., Langhorne, P., Brenkel, I., Regan, L. and Greer, I. (2006). Screening for thrombophilia in high-risk situations: systematic review and cost-effectiveness analysis. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) study. *Health Technol Assess,* 10(11), 1-110.

Wynn, P., Ilyas, N., Kendrick, D., Sutton, A., Coupland, C., Mulvaney, C. and Watson, M. (2010). The Effect of Education and Safety Equipment on Burn and Scald Prevention Practices and on Childhood Thermal Injuries: Update of a Systematic Review and Meta-Analysis. *Injury Prevention,* 16, A242-A242.

Ye, W., Isaman, D. J. M. and Barhak, J. (2012). Use of secondary data to estimate instantaneous model parameters of diabetic heart disease: Lemonade Method. *Information Fusion,* 13(2), 137-145.