EVIDENCE SYNTHESIS AND DECISION MODELLING IN PUBLIC HEALTH

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Felix Adjuah Achana BSc MSc
Department of Health Sciences
University of Leicester

September 2014

EVIDENCE SYNTHESIS AND DECISION MODELLING IN PUBLIC HEALTH

Felix Adjuah Achana BSc MSc

Abstract

This thesis focuses on the challenges of evidence synthesis to inform healthcare decision making within public health. It encompasses both methodological advancement and practical application of existing synthesis methodology, using as an example - accidents prevention in children to illustrate application of the methods within a public health context.

The thesis commences with a systematic review of NICE public health appraisals to identify the barriers to quantitative synthesis of evidence in public health. Then focusing on the prevention of unintentional poisonings in pre-school children, a series of network meta-analyses of the effectiveness evidence are conducted, demonstrating how complex synthesis methodology can be employed to help overcome some challenges of evidence synthesis in a identified in the review of the NICE public health appraisals.

New synthesis methodology is then developed in which the standard network meta-analysis model is first extended to include a covariate for the baseline risk and then to a multiple outcome settings. Baseline risk is a proxy for unmeasured but important patient-level characteristics, which may be modifiers of the treatment effect in a meta-analysis. Thus adjusting for it can account for heterogeneity across different study populations and identify those more likely to benefit from the intervention. The multiple outcome models account for the dependency structure within the data which is important in a decision modelling context, as correlations between effect estimates on different outcomes may have implications for estimating the net benefit associated with treatment.

Finally, a substantive decision analytic model is presented incorporating results from the network meta-analysis and application of the methodology developed to the poison prevention data. The analyses suggest that compared to usual care, more intensive home safety interventions are more effective in preventing medicinal poisonings in pre-school children but are unlikely to be cost-effective for the UK NHS unless policy makers are willing to pay upwards of £75,000 for every QALY gained.

ACKNOWLEDGEMENTS

I would like to thank my primary supervisor Professor Nicola Cooper for offering me the opportunity to undertake this doctoral work and guiding me to completion. I would also like to thank my other supervisors Professors David Jones and Alex Sutton for devoting a substantial amount of their time to helping and guiding me over the past four years. Thanks to Professor Denise Kendrick and the Keeping Children Safe at Home Programme team for providing a rich source of data for the analysis and useful advice and guidance at various stages of the project. I would also like to say a BIG thank you to all members of the Biostatistics group in the Department of Health Sciences and in particular, Dr Sylwia Bujkiewicz, Professor Nuala Sheehan and Mrs Stephanie Hubbard who provided advice and support at various stages during my time in Leicester. Thanks also go to all co-authors and reviewers (anonymous or otherwise) of the papers published as a result of the work of this PhD. My final thanks go to my colleagues and fellow PhD students for the light hearted 'Banta' and the social activities.

TABLE OF CONTENTS

LIST OF FIGURES ..... **v**
LIST OF TABLES ..... **vii**
ABBREVIATIONS ..... **viii**
1. INTRODUCTION **1**
1.1 Thesis aims 1
1.2 Background 2
1.3 Example problem: Accidents prevention in children 4
1.4 Structure of thesis 5
2. REVIEW OF EVIDENCE SYNTHESIS METHODS **8**
2.1 Chapter overview 8
2.2 Pairwise meta-analysis 8
2.2.1 Fixed and random effects models 8
2.2.2 Heterogeneity 11
2.2.3 Meta-regression 11
2.2.4 Publication bias 12
2.3 Network meta-analysis 14
2.3.1 Model for binary outcome data 14
2.3.2 Assessing which intervention is the best 16
2.3.3 Assessing evidence consistency 17
2.3.4 Assessing publication bias in network meta-analysis 18
2.3.5 Convergence diagnostics 18
2.3.6 Goodness fit and model selection 19
2.4 Chapter summary 19
3. REVIEW OF DECISION ANALYTIC MODELLING METHODS **21**
3.1 Chapter overview 21
3.2 Economic evaluation in healthcare 21
3.2.1 Introduction 21
3.2.2 Measuring health outcomes 22
3.2.3 Measuring resource use and unit costs 23

3.2.4 Discounting costs and utilities 24
3.3 Decision analytic modelling methods 24
3.3.1 Cohort versus individual patient-simulation 24
3.3.2 Decision tree models 25
3.3.3 Markov Models 26
3.3.4 Comprehensive decision analysis 28
3.3.5 Incorporating uncertainty 28
3.4 Presenting results of economic evaluations 29
3.4.1 Cost-effectiveness plane 29
3.4.2 Cost-effectiveness acceptability curve 31
3.5 Chapter summary 32
4. SYSTEMATIC REVIEW OF SYNTHESIS METHODS IN PUBLIC HEALTH EVALUATION OF INTERVENTIONS 33
4.1 Chapter overview 33
4.2 Introduction 33
4.3 Systematic review methods 34
4.4 Systematic review results 36
4.4.1 Type of review 36
4.4.2 Included studies – RCTs versus non-randomised studies 37
4.4.3 Quantitative evidence synthesis 40
4.4.4 How evidence from systematic reviews were incorporated into the model 43
4.4.5 Decision analytic modelling methods 43
4.5 Discussion of systematic review findings 47
4.6 Chapter summary 48
5. EFFECTIVENESS OF INTERVENTIONS TO INCREASE UPTAKE OF POISONING PREVENTION MEASURES IN HOUSEHOLDS WITH CHILDREN: NETWORK META-ANALYSES 50
5.1 Chapter overview 50
5.2 Effectiveness evidence 50
5.2.1 Data sources 50
5.2.2 Poisoning prevention outcomes 51
5.2.3 Description of included studies 51
5.2.4 Classification of interventions 56
5.3 Methods 59

5.3.1 Binary outcome data 59
5.3.2 Standard pairwise meta-analysis 59
5.3.3 Network meta-analysis 60
5.3.4 Convergence, goodness- of-fit and model selection 60
5.3.5 Assessment of heterogeneity and evidence consistency 61
5.3.6 Sensitivity analysis 61
5.4 Results 63
5.4.1 Storage of medicines 65
5.4.2 Storage of other household products 66
5.4.3 Safe storage of poisons 69
5.4.4 Possession of a PCC number 70
5.4.5 Safe storage of poisonous plants 71
5.4.6 Assessment of best intervention 72
5.5 Convergence, heterogeneity and consistency assessment results 74
5.5.1 Convergence diagnostics 74
5.5.2 Heterogeneity results 78
5.5.3 Evidence consistency assessment results 79
5.6 Sensitivity analyses results 83
5.6.1 Continuity corrections 83
5.6.2 Sensitivity analysis to prior distributions 84
5.7 Discussion 88
5.7.1 Summary of findings 88
5.7.2 Strengths and limitations 88
5.7.3 Comparisons with existing work 89
5.8 Chapter summary 91
6. ADJUSTING FOR A BASELINE RISK COVARIATES IN NETWORK META-ANALYSIS 92
6.1 Chapter overview 92
6.2 Introduction 92
6.3 Review of baseline risk models for pair-wise meta-analysis 94
6.4 Methods 97
6.4.1 Model with no covariate adjustment 97
6.4.2 Extending the NMA to include a covariate for the baseline risk 98
6.4.3 Models for the baseline risk and treatment by covariate interactions 100

6.4.4 Goodness of fit and model selection 101
6.5 Application examples 101
6.5.1 Functional smoke alarm example 101
6.5.2 Pain relief example 104
6.5.3 Safe storage of medicines example 106
6.6 Results 107
6.6.1 Functional smoke alarm example 107
6.6.2 Pain relief example 108
6.6.3 Safe storage of medicines example 112
6.7 Discussion 112
6.8 Chapter Summary 115
7. MODELLING EFFECTIVENESS ACROSS MULTIPLE OUTCOME MEASURES 116
7.1 Chapter overview 116
7.2 Introduction 116
7.3 Data 119
7.4 Methods 124
7.4.1 Model 1: Univariate NMA 124
7.4.2 Model 2: Multivariate NMA 125
7.4.3 Model 3: Borrowing information outcomes 129
7.4.4 Some limitations when fitting model 3 131
7.5 Implementation 132
7.5.1 Estimating within-study correlations from IPD 132
7.5.2 Formulating prior distributions for the within-study correlations 134
7.5.3 Assessing inconsistency in NMA with multiple outcomes 134
7.5.4 Goodness-of-fit and model selection 135
7.5.5 Model evaluation 135
7.6 Results 136
7.6.1 Univariate and multivariate analyses 136
7.6.2 Model 3: Borrowing strength across outcomes 138
7.6.3 Sensitivity analysis 140
7.7 Discussion 142
7.8 Chapter summary 146
8. MEDICINAL POISONS DECISION MODEL 147

8.1 Chapter overview 147
8.2 Base case analysis 147
8.3 Model structure 149
8.3.1 Intervention model 151
8.3.2 Stage 2: Preschool model 151
8.3.3 Stage 3: Long-term model 152
8.4 Evidence for base case analysis 153
8.4.1 Intervention strategies 155
8.4.2 Baseline prevalence of safety practices 156
8.4.3 Probability of accidental ingestion 157
8.4.4 Probability of inpatient admission following a poisoning injury 158
8.4.5 Probability of moderate, severe and fatal poisoning injury 159
8.5 Costs 159
8.5.1 Intervention costs 160
8.5.2 Healthcare costs of treatment 161
8.6 Utilities 164
8.7 Model evaluation 166
8.7.1 Outcomes of economic evaluation 166
8.7.2 Implementing the analysis 172
8.7.3 Incorporating uncertainty 173
8.8 Results 174
8.8.1 Base case analysis 174
8.8.2 Sensitivity analyses I results 179
8.8.3 Sensitivity analysis II: Baseline risk adjusted 185
8.8.4 Sensitivity analysis III: Using effectiveness estimates from the multivariate analysis 188
8.9 Discussion 191
8.9.1 Summary of findings 191
8.9.2 Strengths and limitations 193
8.9.3 Comparisons with previous work 194
8.10 Chapter summary 194
9. DISCUSSION 196
9.1 Chapter summary 196
9.2 Thesis summary 196

9.3 Strengths and limitations 199
9.4 Future work 202
9.5 Conclusion 204
REFERENCE 206
APPENDICES 215
Appendix I: Extra table of results for Chapter 4 215
Appendix II: Between-study covariance for multi-arm studies reporting multiple outcomes 218
Appendix III: WinBUGS Code for baseline risk adjustment models in Chapter 6 226
Appendix IV: WinBUGS Code for the multivariate NMAs in Chapter 7 228
Appendix V: WinBUGS Code to fit decision analytic model 236
APPENDIX VII: RESEARCH PAPERS 243

LIST OF FIGURES

Figure 2.1: Example of a funnel plot 13
Figure 2.2: An example of a connected network from the example dataset 15
Figure 3.1: Simplified diagram of a decision tree model 25
Figure 3.2: A simplified example of Markov model 27
Figure 3.3: Cost-effectiveness plane (adapted from Welton *et al* ., (2012)) 30
Figure 3.4: An example of CEAC for 2 interventions (20K = £20,000) 31
Figure 4.1: Review of Decision analytic modelling methods in public health 44
Figure 5.1: Network diagrams for poison prevention outcomes 58
Figure 5.2: Posterior density plots for the pooled effect of intervention k relative to usual care, d_k and the between-study standard deviation (sd), σ on the log-odds ratio scale 75
Figure 5.3: History plots for the pooled effect of intervention k relative to usual care, d_k on the and the between-study standard deviation (sd), σ on the log-odds ratio scale 76
Figure 5.4: Autocorrelation plots for the pooled effect of intervention k relative to usual care, d_k and the between-study standard deviation (sd), σ on the log-odds ratio scale 77
Figure 5.5: Autocorrelation plots for the pooled effect of intervention k relative to usual care, d_k and the between-study standard deviation (sd), σ on the log-odds ratio scale 77

Figure 5.6: Diagnostic plots showing evidence of non-convergence in d_7, the effect estimate for low cost/free equipment versus usual care from the random baseline model for safe storage of other household products 78
Figure 5.7: Safe storage of medicines (random effects model) 82
Figure 5.8: Estimated odds ratios (ORs) for home safety interventions compared with usual care 85
Figure 5.9: Estimated odds ratios (ORs) for home safety interventions compared with usual care for safe storage of poisons from sensitivity analysis using different estimates from Reich et al. (2011) 86
Figure 5.10: Estimated odds ratios (ORs) for home safety interventions compared with usual care for safe storage of medicines 87
Figure 6.1: Intervention network for possession of functional smoke alarm, p-operative pain relief data and safe storage of medicines 103
Figure 6.2: NMA adjusting for baseline morphine use assuming independent slopes for different treatment effects and unconstrained baseline risk 109
Figure 7.1: Intervention networks for the poisoning prevention outcome. 123
Figure 7.2: Summary forest plot of intervention effects relative to usual. Outcomes are safe storage of medicines, safe storage of other household chemicals and possession of a PCC telephone number 137
Figure 8.1: Decision model structure. Arrow heads indicate direction of movement of households/individuals through the model 150
Figure 8.2: Decision tree diagram of the intervention model 167
Figure 8.3: Movement of individuals between health-states in preschool model 169
Figure 8.4: Cost-effectiveness of home safety interventions to reduce unintentional poisoning in children under 5years old from the base case analysis 176
Figure 8.5: Cost-effectiveness acceptability curves for the base case analysis 178
Figure 8.6: Cost-effectiveness acceptability curves using estimates from the multiple outcomes NMA (Model 3) for safes storage of medicines described in Chapter 7. 191

LIST OF TABLES

Table 4.1: NICE PH appraisals and summary of evidence synthesis methods and decision modelling used to inform their development 38
Table 4.2: Review of quantitative methods used to synthesise public health evidence for NICE public health appraisal 41
Table 4.3: Review of decision analytic models evaluating the cost-effectiveness of interventions for NICE public health appraisals 45
Table 5.1: Summary of studies evaluating poison prevention measures in households with children 53
Table 5.2: Model fit statistics for the network meta-analysis models 64
Table 5.3: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of medicines 66
Table 5.4: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of other household products 68
Table 5.5: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of poisons 70
Table 5.6: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for possession of PCC number 71
Table 5.7: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of poisonous plants. 72
Table 5.8: Assessment of the best intervention 73
Table 5.9: Heterogeneity statistics from NMA models (log odds ratio scale) 79
Table 5.10: Evidence consistency checks for Safe storage of medicines and other household products 81
Table 6.1: Summary of methods for modelling the relationship between treatment effect and baseline risk in pair-wise meta-analysis with a binary outcome 96
Table 6.2: NMA with baseline risk adjustment applied to functional smoke alarm data 110
Table 6.3: NMA with baseline risk adjustment applied to pain relief data 111
Table 6.4: NMA with baseline risk adjustment applied to safe storage of medicines data 112
Table 7.1: Subset of the poison prevention data displayed in Table 5.1 121
Table 7.2: Estimates of within-study correlations between pairs of log-odds ratios obtained from studies with IPD 133
Table 7.3: Posterior median and 95% credible intervals of the between-study standard deviation and correlation parameters 139
Table 8.1: Base case analysis for safe storage of medicines 148
Table 8.2: Probabilities used in the base case (medicinal poisons) decision analysis 154
Table 8.3: Intervention costs used in base case model 161
Table 8.4: Health sector costs (2012 prices) 162
Table 8.5: Estimates of the NHS cost of treating poisoning injury by severity category 164
Table 8.6: Utilities used in base case analysis 165
Table 8.7: Base case analysis results 175
Table 8.8: Probability that an intervention is the most cost-effective at a willingness-to-pay ratio of £30,000 per QALY from base-case and a number of sensitivity analyses. 180

Table 8.9: Results of sensitivity analysis from the decision analysis carried out in Chapter 8 ... 182
Table 8.10: Sensitivity analysis results incorporating adjusted estimates based on the methods presented in Chapter 6. ... 187
Table 8.11: Sensitivity analysis using effectiveness estimates on safe storage of medicines from multivariate NMAs (model 3) presented in Chapter 7 ... 190

ABBREVIATIONS

CEAC	Cost-effectiveness acceptability curve
CI	Confidence interval
COX-2	Cyclooxegenase-2 inhibitor
CrI	Credible interval
DIC	Deviance Information Criterion
Dres	Deviance residuals
DRG	Diagnosis Related Groups
DALY	Disability Adjusted Life Year
FSA	Functional Smoke Alarm
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio
INB	Incremental Net Benefit
IOM	Institute of Medicine
IPD	Individual Patient Data
KCS	Keeping Children Safe
LY	Life Year
MCMC	Markov Chain Monte Carlo
NHS	National Health Service
NHS/PSS	National Health Service/Personal Social Services
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network Meta-Analysis
NSAIDS	Non-steroidal anti-inflammatory drugs
OR	Odds ratio
PH	Public Health
PSSRU	Personal Social Services Research Unit
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
RE	Random effects
SD	Standard Deviation
UK	United Kingdom
USA	United States of America
WHO	World Health Organisation

1. INTRODUCTION

1.1 Thesis aims

When summarising evidence to inform economic evaluations, it is essential that a structured systematic and coherent approach to synthesis of the evidence is followed, to minimise various forms of biases, including biases associated with arbitrary decisions about which piece(s) of evidence to include in the analysis and to ensure overall transparency of the process. Meta-analysis of well conducted trials (ideally from a systematic review of the evidence), has traditionally been considered as the highest form of evidence, yet use of these methods in public health (PH) evaluation of interventions appear to be limited. This thesis aims to:

a) Review how evidence is currently synthesised (i.e. determine what is currently being done or not done) when summarising evidence to inform economic evaluation within PH,
b) Identify the barriers to quantitative synthesis of evidence in PH evaluations,
c) Demonstrate how complex synthesis methodology (including methods that current exists as well as new methodology developed specifically in this thesis) can be applied in PH context to overcome the barriers identified in (b) and facilitate a more realistic modelling of the data in order to answer the relevant policy questions, and
d) Demonstrate how the methods identified and described in (c) can be integrated with other evidence in a decision analytic modelling framework assessing the cost-effectiveness of interventions in public health.

In the remainder of this chapter, the challenges of evidence synthesis in PH evaluation of interventions and the example problem will be introduced, followed by a chapter by chapter outline of the thesis. The introduction to the challenges of evidence synthesis within PH and the example problem presented in the sections below are based on two recent papers - the first paper (Achana *et al.*, 2014b) has already published in the *Journal of Clinical Epidemiology* whilst the second paper (Achana et al. 2014d) has been submitted to the journal *PLOS ONE* and is currently going through the peer review process.

1.2 Background

Systematic reviews and economic evaluations conducted within a decision modelling framework are two important tools in healthcare evaluation (Novielli *et al.*, 2010; Cooper *et al.*, 2011b). Systematic reviews with or without meta-analyses have been accepted as providing a transparent and consistent way of obtaining research evidence about the clinical and cost-effectiveness interventions in a way that minimizes bias (Higgins and Green, 2011). Decision analytic models offer additional framework through which effectiveness evidence, ideally from a systematic review, may be integrated with other relevant evidence and information on resource utilisation in order to derive comparative estimates of cost-effectiveness. By providing a framework for assessing clinical and cost-effectiveness, these methods enable policy relevant questions, such as which interventions represent the best use of healthcare resources, to be answered (Drummond *et al.*, 2005). For example, the National Institute for Health and Care Excellence (NICE, 1999) produces guidance and recommendations through its technology appraisal programme (www.nice.org.uk/ta) about the clinical and cost-effectiveness of new and existing treatments for use within the National Health Service (NHS) in England. Similarly, the Health Technology Assessment (HTA) programme of the UK National Institute for Health Research (NIHR) commissions and disseminates research information (available online from www.hta.ac.uk) about effectiveness, costs-effectiveness and broader impact of healthcare interventions to help in healthcare planning and delivery throughout the NHS.

A key component of the systematic review is how the evidence, on outcomes such as effectiveness and adverse events, is synthesised. Meta-analysis when used in a systematic review to combine quantitative information from multiple well-conducted randomised controlled trials (RCTs) is considered at the top of the hierarchy of evidence for intervention effectiveness (Sutton *et al.*, 2009). An alternative approach to evidence synthesis, when meta-analysis is considered inappropriate (for example, a significant degree of heterogeneity in the patient populations, study design, methods and other characteristics of the included studies may render meta-analysis inappropriate), is narrative synthesis (also referred to as qualitative synthesis (Rodgers *et al.*, 2009; Eden *et al.*, 2012). In this approach, individual studies identified in the systematic review are summarised using a variety of formats without combining the results of the systematic review quantitatively.

Meta-analysis is widely applied in reviews of clinical effectiveness evidence including, treatments and medical device technologies where the interventions and health outcomes are usually well defined and evaluated in well conducted RCTs (Weatherly *et al.*, 2009). In other fields of healthcare evaluation, however, things may not always be as clear cut. A good example is PH, where interventions are often more complex and less well defined than clinical interventions (Armstrong *et al.*, 2008). There may also be a lack of good quality evidence, particularly from RCTs in PH, for a number of well documented reasons (Petticrew and Egan, 2006; Rosen *et al.*, 2006) including: (i) limited generalisability of the findings of RCTs to the wider population due to highly selected study populations, (ii) a narrow definition of intervention strategies and outcomes, and (iii) a focus on the individual instead of the community that is of interest in PH. Even when feasible, many have argued that RCTs may not always be possible to conduct in PH for other reasons; for example, ethical concerns may be raised regarding not offering the control population a possibly beneficial intervention (Rosen *et al.*, 2006). Also, many of the RCTs conducted in PH tend to be cluster randomised trials and hence have more complex designs that need adjusting for in the meta-analysis. In addition, the best available PH evidence may often come from observational non-randomised studies (Armstrong *et al.*, 2008), despite the increased risk of bias associated with the lack of randomisation. For these reasons, the use of quantitative evidence synthesis methods, such as meta-analysis in PH raises a number of methodological challenges. These include:

i) Increased methodological heterogeneity and risk of bias from including studies with different designs (RCTs, cluster-RCTs, controlled before-and-after studies and other observational non-randomised studies).
ii) The interventions or 'programme' being evaluated is often described in little detail and less clearly defined compared to, for example, pharmaceutical treatments and medical device technologies.
iii) A wide range of outcomes measures (including intermediate and/or surrogate outcomes) are often used and defined inconsistently across studies.

In the next section, an active area of PH research, namely accidents prevention in children will be introduced as the motivation problem and used throughout the thesis to demonstrate how advancements in synthesis methodology (including existing methods as well as

methodology developed specifically in this thesis) can be employed to help overcome some of the challenges listed above present in synthesis of evidence in a PH evaluation context.

1.3 Example problem: Accidents prevention in children

Unintentional childhood injury is a major public health concern. According to a UK Audit Commission report '*better safe than sorry*' (The Audit Commission, 2007), unintentional injury is currently the leading cause of death among pre-school children (i.e. the 0-4 years old age group) in England. Falls, poisoning and thermal (burns and scalds) injuries are by far the most common injuries reported for this age group in hospital emergency departments in the UK (The Audit Commission, 2007). The importance of reducing childhood injuries and their inequalities is emphasised again and again in many UK government reports (Department of Health, 1999; Department of Health, 2002; Children Act, 2004; Department of Health, 2005; Health, 2007; The Audit Commission, 2007) but evidence of a systematic approach to NHS injury prevention in children is lacking (The Audit Commission, 2007). In response to these concerns, the NIHR funded the 'Keeping Children Safe at Home' (KCS) programme of research to look into causes and prevention of childood accidents in the home. KCS is a 5-year multi-centre programme of research lead by Professor Denise Kendrick at the University of Nottingham with participating centres at the Universities of Leicester, Newcastle, Norwich and of the West of England (Bristol). The overall aim of KCS is to identify the most effective and cost-effective strategies for preventing unintentional injury (primarily focusing on falls, burns, scalds and poisonings) in pre-school children (0-4 year olds) at home.

Many of the issues typical of PH appraisals outlined in Section 1.2 above, such as including evidence from studies of different designs and heterogeneity of the interventions and outcome measures, are particularly relevant to accidents prevention in children. Thus, when this area was previously evaluated by NICE PH30 (NICE, 2010b), only narrative summaries of the evidence identified from the systematic review of the clinical and cost-effectiveness evidence were conducted (Pearson *et al.*, 2009) and estimates from individual studies were used to inform the subsequent cost-effectiveness analyses (Pitt *et al.*, 2009). This goes against the tenets of evidence-based decision making where decisions should ideally be made based on all the available evidence relevant to the decision problem (Welton *et al.*, 2012).

In this thesis, accidents prevention in children is revisited as part of the KCS project with a focus on evaluating the clinical and cost-effectiveness of strategies to prevent unintentional poisonings in pre-school children at home. In doing so, data from two published systematic reviews of the effectiveness evidence (Kendrick *et al.*, 2012b; Young *et al.*, 2013) will be used to provide a) an example context for illustrating how recent advances in quantitative synthesis methodology help address the challenges of evidence sythesis in PH outlined in Section 1.2, and b) a basis for further methodological developments undertaken here to deal with specific issues of the heterogeneity in the baseline risk across studies and borrow information across a series of evidence networks. Note that as part of the KCS project, the evaluation of strategies to increase uptake of functional smoke alarms in households with children and ultimately prevent thermals injuries (primarily burns) has been completed (Saramago Goncalves, 2012) and published in a peer-reviewed journal (Cooper *et al.*, 2011a; Saramago *et al.*, 2014). The evaluation of strategies to prevent falls (Hubbard *et al.*, 2014) and other thermal injuries (primarily scalds) is currently being undertaken as a separate subproject under the KCS programme remit.

1.4 Structure of thesis

The remainder of the thesis is structured as follows: Chapters 2 and 3 introduce the methodology of evidence synthesis and economic evaluation (conducted within a decision analytic modelling framework) respectively as applied to healthcare decision making. These methods will be used to evaluate the clinical and cost-effectiveness of poison prevention practices described in Chapters 5 and 8. They will also form the foundation for development of new synthesis methodology in Chapters 6 and 7 briefly outlined below. Before that, Chapter 4 presents the outcome of a systematic review conducted to i) determine the current state of affairs (i.e. what is currently being done/not done) regarding the use of evidence synthesis methods in PH evaluations and ii) establish a baseline or quality bar for development and application of methodology (outlined in this thesis) to PH evaluation of interventions.

Chapters 5, 6 and 7 demonstrate how recent developments in synthesis methodology (some of which has been developed as part of this thesis) over and above current practices established in Chapter 4, can be applied to overcome challenges of evidence synthesis in PH. Specifically, the methodology of network meta-analysis is applied to the poison prevention data in Chapter 5. Network meta-analysis (NMA) enables simultaneous comparison of many interventions while preserving randomisation. Such methods offer some advantages over pairwise meta-analysis, including the possibility to relax the need for seemingly similar but different interventions to be 'lumped' into two treatment groups for the purpose of conducting a pairwise meta-analysis – a particularly present issue in many PH systematic reviews as described in Chapter 4.

Chapters 6 and 7 introduce new synthesis methodology by extending the standard NMA model to deal with some specific issues. The methods developed in these two chapters are quite general in the sense that they are applicable to evaluation of clinical and pharmaceutical interventions as well as evaluation of PH interventions. Chapter 6 concerns the problem of how include a covariate to account for baseline imbalances in the control group event rate (often referred to as baseline risk) in a network meta-analysis. This covariate is typically measured with error and has been of considerable interest to statisticians and clinicians alike as a proxy for unmeasured but important patient- and or study-level characteristics, which may be modifiers of treatment effect and a potential source of heterogeneity. In the PH context as exemplified by the accidents data introduced above, baseline risk meta-regression can be used to account for residual heterogeneity in the definition of the control group intervention which may persist even after interventions are classified into more homogenous treatment packages as will be shown in Chapter 5.

Chapter 7 presents methods for synthesis of evidence across multiple outcomes. The methods allow for appropriate modelling of the correlation structure within the data, which is important when summarising evidence to inform an economic evaluation. Further extensions of the modelling approach are developed to borrow information across studies, treatments and outcomes which can be useful in situations where the evidence base is either sparse or limiting in some important respects, as is often the case in PH evaluations. This type of

analysis can be used to produce firmer estimates of the treatment effects by making use of all available information relevant to the decision problem (including information from closely related evidence networks) which may be beneficial in reducing decision uncertainty.

Chapter 8 presents the development of a probabilistic decision model to evaluate the cost-effectiveness of home safety interventions to prevent unintentional poisoning injury in pre-school children. Results from the analyses carried out in Chapters 5, 6 and 7 will be used to inform parameters of the decision model.

Finally, Chapter 9 concludes the thesis by summarising the important findings from the work presented in this thesis and discusses how recent developments in synthesis methodology (including methods developed in this thesis) can help overcome the challenges of evidence synthesis in PH evaluations. Limitations of the methodology outlined and applied throughout the thesis are also discussed together with opportunities for further work.

2. REVIEW OF EVIDENCE SYNTHESIS METHODS

2.1 Chapter overview

This Chapter introduces statistical methods for summarising evidence from multiple sources. The methods covered mainly concern meta-analysis of summary or aggregate data and their extensions to allow for inclusion of study-level covariates. An introduction to network meta-analysis (NMA) for simultaneous comparison of multiple interventions will also be given. The methods introduced in this chapter be used in the subsequent chapters to synthesise the evidence on intervention effectiveness. They will also serve as foundation for development of new synthesis methodology. Note that methods for meta-analysis of individual patient data (IPD) (Simmonds *et al.*, 2005; Riley *et al.*, 2007c) have not been reviewed because the thesis mainly concerns the synthesis of aggregate or summary level data.

2.2 Pairwise meta-analysis

Glass (Glass, 1976) was one of the first to define meta-analysis as a statistical method for summarising results of several independent studies. Pairwise meta-analysis methods are useful for comparing two interventions with one another when there are multiple sources of evidence on the two interventions. By combining information from all relevant studies, a meta-analysis can provide a more precise estimate of the effect of one intervention relative to another.

2.2.1 Fixed and random effects models

There are two types of meta-analysis models in most common use – fixed effect and random effects models. In a fixed effect model, it is assumed that effect estimates from a set of N studies are estimating the same underlying effect, so that they can be pooled to obtain a summary estimate of all effect sizes as follows (Welton *et al.*, 2012):

$$Y_i \sim \text{Normal}\left(d, \frac{s_i^2}{n_i}\right) \quad i = 1, 2, \cdots, N \tag{2.1}$$

where Y_i, s_i^2 and n_i respectively represent the effect size, within-study variance and number of individuals in the *i*th study, and *d* is the underlying mean effect common to all the studies, and of interest in the meta-analysis. In practice, the s_i^2 s are usually assumed to be known and estimated by the within-study variance (υ_i^2) from the summary data:

$$Y_i \sim \text{Normal}\left(d, \upsilon_i^2\right) \quad i = 1,2,\cdots,N \tag{2.2}$$

In many medical applications, however, the assumption of a common underlying true effect may not always hold due to differences in patient populations, and study location and settings (Sutton *et al.*, 2000; Welton *et al.*, 2012). If there are doubts about the validity of this assumption, then a random effects meta-analysis may be preferable. In the random effects model, the assumption of a single or common true underlying effect is relaxed to allow for between-study variability (also known as heterogeneity):

$$Y_i \sim \text{Normal}\left(\delta_i, \upsilon_i^2\right) \quad i = 1,2,\cdots,N$$

$$\delta_i \sim \text{Normal}\left(d, \sigma^2\right) \tag{2.3}$$

where δ_i is the *true* effect size specific to the *i*th study, assumed to be drawn from a normal distribution with overall population mean *d* and between-study variance σ^2. It has been suggested however, that the mean of a random-effects distribution, as the average of the individual study effects, may not accurately represent the different study populations especially if there is high degree of heterogeneity (Ades *et al.*, 2005; Higgins *et al.*, 2009). Instead, the predictive effect in a new study, d_{new}, which takes into account heterogeneity in the data, has been suggested as representing a more accurate and robust summary of the data than the random effects mean. If fitted within a Bayesian framework (see as discussed in the next paragraph) the predictive effect in a new study can be obtained as follows (Higgins *et al.*, 2009):

$$d_{new} \sim \text{Normal}\left(d, \sigma^2\right)$$

(2.4)

where d and σ^2 represents the same quantities as in equation (2.3), namely the mean of the random effects distribution and the between-study variance respectively.

The models specified above can be fitted using Frequentist or Bayesian methods. Under the Frequentist approach to statistical inference, all the information about the parameters of the model is contained in the data; hence the analysis can be conducted and parameters estimated by finding values of d and σ^2 that maximise the likelihood functions under equations (2.2) and (2.3) (Welton *et al.*, 2012). The Bayesian approach to statistical inference on the other hand, uses Bayes theorem to combine external information (termed *prior beliefs*) with the information contained in the data (termed the *likelihood*) to obtain a *posterior* summary of all the available information upon which inference is based (Ntzoufras, 2009; Lunn *et al.*, 2012; Welton *et al.*, 2012):

$$posterior \propto likelihood \times prior \tag{2.5}$$

Thus in addition to the *likelihood* for the data that can be derived from equations (2.1) to (2.3) above, prior distributions also need to be specified for the parameters d and σ^2 when conducting a meta-analysis within the Bayesian framework. If there were no other available evidence about the parameters external to the data, then flat or 'vague' prior distributions could be specified over plausible ranges supported by the parameters of the model. In that case, any flat or 'vague' prior distribution containing a minimal amount of information will be completely dominated by the data and a Bayesian analysis should produce results close to what will be obtained from a frequentist analysis. For example, in the meta-analysis of binary outcome data, the effect size Y_i would normally be a log (odds ratio) so that the following prior distributions, which are considered to be minimally informative on the log-odds ratio scale for most practical medical applications (Dias *et al.*, 2012; Welton *et al.*, 2012) can be specified for the parameters d and σ:

$$d \sim \text{Normal}\left(0,10^3\right) \tag{2.6}$$

$$\sigma \sim \text{Uniform}\left(0,2\right) \tag{2.7}$$

Unless otherwise stated, all analyses in this thesis will be conducted from the Bayesian framework mainly because of the flexibility with which increasingly complex models can be fitted. In addition to the increased flexibility, the Bayesian approach also allows for the

uncertainty around the between-study variance parameter, σ^2 to be taken into account automatically in the meta-analysis (Higgins *et al.*, 2009; Welton *et al.*, 2012). The software of choice for fitting the Bayesian meta-analyses carried out in this thesis is the WinBUGS software (Lunn *et al.*, 2000) which uses Markov Chain Monte Carlo (MCMC) simulations to obtain posterior summary estimates of the parameters of interest. Again unless otherwise stated, 'vague' or minimally informative prior distributions will be specified for the model parameters, so that the results of the analysis are very close to what would have been obtained if fitted using Frequentist methods.

2.2.2 Heterogeneity

Between-study variability in the treatment effect (i.e. systematic differences in effect sizes across studies) which is more than can be attributed to sampling error alone is termed statistical heterogeneity (Sutton *et al.*, 2000; Borenstein *et al.*, 2009). As stated in the previous section, a random effects model should be fitted, and its magnitude quantified by the parameter σ^2 specified in equation (2.3) if unexplained heterogeneity in the effect estimates is expected across studies. Alternative measures of heterogeneity include the Cochran *Q-statistic* which provides a test of homogeneity of the effect sizes across studies and the I^2-*statistic* which is based on the *Q-statistic* and estimates the proportion of the total variability in the effect sizes that could be attributed to heterogeneity (Higgins and Thompson, 2002).

Statistical heterogeneity is explainable if the variability in the treatment effect is attributable to differences in the characteristics of the studies such as methods, design and patient populations. Statistical heterogeneity that is not explained by observable characteristics as above is said to be residual, and is accounted for in a random effects meta-analysis. When there is evidence of substantial heterogeneity, sub-group analyses or meta-regression methods can be performed to investigate sources of the heterogeneity and if possible adjust for it (Higgins *et al.*, 2009).

2.2.3 Meta-regression

The random effects model described above will account for the heterogeneity between studies but does not explain it. To try and explain or explore potential sources of the heterogeneity, the meta-analysis can be extended to include study and aggregate patient characteristics as study-level covariates as follows:

$$Y_i \sim \text{Normal}\left(\delta_i, \upsilon_i^2\right)$$

$$\delta_i \sim \text{Normal}\left(d + \beta\left(x_i - \bar{x}\right), \sigma^2\right) \tag{2.8}$$

where the regression coefficient β in equation (2.8) gives a measure of the relationship between the treatment effect and the covariate x_i (e.g. mean age) in study *i* centred on mean $\bar{x}$ (e.g. the mean of the mean ages across studies), *d* is the mean treatment effect when $x_i = \bar{x}$ (i.e. at the mean covariate value) and σ^2 gives a measure of the remaining heterogeneity unexplained by the covariate. If subgroup analysis is required then x_i will be categorical covariate with values indicating the subgroup to which the *i*th study belongs. All other parameters have the same interpretation as before. If performed under a Bayesian framework, prior distributions need to be specified for parameters *d*, β and σ. Note that if required, Equation (2.8) can easily be extended to include more than one covariate or subgroups and multiple interaction terms in the meta-analysis. Meta-regression models may, however, lack sufficient power to detect the associations they intend to measure since a typical meta-analysis will involve relatively few studies (Egger *et al.*, 2000). Also, as in any regression analysis, the model is susceptible to confounding by unknown variables and aggregation or ecological bias may arise if the relationship between aggregated study-level characteristics and outcomes do not reflect the true relationship at the individual level (Sutton *et al.*, 2000). For these reasons, the results of a meta-regression should be treated with caution and as associative rather than causative.

2.2.4 Publication bias

Publication bias refers to the tendency for studies that show evidence of a statistically significant effect to be published over and above those that do not (Dickersin *et al.*, 1987; Peters *et al.*, 2008). Such biases, if present, have the potential to distort the results of a meta-analysis, leading to inaccurate and misleading conclusions. Methods exist to detect and adjust

for publication bias; see for example (Egger *et al.*, 1997; Egger *et al.*, 2000; Peters *et al.*, 2008; Moreno *et al.*, 2009). The funnel plot (Figure 2.1) is the simplest of these methods. This is a scatter plot with the effect size on the x-axis and some measure of the precision of the effect sizes such as the inverse of the standard error on the y-axis (Sterne and Egger, 2001).

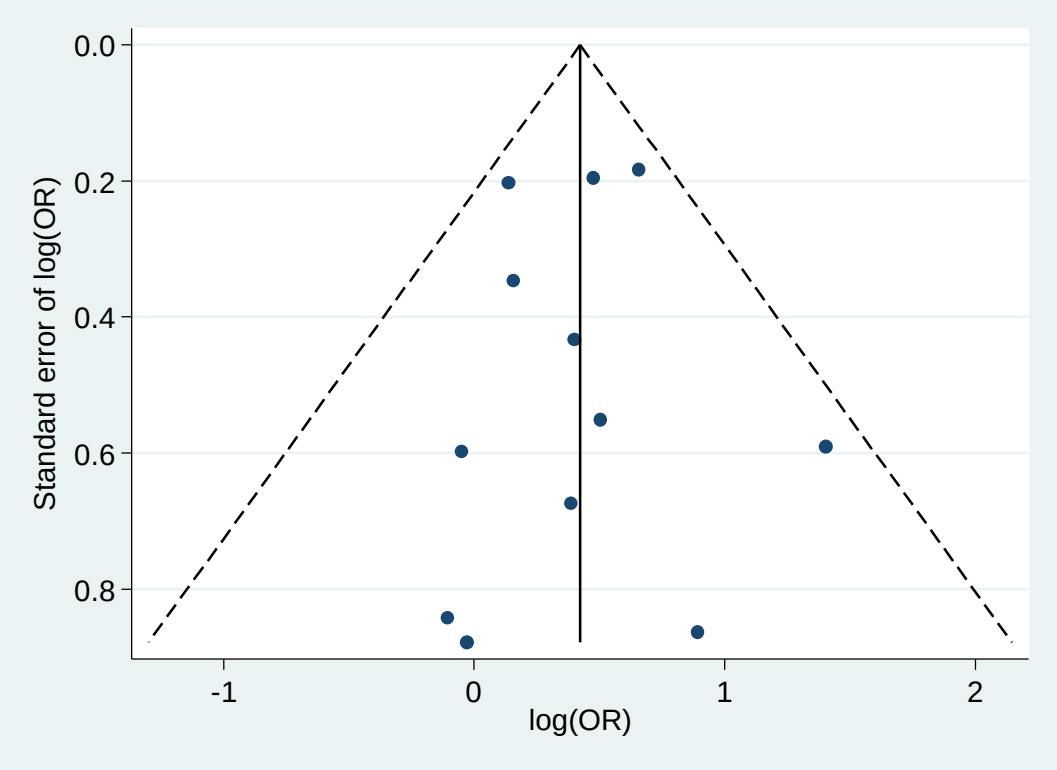


Figure 2.1: Example of a funnel plot
Pseudo 95% confidence limits are represented as dotted lines. Log(OR) refers to log(odds ratio).

A funnel plot should be symmetric with a characteristic funnel shape appearance when no publication bias is present so that greater variability in the effect sizes is observed in the smaller and less precise studies towards the lower part or base of the funnel (Peters *et al.*, 2008). Note that funnel asymmetry can also occur for other reasons, such as small study effects, If there is evidence of funnel plot asymmetry, the extent of bias may be quantified by employing tests for 'publication bias' assessment that have been proposed (Egger *et al.*, 1997; Harbord *et al.*, 2006; Rucker *et al.*, 2008; Peters *et al.*, 2010). However, these tests typically have low power and it is recommended that test of funnel plot asymmetry should only be carried only if there are at least 10 studies in the meta-analysis (Sterne *et al.*, 2011).

Methods used to correct publication bias include trim and fill (Duval and Tweedie, 2000) and several regression based methods described in detail in Moreno et al. (Moreno *et al.*, 2009).

2.3 Network meta-analysis

2.3.1 Model for binary outcome data

Network meta-analysis (Lumley, 2002) also called mixed treatment comparisons (Lu and Ades, 2004; Caldwell *et al.*, 2005) or multiple treatment meta-analysis (Salanti *et al.*, 2008), methods extend standard meta-analysis to enable simultaneous comparison of multiple treatments while maintaining randomisation. These methods enable 'direct evidence' (i.e. studies that directly compared the two treatments under consideration) and 'indirect evidence' (i.e. the remaining studies in the network under the consistency assumption) on pairwise contrasts to be pooled under the assumption that there is consistency (see Section 2.3.3 below for explanation and assessment of evidence consistency) between the direct and indirect evidence, hence they are often referred to as mixed treatment comparisons (Lu and Ades, 2004; Caldwell *et al.*, 2010; Welton *et al.*, 2012). For example, in the following ensemble of evidence on 3 interventions labelled A, B and C taken from Welton *et al.* (Welton *et al.*, 2012), trials comparing interventions A and B (AB trials) will provide direct evidence to estimate effect of B relative to A, denoted as $\hat{d}_{AB}^{dir}$. At the same time, trials of A versus C (AC trials) and those of B versus C (BC trials) will provide indirect evidence to estimate effect of intervention B relative to A, denoted as $\hat{d}_{AB}^{indir}$ through the relationship:

$$\hat{d}_{AB}^{indir} = \hat{d}_{AC}^{dir} - \hat{d}_{BC}^{dir}$$

(2.9)

where $\hat{d}_{AC}^{dir}$ and $\hat{d}_{BC}^{dir}$ represent direct estimates from AC trials and BC trials respectively. The essential requirements for this type of analysis are that: i) the interventions should be linked with each other, forming a connected network of treatments as shown in Figure 2.2, and ii) there should be consistency in the evidence structure when direct and indirect evidence are pooled on pairwise contrast as explained in Section 2.3.3.

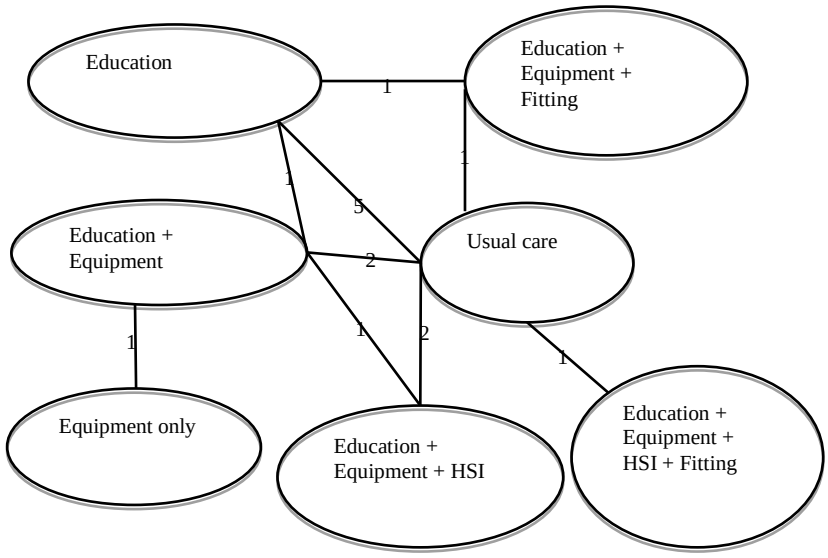


Figure 2.2: An example of a connected network from the example dataset Interventions that have been compared in trials included in the meta-analysis are connected by lines; the number of trials is indicated on the lines

Given binary outcome data from treatment arms of each study included in the meta-analysis, a random effects NMA may be specified using the method of Lu and Ades (Lu and Ades, 2004). It is assumed that the occurrence of r_{ik} events from a total of n_{ik} individuals in the *k*th-arm ($k = A, B, C, \cdots,$) of the *i*th-study follow a binomial distribution with underlying event probability p_{ik}:

$$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ik}) = \begin{cases} \mu_{ib}, & if \quad k = b \\ \mu_{ib} + \delta_{ibk}, & if \quad k > b \end{cases} \qquad \text{for } b = A, B, C, \cdots, \qquad (2.10)$$

$$\delta_{ibk} = d_{bk} + \varepsilon_{ibk}, \; \varepsilon_{ibk} \sim \text{Normal}(0, \sigma_{bk}^2)$$

where d_{AA} = 0 (i.e. the intervention effect in the reference or baseline intervention for the entire network is set to 0) and *k*>*b* implies intervention *k* comes alphabetically after *b*. The parameter μ_{ib} is the effect of baseline intervention *b* (log odds) in study *i* and ε_{ik} denote a random effect indicating that the study-specific effects (log odds ratios) of intervention *k* relative to *b*, δ_{ibk}, are normally distributed with mean d_{bk} and between-study variance σ_{bk}^2. Note that a fixed effect model is obtained if $\sigma_{bk}^2 = 0$.

The fundamental assumption underlying random effects network meta-analysis is that the intervention effects are exchangeable (see Section 2.3.3 for explanation of the concept of exchangeability) across the entire network of trials regardless of whether or not treatments *b* and *k* are included in trial *i* (Lu and Ades, 2004). Validity of this exchangeability assumption means that the pooled intervention effects d_{bk} can further be expressed as functions of basic parameters taken with reference to treatment *A*, (i.e. $d_{bk} = d_{Ak} - d_{Ab}$). Effect estimates from trials with more than 2 treatment groups will be correlated through sharing a common comparator treatment. The correlation may be taken into account by assuming homogenous variances (i.e. $\sigma_{bk}^2 = \sigma^2$ for all *b* and *k*) so that the covariance is equal to $\frac{\sigma^2}{2}$ (Lu and Ades, 2004). Alternatively, heterogeneous variance models have also been proposed (Lu and Ades, 2006). The analysis if conducted within the Bayesian framework, require prior distributions to be specified for the parameters d_{bk} and μ_{ib} as in equation (2.6) and σ as in equation (2.7). Parameter estimation is then by Markov chain Monte Carlo simulation implemented in the WinBUGS software (Lunn *et al.*, 2000).

2.3.2 Assessing which intervention is the best

As explained above, network meta-analysis enables simultaneous comparison of multiple interventions by using all available data in a connected network of studies. The main advantage of NMA is that interventions can be ranked in terms of their efficacy and used to estimate the probability that each intervention is the best option (Welton *et al.*, 2012). This can easily be implemented in WinBUGS using the '**rank'** and '**equals'** functions as follows:

$$r_k = \begin{cases} \mathbf{rank}(\mathbf{d}, k) & \text{if odds ratio} < 1 \text{ confers benefit} \\ (K+1) - \mathbf{rank}(\mathbf{d}, k) & \text{if odds ratio} < 1 \text{ confers harm} \end{cases} \tag{2.10}$$

$$best_k = \mathbf{equals}(r_k, 1) \tag{2.11}$$

where for $k = 1, 2, \cdots, K$, r_k indicates the rank for the *k*th intervention (the most effective intervention is ranked number 1), and $best_k$ indicates the probability that intervention k is the best, *K* is the total number of interventions being evaluated in the NMA and **d** is a vector of mean effects relative to the reference intervention. When reporting the ranking of

interventions in this thesis, the posterior median estimate is preferred to instead of posterior mean (can be a whole number of decimal). This ensures that r_k takes the value 1 if *k* is the best intervention and *K* if *k* is the worst intervention (Welton *et al.*, 2012).

2.3.3 Assessing evidence consistency

NMA assume that the trial-specific effects δ_{ibk} are normally distributed around a common mean with variance σ_{bk}^2. In other words, the δ_{ibk} s are assumed to be exchangeable which is to say they are different but there is no way of predicting the rankings of their magnitude a priori (Welton *et al.*, 2012). This assumption is unlikely to be met if there are evidence loops in the network where the direct and indirect evidence on pairwise contrasts are inconsistent, or do not agree (Dias *et al.*, 2010). Note that evidence loops formed solely by multi-arm trials are excluded from the consistency assessment, since by definition, evidence from a multi-arm trial cannot be inconsistent (Dias *et al.*, 2010). Doubts about the validity of NMA have been expressed because of concerns that direct and indirect evidence from disparate sources may not be consistent and should not be pooled together (Song *et al.*, 2003; Song *et al.*, 2011). Therefore, carrying out checks for evidence inconsistency in NMA is crucial if the results of the analysis are to be trusted.

Methods for assessing evidence inconsistency have been published, for example the papers by Lu and Ades, 2006 (Lu and Ades, 2006), Dias et al., 2011 (Dias *et al.*, 2011b) and the recently released NICE Decision Support Unit Technical Support Document 4 (Dias *et al.*, 2011b). In this thesis, one of these methods, called node-splitting (Dias *et al.*, 2011b), will be used to assess the consistency of the evidence when NMAs are used to compare intervention effectiveness. Briefly, for each pairwise contrasts in a closed loop of evidence, the node-splitting method enable separate estimates of the mean treatment effect based on the direct evidence and indirect evidence to be calculated. The difference between these two estimates can be used to construct a test for inconsistency and derive a 2-sided *p-value* for the null hypothesis that direct and indirect estimates are different. Note that, the test may lack sufficient power to detect inconsistency especially if the number of studies in the meta-analysis is small (Dias *et al.*, 2013). Therefore failure to reject the null hypothesis does not necessarily imply consistency of the evidence. In addition, the node splitting method can

detect inconsistency but does not explain it. That is to say the method does not identify the pairwise comparisons in the network that are inconsistent. Therefore, when inconsistencies are detected in a network of evidence it is important to go back and re-examine the entire data to see if the cause of the inconsistency can be identified from available evidence (Dias *et al.*, 2013). It is also important to consider carefully whether reliable conclusions can be drawn from combining direct and indirect evidence when there is inconsistency in the evidence structure.

2.3.4 Assessing publication bias in network meta-analysis

There are as yet no methods for assessing publication bias in network meta-analysis, therefore assessment of publication bias in this thesis will be carried out using the methods reviewed in Section 2.2.4 for assessing publication bias in pairwise meta-analysis.

2.3.5 Convergence diagnostics

When using MCMC based estimation procedures implemented through the WinBUGS software, it is important to carefully assess and report choice of prior distributions, initial/starting values, number and length of iterations in addition to checking for evidence that convergence of the simulated samples is adequate (Spiegelhalter *et al.*, 2000). This is because the results of the analysis can be sensitive to the choice of prior distributions, initial values, length of 'burn-in' and so on. Convergence can be assessed by examining the history, kennel density, autocorrelation and Brooks-Gelman diagnostic plots available from the WinBUGS menu as follows:

i) The history plot shows successive realisations of the MCMC sampler plotted against the iteration number for each parameter of interest (Lunn *et al.*, 2012). A stable plot with 'fat hairy caterpillar-like' appearance is evidence that the Markov Chain has reached stability (and may have converged) whereas a snake-like appearance may indicate a high degree of autocorrelation or evidence of non-convergence (Lunn *et al.*, 2012).
ii) Posterior kernel density plots are used to assess whether or not the distributional form of model parameters appear as expected, whilst the autocorrelation plots assesses the degree of correlation between successive iterations of the sampler. Rapid or gradual thinning

out towards zero if observed in the autocorrelation plot should indicate low correlation between 40 sequential iterations of the sampler and vice versa. Low autocorrelation would usually indicate faster mixing and convergence whereas higher autocorrelation indicate slow mixing convergence, hence necessitating longer running of the sampler.

iii) The Brooks-Gelman diagnostic plot (Brooks and Gelman, 1998) compares the within and between chain variances from simultaneous running of multiple chains with different starting values to assess evidence of convergence for each parameter being monitored. The green line of the plot represent the normalized width of the central 80% interval of the pooled runs (B), blue line represents the normalized average width of the 80% intervals within the individual runs (W) and red is R where $R= B/W$. Convergence is deemed to have occurred if R has converged to 1 and B and W have converged to stability (Spiegelhalter *et al.*, 2007).

2.3.6 Goodness fit and model selection

The posterior mean residual deviance, defined as the deviance for the fitted model minus deviance for the saturated model, will be used to assess how well the model predictions fit the observed data (McCullagh and Nelder, 1989; Spiegelhalter *et al.*, 2002). Under the null hypothesis that the model fits the data well, the posterior mean residual deviance is expected to be approximately equal the number of unconstrained data points. Therefore models would be judged to provide adequate fit if the residual deviance is close to the number of data points in the model. The fit of alternative models (for example, fixed effect versus random effects model) can be compared using the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002; Spiegelhalter *et al.*, 2014). The DIC is the sum of the posterior mean residual deviance and the effective number of parameters and, as such, provides a measure of model fit that penalizes for model complexity.

2.4 Chapter summary

The statistical methods for summarising evidence from multiple sources were reviewed in this chapter. The NMA methods presented in Section 2.3 especially will used to synthesis the evidence on the effectiveness of interventions to increase uptake of poison prevention

measures presented in Chapter 5.. The NMA model described in Section 2.3 is then extended to include a covariate for the baseline risk in Chapter 6 based on methods developed in Achana *et al.*(Achana et al., 2013) and to multiple outcome settings in Chapter 7 based on methodology developed in Achana et al (Achana *et al.*, 2014a). he results from the analyses in Chapters 5, 6 and 7 are used to inform the cost-effectiveness evaluation of poison prevention practices in Chapter 8. Before that, an introduction to methods for economic evaluation of healthcare interventions is presented next in Chapter 3.

3. REVIEW OF DECISION ANALYTIC MODELLING METHODS

3.1 Chapter overview

Methods for quantitative synthesis of evidence from diverse sources were reviewed in chapter 2. In this chapter, the methods for economic evaluation of healthcare interventions are introduced. Emphasis will be placed on methods for economic evaluations conducted within a decision analytic modelling framework as opposed to the conduct of economic evaluations alongside clinical trial data.

3.2 Economic evaluation in healthcare

3.2.1 Introduction

Healthcare decision makers all over the world are faced with the problem of deciding how best to allocate resources within limited budgetary constraints. Hence clinical and economic dimensions of healthcare provision should be taken into account in making decisions about which interventions to fund if resources are to be efficiently allocated. Economic evaluation of healthcare interventions offers a framework for synthesis of data on clinical outcomes and resource use in order to estimate the costs and benefits associated with two or more competing interventions (Briggs *et al.*, 2006; Gray *et al.*, 2011).

Evaluations may be conducted alongside a randomised controlled trial (RCT) or through a modelling exercise. Economic evaluations conducted through modelling exercises are called decision analytic models. They provide an explicit quantitative approach to synthesis of information from multiple sources and are useful for comparing the cost-effectiveness of competing interventions that may not have been directly considered in a single RCT and also in situations where there may be limited or non-existent trial data on long term costs and effects (Welton *et al.*, 2012; Baio, 2013). For these reasons, modelling is increasingly being employed by decision making bodies such as NICE when deciding which interventions should be funded by the NHS (NICE, 2008; NICE, 2012).

Ultimately the goal of any economic evaluation is to compare the costs and benefits associated with competing interventions. This can be done either through a 'cost-benefit' or a 'cost-effectiveness' analysis. Cost-benefit analysis attempts to assess whether the monetary value of health benefits is greater or less than the costs of obtaining the benefits by expressing both the clinical outcomes and the resource use purely in monetary terms (Gray *et al.*, 2011). Cost-effectiveness analysis, on the other hand, compares costs and effects of two or more interventions using either disease-specific or generic measures of health. Disease specific measures, as the name implies, are specific to a particular disease or health condition. Examples include the number of symptom free days, true positive cases of cancer detected, number of deaths averted, and so on. Generic measures of health (also referred to as utilities) are non-disease specific, thus can be used to measure the health benefit across different disease domains. Examples include Life Year (LY), the Quality Adjusted Life Year (QALY) and the Disability Adjusted Life Year (DALY). Cost-effectiveness analyses in which the health outcomes are expressed in terms of these generic measures are referred to as a cost-utility analysis (Briggs *et al.*, 2006; Welton *et al.*, 2012).

3.2.2 Measuring health outcomes

The Quality Adjusted Life Year, or QALY for short, is one of the most frequently used utility measures as it incorporates both quality and quantity of life when valuing health (Phillips, 2005). As stated above, QALYs are useful when comparing the value of health generated by interventions across different disease domains. Thus, they are the preferred measure of health outcomes for NICE reference case analysis. Generally, QALYs allow health to be valued on a scale of 0 to 1, with 0 being equivalent to death and 1 being equivalent to one-year life expectancy in perfect health. In some instances, negative QALY values can be used to indicate health states considered worse than death (Baio, 2013).

As mentioned in the section above, there are other generic units for measuring health outcomes in economic decision analysis such as the Life-Year. Another is the DALY which specifically takes into account excess morbidity and mortality in a given population and is the preferred measure for valuing health by the World Health Organisation (World Health Organisation, 2014). However, for the reasons stated above, the QALY will be the main

measure of health outcome in cost-effectiveness evaluations presented in Chapter 8 as commended by NICE (NICE, 2012; NICE, 2013).

3.2.3 Measuring resource use and unit costs

In any cost-effectiveness evaluation, careful thought and attention should be given to collection of data on resource use and unit costs because the results of the evaluation can change substantially depending on what is included in the estimation of costs. There are two issues to consider in any costing exercise (Baio, 2013). The first concerns identifying what items to include in the costing analysis. This will be influenced by many factors including the perspective of the analysis and the natural history of the disease. Costs and benefits should be considered from the viewpoint of the decision maker and this is often referred to as the perspective of the economic evaluation (Briggs *et al.*, 2006; Gray *et al.*, 2011). For example, when appraising new technologies for a centrally funded healthcare system such as the UK National Health Service (NHS), NICE recommends that potential impact on resource use, costs and savings should be considered from the perspective of the NHS and personal social services in its base case analysis (NICE, 2013). However, the NICE methods manual for the development of NICE public health guidance (NICE, 2012) suggested that other perspectives such as that of employers, the private sector or the wider society can be considered when evaluating PH interventions if considered more appropriate than a health sector perspective.

The second issue concerns identifying sources of evidence on unit costs and resource use relevant to the decision problem and deciding on how best to incorporate such evidence into the analysis. Evidence on unit costs and resource use may come from a number of sources, including data collected as part of a clinical trial, a costing study, another economic evaluation or routinely published datasets of unit costs. For example, the 'Unit Costs of Health and Social Care' (Curtis, 2012) published annually by the Personal Social Services Research Unit (PSSRU) in the UK and the Diagnosis Related Groups (DRG) payments developed for the United States (US) Medicare system provide routine costing data on clinical activities. Once the data has been collected, it may be necessary to make certain adjustments (for example using to Bank of England's inflation calculator (Bank of England, 2014)) to bring all prices to a common base year. This will account for the effect of inflation

when costs are incurred over different time periods and also adjust for currency differences if information from countries with different currency units is included in the analysis.

3.2.4 Discounting costs and utilities

It is widely recognised that costs and benefits occurring over a period of time are worth more in the present than in the future (Briggs *et al.*, 2006; Gray *et al.*, 2011; Baio, 2013). This is because of the opportunity cost associated spending money in the present rather than waiting (for example, the interest forgone in delaying to make an investment) and a desire by individuals and societies to enjoy benefits in the present rather than in the future (Torgerson and Raftery, 1999). Discounting is thus used in economic evaluations to account for the differential timings in costs and outcomes and hence, ensure the results of the analysis reflect the present value of costs and benefits that have accrued over the time horizon of the analysis (NICE, 2008). This is achieved by discounting costs and benefits in evaluations with longer time horizons (more than 1 year) by a factor of $1/(1+d)^t$, where d is the discount rate and t is the time at which the costs and benefits are realised (Drummond *et al.*, 2005). There is currently ongoing debate about whether or not future benefits should be discounted at all and if so what rate of discount to use (Torgerson and Raftery, 1999; Baio, 2013). The discount rate for both costs and benefits suggested by NICE when appraising healthcare technologies (NICE, 2013) and PH interventions (NICE, 2012) is of 3.5% per annum. This rate will be used to discount costs and health utilities in the decision analytic model developed in Chapter 8.

3.3 Decision analytic modelling methods

3.3.1 Cohort versus individual patient-simulation

Estimates of expected costs and effects required in a cost-effectiveness evaluation can be obtained using cohort-level or patient-level (also called micro-simulation) simulation based methods. In patient-level simulation, costs and effects are first modelled for each individual patient and then averaged across a sufficiently large sample of patients to obtain the expected costs and effects. In a cohort-level model, individual patients are aggregated into a group (i.e.

the cohort) with expected costs and effects estimated for the whole group rather than for individual patients (Davis *et al.*, 2014). Cohort models are currently the most commonly used techniques in decision analytic based cost-effectiveness analysis. As such, cohort models will be the focus of the remainder of this chapter, and of the decision models developed in Chapter 8. The two most commonly used cohort simulation models are decision trees and Markov models. These are discussed next.

3.3.2 Decision tree models

A decision tree is a branching tree system that can be used to estimate the cost-effectiveness of healthcare interventions. Decision trees have nodes and branches. The nodes represent stages in the pathway where choices or events of uncertain outcome occur. There are three types of nodes in a simple decision tree structure: i) decision node, ii) chance node and iii) terminal node (Gray *et al.*, 2011). A decision node is usually indicated by a squared box and represents the choices or options available to the decision maker. Chance nodes are circular and represent events of uncertain outcome while terminal nodes are triangular and represent final outcomes in the decision process (Gray *et al.*, 2011). The branches of the tree represent mutually exclusive pathways or the sequence of events that can occur over time (Briggs *et al.*, 2006; Gray *et al.*, 2011) and indicate the flow of information (usually from left to right) through the model.

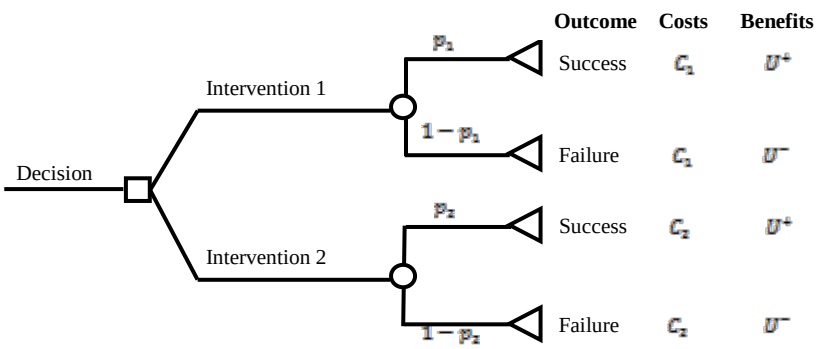


Figure 3.1: Simplified diagram of a decision tree model

Figure 3.1 gives a simplified example of a decision tree model taken from Welton *et al.* (Welton *et al.*, 2012). The options available to the decision maker concern whether to recommend one of two possible interventions represented by the branches of the decision node. In this simplified example, each choice leads to one of two mutually exclusive chance events (i.e. whose occurrence is determined by some underlying event probability). Assessment of cost-effectiveness is then based on comparing the estimated payoffs (i.e. the expected costs and benefits) for each intervention, calculated as a sum of pathway values weighted by the respective pathway probabilities.

In principle, decision trees are simple to construct, easy to understand and more transparent than complicated modelling techniques. However, they are mainly suited to modelling acute conditions where events occur over short time horizons. They are not suitable for modelling chronic disease progression or complex health conditions where the individuals move between different disease/health states and events play over longer time periods as the decision tree quickly become bushy, cumbersome and difficult to handle (Briggs *et al.*, 2006). For these reasons, economic evaluations of chronic and complex conditions are often conducted using Markov models.

3.3.3 Markov Models

Markov models are techniques for analysing uncertain processes that occur over time. In healthcare decision making, they are suited to modelling diseases that are chronic or repetitive in nature in which costs and effects are spread over longer time periods (Gray *et al.*, 2011). The basic idea of a Markov model is that the disease in question is divided into distinct health states, which are chosen to represent important clinical and economic events in the disease process. During the modelling process, over a fixed time period called 'Markovian circle' (Briggs and Sculpher, 1998), patients may stay in the same state or move between states. These movements are governed by transition probabilities, the values of which are determined by the clinical prognosis or natural history of the disease. The length of the 'Markovian circle' is chosen so as to be clinically meaningful for the disease and to ensure patients remain in health states long enough for costs and benefits/clinical effects associated with a given health state to be realised. To evaluate the model, costs and effects

are attached to each state at the end of each circle. The analysis is then run for a sufficient number of cycles (known as the time horizon) to allow for all individuals to reach an 'absorbing state', defined as a state from which once entered individuals cannot leave (for example, death). An example of a Markov model for injury progression in pre-school children following accidental ingestion of a toxic substance is presented in Figure 3.2. There are 4 health states (well, injury, death from injury and death from causes unrelated to injury mechanism being evaluated). The two death states are absorbing states, hence individuals can enter but not leave once in these states. This is indicated in Figure 3.2 by the only arrows coming out of these states looping back unto the same states.

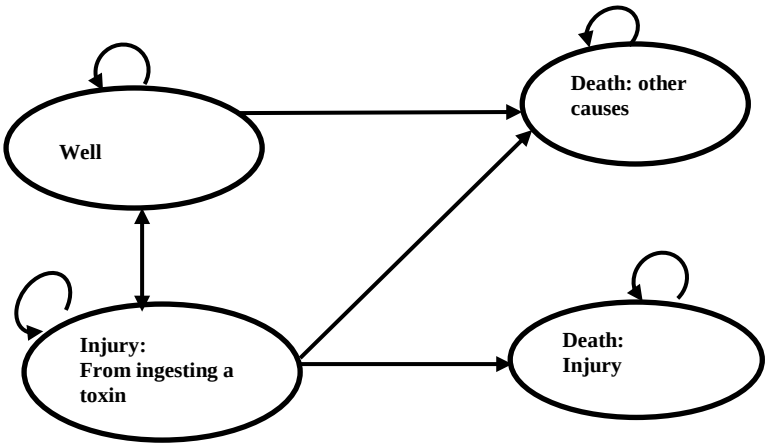


Figure 3.2: A simplified example of Markov model

Markov models have some properties which make them attractive. First of all, structuring the analysis around distinct health states increases flexibility with which chronic disease or complex health processes can be modelled. Secondly, events unfolding over time can be represented as transitions between health states over several Markov cycles. The underlying assumption is that transition probabilities remain constant over time and that transition between states depends only on the current health state of the individual regardless of any previous states the individual may have resided (Briggs *et al.*, 2006). This lack of memory also called the 'Markovian assumption' implies that patients in a given state at a given time have the same prognosis irrespective of pathways taken to arrive at the respective state. This memoryless assumption greatly simplifies the mathematics and computations involved but also limits applicability of Markov models as disease prognosis most often depends on the previous history of the disease (Sun and Faunce, 2008). However, if required, temporary and

tunnel states can be built into the model in order to relax the Markovian assumption and capture short but important health outcomes.

3.3.4 Comprehensive decision analysis

The conventional approach to decision analytic modelling (whether using a decision tree or Markov model) is a two-stage process. In the first stage, meta-analysis to summarise evidence from multiple studies and to derive an estimate of intervention effect(s). The effectiveness estimates are then combined with data on resource use, unit costs and utilities in a second stage analysis to estimate the expected costs and benefits. This process of separating out the evidence synthesis from the rest of the decision analysis effectively ignores important correlations between the effectiveness and cost-effectiveness measures which may have implication for estimating expected costs and effects associated with the interventions being evaluated. As an alternative, Cooper *et al.* (Cooper *et al.*, 2004) proposed a one-stage model, called a comprehensive decision model, in which evidence synthesis and the decision analysis are conducted in a single unified/integrated stage. When implemented within a Bayesian framework, the comprehensive decision modelling approach has the advantage of incorporating uncertainty and the correlations between the parameters and propagating these automatically to the model outputs in a coherent fashion. For these reasons, the models developed in Chapter 8 will utilise the comprehensive decision analysis approach.

3.3.5 Incorporating uncertainty

Uncertainty is an inherent phenomenon in any decision making process that should be handled appropriately to increase confidence in the results of the analysis. Uncertainty in the cost-effectiveness estimates may arise for several reasons, but the two mostly widely reported sources of uncertainty in the health economics literature are i) uncertainties concerning the value of model inputs or parameters and ii) uncertainties about model structure, assumptions and the methods used to evaluate the model (Briggs, 2000; Briggs *et al.*, 2006; Gray *et al.*, 2011).

Uncertainties about the model structure (for example, how many health states to include and the assumptions that go into making these decisions about the model structure) can contribute to uncertainty in the model outputs and results (Gray *et al.*, 2011). This type of uncertainty, collectively termed structural uncertainty, may be handled either through model averaging techniques (Bojke *et al.*, 2009; Jackson *et al.*, 2009) or more commonly through deterministic sensitivity and scenario analyses in which the impact of varying model assumptions and structure can be investigated under different scenarios.

Parameter uncertainty relates to the precision with which input parameters are estimated and incorporated into the decision model. Uncertainty in model inputs such as effectiveness estimates, transition probabilities, unit costs, resource use and utilities can be incorporated in the cost-effectiveness evaluation through probability distributions (Gray *et al.*, 2011; Welton *et al.*, 2012). This type of analysis in which the parameters are characterised by probability distributions is often referred to as a probabilistic decision analysis to distinguish it from a deterministic analysis where point estimates (usually the mean value of each parameter) are included in the model without taking account of the uncertainty around the input parameters (Welton *et al.*, 2012).

3.4 Presenting results of economic evaluations

3.4.1 Cost-effectiveness plane

The ultimate aim of an economic evaluation is to estimate expected costs and effects associated with two or more interventions in order to derive estimates of cost-effectiveness. The estimates of cost-effectiveness in this process are normally presented in terms of a ratio statistic known as the Incremental Cost-Effectiveness Ratio (ICER) (Welton *et al.*, 2012). The ICER represents the cost incurred per additional unit of effectiveness as defined by the equation:

$$ICER = \frac{C^{(t)} - C^{(c)}}{E^{(t)} - E^{(c)}} = \frac{\Delta C}{\Delta E} = \lambda \tag{3.1}$$

where $C^{(t)}$ and $E^{(t)}$ represent the costs and effects (benefits) in the intervention group, $C^{(c)}$ and $E^{(c)}$ represent the corresponding values in the control group, ΔC and ΔE represent

incremental costs and effects, and λ is a Willingness-to-pay or ceiling ratio indicating the maximum amount that a decision maker is willing to pay per additional unit of effectiveness (Welton *et al.*, 2012).

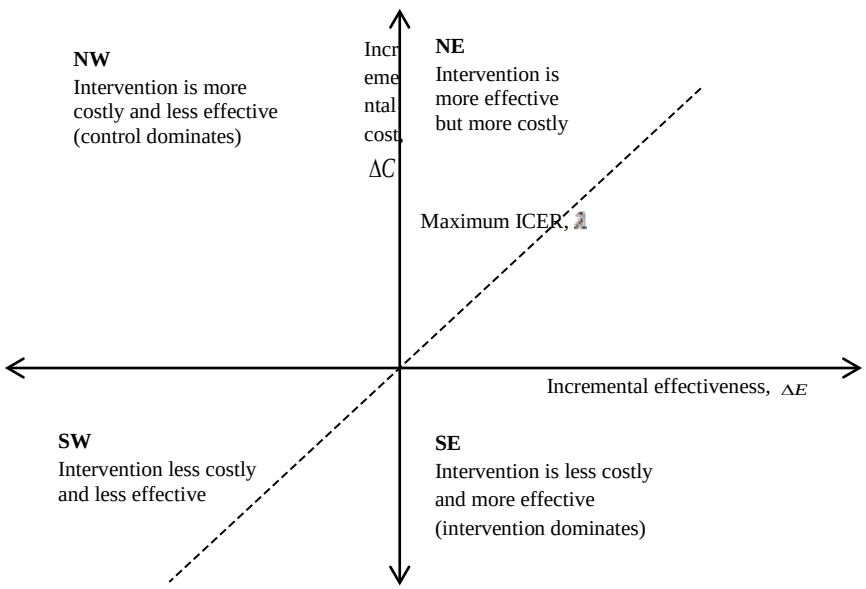


Figure 3.3: Cost-effectiveness plane (adapted from Welton *et al* ., 2012))

One way of presenting the results of a probabilistic decision model is to plot the simulated ICERs on a cost-effectiveness (CE) plane (Figure 3.3). The CE plane is divided into four quadrants with incremental effectiveness, ΔE, displayed along the x-axis and incremental costs, ΔC displayed along the y-axis. It will be relatively uncontroversial to make decisions on the basis of ICERs that lie either in the south-east or the north-west quadrants of Figure 3.3 as these quadrants indicate that the intervention is more effective and cheaper than the control or vice versa. However, in the situation where the ICER falls in the north-east or south-west quadrants, decisions about which intervention is cost-effective will depend on the amount the decision maker is willing to pay and hence which side of the dashed line, represented by λ, that the ICER falls. The intervention is considered cost-effective at the specified value of λ if the ICER lies below (i.e. the right hand side of) the dashed line and cost-ineffective if the ICER lies above the dash line. This criterion can be used to calculate the probability that an intervention is cost-effective at a specified λ in a probabilistic cost-effectiveness analysis as explained in the next section.

3.4.2 Cost-effectiveness acceptability curve

The exact value of λ, the amount a decision maker may be willing to pay for an additional unit of health benefit (for example 1 QALY) may be unknown in practice (Welton *et al.*, 2012). In the UK, NICE currently uses λ values of £20,000 to £30,000 per QALY gained when appraising the cost-effectiveness of health technologies (NICE, 2013). In a probabilistic analysis, this problem can be overcome by calculating the probability that a new intervention is cost-effective for different values of λ and then plotting these probabilities against λ to form a cost-effectiveness acceptability curve (CEAC). An example of such a plot is shown in Figure 3.4. These probabilities can be obtained in two ways. First, the simulated values of a probabilistic ICER are plotted and the required probabilities that the intervention is cost-effective at different values of λ are obtained simply by counting the proportion of the simulations below the dash lines representing different threshold values of interest to the decision maker.

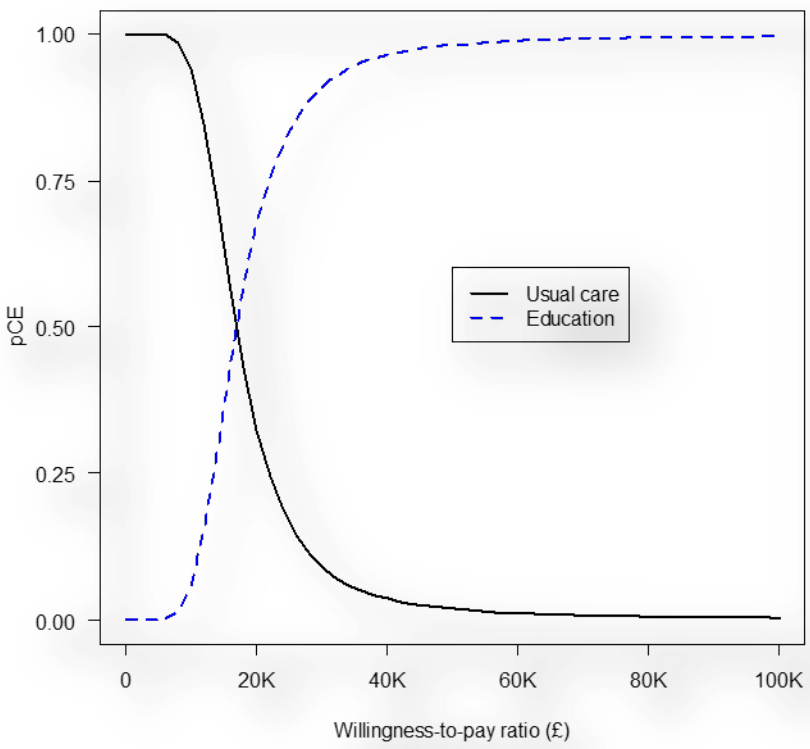


Figure 3.4: An example of CEAC for 2 interventions (20K = £20,000)

Alternatively, the required probabilities can be derived from the Incremental Net Benefit (INB) function which is defined as:

$$INB = \lambda \Delta E - \Delta C > 0 \tag{3.2}$$

The parameters in equation (3.2) have the same definition as in equation (3.1). The probability that the intervention is cost-effective compared to the control intervention at a given λ , denoted as $pCE(\lambda)$ is then given by:

$$pCE(\lambda) = \frac{\text{Number of simulations in which } INB > 0}{\text{Total Number of simulations}} \tag{3.3}$$

Equations (3.2) and (3.3) will be applied to decision analytic models developed in Chapter 8 to calculate $pCE(\lambda)$ and used to plot CEACs for the interventions being evaluated.

3.5 Chapter summary

This Chapter has reviewed the methods for economic evaluation of healthcare interventions within a decision analytic modelling framework. It is the last of the chapters reviewing the methods that will be used to carry out the analyses in the remainder of this thesis. In the remainder of the thesis (specifically Chapters 5 to 8), the ability of new synthesis methodology (including new methods developed in this thesis) to address the challenges of evidence synthesis in a PH context will be demonstrated and critically evaluated with applications to the example problem. Before that, a review of PH appraisals published by NICE is conducted to understand how evidence is currently synthesised in PH evaluation of interventions.

4. SYSTEMATIC REVIEW OF SYNTHESIS METHODS IN PUBLIC HEALTH EVALUATION OF INTERVENTIONS

4.1 Chapter overview

This Chapter presents the outcome of a systematic review of the synthesis and decision modelling methods that were used to inform the development of NICE public health appraisals published between 2006 and 2012. The review was conducted with the aim of identifying the current state of affairs (i.e. what is already done and/or not done) and the barriers to quantitative synthesis of evidence in public health evaluation of interventions. It is anticipated that the review findings will provide the basis for establishing a quality bar to guide application (and where necessary development) of new synthesis methodology with potential to address some of the challenges of evidence synthesis in PH evaluations. The part of the review relating to use of evidence synthesis methods in public health have been published in the *Journal of Clinical Epidemiology* and is given in Appendix VI - Research paper 1(Achana *et al.*, 2014b).

4.2 Introduction

The challenges of evidence synthesis in PH evaluations of interventions were discussed in the introductory chapter. Many of the issues identified there are methodological challenges arising from the heterogeneous nature of PH evidence including studies of different designs, interventions, outcomes and patient populations. Despite these difficulties, there have been increasing calls for PH decision making to be based on the best available evidence whenever possible. For example, a 2004 Department of Health report (Wanless, 2004) on improving health and reducing health inequalities in England called for economic evaluations of PH interventions to ensure judicious use of scarce resources. Following this report, the remit of NICE which already evaluated pharmaceutical interventions, was expanded to include the development of guidance for PH based on sound appraisals of intervention effectiveness and cost-effectiveness (Chalkidou *et al.*, 2008). Consequently, a number of PH appraisals were produced by NICE since 2006 on a wide range of issues including smoking cessation,

alcohol-use, and, particularly of relevance to the example used in the example problem in this thesis, accidents prevention in children.

To help address specific methodological challenges and provide advice on the technical aspects of the appraisal development process, NICE published a methods manual for PH evaluation in 2006 (NICE, 2006), which was subsequently updated in 2009 (NICE, 2009). (A further update was published in September 2012 (NICE, 2012) after this review was completed but the guidance was unchanged). The guidance recommended:

"Meta-analysis may be used to produce a graph if the data (usually from RCTs) are sufficiently homogenous and if there are enough relevant and valid data from comparable (or the same) outcome measures. Where such data are not available, the synthesis may have to be restricted to a narrative overview of individual studies looking at the same question",

"Before pooling or combining the results of different studies, the degree of heterogeneity in the data should be assessed to determine how the results have been affected by the circumstances in which studies were carried out", and

"Publication bias (Dickersin *et al.*, 1987; Sutton *et al.*, 2000) should be critically assessed and reported in the interpretation of the meta-analysis results".

These recommendations match well to the challenges in systematic review/meta-analysis in PH highlighted by the Cochrane Collaboration (Armstrong *et al.*, 2008) and the 2011 Institute of Medicine (IOM) report on standards for systematic reviews (Eden *et al.*, 2012).

In view of the aforementioned challenges facing PH evaluations, and recommendations for synthesis of PH evidence contained in the NICE methods manuals, a review of all NICE PH appraisals published since 2006 was conducted. The conduct and findings of this review forms the subject matter of the remainder of this chapter.

4.3 Systematic review methods

Completed PH appraisals published between 01/03/2006 and 25/09/2012 were identified for inclusion in the review through the NICE website

(http://www.nice.org.uk/Guidance/PHG/Published). Each PH appraisal consisted of a number of articles such as qualitative reviews, epidemiologic reviews, expert opinions; field reports and other similar non-quantitative review reports as well as quantitative systematic reviews of effectiveness and cost-effectiveness, and decision analytic modelling reports. These were retrieved from the 'background information' sections and assessed for eligibility. The 'how this guidance was produced' sections were also searched for relevant articles if none were identified under 'background information'. Articles considered for inclusion in this review were systematic reviews of the quantitative effectiveness and cost-effectiveness evidence and/or decision analytic modelling reports. Qualitative evidence reviews, epidemiologic reviews, field reports, expert opinions and other similar non-quantitative evidence review reports were excluded. In addition, the final appraisal/guidance documents developed for each PH appraisal area were also excluded as these did not contain relevant information on the conduct of the evidence synthesis and decision modelling, which are of interest in this review. All except two (PH1 and PH2) of the appraisals were published after the 2006 NICE methods manual (NICE, 2006) so should have had access to the guidance for quantitative effectiveness evidence synthesis techniques.

Information extracted from the retrieved articles was used to assess the methods used to synthesise the effectiveness evidence and subsequent incorporation of the evidence into the decision models (where developed) that informed the PH appraisal. The assessment criteria for the synthesis methods were:

- Type of systematic review – narrative summary versus meta-analysis;
- Included studies – RCT versus observational (non-randomised) studies;
- Methods used to synthesise the evidence (if undertaken), including specification of the statistical model (including fixed and/or random effects models), heterogeneity and publication bias and the outcome measures used, as well as presentation of results; and
- How evidence from the systematic review was used to inform the cost-effectiveness analysis, if applicable.

Where economic decision models were developed as part of a public health appraisal, the following criteria were used to assess the methodology based on guidelines for good practice in decision analytic modelling (Philips *et al.*, 2004; Philips *et al.*, 2006):

- Model structure – decision tree/Markov processes/other model types;
- Perspective of the base case analysis;
- Interventions under consideration;
- Unit of analysis – individual/household/other;
- Time horizon;
- Quantification (Yes/No) - incorporation of uncertainty in model structure, assumptions and parameters – deterministic /probabilistic sensitivity analysis;
- Discounting (Yes/No);
- Final measures of health outcomes – generic measures such as the QALY and life-year gained or disease specific utility measures.

4.4 Systematic review results

Thirty-nine completed PH appraisals published since 2006 were identified from the NICE PH website. Within these 39 appraisals, 371 potentially relevant articles were retrieved and, after screening the titles and reading the introduction and/or abstract sections, 164 were excluded as they failed to meet the inclusion criteria outlined above. Fifty-two articles, identified as duplicates and supplementary appendices, were combined with the corresponding main report and counted as one article leaving a total of 155 articles for inclusion in this review. No relevant supporting document meeting the inclusion criteria existed for one appraisal (PH36) which evaluated prevention and control of hospital infection. Therefore, the number of included articles per appraisal ranged from 0 to 10 with a median of approximately 4 articles per appraisal.

4.4.1 Type of review

Table 4.1 lists all 39 PH appraisals by summary of the evidence synthesis and cost-effectiveness analyses undertaken to inform each appraisal development. One appraisal (PH36) reported neither effectiveness and cost-effectiveness evidence reviews nor a decision

model, 2 appraisals (PH33 and PH34) reported reviews of evidence but conducted no cost-effectiveness analysis and a 4th appraisal (PH7) reported evidence reviews and decision models but no estimates of cost-effectiveness were presented.

Twenty-nine (74.4 %) of the 39 appraisals contained systematic reviews in which only a narrative summary of the evidence was conducted, another 7 (18 %) conducted both narrative summary and meta-analysis, 2 appraisals (5%) conducted only meta-analysis, and 1 (2.6%) appraisal had no systematic review and hence no evidence synthesis. In the narrative summary approach, the review findings were summarised study by study in the text and through tables. Sometimes forest plots were used to display results of primary studies but no overall mean or pooled result was presented (see PH4 for an example). Eight of the 29 appraisals using only a narrative summary approach did not report the reasons for not pooling the data, 2 included only review level evidence from overview of reviews, and 19 cited heterogeneity as the reason why meta-analysis was not considered appropriate. The reported causes of heterogeneity are presented in Appendix I.

4.4.2 Included studies – RCTs versus non-randomised studies

Two (PH23 and PH38) of the 38 appraisals (containing a systematic review) included evidence from RCTs only in the effectiveness review. The remaining 36 appraisals were informed by reviews of both randomised and observational (non-randomised) evidence identified from individual study reports and/or published systematic review reports. All 38 appraisals (containing a systematic review) graded the quality of primary studies and assessed the applicability of the evidence adhering to the PH appraisal methods guidelines (NICE, 2006; NICE, 2009).

Table 4.1: NICE PH appraisals and summary of evidence synthesis methods and decision modelling used to inform their development

NICE public health appraisal title	Review of the effectiveness and cost-effectiveness evidence and decision analysis for ach appraisal					
	Systematic review of effectiveness (narrative summary)	Systematic review of effectiveness (At least one M-A)†	Cost-effectiveness reviews	Study quality	Decision Model	Source of effectiveness estimate used in decision model‡
Brief interventions and referral for smoking cessation (PH1)	✓	x^{nr}	✓	✓	✓	Published review
Four commonly used methods to increase physical activity (PH2)	✓	x^{nr}	✓	✓	✓	Individual study[1]
Prevention of sexually transmitted infections and under 18 conceptions (PH3)	✓	✓	✓	✓	✓	Published review
Interventions to reduce substance misuse among vulnerable young people (PH4)	✓	x^{o}	✓	✓	✓	Individual study[1]
Workplace interventions to promote smoking cessation (PH5)	✓	x^{m}	✓	✓	✓	Individual study[3]
Behaviour change (PH6)	✓	x^{nr}	✓	✓	✓	Individual study[4]
School-based interventions on alcohol (PH7)	✓	✓	✓	✓	✓	Individual study[1]
Physical activity and the environment (PH8)	✓	$x^{i,m,o}$	✓	✓	✓	Individual study[3]
Community engagement (PH9)	✓	$x^{i,m,o,p}$	✓	✓	x	Not applicable
Smoking cessation services (PH10)	✓	✓	✓	✓	✓	New Meta-analysis
Maternal and child nutrition (PH11)	✓	x^{m}	✓	✓	✓	Individual study[5]
Social and emotional wellbeing in primary education (PH12)	✓	✓	✓	✓	✓	Individual study[5]
Promoting physical activity in the workplace (PH13)	✓	x^{nr}	✓	✓	✓	Individual study[5]
Preventing the uptake of smoking by children and young people (PH14)	✓	$x^{m,o}$	✓	✓	✓	Individual study[1]
Identifying and supporting people most at risk of dying prematurely (PH15)	✓	$x^{i,m,p}$	✓	✓	✓	Individual study[1]
Mental wellbeing and older people (PH16)	✓	$x^{i,m,o}$	✓	✓	✓	Individual study[1]
Promoting physical activity for children and young people (PH17)	✓	$x^{i,m,o}$	✓	✓	✓	Analyst estimate[4]
Needle and syringe programmes (PH18)	✓	x^{nr}	✓	✓	✓	Individual study[3]
Management of long-term sickness and incapacity for work (PH19)	✓	✓	✓	✓	✓	New meta-analysis
Social and emotional wellbeing in secondary education (PH20)	✓	$x^{i,m,o}$	✓	✓	✓	Individual study[2]
Reducing differences in the uptake of immunisations (PH21)	✓	x^{nr}	✓	✓	✓	Individual study[4]
Promoting mental wellbeing at work (PH22)	✓	x^{i}	✓	✓	✓	Individual study[3]
School-based interventions to prevent smoking (PH23)	✓	✓	✓	✓	✓	New Meta-analysis
Alcohol-use disorders - preventing harmful drinking (PH24)	✓	x^{nr}	✓	✓	✓	Published review
Prevention of cardiovascular disease (PH25)	✓	x^{nr}	✓	✓	✓	Individual study[5]
Quitting smoking in pregnancy and following childbirth (PH26)	✓	$x^{i,m,o}$	✓	✓	✓	Published review
Weight management before, during and after pregnancy (PH27)	✓	✓	✓	✓	✓	Not clear[5] reported
Looked-after children and young people (PH28)	✓	$x^{m,o,p}$	✓	✓	✓	Individual study[3]
Strategies to prevent unintentional injuries among under-15s (PH29)	✓	x^{i}	✓	✓	✓	Individual study[3]
Preventing unintentional injuries among under-15s in the home (PH30)	✓	$x^{i,o}$	✓	✓	✓	Individual study[3]
Preventing unintentional road injuries among under-15s: road design (PH31)	✓	x^{i}	✓	✓	✓	Individual study[2]
Skin cancer prevention: information, resources and environmental changes (PH32)	✓	$x^{i,m}$	✓	✓	✓	Individual study[3]

Increasing the uptake of HIV testing among black Africans in England (PH33)	✓	x[m]	✓	x	x	Not applicable
Increasing the uptake of HIV testing among men who have sex with men (PH34)	✓	x[nr]	✓	x	x	Not applicable
Preventing type 2 diabetes - population and community interventions (PH35)	✓	✓	✓	✓	✓	New meta-analysis
Prevention and control of healthcare-associated infections (PH36)		x	x	x	x	Not applicable
Tuberculosis - hard-to-reach groups (PH37)	✓	x[nr]	✓	✓	✓	Individual study[5]
Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38)	✓	✓	✓	✓	✓	New Meta-analysis
Smokeless tobacco cessation - South Asian communities (PH39)	✓	x[s]	✓	✓	✓	Published review

Ticks indicates a systematic review of evidence, meta-analysis or decision model have been conducted whilst x indicate such analyses have not been conducted.

[†]Reported reason why meta-analysis was not done (i=heterogeneity of interventions, m=heterogeneity methods, design and settings, o= heterogeneity of outcome measures, p= heterogeneity of study populations, s= heterogeneity of studies (specific cause not stated), nr= not reported or reported that studies do not support a meta-analysis).

‡Selection of individual study estimate of intervention effect used in decision model (1= used a pre-specified criteria reported in the decision model report, 2=discussion with NICE or estimates selected based on quality grading of evidence using the methods guide manuals, 3= selected studies based on relevance of the intervention to the decision problem, 4= assumption/analyst estimated based on an assumption, 5 =not clearly reported).

4.4.3 Quantitative evidence synthesis

Only 9 of the 39 appraisals (23%) contained one or more systematic review with a meta-analysis (Table 4.2). In total, there were 10 systematic review and/or decision analytic modelling reports with at least one meta-analysis within the 9 appraisal areas (Note: PH10 has two systematic review reports in which a meta-analysis was conducted). Four of the 10 meta-analyses included RCTs only and 6 included both RCT and observational (non-RCT) studies. Six of the 10 meta-analyses were conducted on 'final outcomes'; that is, the main outcome measures on which the corresponding cost-effectiveness analyses were based (e.g. PH10 Smoking abstinence). The remaining 4 meta-analyses were conducted on 'intermediate outcomes' (e.g. PH3 Uptake of Chlamydia screening in schools rather than prevention of Chlamydia).

There was evidence that interventions may have been 'lumped' (Caldwell *et al.*, 2005; Caldwell *et al.*, 2010) into two broad intervention groups to facilitate inclusion of more studies in 7 of the 10 reports with a meta-analysis. For example, in PH23, which investigated the effect of school-based interventions on alcohol consumption, seemingly different interventions (such as lessons delivered by teachers or other professionals as part of the curriculum; peer led education by other pupils; external contributions from, for example, the police, life education centre staff; and implementation of school policy type interventions) were lumped together to form one 'intervention group' which was then compared to the no intervention control in a pairwise meta-analysis.

Seven of the 10 review reports conducted random effects pairwise meta-analysis, one conducted both fixed and random effects analysis, one conducted a random effects mixed treatment comparison (described in Section 2.3) alongside the pairwise analysis and another one did not clearly present the statistical model used. Six of the 10 systematic reviews presented forest plots with heterogeneity statistics displayed on them, 2 (PH3 and PH1) presented forest plots without heterogeneity statistics and one review (PH35) did not present a forest plot. Only one review (PH23) assessed publication bias using funnel plot and Egger's test for asymmetry.

Table 4.2: Review of quantitative methods used to synthesise public health evidence for NICE public health appraisal

Appraisal title	Systematic review report title	Included RCTs only	Main outcome	Description of main outcome	Outcome measure: statistic	Type of synthesis	Model type	Lumping[1] of interventions	Presentation of results	Assessed publication bias	Software	Used result of M-A in decision model
Prevention of sexually transmitted infections and under 18 conceptions (PH3)	Review 2 - Review of evidence for the effectiveness of screening for genital chlamydial infection in sexually active young women and men	No	I	Uptake of proactive chlamydia screening using home-collected specimens	Rate (%)	M-A	Random effects	No	FP/Txt	No	RevMan, Stata	No
School-based interventions on alcohol (PH7)	Alcohol and schools: effectiveness and cost-effectiveness review	No	F	Alcohol use	WMD	M-A	Random effects	Yes	FP/Txt	No	Not stated	No
Smoking cessation services (PH10)	Cut down to quit' with nicotine replacement therapies	Yes	F	6 or more months' sustained abstinence	RR and Cohen's d	M-A	Random effects	Yes	FP/T/ Txt	No	RevMan	Yes
Smoking cessation services (PH10)	Final report	No	F	6 or months' sustained abstinence	Cohen's d	M-A	Fixed & random effects	Yes	FP/T/ Txt	No	RevMan	No
Social and emotional wellbeing in primary education (PH12)	Teesside review	Yes	I	Social problem solving	SMD	M-A	Random effects	Yes	FP/T	No	RevMan	No
Management of long-term sickness and incapacity for work (PH19)	PH19 Management of long-term sickness and incapacity for work: Economic analysis report	No	Y	Number returning to work following sickness	RR	M-A	Random	Yes	FP/T/Txt	No	Revman	Yes
School-based interventions to prevent smoking (PH23)	School-based interventions to prevent smoking: quantitative effectiveness review	Yes	F	smoking uptake	OR	M-A	Random effects	Yes	FP/Txt	Yes	Stata	Yes
Weight management before, during and after pregnancy (PH27)	Weight management before, during and after pregnancy: evidence review	No	I	Number exceeding IoM[2] guidelines for healthy weight gain	RR	M-A	Random effects	Yes	FP/T/ Txt	No	RevMan	No

Preventing type 2 diabetes - population and community interventions (PH35)	PH35 Preventing type 2 diabetes - population and community interventions: report on cost-effectiveness evidence and methods for economic modelling	No	I	Body mass index	WMD	M-A	Not reported	Yes	T/Txt	No	Not reported	Yes
Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38)	Prevention of type 2 diabetes: systematic review & meta-analysis of lifestyle, pharmacological and surgical interventions	Yes	F	Reduce progress to diabetes for people with IGT	HR	M-A & NMA	Random effects	No	FP/TxT	No	RevMan (M-A) WinBUGS (NMA)	Yes

Presentation of results (FP = Forest plot, T=Table, Txt=Text), M-A = pairwise meta-analysis, NMA = network meta-analysis
Main outcomes (I = intermediate outcome, F= Final outcome)
Outcome measure (RR = Risk ratio, OR= Odds ratio, SMD = Standardised mean difference, WMD = Weighted mean difference)
1 = lumping is a term used in the literature(Caldwell *et al.*, 2005; Caldwell *et al.*, 2010) to described the tendency to aggregate or treat seemingly similar but disparate /different interventions as one intervention group in order for example to facilitate inclusion of many studies in a meta-analysis. A classic example is treating different doses of a drug as if they were the same treatment
2 = American Institute of Medicine (IOM) Guidelines on Weight Management in Pregnancy.

4.4.4 How evidence from systematic reviews were incorporated into the model

Thirty-five (89.7%) of the 39 appraisals were informed by cost-effectiveness evaluations contained in one or more decision analytic modelling reports (Table 4.1). Twenty-three (66%) of these used estimates of intervention effectiveness derived from individual studies identified in the systematic review to inform the decision analysis (reasons for using the studies selected are given in Table 4.1), 5 (14%) used previously published systematic review results, another 5 (14%) used estimates from a meta-analysis of studies identified in the systematic review, 1 used expert opinion/analyst estimate and another one did not clearly report the source(s) of the intervention effect.

4.4.5 Decision analytic modelling methods

The characteristics of the economic decision models identified in the review for each appraisal are summarised in Figure 4.1. A detailed summary of the models developed for each appraisal are given in Table 4.3. Five of the 39 appraisals (PH9, PH32, PH33, PH34 and PH36) reported no economic decision analysis, 35 presented one economic decision model, one appraisal (PH21) presented two models, and another appraisal (PH32) presented 3 models.

Thirty-one of the 35 appraisals with at least one decision model, had either decision tree (n=16), Markov (n=13) or a combination of decision tree and Markov (n=2) structures. Of the remaining 4 appraisals, 2 were infectious disease models (PH18 and PH21) whilst another 2 appraisals did not clearly report the model structure (PH25 and PH26).

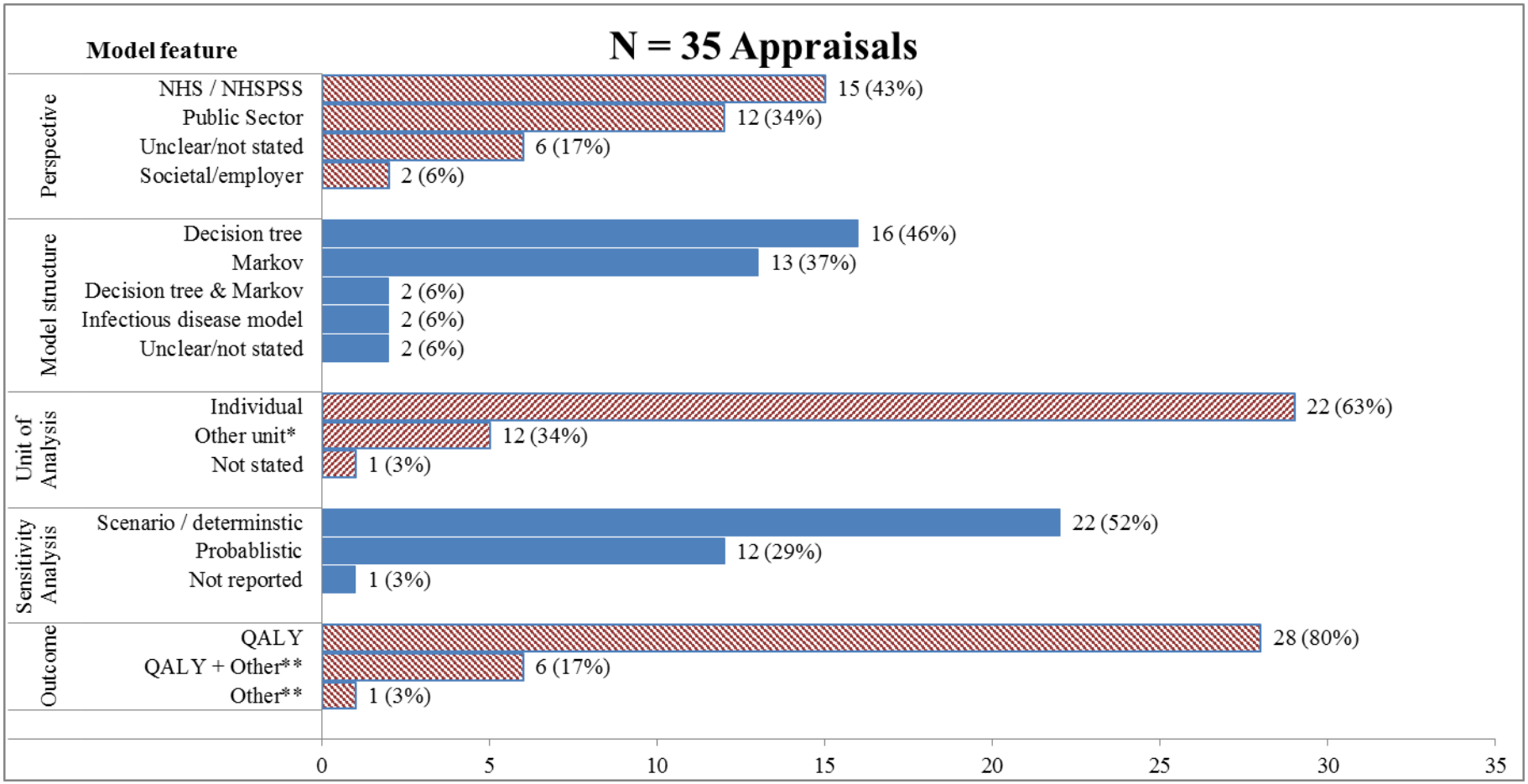


Figure 4.1: Review of Decision analytic modelling methods in public health
***Other unit of analysis refers to 'traveller community' in PH 19 and 'household' in PH29 and PH30,**
****Other measures of health (cost per case averted, net present value or life years gained).**

Table 4.3: Review of decision analytic models evaluating the cost-effectiveness of interventions for NICE public health appraisals

		Review of the Decision analytic modelling methods								
Year	Appraisal reference and title	Decision model	Perspective Base case	Decision tree	Markov / Other	Proba-bilistic	Unit of analysis[5]	Time horizon	SA[6]	Health outcome[7]
2006	PH1(Brief interventions and referral for smoking cessation)	✓	NHSPSS	✓			I	Lifetime	U,T	QALY
	PH2 (Four commonly used methods to increase physical activity)	✓	NHSPSS	✓			I	Lifetime	U	QALY
2007	PH3 (Prevention of sexually transmitted infections and under 18 conceptions)	✓	NHS	✓			I	1 year	T	QALY
	PH4 (Interventions to reduce substance misuse among vulnerable young people)	✓	Societal	✓			I	Lifetime	U	QALY
	PH5 (Workplace interventions to promote smoking cessation)	✓	NHS		✓		I	Lifetime	M	QALY
	PH6 (Behaviour change)	✓	NHS		✓		I	20 years	U,M,T	QALY
	PH7 (School-based interventions on alcohol)	✓	NHS	✓			I	NS	None	C/C
2008	PH8 (Physical activity and the environment)	✓	NHS	✓			I	Lifetime	U	QALY
	PH9 (Community engagement)	x								
	PH10 (Smoking cessation services)	✓	Not clear		✓		I	Lifetime	U	QALY
	PH11 (Maternal and child nutrition)	✓	None	✓			I	Lifetime	U	QALY
	PH12 (Social and emotional wellbeing in primary education)	✓	None	✓		✓	I	3 years	P	QALY
	PH13 (Promoting physical activity in the workplace)	✓	Employer	✓			I	Lifetime	U	QALY
	PH14 (Preventing the uptake of smoking by children and young people)	✓	NHS		✓		I	Lifetime	U	QALY,LY
	PH15 (Identifying and supporting people most at risk of dying prematurely)	✓	PS	✓			I	20years	U	QALY
	PH16 (Mental wellbeing and older people)	✓	PS	✓			I	1year	U, T	QALY
2009	PH17 (Promoting physical activity for children and young people)	✓	PS	✓			I	NA	U,T	QALY
	PH18 (Needle and syringe programmes)	✓	NHSPSS		IDM[3]	✓	I	20 years	U,T,S	QALY
	PH19 (Management of long-term sickness and incapacity for work)	✓	NHSPSS		✓		Tra[4]	Lifetime	U,T,P	QALY
	PH20 (Social and emotional wellbeing in secondary education)	✓	PS		✓	✓	I	Lifetime	U,M,T,P	QALY
	PH21 (Reducing differences in the uptake of immunisations)[2]	✓	NHSPSS		IDM	✓	I	Lifetime	None	QALY
	PH22 (Promoting mental wellbeing at work)	✓	NHS		✓	✓	I	Lifetime	None	QALY
2010	PH23 (School-based interventions to prevent smoking)	✓	PS	✓			I	Lifetime	U,T	QALY
	PH24 (Alcohol-use disorders - preventing harmful drinking)	✓	NHSPSS		✓	✓	I	NS	U,P	LY,QALY
	PH25 (Prevention of cardiovascular disease)	✓	None	NS	NS	NS	NS	10 years	None	QALY,LY
	PH26 (Quitting smoking in pregnancy and following childbirth)	✓	None	NS	NS		I	Lifetime	U	QALY

	PH27 (Weight management before, during and after pregnancy)	✓	NHSPSS	✓	✓	✓	I	15 years	U,P	QALY
	PH28 (Looked-after children and young people)	✓	NHSPSS		✓	✓	I	Lifetime	U,P	QALY
	PH29 (Strategies to prevent unintentional injuries among under-15s)	✓	PS	✓			H	30 years	U	QALY,NPV
	PH30 (Preventing unintentional injuries among under-15s in the home)	✓	PS	✓	✓	✓	H	100 years	U,P	QALY
	PH31 (Preventing unintentional road injuries among under-15s: road design)	✓	PS	✓		✓	I	95 years	U,P	QALY,NPV
2011	PH32 (Skin cancer prevention: information, resources & environmental changes)[2]	✓	PS	✓		✓	I	Lifetime	U, P	QALY
	PH33 (Increasing the uptake of HIV testing among black Africans in England)	x								
	PH34 (Increasing the uptake of HIV testing among men who have sex with men)	x								
	PH35 (Preventing type 2 diabetes - population and community interventions)	✓	PS		✓		I	80 years	U	QALY
	PH36 (Prevention and control of healthcare-associated infections)	x								
2012	PH37 (Tuberculosis - hard-to-reach groups)	✓	PS		✓		I	20 years	T	QALY
	PH38 (Preventing type 2 diabetes - risk identification and interventions for individuals at high risk)	✓	PS		✓	✓	I	Lifetime	P	QALY, C/C
	PH39 (Smokeless tobacco cessation - South Asian communities)	✓	Not clear		✓		I	Lifetime	U	QALY

1=No cost-effectiveness analysis presented;
2=More than 1 models presented
3=IDM=Infectious disease model;
4=traveller community;
5=Unit of analysis (I=individual; H=household; stated given if other);
6=Sensitivity analysis (U=unidirectional, M=multidirectional, P=probabilistic, T=threshold);
7=outcomes (QALY = quality adjusted life year; C/C = cost per cases averted; LY= life year; NPV = net present value)

Other abbreviations
D = Deterministic model;
P=Probabilistic model;
NS=Not stated;
PS = Public sector

Infectious disease modelling was used to evaluate the cost-effectiveness of interventions linked to needle and syringe programmes for injecting drug users (PH18) and of interventions to reduce differences in the uptake of childhood immunisations (PH21).

Twenty-two (54%) of the 35 appraisals were informed by a deterministic decision analysis, 12 (29%) were probabilistic and one appraisal (3%) reported no sensitivity analysis. The most preferred measure of health outcomes was the QALY with 28 of the 35 appraisals (80%) using only QALYs to measure cost-effectiveness, 6 (17%) using both QALYs and Life Years (LY) gained or Net Present Value (NPV) and one appraisal (3%) based cost-effectiveness on Cases averted.

4.5 Discussion of systematic review findings

This review of completed NICE public health appraisals illustrates the current situation regarding the use of evidence synthesis methods to inform public health decision making in the UK. It identified that effectiveness evidence was mostly synthesised using narrative summaries and that quantitative synthesis was not carried out for the majority of evaluations in PH systematic reviews. Of the 39 appraisals published since 2006, only 9 (23%) were informed by at least one systematic review with a meta-analysis. The other 30 appraisals may have refrained from meta-analysis due to a lack of randomized trials, or heterogeneity in study design, i.e. a mix of RCTs and non-RCTs. Moreover, systematic reviews opting for a quantitative summary tended to use the simplest methods such as fixed or random effects pairwise meta-analysis which only enables comparison between two interventions at any one time and thus potentially limiting the scope of the analysis and the utility of the findings. These findings would seem to indicate that, despite great advances in quantitative synthesis techniques, application in PH evaluation is still very much in its infancy and appears to lag behind other areas of healthcare such as the evaluation of clinical interventions. There are several reasons for this, including the heterogeneous nature of PH evidence arising from i) variations in many aspects of study design, ii) the exact nature of the interventions; iii) varying/differential outcome measures; iv) the wider scope of many PH research questions; and v) the quantitative skills of the researchers including familiarity with Bayesian software and modelling techniques (Mills *et al.*, 2011).

The review found nearly 80% of PH NICE appraisals did not attempt a quantitative synthesis at all due to what investigators believe are insurmountable problems owing to the heterogeneous nature of the evidence base. NICE guidance states that ‘‘Meta-analysis may be used to produce a graph if the data (usually from RCTs) is sufficiently homogenous’’ (Section 5.4.4.2 of NICE guidance 2012). Often the evidence from RCTs is limited and the best available evidence is from the non-RCTs. However, provided reviewers quality assess non-RCTs (as they would RCTs) to identify well-conducted studies, to limit confounding by selection bias, then meta-analysis can be considered. In addition, exploring heterogeneity and attempting to account for it should be part of the analysis and greater awareness of modern methods, and greater expertise in using them, will yield fruit for future PH reports. Underlying this desire for PH reviews to become more quantitative, in the face of the challenges encountered, is a firm belief that a structured and transparent description and analysis of the decision question is desirable. However, there are several other reasons why conducting a meta-analysis may not be advisable; for example: (i) a small number of studies may mean that statistical heterogeneity is underestimated; (ii) some studies are too biased to draw a valid conclusion from them; (iii) there is evidence of publication bias; and (iv) incomplete or selective reporting of outcomes.

This review is limited to only considering NICE PH appraisals in the review and does not claim to have all the answers to all evidence synthesis challenges that exist in PH evaluation. For example, none of the above considers directly the influence of the study quality/validity of the individual studies going into an analysis – although others are doing work in other contexts that could be adapted, for example including different, both observational and randomised, evidence (Turner *et al.*, 2009). Public health evaluations are notoriously messy and complex, with many factors to consider. But if a decision has to be made, explicit, transparent and appropriate analysis of the data should be preferred to current alternatives.

4.6 Chapter summary

This chapter reviewed completed published NICE public health appraisals to illustrate the current situation regarding the use of evidence synthesis methods to inform public health

decision making in the UK. There was evidence from the review that effectiveness evidence was mostly synthesised using narrative summaries and that quantitative synthesis was not carried out for the majority of evaluations in PH systematic reviews. In the next three chapters, more complex synthesis models will be presented that offer the opportunity to model the types of data commonly available in PH appraisals more appropriately rather than carrying out less focused and detailed reviews of the literature.

5. EFFECTIVENESS OF INTERVENTIONS TO INCREASE UPTAKE OF POISONING PREVENTION MEASURES IN HOUSEHOLDS WITH CHILDREN: NETWORK META-ANALYSES

5.1 Chapter overview

There is evidence from previous systematic reviews and meta-analyses (DiGuiseppi *et al.*, 2001; Kendrick *et al.*, 2012c) that home safety interventions are effective in promoting a range of poison prevention behaviours in households with children. However, these two meta-analyses compared any intervention against a "usual care or no intervention" which potentially limits the usefulness of the results in a cost-effectiveness evaluation. In this chapter, network meta-analysis will be used to re-analyse the data from the two systematic reviews and evaluate the effectiveness of different poison prevention strategies in households with children. The effectiveness estimates from the analysis presented in this chapter will be used to inform a decision analytic model developed in chapter 8 in order to determine the most cost-effective strategy for preventing unintentional poisoning injuries in children under 5 years of age. The analyses and results presented here have been submitted to the Journal *PLOS ONE* and the submission is currently undergoing peer-review (Appendix VI – Research paper 2).

5.2 Effectiveness evidence

5.2.1 Data sources

The main source of effectiveness evidence is a Cochrane systematic review of home safety education and provision of safety equipment for injury prevention in children. This review was first published in 2007 (Kendrick *et al.*, 2007) and updated in 2012 (Kendrick *et al.*, 2012b). The remaining evidence came from an overview of reviews (Young *et al.*, 2013) conducted to identify newly published studies and studies which did not meet the inclusion criteria for the Cochrane review but judged suitable for inclusion in the overview of reviews (Young *et al.*, 2013) and the analyses described in this chapter (e.g. studies which compared

2 active interventions). Both reviews included randomised controlled trials (RCTs), controlled before-and-after studies, and other non-randomised study designs if they reported home safety interventions of interest.

5.2.2 Poisoning prevention outcomes

The Cochrane review identified 6 poisoning prevention practices or behaviours in the literature, of which 5 (safe storage of medicines, safe storage of other household products, safe storage of poisons, safe storage of poisonous plants and possession of poison control centre (PCC) telephone number) are included in the analysis reported here. Possession of syrup of ipecac was excluded (together with studies reporting only this outcome) because its use is no longer recommended (Benson *et al.*, 2013; Hojer *et al.*, 2013). Safe storage of poisons refers to the combined endpoint of medicinal and non-medicinal substances in studies that reported just one combined outcome for storage of medicinal and non-medicinal substances instead of two separate outcomes. Safe storage was defined in the Cochrane review and subsequent related publications as storing potentially toxic substances (medicinal or non-medicinal) at adult eye level and/or in locked cupboards/drawers/cabinets where they are inaccessible to children (Kendrick *et al.*, 2008).

5.2.3 Description of included studies

Twenty-four primary studies were identified from the 2012 review and 3 from the overview of reviews. One study (Minkovitz *et al.*, 2003) divided patients into a randomised and a quasi-randomised study groups and analysed the two groups separately. This study was included in the analysis reported here as two separate studies, thus increasing the total number to 28. Of the 3 studies not included in the Cochrane review, one (Reich *et al.*, 2011) had been published since the Cochrane review and the other two (Minkovitz *et al.*, 2003; Johnston *et al.*, 2006) reported an intervention (i.e. healthy steps for young children program) that was not considered suitable to be included in either of the two intervention groups considered in the Cochrane review. The healthy steps program was considered to be different from the generic safety education considered in the other studies because it provided intense parental training in managing all aspects of a child's development, including safety education. The advantage of the analysis reported in this chapter is that this intervention can

be evaluated together with the other interventions by including it as a separate strategy in the network meta-analysis.

Table 5.1 presents summary characteristics of the 28 studies. One study (Babul *et al.*, 2007) had 3 intervention arms (i.e. 3-arm study). The remaining 27 studies were either two-arm studies or were converted into two-arm if two or more arms were considered similar enough to be combined into 2-arm studies. For example, Nansel *et al.* (Nansel *et al.*, 2008) is a 3-arm study but two of the arms (i.e. the tailored injury prevention information arm and the tailored injury prevention plus tailored summary information for providers arm) were combined be combined as one intervention in the analyses reported here. Twenty-one (75%) of the 28 studies were conducted in the USA, 5 (18%) in Europe (of which 3 were conducted in the UK), 1 study in Canada and another in South Africa. Nineteen (68%) were RCT and 9 (32%) were observational non-RCT studies. There were 10 (35%) studies with a cluster design, of which 5 were randomised and 5 non-randomised studies.

Table 5.1: Summary of studies evaluating poison prevention measures in households with children

Comparison	Study	Study quality	List of free and or low cost equipment provided as part of intervention scheme	Safe storage of medicines	Safe storage of other household products	Safe storage of poisons	Safe storage of poisonous plants	Possession of pcc[2] telephone number
Usual care (1) vs. Education (2)	Kelly (1987)[a], RCT, USA	A=U,B=Y,F=N	Not applicable	54/54 55/55	43/54 49/55			
	Nansel (2002)[b], RCT, USA	A=Y,B=U,F=Y	Not applicable	83/89 79/85	65/89 66/85			59/89 63/85
	Kelly (2003)[c], Cluster-RCT, USA	A=U,B=Y,F=Y	Not applicable					45.56/136.68[c] 112.95/137.63[c]
	McDonald (2005), RCT, USA	A=Y,B=U,F=N	Not applicable	6/60 4/57	3/57 6/61			
	Gielen (2007), RCT, USA	A=Y,B=N,F=Y	Not applicable	178/271 188/249	44/62 57/73	222/333 245/322		
	Nansel (2008), Non-RCT, USA	A=U,B=N,F=N	Not applicable	72/74 140/144	59/73 117/144			50/59 90/119
	Reich (2011)[c], RCT, USA	A=U,B=N,F=N	Not applicable			Log-OR(SE) = -0.192(0.2863) [d]		
Equipment only (1) vs. Education + Equipment (3)	Woolf (1987), Cluster-RCT, USA	A=U,B=Y,F=N	Sticker with poison control centre telephone number, bottle of ipecac					29/143 47/119
	Woolf (1992), Cluster-RCT, USA	A=U,B=Y,F=N	Kitchen cabinet locks, a coupon for purchase of syrup of ipecac, and two telephone stickers with the telephone number of the poison centre.		60/151 89/150			59/151 117/150
	Clamp (1998), RCT, UK	A=U,B=N,F=Y	Smoke alarm, 2 window locks, 3 cupboard locks, 6 socket covers and a door slam device.	68/82 79/83	49/82 59/83			
Usual care (1) vs. Education + Equipment (3) vs. Education + Equipment + Home Safety inspection (4)	Babul (2007), RCT, Canada	A=Y,B=N,F=N	Smoke alarm, a coupon for 50% savings on a safety gate, corner cushions, cabinet locks, blind cord windups, water temperature card, doorstoppers, electrical outlet covers and a poison control sticker.	147/149 171/173 160/163			112/147 136/172 123/160	
Usual care (1) vs. Education + Equipment + Home Safety inspection (4)	Kendrick (1999), Cluster non-RCT, UK	B=N,F=N,C=Y	Stair gates, fireguards, cupboard locks and smoke alarms.		317/367 322/363			
	Sangvai (2007), RCT, USA	A=Y,B=Y,F=N	Smoke detectors, gun locks, cabinet locks and water temperature cards.			3/10 13/16		
	Swart (2008), Non RCT, South Africa	A=U,B=Y,F=Y	Child-proof locks and paraffin container safety caps.	70.26/79.58[c] 74.07/80[c]	46.86/57.96[c] 50.87/58.27[c]			
	Hendrickson (2002), USA, RCT	A=N,B=N,F=Y	Full publication or report not available to extract information on equipment provision.		14/40 34/38			8/40 34/38
Usual care (1) vs. Education + Equipment (3)	Watson (2005), Cluster-RCT, UK	A=Y,B=N,F=Y	Stair gates, fire guards, smoke alarms, cupboard locks and window locks.	683/738 712/762	327/669 368/693			

Usual care (1) vs. Education + Home Safety inspection (6)	Petridou (1997), Cluster non-RCT, Greece	B=N,F=Y,C=Y						67.26/100.12[c] 71.08/97.83[c]
Usual care (1) vs. Education + Equipment + Home Safety inspection + Installation (7)	Schwarz (1993), Cluster non-RCT, USA	B=N,F=N,C=Y	Smoke detectors, batteries, bathwater thermometer, nightlight, a bottle of syrup of ipecac, a sticker for the telephone with emergency telephone numbers.	88.42/248.37[c] 128.16/248.37[c]				
	Phelan (2011), RCT, USA	A=Y,B=N,F=Y	Full publication or report not available to extract information on equipment provision.			17/149 2/150		16/138 71/139
Usual care (1) vs. Education + Home visit (8)	Minkovitz (2003a)[e], RCT, USA	A=Y,B=N,F=Y	Not applicable			463/761 523/832		
	Minkovitz (2003b)[e], Cluster non-RCT, USA	B=N,F=Y,C=Y	Not applicable			596/955 754/1189		
	Johnston (2006), non-RCT, USA	B=N,F=Y,C=Y	Not applicable			155/232 71/91		82/91 222/232
Education (2) vs. Education + Equipment (3)	Posner (2004), RCT, USA	A=Y,B=Y,F=N	Not applicable	14/47 19/49	22/47 34/49		9/16 11/16	27/47 35/49
	Bulzachelli (2009), Non-RCT, USA	A=U,B=N,F=N	Not applicable			5/49 10/105		
Education (2) vs. Education + Equipment (5)	Sznajder (2003), RCT, France	A=Y,B=N,F=Y	Cupboard and drawer latches, door handle covers, table protection corners, electric outlet covers, a non-skid bathtub mat, a smoke detector, and a phone sticker with the number of the poison control centre.	44/49 43/45	32/41 40/48		48/49 41/48	
Education+ equipment (3) vs. Education + Equipment + Home Safety inspection (4)	Gielen (2002)[c], Cluster-RCT, USA	A=U,B=U,F=N	Safety products (e.g., ipecac syrup, cabinet latches, safety gates, smoke alarms, batteries, and hot water thermometers) are sold at 10% to 15% below retail cost in a homelike environment where their use can be demonstrated.			6.87/56.93[c] 5.89/58.89[c]		
Education+ equipment (3) vs. Equipment only (9)	Dershewitz (1977), RCT, USA,	A=U,B=Y,F=N	Electric outlet covers and three kindergards, which are easily installed plastic locking devices intended to prevent children from getting into cabinets.	22/102 20/104	1/101 0/104			
Education + Equipment + home Safety inspection (4) vs. Education + equipment + home safety inspection + Fitting (7)	King (2001), RCT, USA	A=Y,B=Y,F=Y	Coupons from a national retail store for a $10 discount per item (to a maximum of $50) when purchasing recommended safety devices.		261/469 273/482			

Abbreviations: A = adequate allocation concealment; B = blinded outcome assessment; C, prevalence of confounders does not differ by more than 10% between treatment arms; CBA, controlled before-and-after study;

F = at least 80% participants followed up in each arm; NMA, network meta-analysis; RCT, randomized clinical trial; U = unclear; Y= yes.
[1]Figures are number of events/total number households in the intervention with lowest code followed by the intervention with the highest code
[2]PCC = Poison control centre
[a] Study was excluded from analysis for safe storage of medicines because both treatment and control arms reported 100% event rate.
[b] Two intervention arms were combined (tailored advice and tailored advice + care provider feedback)
[c] Figures adjusted for the effect of clustering using ICC and method reported Kendrick et al (2012)
[d] Combined from two log-odds ratios for Education book vs. No Book (OR=0.80, SE=0.41) and Education Book vs. Non-Education Book (OR=0.85, SE=0.40) reported in Reich et al. (2011)
[e] Minkovitz (2003) included as two separate studies (reason given in the results section)

5.2.4 Classification of interventions

The Cochrane review compared a generic home safety intervention versus no intervention using pairwise meta-analysis methods. Interventions trialled by individual studies were therefore placed in an intervention and a control group to facilitate inclusion of studies in the analysis. The net effect of this was that seemingly similar but different strategies were grouped together as one intervention, potentially increasing the heterogeneity in intervention definition across studies. In addition, studies reporting interventions that could not be fitted into one of the two groups above (for example Minkovitz and Johnson et al.'s studies) were excluded from the Cochrane systematic review analysis. To minimise the heterogeneity and include all the relevant evidence available for the NMA reported here, the Cochrane review interventions were reclassified into 9 relatively homogenous intervention packages as follows:

1) Usual care - including usual safety education or no education (UC)
2) Education - more than usual safety education (E)
3) Education + provision of free/low cost equipment (E+FE)
4) Education + provision of free/low cost equipment + home safety inspection (E+FE+HSI)
5) Education + provision of free/ low cost equipment + fitting (E+FE+F)
6) Education + home safety inspection (E+HSI)
7) Education + free/low cost equipment + home safety inspection + fitting (E+FE+HSI+F)
8) Education + home visit - Healthy Steps for Young Children program (E+HV) and
9) Free/low cost equipment only (FE).

Usual safety education as included in the intervention (1) above is defined to include the level of standard safety education for all injury types (scalds, falls and poisonings) in the home and is not just limited to the prevention of poisonings. Figure 5.1 displays the network diagrams (one for each of the 5 outcomes listed in Section 5.2.2) showing the comparisons between interventions made by the individual studies. The network diagrams give a visual description of the evidence available for each outcome. The oval circles represent interventions with lines linking any two interventions indicating treatment pairs that have been compared head-to-head in at least one study. The number on each line shows the number of such pairwise comparisons. Interventions that have not directly been compared in a study are therefore not linked directly. For example, there are 5 head-to-head studies comparing usual care versus education in the safe storage of medicines network (Panel A) but there is no direct line

linking usual care to low cost/free equipment, since there were no direct comparisons between these interventions.

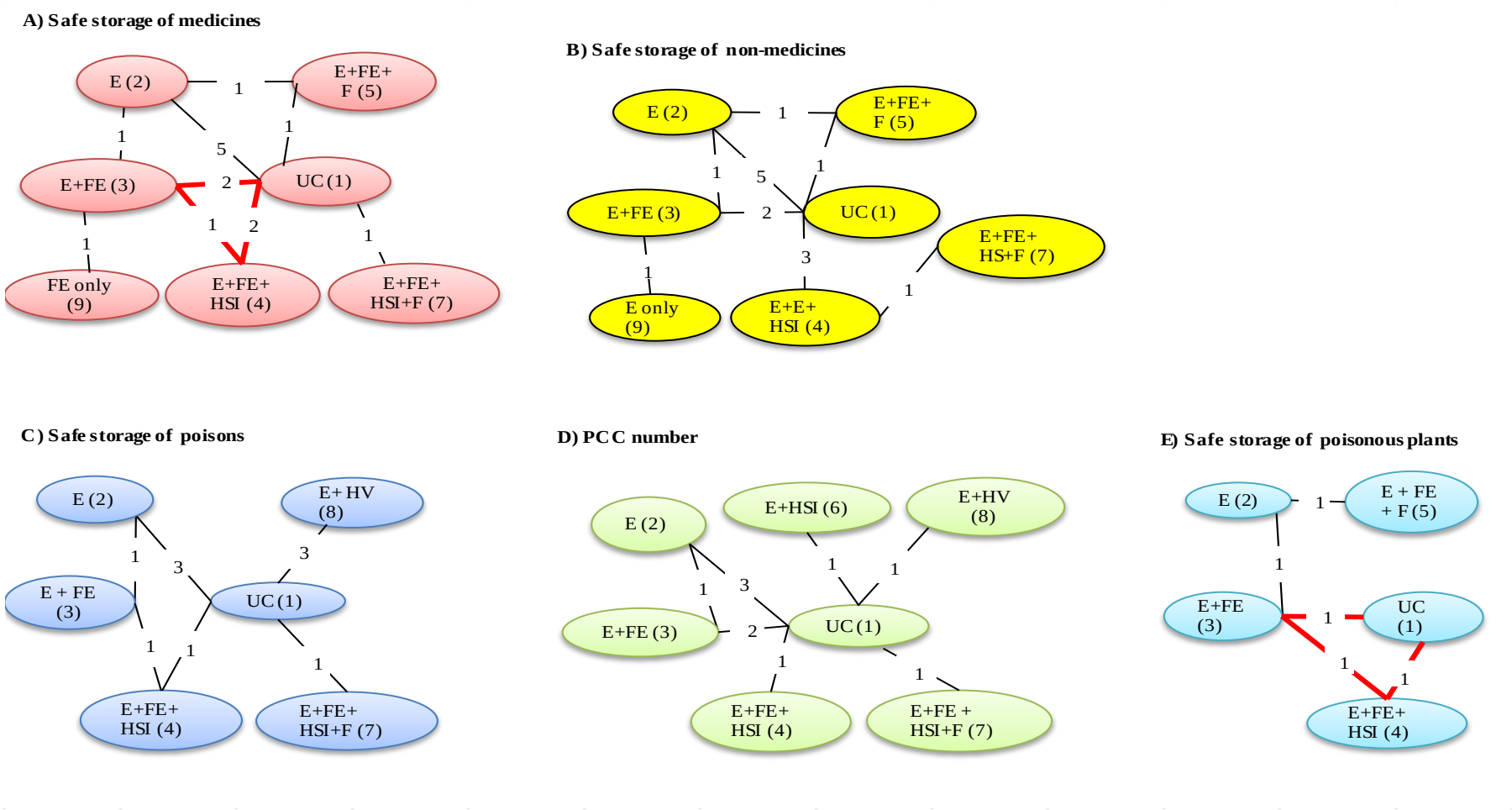


Figure 5.1: Network diagrams for poison prevention outcomes
PCC = poison centre control number. Nodes/oval circles represent intervention with the intervention number in brackets. E = education, F = Fitting, FE = low cost/free equipment, HSI = Home safety inspection, HV = Home visit

Also, the 3-arm Babul study (Babul *et al.*, 2007) which compared usual care versus 'education plus low cost/free equipment' versus 'education plus low cost/free equipment and home safety inspection' is indicated by the bold red lines in the networks for the safe storage of medicines (panel A) and poisonous plants (panel E).

5.3 Methods

5.3.1 Binary outcome data

All the poisoning prevention outcomes described above are binary as they consider whether or not a household safely stores (i.e. out of reach of children) potentially toxic substances (e.g. medicines, other household products, etc.). The data available on each outcome are the number of events (e.g. the number of households with safe storage of medicines) and the total number of households in the intervention and control arms of studies that reported the respective outcome (Table 5.1). The analysis reported in this chapter evaluates the effectiveness of the interventions for each of these outcomes separately using existing synthesis methodology described in Chapter 2 (i.e. univariate pairwise and network meta-analysis models without taking into account the potential correlation between outcomes measures). In Chapter 7, new synthesis methodology will be developed as part of this thesis and used to combine evidence on multiple interventions across multiple poison prevention measures in order to account for correlations between the effect estimates on different outcomes.

5.3.2 Standard pairwise meta-analysis

Standard pairwise meta-analyses were used to compare interventions if head-to-head data from two or more studies was available. Effect sizes were calculated from the available data and pooled across studies inverse variance weighting methods as described in Chapter 2 Section 2.2.1. Both fixed and random effects models were fitted. However, due to the difficulty of obtaining reliable estimate of the between-study variance when the number of studies is small, only fixed effect models were fitted if fewer than 5 studies were available for the meta-analysis. The models were fitted in Stata version 12 (Stata Corporation, 2012) using the *metan* command. Pooled estimates of the intervention effects are presented as odds ratios (OR) and 95% confidence intervals. Statistical heterogeneity was assessed using the χ^2-test

and quantified using the I^2-statistic and the between-study variance, σ^2, which was defined in equation (2.3) of Chapter 2.

5.3.3 Network meta-analysis

The NMA models were fitted using the method of Lu and Ades (Lu and Ades, 2004; Caldwell *et al.*, 2005) with usual care intervention taken as the reference or baseline intervention in all the networks. The model equation is given by equation (2.10) of Chapter 2. Analyses were conducted within a Bayesian framework using Markov Chain Monte Carlo (MCMC) simulations in WinBUGS version 1.4.3 (Lunn *et al.*, 2000). The WinBUGS code was taken from the NICE Technical Support Document 2 (Dias *et al.*, 2011a). Initially, as suggested in this technical document, minimally informative prior distributions were specified for parameters of the model as indicated in Chapter 2 and repeated here for clarity:

$$d_{Ak}, \mu_{ib} \sim \text{Normal}\left(0, 10^3\right) \text{ and } \sigma \sim \text{Uniform}\left(0, 2\right)$$

To investigate the effect these priors have on parameter estimates, sensitivity analyses were conducted using alternative 'vague' prior distributions as presented later in Section 5.6.2. Parameter estimation was based on the posterior median and the 95% credible intervals. These were obtained after running 3 MCMC chains for 40 000 iterations using different starting values. The first 10 000 iterations were discarded as 'burn-in' samples to ensure that the starting values did not influence the samples on which inference is based (Spiegelhalter *et al.*, 2007). For each parameter of interest (namely, the pooled effect of interventions relative to one another and the between-study standard deviation), the history, density and Brooks-Gelman diagnostic plots (Brooks and Gelman, 1998) were examined for evidence that the MCMC simulations were stable and convergence was satisfactory. The probability each intervention was the best and the relative rankings of interventions were obtained as described in Section 2.3.2.

5.3.4 Convergence, goodness- of-fit and model selection

As described in Section 2.3.5 of Chapter 2, convergence of the MCMC samples were assessed by examining the history, kennel density, autocorrelation and Brooks-Gelman diagnostic plots available from the WinBUGS menu. The posterior mean residual deviance

was used to assess how well the model predictions fit the observed data (McCullagh and Nelder, 1989; Spiegelhalter *et al.*, 2002). Under the null hypothesis that the model fits the data well, the posterior mean residual deviance is expected to be approximately equal the number of unconstrained data points. Therefore models were judged to provide adequate fit if the residual deviance is close to the number of data points in the model. The Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) was used to discriminate between the fit of alternative model candidates (for example, fixed effect versus random effects model) and to select the best fitting model.

5.3.5 Assessment of heterogeneity and evidence consistency

Statistical heterogeneity was quantified using the between-study standard deviation, σ, in the random effects NMA models. The degree of heterogeneity was assessed as reasonable, high or extremely high based on guidelines for interpreting σ on the log-odds ratio scale suggested by Spiegelhalter *et al.* (Spiegelhalter *et al.*, 2004). These guidelines state that values of σ from 0.1 to 0.5 may be considered as indicating a reasonable degree of heterogeneity in most situations, 0.5 to 1 as high and values above 1 as very extreme heterogeneity. The consistency of each network in Figure 5.1 was checked using the method of node-splitting (Dias *et al.*, 2010) introduced in Chapter 2 Section 2.3.3. Briefly node-splitting involves first calculating two estimates of the intervention effect – one based on direct evidence and the other based on the indirect evidence for each pairwise comparison that have both sources of evidence. The direct and indirect estimates are then compared with each other for evidence of conflict. The whole process is repeated for all relevant pairs of interventions in the network forming part of closed loops of evidence and hence having both sources of evidence.

5.3.6 Sensitivity analysis

As explained in Chapter 2 Section 2.3.5, the results of analyses conducted within the Bayesian framework and using MCMC simulations to estimate model parameters can be sensitive to specification of prior distributions and choice of starting values, among other issues. In particular, it is well known that priors intended to be 'vague', or minimally informative can be problematic to specify for the variance and standard deviation terms

(Lambert *et al.*, 2005). Because of this, it has been suggested good practice, when carrying out MCMC based Bayesian analysis (Spiegelhalter *et al.*, 2000; Lambert *et al.*, 2005; Dias *et al.*, 2011a; Lunn *et al.*, 2012), to investigate sensitivity of results to model assumptions and specification of 'vague' prior distributions.

Three groups of sensitivity analyses were therefore conducted. As the quality of included studies varied, the first group of sensitivity analyses were conducted to investigate the effect of study quality by restricting the analyses to data from RCTs only. It was not possible to repeat the analysis for safe storage of poisonous plants by excluding non-randomised studies because only 3 studies provided data for this outcome.

The second group of sensitivity analyses explored the effect of using continuity corrections to facilitate inclusion of one study with zero events in one arm (Dershewitz and Williamson, 1977) and another study with 100% event rate in intervention-arm (Kelly *et al.*, 1987). This is because, although the Bayesian approach has the advantage of not requiring artificial cell corrections to accommodate study arms with zero or 100% event rates, occasionally the model may fail to run, especially when the data is sparse (Dias *et al.*, 2011a) as is the case for the poisoning prevention data (Table 5.1). Two solutions to this problem suggested for the NMA context in NICE Technical Support Document 2 (Dias *et al.*, 2011a) are to: i) apply the usual continuity correction by for example adding 0.5 and 1 to numerator and denominators of the affected studies respectively, and ii) assume a model for the baseline effects, μ_{iA} in order to make them identifiable.

Finally, the third group of sensitivity analyses were carried out to investigate influence of 'vague' prior distributions on estimates of intervention effects and the between-study variance parameters. For this, the Normal (0, 10^3) prior distributions on the baseline effects, μ_{iA} and d_k, the pooled effect of intervention k relative to usual care were replaced with Normal (0, 10^6). Sensitivity to prior distribution for the between-study standard deviation σ was also assessed by replacing σ ~ Uniform (0, 2) with the following prior distributions suggested as being weakly informative (Lambert *et al.*, 2005): i) σ ~ Uniform (0, 100), ii) σ ~ Normal (0, 100) truncated at 0 and iii) σ^2 ~ Inverse-Gamma (0.001, 0.001).

5.4 Results

Table 5.2 displays the goodness of fit and DIC statistics from the NMA models fitted to each of the 5 poison prevention outcomes. These statistics were used to discriminate between fixed and random effects models on the bases of goodness of fit to the data. The statistics in Table 5.2 are therefore referred to in the sections below where the results of the NMA for each outcome are presented. In addition and for each outcome, the results of the pairwise meta-analyses and those from the NMA for each outcome are presented alongside each other to aid comparison of effect estimates obtained using only direct evidence (i.e. evidence from studies that have directly compared the two interventions) and those from using both direct and indirect evidence (NMA).

Table 5.2: Model fit statistics for the network meta-analysis models

	Outcomes									
	Safe storage of medicines		Safe storage of other household products		Safe storage of Poisons		Possession of PCC number		Safe storage of poisonous plants	
Statistic	FE model	RE model	FE model	RE model	FE model	RE model	FE model	RE model	FE model	RE model
N	24	24	30	30	19	19	20	20	6	
DRes	24.3	23.5	43	30.9			59.2	19.5	6.6	
DIC	144.5	146	199	193.3	126.8	125.1	168	132	41.7	

DRes = Residual deviance and is used compare fit of model to the data. A better fitting model should have residual deviance close to the number of data points
DIC = Deviance Information Criterion, is used to choose between models. Model with the lowest DIC is preferred based on a difference of about 5 DIC points between models being considered important
FE= Fixed effect, RE=Random effects model
N= number of data points in model
PCC= Poison centre control telephone number

5.4.1 Storage of medicines

Fifteen of the 29 studies investigated the effectiveness of 7 interventions to promote safe storage of medicines (Figure 5.1 plot A). Eleven (85%) studies were RCTs and 2 (15%) were non-RCTs (Table 5.1). One study (Kelly *et al.*, 1987) reported a 100% event rate for this outcome in both arms. The model failed to run when this study was included. The study was therefore excluded from the analysis. To investigate the problem further, however, a sensitivity analysis was conducted by adding 0.5 and 1 to the numerators and denominators of Kelly *et al.*'s study as suggested in the NICE technical guidance for network meta-analysis (Dias *et al.*, 2011a).

Table 5.2 presents model fit statistics for safe storage of medicines. The residual deviance was 24.29 and 23.36 in the fixed effect and random effects models respectively. These are very close to 24 (the number of data points in the model), indicating that both models fitted the data well. Both fixed and random effects models also have comparable DIC values (DIC = 144.48 for the fixed effect and 146.02 for the random effects model) and therefore very little to choose between them based on the DIC statistic. The random effects results are however preferred for making inference as they are slightly more conservative due to the fact that the estimated between-study heterogeneity of 0.269 (95% CrI 0.009 to 1.034) (see Section 5.5.2, Table 5.9 for the heterogeneity statistics), although reasonable on the log-odds ratio scale (Spiegelhalter *et al.*, 2004), was not zero as assumed under a fixed effect model. Table 5.3 presents the random effects estimates alongside those from the pairwise meta-analysis (fixed effect estimates were also obtained but are not presented). The pairwise meta-analysis results are included in Table 5.3 to allow comparison between effectiveness estimates from NMA and the corresponding estimates obtained using only direct evidence (pairwise meta-analysis) where available. The results show that home safety interventions increase safe storage of medicines with 'education plus low cost/free equipment' the most likely to be effective (probability best = 0.39) (see Table 5.8), with an estimated odds ratio compared to usual care of 2.51 (95% CrI: 1.01 to 6.00). When the effect of study design on the NMA results was assessed, by repeating the above analysis using only data from the 11 RCTs, the results were similar, although for this analysis the network was limited to only 6 interventions (i.e. excluding the intervention 'education plus low cost/ free equipment and home safety inspection').

Table 5.3: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of medicines
Pairwise meta-analysis results are presented above the diagonal line whilst NMA results are presented below the diagonal line

Inter-vention*	(1) UC	(2) E	(3) E+FE	(4) E+FE+HSI	(5) E+FE+F	(7) E+FE+ HSI+F	(8) FE
(1) UC		1.63 (1.17, 2.26)‡	1.16 (0.20, 8.36)	1.33 (0.53, 3.31)	1.15 (0.77, 1.71)	1.93 (1.35, 2.76)	
(2) E	1.39 (0.73, 2.28)		1.27 (0.59, 2.71)		2.44 (0.45, 13.28)		
(3) E+FE	2.51 (1.01, 6.00)	1.85 (0.77, 4.60)		0.62 (0.10, 3.78)		0.69 (0.35, 1.35)	0.86 (0.43, 1.67)
(4) E+FE+HSI	1.41 (0.46, 3.89)	1.02 (0.31, 3.34)	0.54 (0.15, 1.90)				
(5) E+FE+F	1.31 (0.64, 3.47)	0.94 (0.43, 3.06)	0.52 (0.17, 1.88)	0.95 (0.28, 4.09)			
(7) E+FE+ HSI +F	1.93 (0.76, 5.12)	1.38 (0.52, 4.79)	0.77 (0.22, 2.80)	1.37 (0.36, 6.12)	1.48 (0.37, 4.74)		
(9) FE	2.13 (0.51, 8.42)	1.53 (0.40, 6.72)	0.83 (0.28, 2.53)	1.51 (0.30, 8.86)	1.60 (0.28, 7.19)	1.08 (0.20, 5.72)	

Pairwise estimates (Above the diagonal line)

NMA estimates (Below the diagonal line

*Interventions components (UC = Usual care, E = Education, FE = Free/low cost equipment, HSI = Home safety inspection, F = Fitting of equipment)
Blank cells indicate that no direct evidence on the specific pairwise comparison was available.
NMA estimates are from random effects model. ‡ indicate estimate is from a random effects pairwise meta-analysis model. All other pairwise results are individual study or fixed effect estimates
The intervention with the lowest number is always the comparator, e.g. 1.39 (0.73, 2.28) is the NMA estimate for Education (2) vs. Usual care (1) whereas 1.63 (1.17, 2.26) is corresponding estimate from pairwise meta-analysis

5.4.2 Storage of other household products

Sixteen studies with 7 interventions were included in the network for safe storage of other household products (Figure 5.1B), of which 11 (73%) studies were RCTs and 4 (27%) were non-RCTs (Table 5.1). One study (Dershewitz and Williamson, 1977) reported zero events (i.e. none of households surveyed had reported safe storage of other household products) in the provision of 'low cost/free equipment' intervention arm. The model failed to run when this study was included in the analysis, possibly due to the fact that only this study in the network had investigated the effect of 'low cost/free equipment' on safe storage of other

household products. To facilitate inclusion of this study in the analysis, a continuity correction was applied by adding 0.5 and 1 to the denominator and numerator of the affected study respectively. The data were re-analysed as part of the sensitivity analysis (results presented in Section 5.6.1) without the continuity correction by: i) firstly placing a model on the baseline effects, μ_{iA} as suggested in NICE technical support document (Dias *et al.*, 2011a) which facilitated inclusion of the affected study; and ii) secondly excluding the affected study all together.

The model fit statistics (Table 5.2) for other household products indicated that the fixed effects model fitted the data poorly (as the residual deviance of 43 is not close to 30, the number of data points in the model, DIC=199). The corresponding statistics for the random effects model indicates a better fit (residual deviance of 30.94 is very close to 30, DIC=193). The results of random effects NMA model are presented for storage of other household products in Table 5.4 together with the pairwise results. These show that home safety interventions increased safe storage of other household products but only 'education plus low cost/free equipment plus home safety inspection' (OR 2.53, 95% CrI 1.11 to 7.20) showed significant improvement when compared to 'usual care'.

Table 5.4: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of other household products
Pairwise meta-analysis results are presented above the diagonal line whilst NMA results are presented below the diagonal line

Inter-vention*	(1) UC	(2) E	(3) E+FE	(4) E+FE+HSI	(5) E+FE+F	(7) E+FE+ HSI+F	(8) FE
(1) UC		1.36 (0.93, 1.98)	2.01 (1.38, 2.92)	1.78 (1.23, 2.57)	1.18 (0.96, 1.47)		
(2) E	1.27 (0.68, 2.47)		2.58 (1.12, 5.94)		1.41 (0.49, 4.06)		
(3) E+FE	2.26 (0.94, 5.60)	1.78 (0.67, 4.72)				1.04 (0.81, 1.35)	0.32 (0.01, 7.96)
(4) E+FE+ HSI	2.53 (1.10, 7.13)	1.99 (0.71, 6.71)	1.12 (0.34, 4.45)				
(5) E+FE+F	1.33 (0.47, 4.30)	1.06 (0.35, 3.47)	0.59 (0.15, 2.47)	0.54 (0.12, 2.10)			
(7) E+FE+ HSI +F	2.60 (0.55, 15.68)	2.06 (0.38, 13.61)	1.15 (0.19, 8.62)	1.04 (0.26, 4.19)	1.93 (0.28, 15.24)		
(9) FE	0.37 (0, 15.10)	0.29 (0, 12.31)	0.17 (0, 6.05)	0.14 (0, 6.34)	0.27 (0, 12.87)	0.13 (0.00, 7.67)	

Pairwise estimates (Above the diagonal line)

NMA estimates (Below the diagonal line

*Interventions components (UC = Usual care, E = Education, FE = Free/low cost equipment, HSI = Home safety inspection, F = Fitting of equipment)

Blank cells indicate that no direct evidence on the specific pairwise comparison was available.

NMA estimates are from random effects model, pairwise results are individual study or fixed effect estimates

The intervention with the lowest number is always the comparator, e.g. 1.27 (0.68, 2.47) is the NMA estimate for Education vs. Usual care whereas 1.36 (0.93, 1.98) is corresponding estimate from pairwise meta-analysis

The effect of study design on the NMA results was assessed by repeating the above analysis using only data from the 11 RCTs limiting the network to 6 interventions (i.e. excluding education only). The results changed slightly but the most intensive intervention was still most likely to be the most effective (probability best = 0.56) closely followed by the intervention 'education plus low cost/ free equipment and home safety inspection' (probability best = 0.44) (Table 5.8).

5.4.3 Safe storage of poisons

Ten studies provided data on effectiveness of 5 interventions to increase safe storage of poisons in households with children (Figure 5.1C). Six (67%) studies were RCTs and 3 (33%) were non-RCTs (Figure 5.1C). Nine of the 10 studies reported arm-level outcome data (i.e. number of households with a PCC number/total number of households in each treatment arm). Therefore, the likelihood for the 9 studies is specified using equation (2.8). The remaining study by Reich et al (Reich *et al.*, 2011) reported two ORs and standard errors (i.e. OR1= education book vs. no book intervention and OR2= education book vs. non-educational book intervention; see Table 5.1). The two ORs were combined using a fixed effect meta-analysis. The combined log OR was given a normal likelihood and included in the analysis as follows:

$$\log(OR_{ibk}) \sim \text{Normal}(\delta_{ibk}, se^2_{ibk}) \tag{5.1}$$

where $\log(OR_{ibk})$ is the combined treatment effect with variance se^2_{ibk} from Reich et al (2011) and δ_{ibk} is the study-specific effect of intervention *k* relative to intervention *b*. The model fit statistics (Table 5.2) indicated that both fixed and random effects models fitted the data well. However, the random effects model results were preferred again as they were slightly more conservative than the fixed effect results (Table 5.5). There was evidence to suggest that compared to usual care, the most intensive intervention, i.e. 'education plus low cost/free equipment plus home safety inspection plus fitting' (OR 11.24, 95% CrI 1.92 to 114.70) was

Table 5.5: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of poisons
Pairwise meta-analysis results are presented above the diagonal line whilst NMA results are presented below the diagonal line

Inter-vention*	(1) UC	(2) E	(3) E+FE	(4) E+FE+HSI	(7) E+FE+ HSI+F	(8) E+HV
(1) UC		1.50 (1.09, 2.05)		10.11 (1.60, 64.01)	9.53 (2.16, 42.03)	0.98 (0.86, 1.12)
(2) E	1.31 (0.65, 2.82)		0.93 (0.30, 2.87)			
(3) E+FE	2.35 (0.56, 11.50)	1.79 (0.45, 7.92)		0.81 (0.25, 2.60)		
(4) E+FE+ HSI	3.26 (0.98, 11.84)	2.69 (0.59, 14.85)	1.50 (0.38, 6.58)			
(7) E+FE+ HSI+F	11.24 (1.92, 114.70)	8.61 (1.25, 94.30)	5.01 (0.82, 42.85)	3.53 (0.53, 31.75)		
(8) E+HV	0.92 (0.42, 1.74)	0.69 (0.22, 1.73)	0.43 (0.13, 1.25)	0.25 (0.04, 1.25)	0.09 (0.01, 0.34)	

Pairwise estimates (Above the diagonal line)

NMA estimates (Below the diagonal line

Blank cells indicate that no direct evidence on the specific pairwise comparison was available.
The intervention with the lowest number is always the comparator, e.g. 1.31 (0.65, 2.82) is the estimate for NMA estimate for Education (2) vs. Usual care (1) whereas 1.50 (1.09, 2.05) is corresponding estimate from pairwise meta-analysis
NMA estimates are from random effects model. Pairwise results are individual study or fixed effect estimates
*Interventions components (UC = Usual care, E = Education, FE = Free/low cost equipment, HSI = Home safety inspection, F = Fitting of equipment, HV=Home visit)

effective in promoting safe storage of poisons in the home. Repeating the analysis using only data from the 6 RCTs identified both education and low/free equipment (Probability best 0.38), and education, low cost/free equipment, home safety inspection and installation (Probability best 0.36) to be the most effective at promoting the number of households with storage of poisons compared to usual care intervention (Table 5.8).

5.4.4 Possession of a PCC number

Ten studies with 7 interventions were included in the network for possession of a PCC number (Figure 5.1C), of which 7 (70%) were RCTs and 3 (30%) were non-RCTs (Table 5.1). The model fit statistics indicated that only the random effects NMA model fitted the data well (Table 5.2), and therefore only the results from this model are presented alongside the pairwise results (Table 5.6). There was evidence that compared to usual care; education

plus low cost/free equipment and home safety inspection (OR 39.25, 95% CrI 2.35 to 724.40) increased the number of households with a PCC number.

Table 5.6: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for possession of PCC number
Pairwise meta-analysis results are presented above the diagonal line whilst NMA results are presented below the diagonal line

Intervention	(1) UC	(2) E	(3) E+FE	(4) E+FE+HSI	(6) E+HSI	(7) E+FE+HSI+F	(8) E+HV
(1) UC		2.69 (1.91, 3.78)	3.89 (2.69, 5.63)	34 (9.33, 123.97)	1.30 (0.71, 2.39)	7.96 (4.29, 14.77)	2.44 (0.96, 6.21)
(2) E	2.04 (0.50, 7.82)		1.85 (0.79, 4.32)				
(3) E+FE	3.81 (0.77, 18.75)	1.87 (0.32, 11.50)					
(4) E+FE +HSI	39.25 (2.35, 724.4)	19.43 (0.84, 499.8)	10.36 (0.40, 279.1)				
(6) E+HSI	1.31 (0.09, 17.89)	0.64 (0.03, 12.52)	0.34 (0.02, 7.50)	0.03 (0.00, 1.55)			
(7) E+FE+HSI+F	8.14 (0.60, 114.1)	3.9 (0.21, 78.45)	2.12 (0.10, 46.29)	0.21 (0.00, 10.12)	6.21 (0.15, 266.3)		
(8) E+HV	2.41 (0.16, 37.34)	1.19 (0.06, 25.55)	0.63 (0.03, 14.77)	0.06 (0.0, 3.22)	1.85 (0.04, 81.59)	0.30 (0.01, 13.40)	

Pairwise estimates (Above the diagonal line)

NMA estimates (Below the diagonal line

*Interventions components (UC = Usual care, E = Education, FE = Free/low cost equipment, HSI = Home safety inspection, F = Fitting of equipment, HV=Home visit)
Blank cells indicate that no direct evidence on the specific pairwise comparison was available.
NMA estimates are from random effects model. Pairwise results are individual study or fixed effect estimates
The intervention with the lowest number is always the comparator, e.g. 2.04 (0.50,7.82) is the NMA estimate for Education (2) vs. Usual care (1) whereas 2.69 (1.91, 3.78) is corresponding estimate from pairwise meta-analysis

5.4.5 Safe storage of poisonous plants

Three studies, one of which is the 3-arm (Babul *et al.*, 2007) study, provided data on 5 interventions for safe storage of poisonous plants (Figure 5.1E). Only a fixed effect NMA model was fitted due to the small number of studies (i.e. only 3 studies) available for this

outcome. There was no evidence that home safety education interventions are effective in increasing safe storage of poisonous plants (Table 5.7).

Table 5.7: Estimated odds ratios (95% confidence/credible intervals) from pairwise and random effects network meta-analysis model for safe storage of poisonous plants. Pairwise meta-analysis results are presented above the diagonal line whilst NMA results

Intervention	(1) UC	(2) E	(3) E+FE	(4) E+FE+HSI	(5) E+FE+F
(1) UC			1.18 (0.70, 2.0)	1.04 (0.61, 1.76)	
(2) E	0.68 (0.13, 3.16)				0.12 (0.01, 1.03)
(3) E+FE	1.18 (0.70, 2.01)	1.74 (0.40, 8.44)		0.88 (0.52, 1.48)	
(4) E+FE+HSI	1.04 (0.60, 1.77)	1.53 (0.32, 8.16)	0.88 (0.52, 1.48)		
(5) E+FE+F	0.05 (0.00, 0.84)	0.08 (0, 0.62)	0.04 (0, 0.66)	0.05 (0, 0.81)	

Pairwise estimates (Above the diagonal line)

NMA estimates (Below the diagonal line

*Interventions components (UC = Usual care, E = Education, FE = Free/low cost equipment, HSI = Home safety inspection, F = Fitting of equipment, HV=Home visit)
Blank cells indicate that no direct evidence on the specific pairwise comparison was available.
The intervention with the lowest number is always the comparator, e.g. 0.68 (0.13, 3.16) is the NMA estimate for Education vs. Usual care with corresponding no estimate from the pairwise meta-analysis, hence the cell is blank
NMA estimates are from random effects model. Pairwise results are individual study or fixed effect estimates

5.4.6 Assessment of best intervention

Table 5.8 presents estimates of the probability that each intervention is the 'best' for each poison prevention measure. Relative ranking of the interventions are also presented with the most effective intervention ranked as number one. Overall, the estimated probabilities were generally low across all injury prevention outcomes, suggesting that no one intervention completely dominated the others when it comes to promoting poisoning prevention behaviours in the home. The only possible exception was 'education plus low cost/free equipment plus home safety inspection' which had an 82% probability of being the most effective intervention for safe storage of poisons and 'education plus low cost/free equipment, with a 68% probability of being the most effective intervention for increasing uptake of PCC number.

Table 5.8: Assessment of the best intervention

Intervention	Safe storage of medicines		Safe storage of other household products		Safe storage of poisons		Safe storage of poisonous plants		Possession of PCC number	
	Prob(Best)[1]	Rank[2]	Prob(Best)	Rank	Prob(Best)	Rank	Prob(Best)	Rank	Prob(Best)	Rank
(1) UC	0.00	6 (7, 3)	0.00	6 (7, 4)	0.00	5 (6, 3)	0.16	3 (4, 1)	0.00	6 (7, 4)
(2) E	0.04	4 (6, 1)	0.01	5 (7, 2)	0.00	4 (6, 2)	0.21	4 (4, 1)	0.01	5 (7, 2)
(3) E+FE	0.28	2 (6, 1)	0.22	3 (5, 1)	0.04	3 (6, 1)	0.43	2 (4, 1)	0.04	3 (7, 1)
(4)E+FE+HSI	0.10	5 (7, 1)	0.22	2 (5, 1)	0.14	2 (5, 1)	0.20	2 (4, 1)		
E+FE+F	0.08	5 (7, 1)	0.05	4 (7, 1)			0.00	5 (5, 5)	0.68	1 (5, 1)
E+HSI									0.03	5 (7, 1)
E+FE+HSI+F	0.27	3 (7, 1)	0.37	2 (7, 1)	0.82	1 (3, 1)			0.19	2 (7, 1)
E+HV					0.00	6 (6, 3)			0.06	4 (7, 1)
FE	0.23	3 (7, 1)	0.14	7 (7, 1)						

Prob(Best)[1] = probability intervention is the best.
Rank[2] = posterior median estimate and (95% credible intervals) – rationale for using median estimate instead of the mean is given in Section 2.3.2.
Blank cells indicate none of the studies have considered that intervention/outcome combination.
Interventions components.
UC = Usual care
E = Education
FE = Free/low cost equipment
HSI = Home safety inspection
F = Fitting of equipment
HV = Home visit

5.5 Convergence, heterogeneity and consistency assessment results

5.5.1 Convergence diagnostics

Diagnostic plots for d_k, the pooled effect of intervention k relative to usual care on the log-odds ratio scale, and σ, the between-study standard deviation, from the NMA for safe storage of medicines are presented in Figures 5.2 to 5.5. Similar plots were also obtained for the other outcomes but these are not reported in this thesis. The density plots (Figure 5.2) for d_k look reasonably smooth with the characteristic bell-shaped appearance of a parameter that is assumed to follow a normal distribution. The density plot for σ (Figure 5.2) is truncated at zero because standard deviations and variances are defined on the positive scale. The history plots (Figure 5.3) looked reasonably stable with a 'fat caterpillar' appearance, indicating reasonable degree of convergence. The autocorrelation plots (Figure 5.4) show successive iterations of the two parameters seem to be sampled from independent posterior distributions leading to good mixing and faster convergence. Examination of the Brooks-Gelman diagnostic plot (Figure 5.5) shows *R* has converged to 1, and *B* and *W* have converged to stability, given further evidence of convergence of samples. Overall, the diagnostic plots seem to indicate that the posterior estimates for d_k and σ were obtained from samples with a reasonable degree of convergence.

In contrast to the plots above, Figure 5.6 displays plots for d_7, the log (OR) for 'low cost/free equipment' versus 'usual care' from the sensitivity analysis (baseline effects, μ_{iA} were assumed to be normally distributed as explained in Section 5.4.2) for safe storage of other household products. These show the samples have not converged even after running a large number of iterations (200,000 samples). There was insufficient information in the data to reliably estimate d_7 because the only study (Dershewitz and Williamson, 1977) which considered the equipment only intervention reported zero events for safe storage of other household products. This is supported by the fact that reasonably stable diagnostic plots were observed in the alternative models where a continuity correction was applied to the affected study (diagnostic plots not presented).

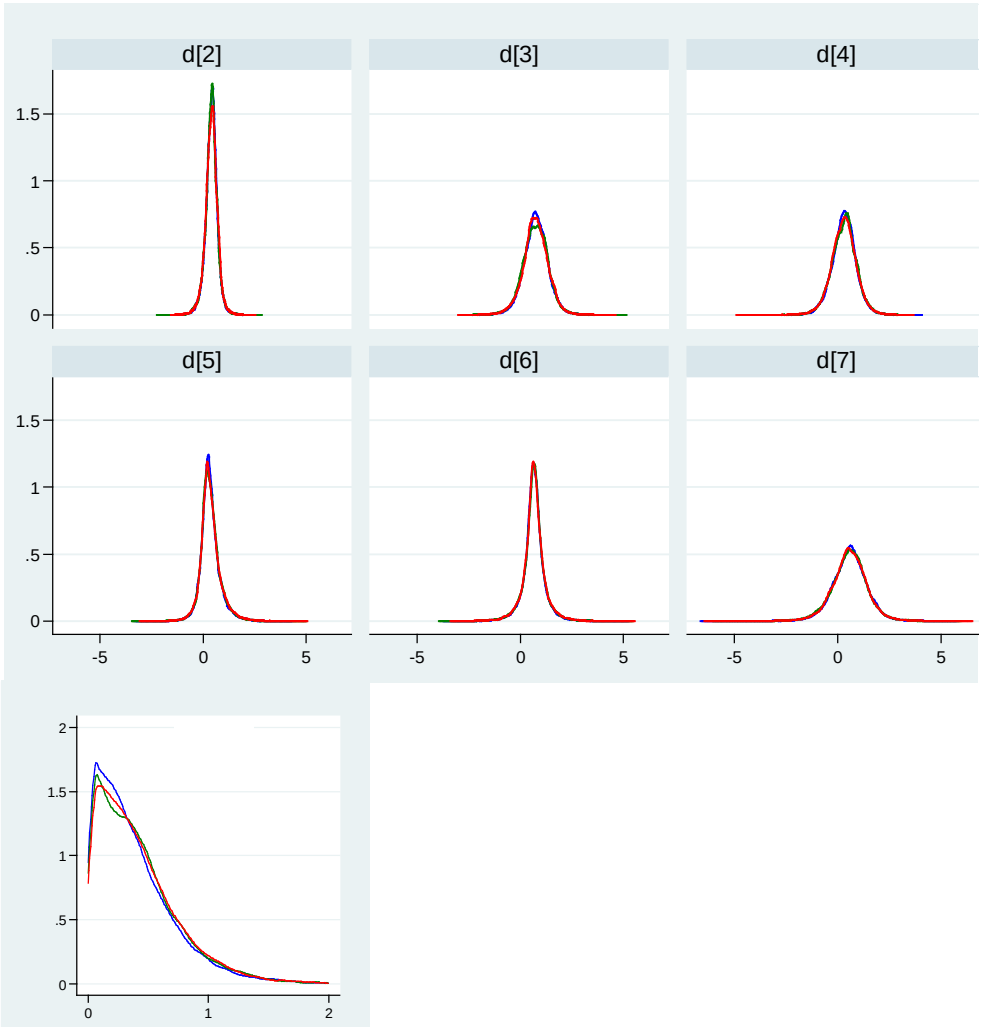
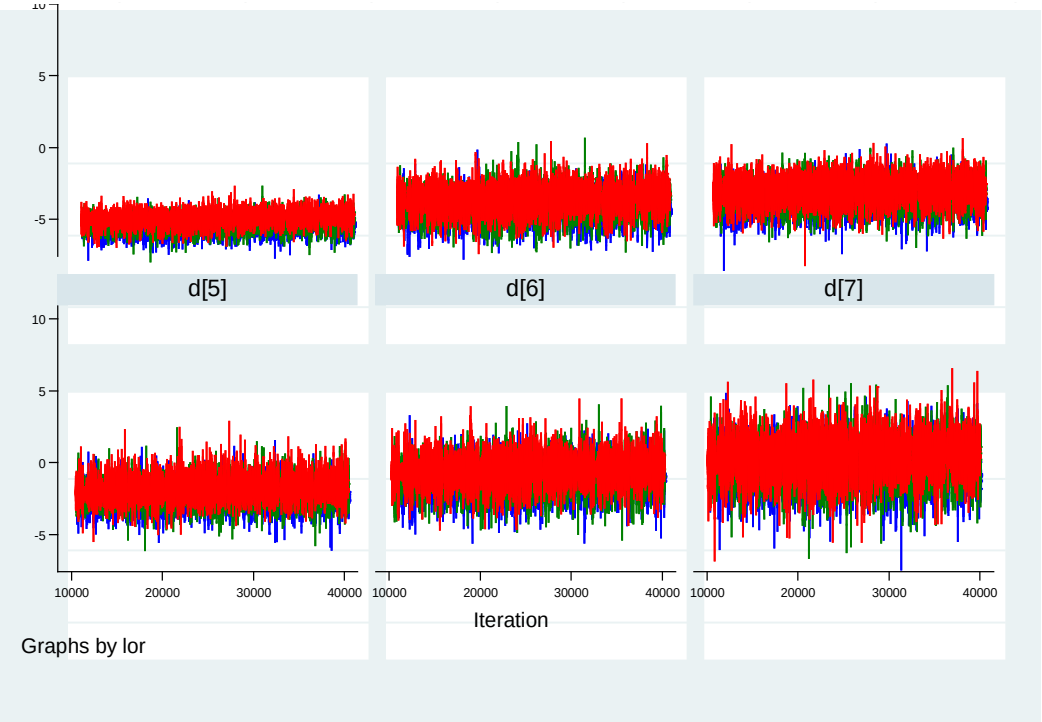
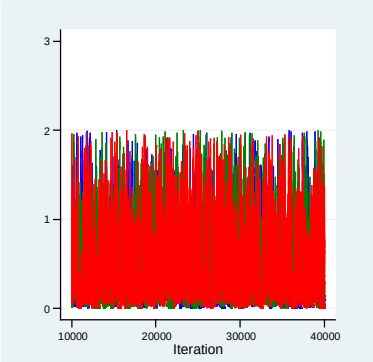


Figure 5.2: Posterior density plots for the pooled effect of intervention *k* relative to usual care, d_k and the between-study standard deviation (sd), σ on the log-odds ratio scale


d[5]
d[6]
d[7]
10
5
0
-5
10000
20000
30000
40000
Iteration
Graphs by lor


3
2
1
0
10000
20000
30000
40000
Iteration

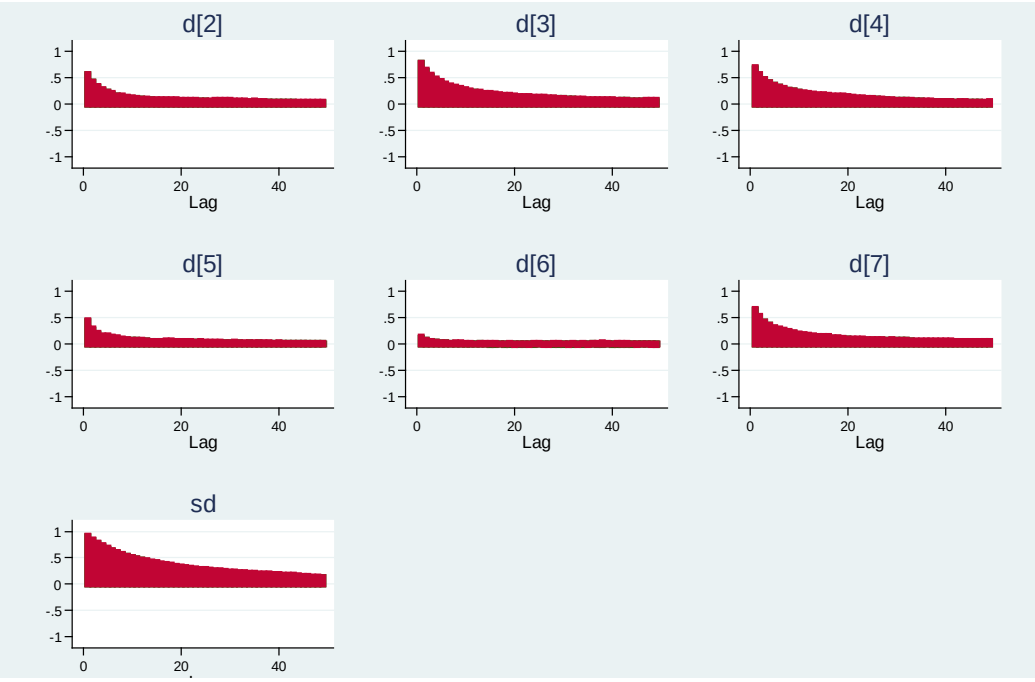


Figure 5.4: Autocorrelation plots for the pooled effect of intervention *k* relative to usual care, d_k and the between-study standard deviation (sd), σ on the log-odds ratio scale

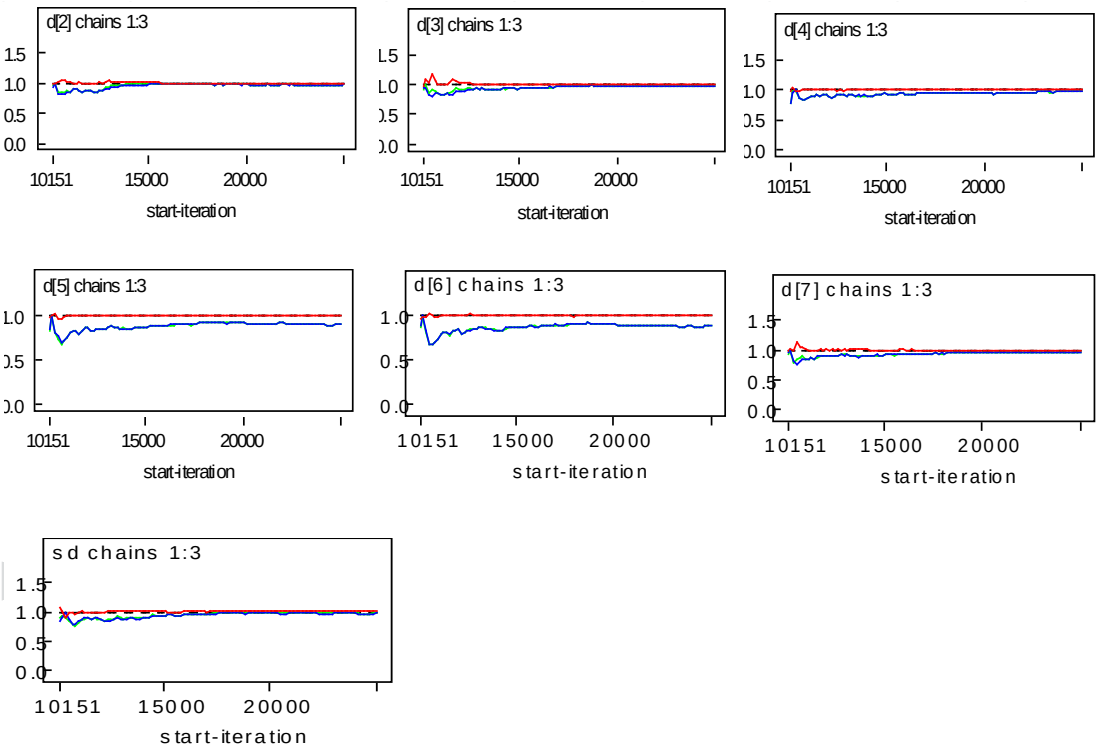


Figure 5.5: Autocorrelation plots for the pooled effect of intervention *k* relative to usual care, d_k and the between-study standard deviation (sd), σ on the log-odds ratio scale

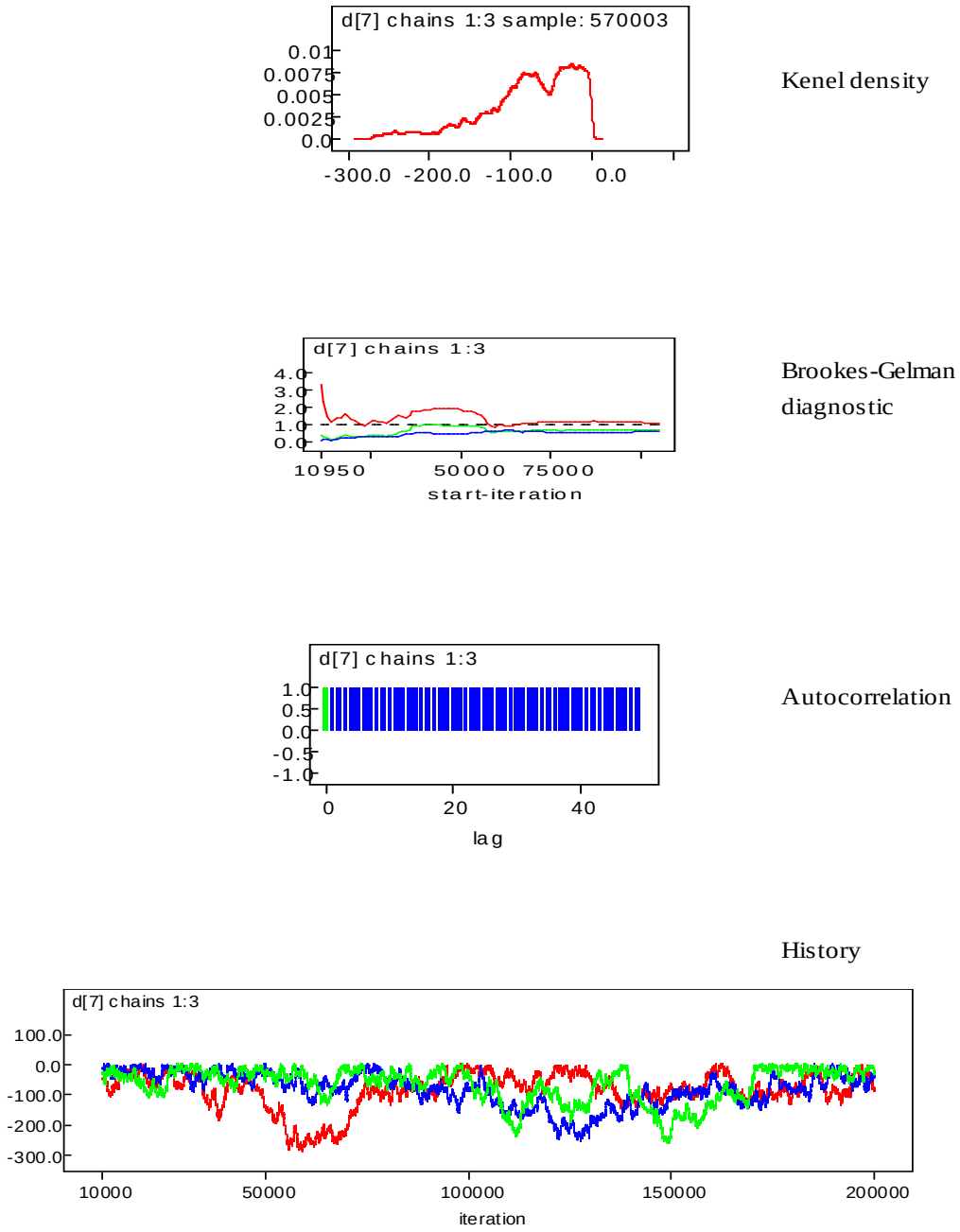


Figure 5.6: Diagnostic plots showing evidence of non-convergence in d_7, the effect estimate for low cost/free equipment versus usual care from the random baseline model for safe storage of other household products

5.5.2 Heterogeneity results

Estimates of the between-study standard deviation σ, on the log-odds ratio scale are presented in Table 5.9. Based on the criteria for interpreting σ outlined in Spiegelhalter *et al.* (Spiegelhalter *et al.*, 2002), these figures would seem to indicate a reasonable degree of heterogeneity in the intervention effects for safe storage of medicines and safe storage of poisons, high degree of heterogeneity for safe storage of other household products and extremely high heterogeneity for possession of a PCC number. The credible intervals, however, show that there is great uncertainty in the estimation of σ possibly as a result of the relatively small number of studies providing direct evidence on pairwise contrast in each network (Figure 5.1).

Table 5.9: Heterogeneity statistics from NMA models (log odds ratio scale)

Outcome	No. of studies	Posterior median of the between-study standard deviation, σ and 95% CrI in brackets
Safe storage of medicines	13	0.269 (0.009 to 1.034)
Safe storage of other household products	15	0.561 (0.128, 1.270)
Safe storage of poisons	10	0.361 (0.029, 1.436)
Possession of a PCC number	10	1.165 (0.574, 1.926)
Safe storage of poisonous plants	3	Fixed effects model fitted, hence $\sigma = 0$ assumed

5.5.3 Evidence consistency assessment results

As stated in Section 5.3.5 (assessment of heterogeneity and evidence consistency), only the relative effects for interventions that form part of a closed loop of evidence in the network (excluding loops formed by multi-arm studies) have both direct and indirect evidence and hence needed to be checked for inconsistency.

There are 3 closed loops of evidence in the network for safe storage of medicines (Figure 5.1A). One loop (indicated by thick red lines in Figure 5.1A) contains evidence from the 3-arm Babul *et al.* study. Relative effects between the interventions contained in this loop cannot be inconsistent since by definition there can be no inconsistency in the evidence from

a multi-arm study (Dias *et al.*, 2010). Therefore only the 5 relative effects from the two remaining loops have 'direct' and 'indirect' sources of evidence and needed to be checked for inconsistency.

The network for safe storage of other household products has two closed loops of evidence and no multi-arm trial (Figure 5.1B). Therefore 5 relative effects needed to be checked for inconsistency in the network for this outcome. The networks for safe storage of poisons (Figure 5.1C) and possession of PCC telephone number (Figure 5.1D) each has one closed loop of evidence. Hence only relative effects between interventions in each loop needs to be checked for inconsistency. The network for storage of poisonous plants has one closed loop formed by the three-arm Babul study (Babul *et al.*, 2007). That means there can be no inconsistency in the evidence structure for this outcome since multi-arm studies are assumed to provide consistent evidence on all treatment pairs (Dias *et al.*, 2010).

Estimated log odds ratios for the relative effects that have both direct and indirect evidence are presented in Table 5.10. Estimates based on the direct and indirect were obtained separately using the method of node splitting described in Section 2.3.3. The combined estimates are the estimates from the NMA model for the respective outcome. The results showed no evidence of inconsistency between the direct and indirect evidence in all networks (i.e. all the p-values in Table 5.10 were not statistically significant at the 5% significance level). This can be seen in Figure 5.7, where the posterior densities of the estimated intervention effect based on the direct, indirect and the combined evidence for the intervention pairs of interest in storage of medicines network show a degree of overlap when plotted side by side. However, it should be noted that relatively small number of studies were available in each network, which means that the analyses may have limited power to detect any inconsistencies in the evidence even if they exist.

Table 5.10: Evidence consistency checks for Safe storage of medicines and other household products
Posterior mean (standard error) of the Log ORs using the full NMA network, direct and indirect evidence on each pairwise contrast

Pair-wise contrast	Combined evidence from NMA model	Direct evidence	Indirect evidence	Inconsistency estimate†	p-value*
Safe storage of medicines					
Usual care (1) vs. Education (2)	0.40 (0.30)	0.50 (0.34)	0.79 (0.86)	0.89 (0.97)	0.306
Usual care (1) vs. Education + Free/low cost Equipment (3)	0.74 (0.61)	0.48 (1.23)	0.38 (0.36)	-0.31 (1.46)	0.820
Usual care (1) vs. Education + Equipment + Fitting (5)	0.35 (0.49)	0.14 (0.56)	1.55 (1.15)	-1.41 (1.27)	0.232
Education (2) vs. Education + Free/low cost Equipment (3)	0.34 (0.58)	0.41 (0.79)	0.12 (1.20)	0.29 (1.44)	0.778
Education (2) vs. Education + Equipment + Fitting (5)	-0.05 (0.54)	1.24 (1.09)	-0.32 (0.68)	1.56 (1.27)	0.189
Safe storage of non-medicines					
Usual care (1) vs. Education (2)	0.24 (0.32)	0.36 (0.38)	-0.24 (0.73)	0.61 (0.83)	0.408
Usual care (1) vs. Education + Free/low cost Equipment (3)	0.82 (0.43)	0.68 (0.54)	1.29 (0.91)	-0.61 (1.05)	0.499
Usual care (1) vs. Education + Equipment + Fitting (5)	0.30 (0.54)	0.17 (0.76)	0.60 (1.01)	-0.43 (1.27)	0.698
Education (2) vs. Education + Equipment (3)	0.58 (0.48)	0.96 (0.85)	0.34 (0.67)	0.61 (1.07)	0.508
Education (2) vs. Education + Equipment + Fitting (5)	0.07 (0.56)	0.31 (0.93)	-0.12 (0.83)	0.43 (1.25)	0.690
Safe storage of poisons					
Usual care (1) vs. Education (2)	0.34 (0.52)	0.20 (0.46)	2.72 (1.61)	-2.52(1.68)	0.118
Usual care (1) vs. Education + Equipment + Home safety inspection (4)	1.36 (0.93)	2.39 (1.22)	-0.07(1.32)	2.46 (1.77)	0.107
Education (2) vs. Education + Free/low cost Equipment (3)	0.60 (0.84)	-0.01 (0.88)	2.61(1.51)	-2.62 (1.76)	0.124
Education + Equipment (3) vs. Education + Equipment + Home safety inspection (4)	0.42 (0.82)	-0.28 (0.82)	2.44(1.50)	-2.74 (1.71)	0.083
Possession of a PCC number					
Usual care (1) vs. Education (2)	0.71 (0.68)	0.70 (0.81)	0.72 (1.72)	-0.02 (1.89)	0.989
Usual care (1) vs. Education + Equipment (3)	1.33 (0.78)	1.34 (0.97)	0.69 (0.81)	0.04 (1.90)	0.986
Education (2) vs. Education + Equipment (3)	0.63 (0.879)	0.63 (1.43)	0.63 (1.26)	0.00 (1.91)	0.992

†inconsistency estimate = direct estimate – indirect estimate of the treatment effect (log-OR)
*p-value = $2 \times (probability\ of\ direct\ estimate > indirect\ estimate)$ which gives the 2-sided probabilities that the direct and indirect evidence are different

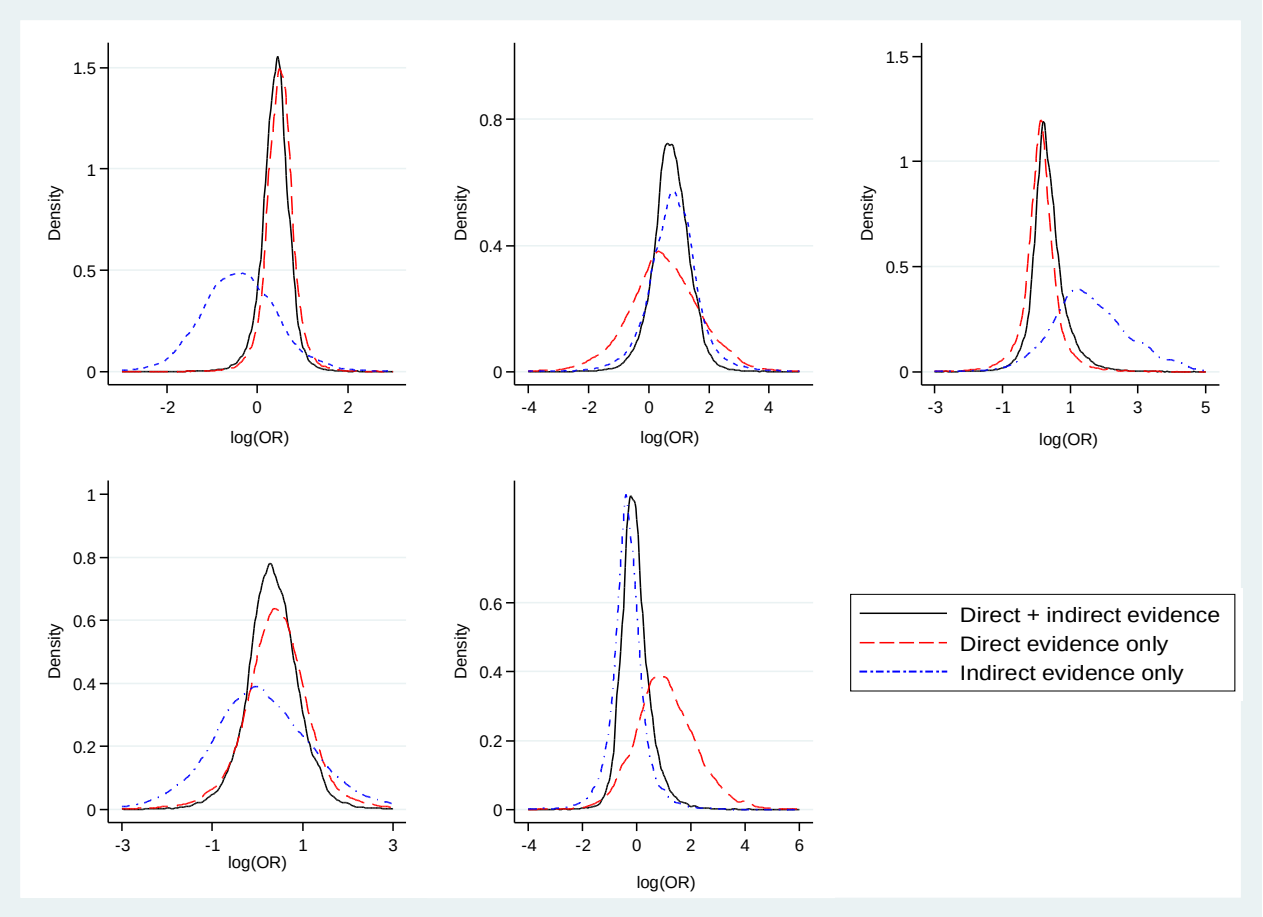


Figure 5.7: Safe storage of medicines (random effects model)

Posterior density plots of the log-odds ratio showing the distribution of estimates based on the direct evidence, indirect evidence and the combined evidence. Panel A = education versus usual care

5.6 Sensitivity analyses results

5.6.1 Continuity corrections

As stated in the results section for safe storage of medicines (Section 5.4.1) and safe storage of other household products (Section 5.4.2), continuity corrections were used to facilitate inclusion of Kelly *et al.* (Kelly *et al.*, 1987) in the NMA for safe storage of medicines and Dershewitz and Williamson's study (Dershewitz and Williamson, 1977) in the model for other household products. Sensivity analyses were thus conducted by re-analysing the data with the two studies excluded from the syntheses for their respective outcomes. Figure 5.8A is a summary forest plot showing the intervention effects relative to 'usual care' from the model for safe storage of medicines with and without the continuity correction applied to the study by Kelly et al (Kelly *et al.*, 1987). The plot shows that including Kelly et al.'s study had minimal impact on effect estimates with both models producing virtually identitical estimates.

For the other household products analysis, in addition to conducting a sensitivity analysis using a continuity correction to facilitate inclusion of the study by Dershewitz and Williamson, a third model was fitted in which the baseline effects, μ_{iA}, were assumed to follow a normally distribution (i.e. a NMA model with random baseline effects). Figure 5.8B displays the summary forest plot results of these sensitivity analysis for other household products. Putting random effects on μ_{iA} enabled the model to run but led to problems of non-convergence for the paramater d_7, the effect estimate for 'low cost/free equipment' versus 'usual care' as shown in the diagnostic plots in Figure 5.6. As a result no reliable estimates were obtained for 'provision of low cost/free equipment' from this model [note that effect estimates were not available for 'low cost/free equipment' from the random baseline model and the model that excluded Dershewitz and Williamson (1977) because only this study compared that particular intervention (i.e. low cost/free equiment)]. Finally, the forest plot in Figure 5.9 shows that using different estimates of the intervention effect from Reich et al. (Reich *et al.*, 2011) resulted in small changes in the pooled intervention effects for safe storage of poisons but not enough to change the conclusions of the analysis.

5.6.2 Sensitivity analysis to prior distributions

Sensitivity analyses to specification of alternative prior distribution for the baseline effects, μ_{iA}, and the pooled effect of intervention k relative to usual care d_k were performed but not reported. The results from these analyses suggest that pooled estimates of the intervention effects were not sensitive to prior distributions placed on μ_{iA} and d_k. The prior distributions for the variance terms (i.e. σ and σ^2) were however quite influential on estimates of the uncertainty around pooled intervention effects. Figure 5.9 presents summary forest plots of effect estimates compared to usual care from models with different prior distributions for σ for safe storage of medicines. In Figure 5.9, the two uniform priors - σ ~ Uniform(0,2) and σ ~ Uniform(0,100) - had the widest intervals with greatest uncertainty around parameter estimates followed by σ ~ N(0,1000)I(0,) with σ^2 ~ Inverse-Gamma(0.0001, 0.0001) producing the narrowest and more precise credible intervals. The prior for σ is less influential when the heterogeneity is low and highly influential when the degree of heterogeneity is high.

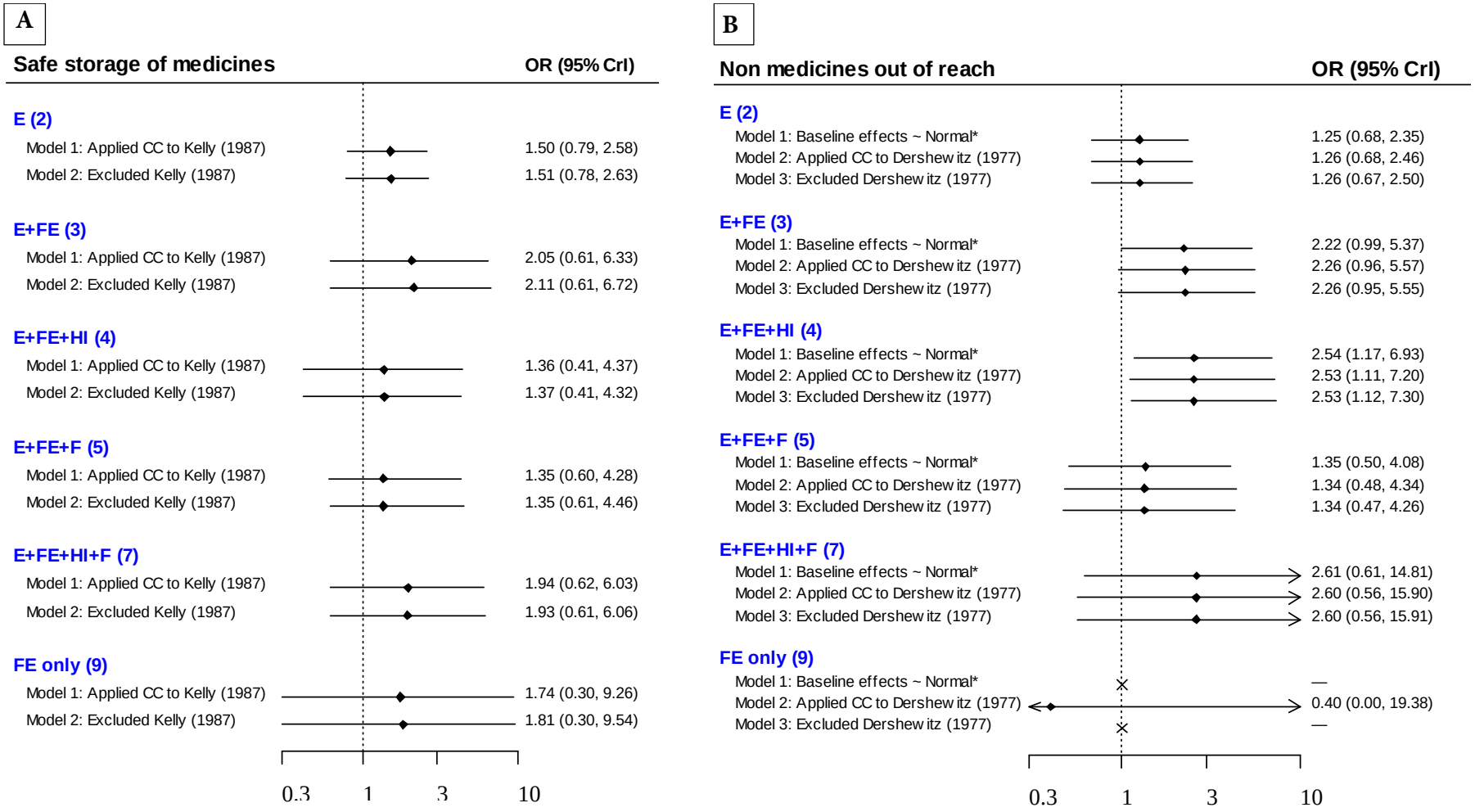


Figure 5.8: Estimated odds ratios (ORs) for home safety interventions compared with usual care
CC = continuity correction, E=Education, F= Fitting, FE= Free/low cost Equipment, HI = Home safety inspection. *Model for baseline effects, $\mu_{iA} \sim \text{Normal}(\bar{\mu}, \sigma_{\mu}^{2})$

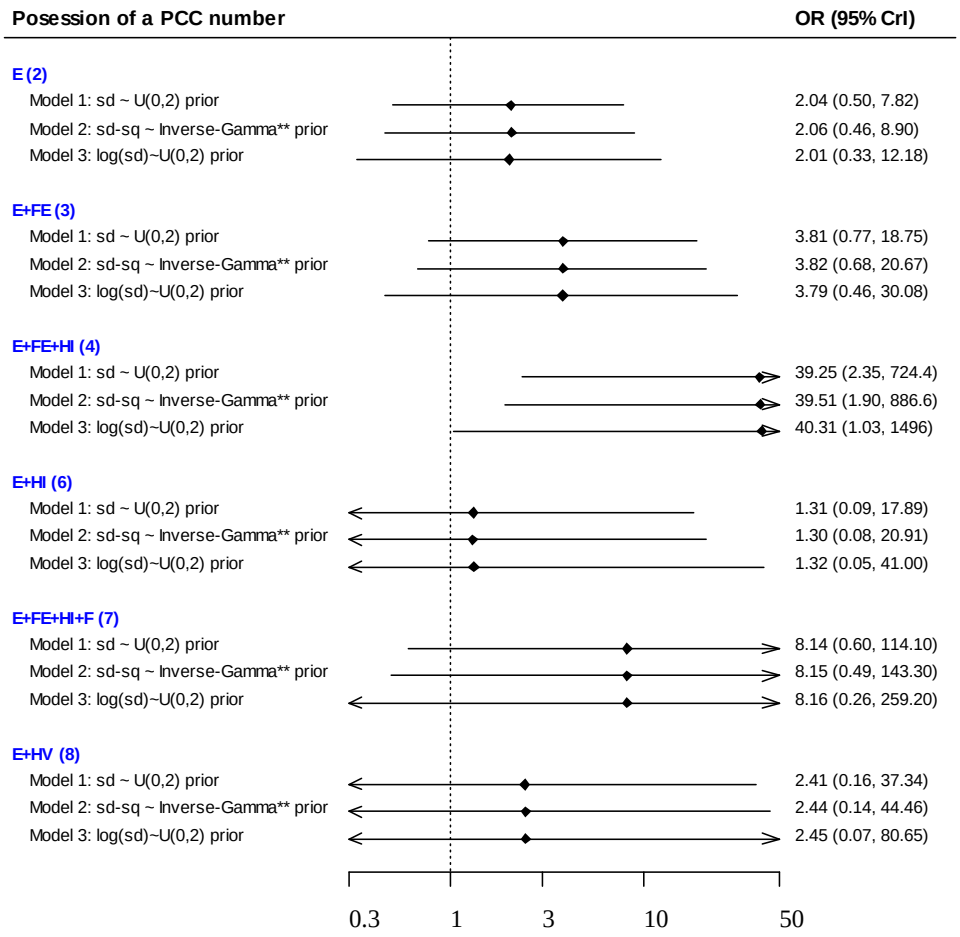


Figure 5.9: Estimated odds ratios (ORs) for home safety interventions compared with usual care for safe storage of poisons from sensitivity analysis using different estimates from Reich et al. (2011)

.

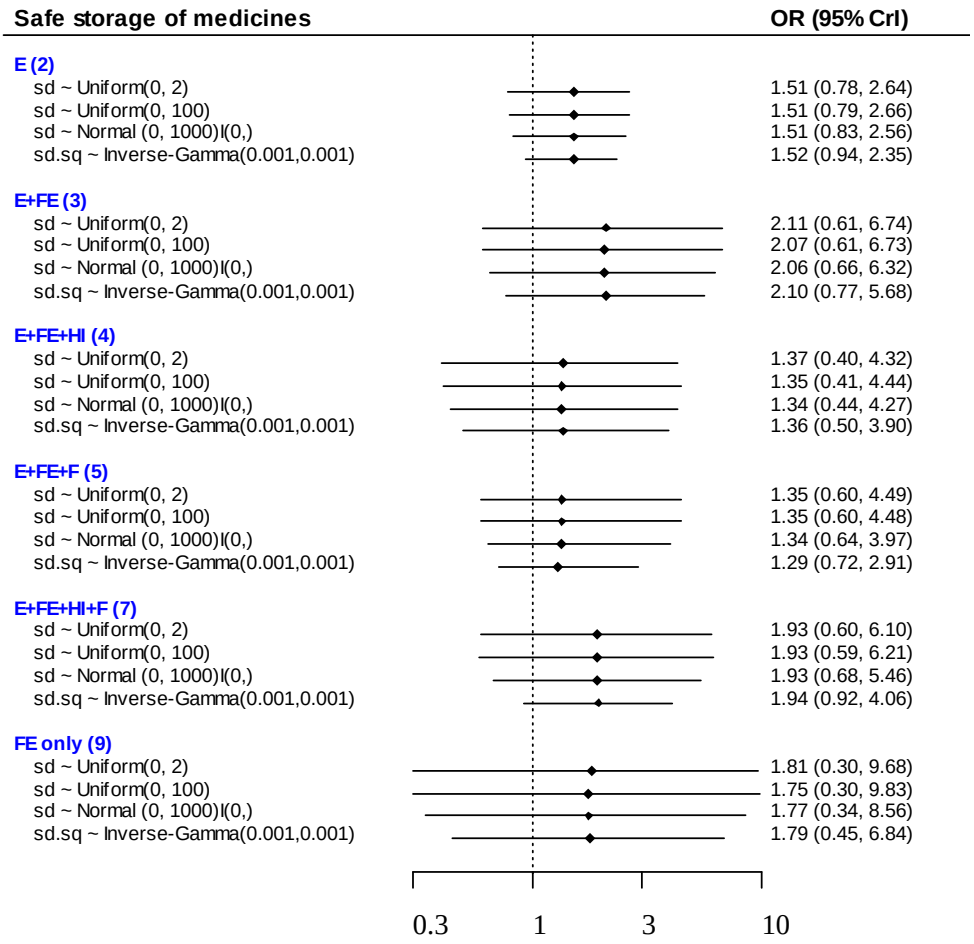


Figure 5.10: Estimated odds ratios (ORs) for home safety interventions compared with usual care for safe storage of medicines
Sensitivity analysis to prior distribution for heterogeneity parameter, σ, sd refers to σ and sd.sq = σ^2. E=Education, F= Fitting

5.7 Discussion

5.7.1 Summary of findings

In this chapter, NMAs were used to compare the different interventions with one another for promoting poison prevention behaviours by households with children. This analysis has allowed comparisons of strategies not addressed within any of the individual primary studies. The findings showed that more intensive interventions are more effective than education alone for each of the poison prevention practices being evaluated. Education plus low cost /free equipment was most effective in promoting safe storage of medicines, 'education plus low cost/free equipment plus home safety inspection and fitting' was most effective in promoting safe storage of other household products and poisons, and 'education plus low cost/free equipment plus home safety inspection' was most effective in promoting possession of a PCC number. There was no evidence that any of the interventions was more effective than the others at promoting safe storage of poisonous plants.

5.7.2 Strengths and limitations

NMA is a useful synthesis tool for comparing multiple injury prevention interventions which are often complex and multi-faceted, and where the number of studies evaluating the same comparisons is small. NMA enables interventions to be ranked in terms of their effectiveness in promoting safety practices providing results which are more likely to be useful to policymakers, service commissioners and providers when making choices between multiple alternatives than multiple pairwise meta-analyses.

No evidence of inconsistency between direct evidence and indirect evidence was found in the analyses, although the power to detect inconsistency will have been limited by sparse data, particularly for analyses involving very few studies. The inclusion of non-randomised study designs allowed greater number of studies to be included in the analysis, but also resulted in the inclusion of studies with greater potential for bias. Sensitivity analyses restricting analyses to RCTs produced similar results suggesting the findings were robust to exclusion of non-randomised studies. The quality of studies included in the analyses (assessed in terms of

allocation concealment (RCTs only), blinded outcome assessment, balance of confounders (non-RCTs only) and completeness of follow-up) was variable. It was not possible to explore the impact of the individual measures of quality on the results since such an analysis would be extremely limited due to the large number of parameters being estimated in the NMA relative to the number of studies and may even lead to disconnected networks.

Although NMA allows interventions to be classified into more categories than standard pairwise meta-analysis, there is, inevitably, still some "lumping" of interventions within these categories. For example, education may differ in intensity across studies; that is, from a leaflet or brochure distributed by post, to intensive face-to-face classes teaching home safety. Subcategorising the interventions further, to avoid "lumping", is reliant on detailed information being reported in the primary study publications. However, in the case of poison prevention education, insufficient detail was often reported to enable further sub categorisation.

5.7.3 Comparisons with existing work

The findings from the analyses carried out in this chapter are consistent with findings from the two previous pairwise meta-analyses. DiGuiseppi found interventions promoting "child-proofing" the home delivered in clinical settings had a modest effect (odds ratio 1.8, statistical significance not reported) on safe storage of cleaning products substances (DiGuiseppi and Higgins, 2000). The second meta-analysis by Kendrick *et al.* (Kendrick *et al.*, 2012c), found that education, with or without the provision of safety equipment was effective in increasing safe storage of medicines (OR 1.53, 95% CI 1.27-1.84), safe storage of household products (OR 1.55, 95% CI 1.22- 1.96) and, increasing availability of poison control centre numbers (OR 3.30, 95% CI 1.70- 6.39). These findings extend those from the previous meta-analyses by demonstrating which elements of multifaceted interventions are most effective. Furthermore, one of the previous meta-analyses failed to find significant effects of education, with or without the provision of safety equipment on keeping (unspecified) poisons (OR 0.57, 95%CI 0.31-1.07) or plants out of reach (OR 1.18, 0.40-

3.48), but the analyses reported in this chapter demonstrated that some poison prevention interventions are effective in promoting these safety practices.

The effect sizes in the NMA for safe storage of medicines, other household products and availability of the poison control centre number are all larger than the effect sizes found in the pairwise meta-analyses previously reported (DiGuiseppi *et al.*, 2001; Kendrick *et al.*, 2012c). It is likely that, by reducing clinical heterogeneity of interventions, the NMAs may explain some of the statistical heterogeneity in effect sizes found in previous pairwise meta-analyses. The findings also suggest pairwise meta-analyses combining all interventions, (which include less intensive, and as it has been shown, less effective interventions) may underestimate the effect of more intensive interventions.

4.4 Implications for practice and research

The findings from the analysis in this chapter suggest that the "best" interventions for increasing a range of poison prevention practices are the more intensive interventions. These include, at a minimum, education and providing equipment, but for some poison prevention practices the most effective intervention requires education, equipment provision and fitting and home safety inspection. The most effective intervention varied by poison prevention practice, so commissioners and providers of poison prevention interventions should tailor the interventions they commission or provide to the poison prevention practices they wish to promote. Knowing which interventions are most effective is important, but is only part of the information required to commission or provide poison prevention and cost-effectiveness is an essential part of any decision making process. The effect sizes from this NMA will be used in subsequent decision analyses to determine the most cost effective interventions for increasing poison prevention practices in Chapter 8. Such an analysis is vital to determine which interventions provide best value for money, as more intensive interventions, which have shown to be the most effective, will also be the most expensive.

Despite 28 studies being included in at least one of the NMAs, the maximum number included in any single NMA was 15 and many comparisons contained only a small number of

studies. Further studies are therefore required to increase precision of effect estimates, to increase power to explore effects by study quality, and to check for inconsistency between direct and indirect evidence of effectiveness. In addition, a more detailed description of the intervention in future studies, in particular of the content of the educational elements of interventions, would be helpful in allowing a finer sub-categorisation and exploration of individual educational components. Methods to incorporate individual level data into NMA analyses are now available (Saramago *et al.*, 2012), and these would be useful for exploring whether the effect of interventions vary by characteristics of study population (e.g. deprivation) and the potential impact of interventions on inequalities in prevention practices.

5.8 Chapter summary

The NMAs demonstrated that the most effective interventions varied by poison prevention practice but overall the more intensive interventions were more effective than education alone for each poison prevention practice. These analyses were carried out by fitting the standard NMA model to each poison prevention outcome separately and will inform the base case of the decision analytic model in Chapter 8. Before that, further modelling extensions of the standard NMA model will be presented first to allow for a baseline risk covariate to be taken into account in the analysis (Chapter 6) and secondly to extend the NMA model to multiple outcomes setting (Chapter 7) in order to account for correlations between the effectiveness estimates and borrow strength across outcomes.

6. ADJUSTING FOR A BASELINE RISK COVARIATES IN NETWORK META-ANALYSIS

6.1 Chapter overview

When summarising evidence to inform an economic evaluation, it is important that potential sources of heterogeneity are explored, to account for variation in the intervention effect across different populations and identify those more likely to benefit from the intervention. In the previous chapter, network meta-analysis (NMA) models were used to synthesise evidence from the example data and compare the effectiveness of poison prevention strategies in households with children. This chapter presents methods that extend the standard NMA model to account for baseline imbalances in the '*non-active*' intervention group event rate. The non-active intervention (i.e. usual care intervention in the poison prevention data described in Chapter 5) arm may represent quite different strategies in different studies. For example, 'usual care' may be 'no safety education' in one study but 'standard or usual safety education' in another. Such differences in the definition of the 'usual safety education' arise due to differences in study protocols, safety practices in the countries where studies were conducted and so on. The methods presented in this chapter can be used to account for residual heterogeneity in the definition of the non-active intervention group and help reduce both heterogeneity and possibly inconsistency in a network meta-analysis. The methods presented here have been published in '*Statistics in Medicine*' (Appendix VI - Research paper 3) (Achana *et al.*, 2013).

6.2 Introduction

In meta-analyses of clinical trials, differences in patient or trial/study level characteristics often give rise to variation in treatment effect estimates between studies - also called heterogeneity (Sutton *et al.*, 2000). Between-study variance in the treatment effects is usually taken into account through including a parameter for the residual heterogeneity in a random effects meta-analysis (Sutton *et al.*, 2000; Borenstein *et al.*, 2009). A random effects model quantifies the degree of heterogeneity but does not explain it. To explain the source of the heterogeneity, patient and study level characteristics are sometimes included in the analysis as covariates (Sutton *et al.*, 2000; Borenstein *et al.*, 2009). A trial-level covariate of interest

as a possible source of heterogeneity is the 'baseline risk' or the underlying risk of the disease. The baseline risk reflects the burden of disease in a study population and defines the average risk of a patient to experience the outcome of interest if they have not been treated (Higgins and Green, 2011). It is potentially an important proxy for the distribution of patient-level characteristics such as age, sex, medical history and disease severity that collectively influence a patient's response to treatment (Thompson *et al.*, 1997). In addition to heterogeneity, baseline imbalances between trials may also give rise to inconsistency (i.e. variability in the treatment effect between pair-wise contrasts (Cooper *et al.*, 2009) in a NMA). Therefore, adjusting for baseline covariates may have the benefit of reducing both heterogeneity and inconsistency in NMA and improve the overall model fit.

Various measures have been used for the baseline risk in meta-analyses. Examples include the observed event rate in the placebo or non-active intervention arm, the observed placebo arm log odds and the average of the observed event rates in the placebo and treatment arms (Brand and Kragt, 1992; Sharp *et al.*, 1996; Walter, 1997). However, including observed measures of baseline risk in a meta-regression can be problematic because of the measurement error in both response (i.e. treatment effect) and explanatory variables, and functional relationship between the two (Thompson *et al.*, 1997). The problem has received considerable attention in the literature with several authors proposing alternative model based solutions. Examples include the methods of McIntosh (McIntosh, 1996), Walter (Walter, 1997), Thomson et al. (Thompson *et al.*, 1997), Sharp and Thomson (Sharp and Thompson, 2000), Arends et al. (Arends *et al.*, 2000) and van Houwelingen et al. (van Houwelingen *et al.*, 2002). The objective is to extend these methods applicable to pairwise meta-analysis to network meta-analysis (NMA) where it may be of interest to adjust for baseline imbalance in the underlying risk across studies. The main reason for so doing is to reduce between-study heterogeneity and possible inconsistency in the direct and indirect trial evidence on pairwise comparisons. This objective is complicated by missing data, due to the fact that not all studies in a network may have a placebo or non-active treatment control, and thus an observed covariate value.

The remainder of the chapter is structured as follows: A review of the pair-wise meta-analysis methods for investigating the relationship between treatment effect and baseline risk is

presented in Section 6.2. Section 6.3 presents an approach which primarily extends the methods of Thompson et al. (Thompson *et al.*, 1997; Sharp and Thompson, 2000) and Arends et al. (Arends *et al.*, 2000) from pair-wise to NMA where it is of interest to adjust for the baseline risk. The methods presented complements previous general multivariate meta-regression models suggested for NMA (see Cooper et al. (Cooper *et al.*, 2009), Salanti et al. (Salanti *et al.*, 2008; Salanti *et al.*, 2009) and Stijnen et al. (Stijnen *et al.*, 2010)) by allowing for:

i) Alternative distributional assumptions to be made about the nature of the *''true''* unobserved baseline risk measure,

ii) The inclusion of trials without a non-active treatment control and hence no baseline risk measure and

iii) Allows for the $treatment \times covariate$ interactions to be exchangeable or even different (i.e. as many regression coefficients as there are treatment effects).

The methods are applied in Section 6.5 using data from two published systematic review and NMAs (McDaid *et al.*, 2010; Cooper *et al.*, 2011a) and the poison prevention data on safe storage of medicines outcome described in Chapter 5. The first example has a binary outcome and examines effectiveness of home safety education interventions to promote ownership of functional smoke alarm in households with children (Cooper *et al.*, 2011a) . The second example has a continuous outcome measure and examines the effectiveness of analgesic treatments in reducing post-operative morphine consumption in adult patients following major surgery (McDaid *et al.*, 2010). The results of these applications are presented in Section 6.6. The chapter concludes with a discussion of the findings from the example datasets, the strengths and limitations of the approach and a summary.

6.3 Review of baseline risk models for pair-wise meta-analysis

Sharp and Thompson (Sharp and Thompson, 2000) and Arends *et al.* (Arends *et al.*, 2000) both present good introductions to baseline risk adjustment and detailed review of available methods for pair-wise meta-analysis. Table 6.1 summarises the important features of six of the methods that are considered relevant to the modelling approach developed in this chapter for NMA. A common feature in these methods is to model the relationship of interest in three parts, although this was only stated explicitly in Arends et al. (Arends *et al.*, 2000). This

involves specifying in any order i) an appropriate likelihood for the data, ii) a regression model relating the *'true'* treatment effect as explanatory variable and the *'true'* baseline risk as the covariate and iii) a model for the distribution of the baseline risk across studies.

Differences between approaches have mostly arisen from slightly different strategies adopted for each part of the model. For example, Thompson et al. (Thompson *et al.*, 1997) Arends et al. (Arends *et al.*, 2000) and Sharp and Thompson (Sharp and Thompson, 2000) assumed a binomial likelihood for a binary outcome whereas McIntosh (McIntosh, 1996), Walter (Walter, 1997) and van Houwelingen et al. (van Houwelingen *et al.*, 2002) used a normal distribution to model a binary outcome measure (e.g. log odds or log odds ratio). Approximating a log odds ratio with a normal distribution can be mathematically and computationally convenient but the normality assumption may be inappropriate if there are trials in the meta-analysis with zero or small numbers of events (Sharp and Thompson, 2000). Secondly, except for the method of Walter (Walter, 1997), all the other methods assumed random study-specific effects. Walter's (Walter, 1997) model is fixed effect in that no allowance is made for any residual heterogeneity other than that explained by the baseline risk, although expecting residual heterogeneity is more realistic in most applications where it is of interest to adjust for the baseline risk.

Finally, the approaches outlined in Table 6.1 make different assumptions about the distribution of the *'true'* unobserved baseline risk across studies (McIntosh, 1996; Ghidey *et al.*, 2011). Some models assumed a vague or minimally informative normal prior distribution (e.g. Thompson et al. (Thompson *et al.*, 1997), Sharp and Thompson (Sharp and Thompson, 2000) and also in Arends et al. (Arends *et al.*, 2000)); a common parametric formulation is to assume that the baseline risk is normally distributed across trials as in McIntosh (McIntosh, 1996), van Houwelingen et al. (van Houwelingen *et al.*, 2002) and also Arends et al. (Arends *et al.*, 2000). Additionally, Arends et al. (Arends *et al.*, 2000) also proposed a more flexible model for the distribution of the baseline risk comprising a mixture of two normal distributions with different means but common between-study variance.

Table 6.1: Summary of methods for modelling the relationship between treatment effect and baseline risk in pair-wise meta-analysis with a binary outcome

Method	Outcome data	Likelihood model	Distribution of baseline risk	Method of estimation	Further notes
Method 1 (Walter, 1997)	Arm level	**Two normal distributions** (Observed treatment and control group log-odds with normal errors	None, RE	ML or WSL with bias correction	Gives only fixed effects results as no allowance for excess heterogeneity. Narrow standard errors for regression slope β
Method 2 (McIntosh, 1996)	Trial level	**Bivariate normal (BVN) approximation: -** Log-OR & control group log-odds assumed bivariate normal with known variance and covariance-matrix estimated from data)	Normal (RE on baseline risk)	ML & Bayesian	BVN assumption may be inappropriate if there are trials with small number of events. May result in more extreme estimates of slope β and lower estimates of between-study heterogeneity, σ^2
Method 3 (van Houwelingen *et al.*, 2002)	Arm level	**Bivariate normal (BVN) approximation** for binary outcome data (treatment & Control group log-odds *or* observed treatment effect & control group log-odds assumed BVN with known covariance-matrix estimated from data)	Normal (RE on baseline risk)	EM algorithm & SAS Proc Mix	Normality of baseline risk across trials may be hard to justify
Method 4a (Thompson *et al.*, 1997), 2000(Sharp and Thompson, 2000)) ; (Arends *et al.*, 2000)	Arm level	**Exact binomial model** (Observed number of events in each treatment-arm assumed binomial)	Fixed, flat prior	Bayesian	Eliminates need for zero-cell corrections Useful in situations where trials with small sampled sizes are included in the meta-analysis
Method 4b (Arends *et al.*, 2000)	Arm level	**Exact binomial model** (Observed number of events in each treatment-arm assumed binomial)	Normal (RE)	Bayesian	
Method 4c (Arends *et al.*, 2000)	Arm level	**Exact binomial model** (Observed number of events in each treatment-arm assumed binomial)	Mixture of two normal distributions	Bayesian	Eliminates need for zero-cell corrections Useful in situations where trials with small sampled sizes are included in the meta-analysis Flexible distributional assumptions for the baseline risk measure

ML = Maximum Likelihood, WLS = Weighted least square, Log-OR = Log-odds ratio; MA=Meta-analysis, FE=Fixed effects, RE=Random effects

Whether or not to assume a model for the baseline risk is a much debated issue with as yet no clear consensus among methodologists (van Houwelingen and Senn, 1999; Arends *et al.*, 2000; Sharp and Thompson, 2000; van Houwelingen *et al.*, 2002). More recently, Ghidey et al. (Ghidey *et al.*, 2007; Ghidey *et al.*, 2011) proposed semi-parametric models for the distribution of the baseline risk as well as models that do not make any distributional assumptions. In the next section, methods for the baseline risk adjustment in NMA are presented that incorporate the different assumptions about the baseline risk distribution across studies in order to assess the effect of these assumptions on parameter estimates.

6.4 Methods

6.4.1 Model with no covariate adjustment

Suppose in a meta-analysis of $i = 1, 2, \cdots, N$ studies, we have $k = A, B, \cdots, N_T$ interventions being compared with one another where N_T is the total number of interventions. Take intervention *A* as the overall baseline or reference intervention of the entire network. For a binary outcome, we assume r_{ik} events occur out of n_{ik} patients in arm *k* of study *i* according to a binomial distribution with underlying event probability p_{ik}. Standard random effects NMA for a binary outcome with no covariate previous specified in equation 2.9 of Chapter 2 is restated below:

$$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ik}) = \begin{cases} \mu_{ib}, & if \quad k = b \\ \mu_{ib} + \delta_{ibk}, & if \quad k > b \end{cases} \qquad \text{for } b = A, B, C, \cdots, \tag{6.1}$$

$$\delta_{ibk} = d_{bk} + \varepsilon_{ibk}, \ \varepsilon_{ibk} \sim \text{Normal}(0, \sigma_{bk}^2)$$

where $d_{AA} = 0$ (i.e. the intervention effect in the reference or baseline intervention for the entire network is set to 0) and $k>b$ implies intervention *k* comes alphabetically after *b*. The parameter μ_{ib} is the effect of baseline intervention *b* (log odds) in study *i* and ε_{ibk} denote a random effect indicating that the study-specific effects (log odds ratios) of intervention *k* relative to *b*, δ_{ibk}, are normally distributed with mean d_{bk} and between-study variance σ_{bk}^2.

As described in Chapter 2 Section 2.3.1, multi-arm studies can be included in the analysis under the homogenous variance assumption (i.e. $\sigma_{bk}^2 = \sigma^2$ for all *b* and *k*) so that the covariance is equal to $\frac{\sigma^2}{2}$(Lu and Ades, 2004). All other parts of the model including specification of prior distributions and estimation of parameters are as described in Section 2.3.1 of Chapter 2.

6.4.2 Extending the NMA to include a covariate for the baseline risk

Using the *'true'* but unobserved non-active control or placebo group log-odds, μ_{iA} (i.e. for *b = intervention A*) in study *i* as a measure of the baseline risk, the study-specific intervention effects in equation (6.1) can be made to depend on the baseline risk through the following regression:

$$\delta_{ibk} = d_{bk} + \beta_{bk}\left(\mu_{iA} - \overline{\mu}\right) + \varepsilon_{ibk} \tag{6.2}$$

where $\varepsilon_{ibk} \sim \text{Normal}\left(0, \sigma_{bk}^2\right)$, δ_{ibk} and σ_{bk}^2 are defined as in equation (6.1), d_{bk} is the mean effect of treatment *k* relative to baseline intervention *b* adjusted for the baseline risk and β_{bk} is the change in the log odds ratio of an event per unit change in the baseline risk for intervention *k* relative to *b* at the mean baseline risk across studies. The baseline risk covariate is centred on $\overline{\mu}$, the mean log odds in the non-active control group (*treatment A*), to improve convergence of the model (Draper and Smith, 1998). For trials with an active intervention control (i.e. baseline intervention $b \neq A$), the following substitution $d_{bk} = d_{Ak} - d_{Ab}$ is made under evidence consistency in equation (6.2):

$$\delta_{ibk} = \left(d_{Ak} - d_{Ab}\right) \times \left(\beta_{Ak} - \beta_{Ab}\right) \times \left(\mu_{iA} - \overline{\mu}\right) + \varepsilon_{ibk} \tag{6.3}$$

where ε_{ibk} and all other variables are as defined in equations (6.1) and (6.2). Although intervention *A* is not actually included in trial *i* of equation (6.3), the fundamental assumption on exchangeability means that intervention arms can be assumed to be missing at random without loss relative potency of the intervention. This assumption makes it possible to imagine that there would still be a baseline risk in studies without intervention *A* and hence, borrow strength from other studies. Therefore, no new parameters are needed for including

for example a B *versus* C study, and all other aspects of the model will remain the same. For multi-arm studies, the model takes the form of a multivariate regression to accommodate the within-study correlations between effect estimates arising from such studies. The multivariate form of equation (6.2) with bold characters denoting vectors and matrices is given by:

$$\boldsymbol{\delta}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \tag{6.4}$$

where $\boldsymbol{\varepsilon}_i \sim \text{Normal}(0, \boldsymbol{\Sigma})$, $\boldsymbol{\delta}_i$ is a vector of study-specific effects in study *i* with elements $(\delta_{i1}, \delta_{i2}, \cdots, \delta_{iNT_{i-1}})$, NT_i is the total number of treatment effects in trial *i*, and $\boldsymbol{\Sigma}$ is a variance-covariance matrix. The design matrix $\mathbf{X}_i$ contain the covariate information with entries indicating the intervention effects being estimated in trial *i* and $\boldsymbol{\beta}$ is a vector of regression coefficients including the intercept and slope terms. Following other formulations (Salanti *et al.*, 2008) as an example, consider a network of 4 studies and 3 interventions labelled *A*, *B* and *C* in which study 1 is AB (i.e. A versus B study), study 2 is AC, study 3 is ABC and study 4 is BC (i.e. no non-active control). With intervention A taken as the overall baseline treatment and assuming homogenous variances (i.e. $\sigma_{bk}^2 = \sigma^2$ for all *b* and *k*), equation (6.4) can be written in full for this network as:

$$\begin{pmatrix} \delta_{1AB} \\ \delta_{2AC} \\ \delta_{3AB} \\ \delta_{3AC} \\ \delta_{4BC} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \mu_{1A} - \bar{\mu} & 1 \\ 0 & 1 & 0 & \mu_{2A} - \bar{\mu} \\ 0 & 1 & \mu_{3A} - \bar{\mu} & 0 \\ 1 & 0 & 0 & \mu_{3A} - \bar{\mu} \\ -1 & 1 & -(\mu_{4A} - \bar{\mu}) & \mu_{1A} - \bar{\mu} \end{pmatrix} \begin{pmatrix} d_{1AB} \\ d_{2AC} \\ \beta_{AB} \\ \beta_{AC} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1AB} \\ \varepsilon_{2AC} \\ \varepsilon_{3AB} \\ \varepsilon_{3AC} \\ \varepsilon_{4BC} \end{pmatrix} \tag{6.5}$$

where $$\begin{pmatrix} \varepsilon_{1AB} \\ \varepsilon_{2AC} \\ \varepsilon_{3AB} \\ \varepsilon_{3AC} \\ \varepsilon_{4BC} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & \sigma^2/2 & 0 \\ 0 & 0 & \sigma^2/2 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix} \right)$$

and $\boldsymbol{\beta} = (d_{AB} \quad d_{AC} \quad \beta_{AB} \quad \beta_{AC})^T$ is the 4×1 matrix of regression coefficients representing the pooled effects of interventions *B* and *C* relative to intervention *A* and the effect of baseline risk on intervention effect estimates. All that remains is to specify models for the

distribution of the *''true''* baseline risk across trials and distribution of the regression coefficients. These are presented in the next section.

6.4.3 Models for the baseline risk and treatment by covariate interactions

As stated in the review of previous models presented in Section 6.3, there is no consensus in the literature about what form of distribution the baseline risk should take. The following models were specified for the distribution of the baseline risk across studies following the example in Arends et al. (Arends *et al.*, 2000):

1. Model 1 assumes that baseline risk is independent or unconstrained so that each study has its own baseline risk measure. This is equivalent to specifying a vague normal prior distribution for the baseline risk across studies: $\mu_{iA} \sim \text{Normal}(0,10^3)$.
2. Model 2 assumes that the baseline risk from each study is drawn from a normal distribution with common mean and between-study variance: $\mu_{iA} \sim \text{Normal}(\bar{\mu}, \sigma_\mu^2)$.

   Prior distributions are specified for $\bar{\mu}$ and σ_μ:

   $\bar{\mu} \sim \text{Normal}(0,10^3)$ and $\sigma_\mu \sim \text{Uniform}(0,100)$.
3. Model 3 assumes the baseline risk is drawn from a mixture of two normal distributions with a common between-study variance:

   $\mu_{iA} \sim \text{Normal}(\bar{\mu}_1, \sigma_\mu^2) \times p_1 + \text{Normal}(\bar{\mu}_2, \sigma_\mu^2) \times (1 - p_1)$ with prior distributions:

   $\bar{\mu}_1, \bar{\mu}_2 \sim \text{Normal}(0,10^3)$, $\sigma_\mu \sim \text{Uniform}(0,100)$ and $p_1 \sim \text{Dirichlet}(\alpha_c = 1)$ for $c = 1,2$.

Similar to the models for the distribution of the baseline risk, models were also specified for the treatment by covariate interactions following the example in Cooper *et al.* (Cooper *et al.*, 2009):

A. Common *treatment* × *covariate* interactions: $\beta_{Ak} = \beta$, $\beta \sim \text{Normal}(0,10^3)$.

B. Exchangeable *treatment* × *covariate* interactions: $\beta_{Ak} \sim \text{Normal}(\beta, \sigma_\beta^2)$

and $\sigma_\beta \sim \text{Uniform}(0,100)$.

C. Independent and unrelated *treatment* × *covariate* interactions: $\beta_{Ak} \sim \text{Normal}(0,10^3)$.

This implies that a total 9 models can be fitted based on the combination of assumptions about distribution of the baseline risk and the *treatment* × *covariate* interaction terms:

Model A1: Unconstrained baseline risk and common slope.

Model A2: Normal distribution for baseline risk and common slope.

Model A3: Mixture distribution for baseline risk and common slope.

Model B1: Unconstrained baseline risk and exchangeable slopes

Model B2: Normal distribution for risk and exchangeable slopes

Model B3: Mixture distribution for baseline risk and exchangeable slopes.

Model C1: Unconstrained baseline risk and independent slopes.

Model C2: Normal distribution for baseline risk and independent slopes.

Model C3: Mixture distribution for baseline risk and independent slopes.

6.4.4 Goodness of fit and model selection

In the applications which follow, adequacy of model fit to the data was assessed through the residual deviance and the Deviance Information Criteria (DIC) is used to select the best fitting model as described in Chapter 2 Section 2.3.5.

6.5 Application examples

6.5.1 Functional smoke alarm example

The data comes from a published NMA (Cooper *et al.*, 2011a) and consists of 20 randomised and non-randomised studies that evaluated the effectiveness of home safety education to increase ownership of functioning smoke alarm (FSA) systems in households with children. The outcome of interest is whether or not a household had a FSA. Thus, each study supplied arm level data on the number of households with a FSA and the total number of households surveyed. The FSA data are used here to illustrate application of the method to binary outcome data. The full data are displayed in FSA NMA paper (Cooper *et al.*, 2011a) with

Figure 6.1A displaying a network diagram for the 7 interventions and 40 data points from the 20 studies. The baseline or non-intervention arm is the usual care intervention.

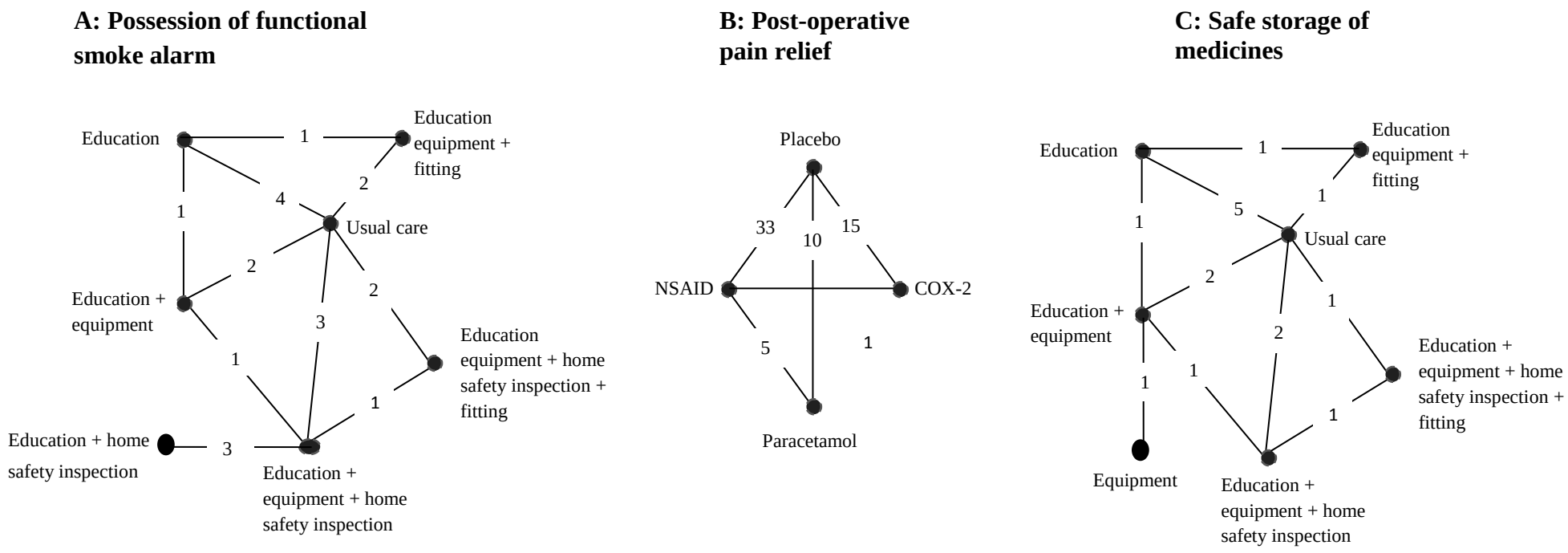


Figure 6.1: Intervention network for possession of functional smoke alarm, p-operative pain relief data and safe storage of medicines

Seven of the 20 studies did not have a usual care intervention and therefore had no baseline risk covariate. Baseline functioning smoke alarm ownership in the remaining 13 studies ranged from about 3% to about 96%. Evidence of significant inconsistency was also detected (Cooper *et al.*, 2011a) using the method of node-splitting (Dias *et al.*, 2010). Hence it is of interest to know whether baseline differences in FSA ownership across studies can explain the heterogeneity and inconsistency. For this example where the outcome is binary, a binomial likelihood was assumed for the arm-level data and NMA without covariate adjustment (model 0) fitted based on the model defined by equation (6.1). The relationship between intervention effect and baseline FSA ownership was then investigated using the methods described in Section 6.4.2. The covariate was centred on the observed mean baseline log odds of 0.81 (calculated outside WinBUGS) for FSA ownership in the 13 studies with a usual care arm. In total, 10 models were fitted (the 9 models described in Section 6.4.3 in addition to the unadjusted model) using Markov Chain Monte Carlo (MCMC) simulation in the WinBUGS software (Spiegelhalter *et al.*, 2007). The following prior distributions were used and intended to be minimally informative:

$$\sigma, \sigma_{\mu}, \sigma_{B} \sim \text{Uniform}(0,100)$$

$$\beta_{Ak}, B, d_{Ak}, \mu_{ib}, \bar{\mu} \sim \text{Normal}(0,10^{3})$$

Models were run for 100 000 iterations, discarding the first 30 000 as burn-in samples after checking the history, and autocorrelation plots for evidence that convergence of samples is adequate. There was evidence of poor convergence for the models that assumed separate/independent *treatment* × *covariate* interactions (model C1, model C2 and model C3). This may be due to i) 7 out of 20 studies not having a usual care intervention arm and hence ii) there being relatively few data points compared to the number of parameters which needed to be estimated. Therefore parameter estimates from models C1 to C3 are not presented in the results in Section 6.6.1.

6.5.2 Pain relief example

The second dataset consists of 56 RCTs with 116 data points from a published Health Technology Assessment (HTA) report (McDaid *et al.*, 2010). This HTA examined effectiveness of 3 non-opioid analgesics (paracetamol, NSAIDs or COX-2 inhibitors) and placebo in reducing morphine consumption following major surgery in adults. The outcome

of interest is the amount of morphine in milligrams (mg) consumed over a 24 hour period (continuous outcome). Each study provided arm-level information on the number of patients together with the mean 24 hour morphine consumption and its standard deviation (SD). The treatment network is given in Figure 6.1B. The dataset is available from the HTA report (McDaid *et al.*, 2010). Two of the trials have no placebo group (i.e. have compared two active treatments) and 4 are 3-arm trials. There is considerable variability in the 24 hour morphine consumption in the placebo arm of trials ranging from a low of 8.6 mg (SD 5.2 mg) to a high of 142 mg (SD 80mg). The average across the placebo group is 45.26 mg. Therefore, a sensitivity analysis was conducted in the original report (McDaid *et al.*, 2010) to investigate the effect of this baseline imbalance in morphine use on the treatment effects estimates. To include the two studies that did not have placebo, the original analysis in the published report was first carried out without these studies in order to derive an estimate of baseline morphine consumption for the two trials. The derived estimates were then included in the sensitivity analysis that adjusted for the baseline morphine use. In the analysis carried out in this chapter, however, exchangeability of the baseline effects across studies was assumed, which allowed for trials without a placebo arm and thus baseline risk measure to be included in the analysis.

Since 24 hour morphine consumption is a continuous outcome, the binomial likelihood and logistic regression model in equation (6.1) was replaced with a normal distribution for the observed arm-specific outcome (i.e. mean 24 hour morphine), Y_{ik} in treatment arm *k* of trial *i*:

$$Y_{ik} \sim \text{Normal}\left(\theta_{ik}, S_{ik}^2\right)$$

$$\theta_{ik} = \begin{cases} \mu_{ib}, & \text{if} \quad k = b \\ \mu_{ib} + \delta_{ibk}, & \text{if} \quad k > b \end{cases} \qquad \text{for } b = A, B, C, \cdots, \tag{6.6}$$

$$\delta_{ibk} = d_{bk} + \beta_{bk}\left(\mu_{iA} - \overline{\mu}\right) + \varepsilon_{ibk}$$

where $\varepsilon_{ibk} \sim \text{Normal}\left(0, \sigma_{bk}^2\right)$, θ_{ik} is the *'true'* unobserved mean morphine consumption in treatment arm *k* of trial *i* with variance, S_{ik}^2 assumed known but estimated from the data (van Houwelingen *et al.*, 2002). The baseline morphine consumption μ_{iA} was centred on 45.26 mg, the average consumption across the placebo arms to improve convergence. All other aspects

of the modelling assumptions and model fit remain the same as in example 1 except for the minimally informative prior distributions specified as follows:

$$\sigma, \sigma_{\mu}, \sigma_{B} \sim \text{Uniform}(0,100)$$

$$\beta_{Ak}, B, d_{Ak}, \mu_{ib}, \bar{\mu} \sim \text{Normal}(0,10^{3})$$

The MCMC simulations were run using WinBUGS for 100 000 iterations, discarding the first 30 000 as burn-in samples after checking the history, and autocorrelation plots for evidence that convergence of samples is adequate. The results are presented in Section 6.6.2 below.

6.5.3 Safe storage of medicines example

The poison prevention data were not available in complete form when the methods for including a baseline risk covariate in NMA were being developed. This is because the overview of reviews (Young *et al.*, 2013) described in Section 5.2.1 was still being conducted to identify studies not included in the Cochrane home safety systematic review update (Kendrick *et al.*, 2012b). It was therefore decided to use the FSA and Pain relief data described above to illustrate application of the NMA with baseline risk models developed in this Chapter. The models were latter applied to the subset of the poison prevention data relating to the safe storage of medicines outcome (see Table 5.1) when the full data became available. Safe storage of medicines was used to illustrate the methods because the decision analytic model in Chapter 8 investigated the cost-effectiveness of strategies for preventing poisonings caused by accidental exposure to medicines. This also allowed the effect of adjusting effectiveness estimates for baseline risk on the cost-effectiveness of home safety interventions to be investigated in a sensitivity analysis. Only the model with unconstrained baseline risk covariate and assuming a common regression slope (Model 1A) was fitted and the results presented in Section 6.6.3 below.

6.6 Results

6.6.1 Functional smoke alarm example

In this example, the log odds ratio was regressed on the *'true'* control group log odds (usual care intervention) taken as a measure of baseline risk. Table 6.2 displays estimates of the residual heterogeneity σ, $treatment \times covariate$ interactions (regression slopes) and model fit statistics excluding the three models (C1, C2 and C3) which showed evidence of non-convergence. Firstly, different assumptions about the distribution of the baseline risk did not seem to affect estimates of the $treatment \times covariate$ interaction terms in this case. The slopes of the regression lines are slightly steeper when minimally informative prior distributions were assumed for the baseline risk (models A1 and B1) than in models that assumed a normal baseline distribution (model A2) or a mixture of two normal distributions (model B2). Secondly, the posterior credible intervals for the slope terms included zero in all models, indicating that none of these are statistically significant. Therefore, baseline imbalance in smoke alarm distribution across studies was not significantly related to effectiveness of home safety education to promote FSA ownership in households with children (provided this analysis is powered appropriately for effects under investigation). Consequently the heterogeneity and also the inconsistency were not significantly reduced in all models that adjusted for the baseline risk compared with the unadjusted model (Table 6.2).

Using the posterior mean residual deviance as a measure of model fit to the data (Table 6.2), both adjusted and unadjusted models predicted values close to the 40 unconstrained data points in the FSA data, indicating that these models fit the data equally well. Since baseline risk appears to be unrelated to intervention effect, there was very little difference to choose between these models; hence only the results from the common slope or *treatment* × *covariate* interaction models are reported for convenience. Posterior median estimates of the slope is -0.08 (95% Credible interval (CrI); -0.41 to 0.28) from model A1, -0.03 (95% CrI; -0.41 to 0.35) from model A2 and -0.03 (95% CrI; -0.39 to 0.34) from model A3, which all indicate non-significant decrease in intervention effectiveness with increasing baseline FSA ownership.

6.6.2 Pain relief example

For pain relief data, the treatment effect, expressed as the mean difference in 24 hour morphine use was regressed on the *'true'* but unobserved 24 hour mean morphine consumption in the placebo group (taken as baseline risk measure). There were no problems with convergence of the MCMC simulations and all 9 models described in Section 6.4.3 were fitted in addition to the unadjusted model. Parameter estimates of interest and model fit statistics are presented in Table 6.3. Firstly, estimates of the regression slopes from the 9 adjusted models were all negative, suggesting evidence of increasing treatment effect with increasing baseline morphine consumption. Estimate of the common regression slope is -0.34 (95% CrI; -0.41 to -0.27) for the unconstrained baseline model (model A1), and -0.31 (95% CrI; -0.38 to -0.23) for models with normal (model A2) and mixture of two normal distributions (model A3) for baseline risk. Similar estimates of the relationship between treatment effects and baseline risk were also obtained from the independent and exchangeable slope models, but only the estimates for NSAIDS and COX-2 were statistically significant. Again, the three modelling assumptions about the distribution of the baseline risk seem to have very little impact on $treatment \times covariate$ interactions. Figure 6.2 plots treatment effects versus baseline 24-hour morphine use from the model with independent/separate slopes (model C1) for paracetamol, NSAIDS and COX-2. The plot shows: i) evidence of increasing effectiveness with increasing baseline morphine use for all three classes of analgesics; ii) NSAIDS and COX-2 are increasingly more effective than paracetamol at higher baseline morphine use, and iii) little difference between NSAIDS and COX-2. The vertical distance between the line of no effect and each treatment regression line gives an estimate of the treatment effect relative to placebo at a given baseline morphine consumption. Similarly, the relative effectiveness of any two analgesics at a given baseline morphine consumption can be obtained from the plot as the vertical distance between the two regression lines. Secondly, adjusting for the baseline risk reduced the residual heterogeneity and improves the overall model fit. From Table 6.3, the posterior mean estimate of the residual heterogeneity σ is 5.44 mg (95% CrI; 4.5 to 5.98) in the unadjusted model and 3.48mg (95% CrI; 3.24 to 4.57) in model C1, the adjusted model with the least reduction in heterogeneity. Compared to the unadjusted models there is at least a 40% reduction in between-study heterogeneity.

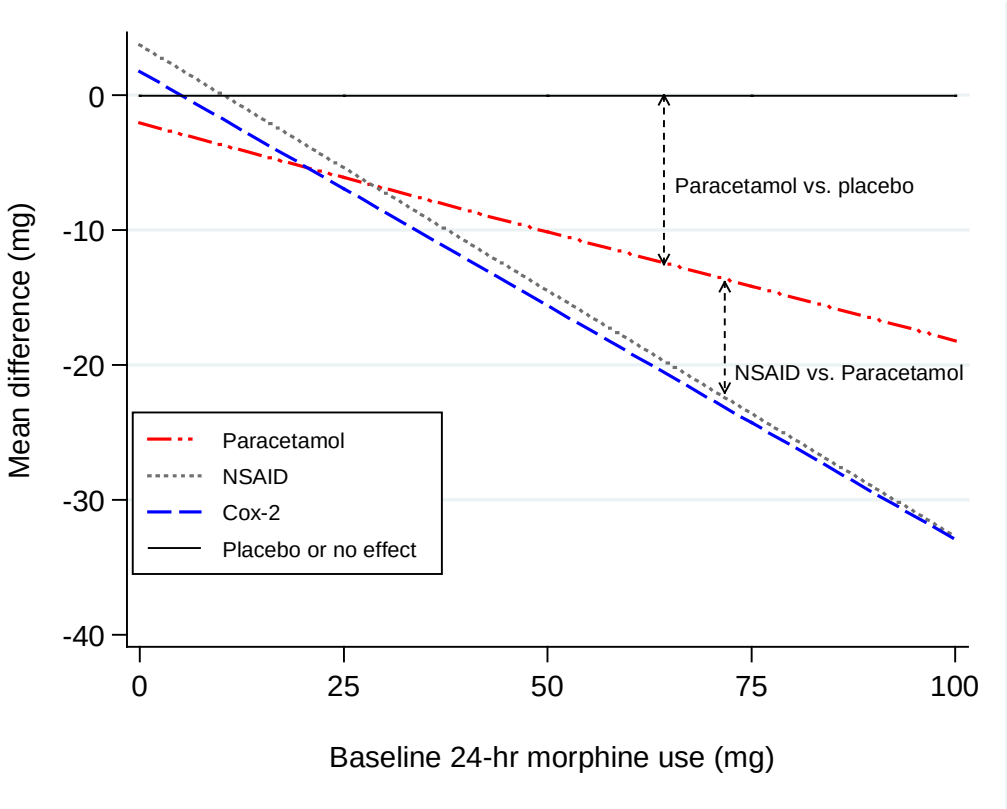


Figure 6.2: NMA adjusting for baseline morphine use assuming independent slopes for different treatment effects and unconstrained baseline risk

Table 6.2: NMA with baseline risk adjustment applied to functional smoke alarm data

Regression slopes	Model 0: Unadjusted model	Model A1: Unconstrained baseline; common slope	Model A2: Baseline normally distributed; common slope	Model A3: Baseline mixture of two normal; common slope	Model B1: Unconstrained baseline; Exchangeable slopes	Model B2: Baseline normally distributed; exchangeable slopes	Model B3:Mixture model; Exchangeable slopes
Common β	-	-0.08 (-0.41, 0.28)	-0.03 (-0.41, 0.35)	-0.03 (-0.39, 0.34)	-	-	-
Education (β_2)					-0.13 (-0.42, 0.22)	-0.09 (-0.39, 0.25)	-0.09 (-0.40, 0.27)
Education + Equipment (β_3)					0.19 (-0.59, 1.38)	0.26 (-0.57, 1.54)	0.25 (-0.57, 1.50)
Education + Equipment + HIS (β_4)					-1.08 (-2.75, 0.201)	-1.20 (-2.81, 0.163)	-1.16 (-2.63, 0.28)
Education + Equipment + Fitting (β_5)					0.26 (-0.32, 1.05)	0.35 (-0.25, 1.20)	0.35 (-0.28, 1.22)
Education + HIS (β_6)					-0.07 (-3.34, 3.02)	-0.07 (-3.45, 3.08)	-0.08 (-3.55, 2.54)
Education + Equipment + Fitting + HSI (β_7)					0.09 (-1.69, 2.48)	0.19 (-2.01, 2.66)	0.165 (-1.98, 2.38)
Mean random effects β	-	-	-	-	-0.09 (-1.55, 1.28)	-0.07 (-1.58, 1.42)	-0.07 (-1.63, 1.30)
Residual heterogeneity, σ	0.77 (0.34,1.47)	0.83 (0.39, 1.56)	0.84 (0.40, 1.59)	0.84 (0.40, 1.59)	0.59 (0.16, 1.35)	0.57 (0.167, 1.30)	0.59 (0.14, 1.38)
SD for random effects β	-	-	-	-	0.88 (0.07, 3.31)	1.02 (0.12, 3.45)	0.97 (0.09, 3.41)
Model fit statistics							
Residual deviance ($\bar{D}$)	41.72	41.49	40.86	40.99	40.33	39.98	40.04
Effective number of parameters (pD)	35.28	35.85	35.70	36.65	35.53	34.97	35.091
Deviance information criteria (DIC)	77.00	77.34	76.56		75.86	74.95	75.131

β_2= interaction term for Education relative to Usual care
SD = standard deviation in treatment effect estimate
Figures are posterior median and 95% credible intervals in brackets.

Table 6.3: NMA with baseline risk adjustment applied to pain relief data

Parameter	Model 0: Unadjusted model	Model A1: Unconstrained baseline; common slope	Model A2: Baseline normally distributed; common slope	Model A3: Baseline mixture of two normal; common slope	Model B1: Unconstrained baseline; exchangeable slopes	Model B2: Baseline normally distributed; exchangeable slopes	Model B3: Mixture model; exchangeable slopes	Model C1: Unconstrained baseline; separate slopes	Model C2: Baseline normally distributed; separate slopes	Model C3: Mixture model; separate slopes
Common β		-0.34 (-0.41, -0.27)	-0.31 (-0.38, -0.23)	-0.31 (-0.38, -0.23)	-	-	-	-	-	-
Paracetamol (β_2)					-0.22 (-0.39, 0.003)	-0.18 (-0.35, 0.04)	-0.19 (-0.36, 0.01)	-0.16 (-0.36, 0.04)	-0.15 (-0.34, 0.06)	-0.13 (-0.34, 0.10)
NSAIDS (β_3)					-0.36 (-0.44, -0.28)	-0.34 (-0.42, -0.25)	-0.33 (-0.42, -0.25)	-0.36 (-0.45, -0.28)	-0.34 (-0.43, -0.26)	-0.34 (-0.43, -0.26)
COX-2 (β_4)					-0.34 (-0.450, -0.18)	-0.27 (-0.43, -0.09)	-0.27 (-0.42, -0.09)	-0.35 (-0.53, -0.19)	-0.25 (-0.44, -0.06)	-0.26 (-0.44, -0.05)
Random effects mean (β)					-0.30 (-0.93, 0.34)	-0.26 (-1.00, 0.47)	-0.27 (-0.98, 0.48)			
Residual heterogeneity, σ	5.44 (4.50, 5.98)	3.19 (2.15, 4.47)	3.19 (2.14, 4.51)	3.22 (2.15, 4.56)	3.20 (2.15, 4.49)	3.13 (2.06, 4.50)	3.13 (2.04, 4.50)	3.28 (2.20, 4.57)	3.16 (2.06, 4.51)	3.20 (2.06, 4.53)
SD for random effects (β)	-	-	-	-	0.35 (0.01, 2.18)	0.39 (0.01, 2.37)	0.36 (0.01, 2.32)	-	-	-
Model fit statistics										
Residual deviance ($\bar{D}$)	124	119.5	121.90	121.1	117.60	121.10	120.30	116.40	120.60	119.70
Effective number of parameters (*pD*)	90.63	84.11	81.97	82.31	85.27	82.58	82.57	85.56	82.58	82.96
Deviance information criteria (*DIC*)	214.63	202.61	203.87	203.41	202.87	203.68	202.87	201.964	203.18	202.66

β = common regression slope or mean of random slope
β_2= interaction term for paracetamol relative to placebo
SD = standard deviation in treatment effect estimate
Figures are posterior median and 95% credible intervals in brackets.

6.6.3 Safe storage of medicines example

Table 6.4 presents the results from applying Model 1A to the safe storage of medicines example. The estimated regression coefficient β was -0.152 (95% CrI 0.359 to 0.115) suggesting that no evidence of an association between intervention effects and baseline rate of safety practice in households with children. Consequently, estimates of the between-study variance σ and the model fit statistics were similar in the adjusted and unadjusted models suggesting adjusting for baseline risk did reduce the heterogeneity in effect estimates across studies or improve the fit of the model to the data.

Table 6.4: NMA with baseline risk adjustment applied to safe storage of medicines data Figures are posterior median and 95% credible intervals in brackets

Regression slopes	**Model 0:** Unadjusted model	**Model A1:** Unconstrained baseline; common slope
Common β	-	-0.1634 (-0.401 to 0.127)
Residual heterogeneity, σ	0.269 (0.009 to 1.034)	0.248 (0.009 to 1.034)
Model fit statistics		
Residual deviance ($\bar{D}$)	22.99	23.31
Deviance information criteria (DIC)	148.69	150.0

6.7 Discussion

The work described in this chapter shows how methods for baseline risk covariate adjustment can be extended from pair-wise meta-analysis to NMA when it is of interest to account for differences in underlying risk across trial populations. This type of analysis can help identify potential treatment effect modifiers which may give rise to heterogeneity in effect estimates and/or inconsistency in the direct and indirect evidence on pair-wise contrasts in a network of trials. The pain relief example shows how adjusting for baseline risk can greatly reduce heterogeneity and improve overall model fit. Similar results and conclusions have been reported before, for example, by Lu et al. (Lu *et al.*, 2007) in a NMA at multiple follow-up times where the baseline effects were adjusted at different follow-up points. However, there

was no evidence of baseline effect in the two accident prevention examples (i.e. FSA and storage of medicines data), and the inconsistencies identified in Cooper *et al.* (2011) were not resolved by FSA baseline risk adjustment model. In meta-analyses of studies evaluating complex and/or public health interventions such as the accidents prevention data, interventions may not always be clearly defined and studies are often of variable quality, and conducted in populations with different characteristics. These factors can introduce heterogeneity in both meta-analyses of clinical trials and studies of non-complex and/or public health interventions, but the problem is more pronounced in public health. The FSA network included both RCTs and non-randomised observational studies both of which are of variable quality. Although care was taken to categorise the interventions appropriately, 'lumping' of interventions within categories could not be completely ruled out (Cooper *et al.*, 2011a). Lumping of interventions creates relative contrasts that are unevenly distributed across contrast and has been cited as a possible source of heterogeneity and inconsistency in NMA (Caldwell *et al.*, 2010).

The main advantage of the approach described in this chapter is that the models can be easily implemented by making simple modifications to freely available WINBUGS code for NMA (Dias *et al.*, 2011a) (see code in Appendix III). Specifying the models in WinBUGS, and analysing them using Markov Chain Monte Carlo simulation is beneficial as it allows the adjustment to be carried out without excluding trials with missing placebo or no treatment control group (and hence no baseline risk covariate). The imputation step is implemented automatically in WinBUGS through the model jointly specified by the likelihood and prior distribution placed on the 'baseline risk' (described section 6.4.3). Since parameters are considered as random variables within the Bayesian framework requiring a distribution (Ntzoufras, 2009), the 'missing covariate' is treated as any other unknown parameter to be estimated under exchangeability (see Mason (2009), page. 117). Alternatively, the analysis can also be carried outside of a Bayesian framework using multivariate meta-analysis methods (for example Stijnen et al. (Stijnen *et al.*, 2010)) fitted in standard statistical software or self-written programs. However, validity of the results obtained from either classical or Bayesian analyses will depend on appropriateness of the assumption that the non-active intervention arm of studies without a baseline risk are missing at random.

Fitting models with separate and/or exchangeable regression slopes described in Cooper et al. (Cooper *et al.*, 2009), in addition to the common slope model can be useful for assessing the appropriateness of these assumptions. For example, the common slope assumption can be tested by first calculating the difference between estimates of any two slopes in the separate slope model followed by a *probability* that this difference is greater than zero using the step function in WinBUGS (Spiegelhalter *et al.*, 2007). A two-sided P-value can then be derived using the formula $2 \times minimum(probability, 1 - probability)$. However, as shown by the FSA example, fitting models with separate/independent slopes may not always be feasible, possibly because of limited availability of data. In those circumstances, the exchangeable slope or even common slope models can be considered as a compromise (Cooper *et al.*, 2009). Under the exchangeable regression slope assumption, power is improved by borrowing strength across regression slopes which shrinks treatment effect estimates towards each other. This can have policy implications especially in a decision making context where manufacturers of alternative interventions may see the effectiveness of their products "shrink" towards that of the competitor. Also the exchangeable slope assumption can reduce heterogeneity in the effect estimates (σ), but the regression slopes themselves can be quite variable as illustrated by the pain relief example where the σ_B*'s* are larger than σ. This shows that the regression slopes are much more variable than the treatment effects and therefore a common regression coefficient may not be the best model for this example.

Finally going back to the review of pair-wise meta-analysis models presented in Section 6.3, a much-debated issue in modelling the relationship between treatment effect and baseline risk has been whether or not to assume a parametric distribution for the baseline risk and what form if any such a distribution should take. Ghidey et al.(Ghidey *et al.*, 2011) examined the issue in a published methods review paper using real and simulated data for pair-wise meta-analysis. The simulated results found no difference between models that assumed normality for the baseline risk and those that did not with both models producing robust/unbiased estimates of the regression slope when the baseline risk is normally distributed across studies (Ghidey *et al.*, 2011). However, the estimate of the regression slope was found to be less biased under the functional modelling approach when normality of the baselines was violated but the relative difference in bias was small.

The results from the approach outlined in this chapter for network meta-analysis appear consistent with the findings from Ghidey et al.(Ghidey *et al.*, 2011) and also with Arends et al (Arends *et al.*, 2000). Estimates of regression slopes from both FSA and pain relief examples were slightly less negative, and tended to shrink towards zero in models that assumed normally distributed baselines (models A2, A3, B2, B3 in Table 6.2 and models A2-A3, B2-B3 and C2-C3 in Table 6.3) compared to the unconstrained or minimally informative prior distributions for the baseline risk (models A1-C1). The effect of different distributional assumptions about the baseline risk were however, very minimal as both unconstrained and normally distributed baseline risk models produced practically identical estimates of regression slopes.

6.8 Chapter Summary

In this chapter, the standard NMA model was extended to take account of differences in the underlying risk across trial populations which may be an important source of heterogeneity in the meta-analysis. Application of the methods in a PH context was demonstrated using two examples from the accidents data on the effectiveness of interventions to increase uptake of functional smoke alarms and safe storage of medicines in households with children. The next chapter tackles another extension of the standard NMA model, this time to enable simultaneous evaluation of multiple interventions across multiple outcomes. This type of analysis is appealing because many studies and systematic reviews usually focus on broad health effects and therefore typically report several outcome measures.

7. MODELLING EFFECTIVENESS ACROSS MULTIPLE OUTCOME MEASURES

7.1 Chapter overview

It is useful, especially in a PH evaluation context where the evidence base is often scarce or limited in one way or another, that all available evidence relevant to the decision problem is taken into account when summarising evidence to inform an economic evaluation. This chapter extends the standard NMA model from the single outcome analyses presented in Chapter 5 to multiple outcome settings. These multiple outcome NMA models appropriately account for the correlation structure within the data, which is important in a decision modelling context, as correlations between effect estimates on different outcomes may have implications for estimating the net benefit associated with treatment. The methods are illustrated using the following three outcomes of the poison prevention data described in Chapter 5 (Table 5.1): i) safe storage of medicines, ii) safe storage of other household products, and iii) possession of a poison centre control telephone number. The chapter starts with a brief introduction to multivariate meta-analysis followed by a description of part of the example data that is used to illustrate the methodology developed here. Next, the methodology is described starting with a revisit to the standard NMA model introduced in Chapter 2 followed by extensions to the multiple outcomes setting. The results of applying the methods to the example problem are then presented together with a discussion and conclusions. The methods presented here have been published in '*BMC Medical Research Methodology'* (Achana *et al.*, 2014a) (Appendix VI - Research paper 4).

7.2 Introduction

One area of meta-analysis that has seen significant methodological development is the application of multivariate statistical methods for the comparison of treatments on two or more endpoints (usually known as multivariate meta-analysis) (Berkey *et al.*, 1998; Arends *et al.*, 2003; Nam *et al.*, 2003; Riley *et al.*, 2007b; Riley *et al.*, 2008; Jackson *et al.*, 2011). These methods are appealing because many studies and systematic reviews focus on broad health effects and therefore typically report several outcome measures (Berkey *et al.*, 1998;

Nam *et al.*, 2003; Kendrick *et al.*, 2007). In such instances, the multivariate approach offers some advantages over separate univariate analyses, including the ability to account for the inter-relationship between outcomes and to borrow strength across studies as well as across outcomes (Bujkiewicz *et al.*, 2013) through modelling the correlation structure (Riley *et al.*, 2007a; Riley *et al.*, 2007b). This can potentially reduce outcome reporting bias (Kirkham *et al.*, 2012) and the uncertainty with which intervention effects are estimated.

Additionally, in a decision making context where the synthesis is meant to inform a health economic evaluation, accounting for the correlations between effect estimates on different outcomes is important as the dependence between outcomes may have implications for estimating quality of life or economic consequences associated with treatment (Ades *et al.*, 2010). An example is the situation where a particularly effective treatment for a disease condition is associated with a large side effect profile. Ignoring information about the inter-relationships between beneficial and 'side effect' endpoints in such instances may have implications for quantifying the benefits associated with treatment.

When summarising effectiveness evidence, correlations between the effectiveness estimates typically arise at either within-study and/or between-study levels. At the within-study level, correlations arise mainly due to differences in patient-level characteristics. They are rarely reported in the published literature and usually have to be estimated from external sources such as individual patient level data if available or elicited from expert opinion (Riley *et al.*, 2008; Riley, 2009; Efthimiou *et al.*, 2014). At the between-study level, correlations arise from i) differences in the distribution of patient-level characteristics across studies, in which case they will be related to the within-study correlations and/or ii) differences in the distribution of other study-level characteristics such as study design, population and baseline disease severity (Rodgers *et al.*, 2009). The within-study correlations thus give an indication of the association between multiple endpoints within a study while the between-study correlations indicate how the underlying true study-specific effects on different outcomes vary jointly across studies.

A second area of rapid methodological development is NMA. NMA methods extend standard pairwise meta-analysis to enable simultaneous comparison of multiple treatments while maintaining randomisation of individual studies (Caldwell *et al.*, 2005) – see Chapter 2 Section 2.3). In brief, the method enables 'direct' evidence (i.e. evidence from studies directly comparing two interventions of interest) and 'indirect' evidence (i.e. evidence from studies that do not compare the two interventions directly) to be pooled under the assumption of evidence consistency (Dias *et al.*, 2010). Estimates of intervention effects can then be obtained, including effects between treatments not directly compared within any one individual study (Caldwell *et al.*, 2010). NMA methods thus provide a coherent framework for appraising all available evidence relevant to a specific decision problem. The results from such analyses are increasingly being used to inform economic evaluations in healthcare decision making where coherent decisions (about judicious use of scarce resource) need to be made based on sound appraisal of all available evidence.

Approaches to extend NMA methodology to multiple outcome settings have been proposed in the literature (Lu *et al.*, 2007; Welton *et al.*, 2008; Ades *et al.*, 2010; Hong *et al.*, 2013), initially focusing on mutually exclusive competing risk outcomes (Ades *et al.*, 2010) or a single outcome measured at multiple time points (Lu *et al.*, 2007; Dakin *et al.*, 2011). More recently, Efthimiou *et al.* (Efthimiou *et al.*, 2014) proposed a method for modelling multiple correlated outcomes in networks of evidence with binary outcome measures. The proposed method accounts for both the within-study and between-study correlation structure and includes a strategy for eliciting expert opinion to inform the within-study correlations. In these methods, on one hand, however, either the within-study correlations are assumed to be zero (Dakin *et al.*, 2011) or the likelihood factorised (Ades *et al.*, 2010) so that the within-study correlations do no need to be explicitly included in the model. In Efthimiou *et al.* (Efthimiou *et al.*, 2014) proposed a method, the within-study correlations are incorporated at the level of the study-specific effects which greatly increases the complexity of the model when multi-arm studies are included in the analysis. The methodology described in the reminder of this chapter contributes to this growing literature on the simultaneous evaluation of correlated outcomes in two ways:

i) Firstly, the within-study level model or likelihood for the data is developed at the treatment arm-level (rather than at the study specific treatment effect level as

explained above) in the first stage analysis labelled as model 2 in the remainder of this chapter. This greatly simplifies the likelihood for multi-arm studies since treatment arms can be considered to be independent as consequence of randomisation (Section 7.4.2).

ii) Secondly, in the second stage (labelled as model 3 in the remainder of the chapter), additional information is borrowed across outcomes based on ideas for combining evidence across human and animal studies originally proposed by DuMouchel and Harris (DuMouchel and Harris, 1983) and also revisited by Jones *et al.* (Jones *et al.*, 2009).

As will be explained later in this Chapter (Section 7.4.3, the proposed second stage analysis allows a) disconnected treatments to be incorporated as nodes in a network of evidence and b) prediction of intervention effects for outcomes where evidence from primary studies is either sparse or not directly available from any one study included in the analysis. The motivating application area here is injury prevention in children where a broad array of outcomes and intervention packages have been evaluated with the aim of increasing safety practices around the home (to ultimately reduce household injuries).

7.3 Data

The example data comes from two published systematic reviews (Kendrick *et al.*, 2012b; Young *et al.*, 2013) of the evidence on home safety education and provision of safety equipment for injury prevention in children (see Table 5.1 of Chapter 5 description of the full set of studies identified from the two reviews). The models developed in this chapter are applied to a subset of the review evidence comprising 22 studies that provided information on 3 of the 5 poison prevention outcomes described in Chapter 5:

a) Safe storage of medicines
b) Safe storage of other household products (e.g. cleaning products) and
c) Possession of a poison control centre (PCC) telephone number.

Of the remaining two poison prevention outcomes that were not considered appropriate for inclusion in the multivariate models developed here, there was insufficient data to on the one outcome (safe storage of poisonous plants). The other outcome considered in Chapter 5 but excluded from the multivariate models developed in this Chapter is safe storage of poisons.

This is a composite of safe storage of medicines and safe storage of other household products and hence is not suitable for inclusion in a model where the two components are considered as separate outcomes. Table 7.1 presents the data from 22 studies, 13 of which considered at least two of the three outcomes listed above that were included in models developed in this chapter. Of these, 8 considered storage of medicines and storage of other household products, 2 considered storage of other household products and possession of a PCC telephone number, and 3 considered all three outcome measures. Individual patient data (IPD) were available for 8 of the 13 studies, of which 7 were in a format suitable for the analysis reported here as explained by the footnotes in Table 7.1. The interventions trialled in the 22 studies were classified into 9 relatively homogenous treatment packages:

(1) Usual care (UC)
(2) Education (E)
(3) Education + provision of free/low cost equipment (E+FE)
(4) Education + provision of free/low cost equipment + home safety inspection (E+FE+HSI)
(5) Education + provision of free/low cost equipment + fitting of equipment (E+FE+F)
(6) Education + home safety inspection (E+HSI)
(7) Education + provision of free/low cost equipment + home safety inspection + fitting of equipment (E+FE+HSI+F)
(8) Education + home visit (E+HV)
(9) Provision of free/low cost equipment (FE)

Figure 7.1 shows the same network diagrams presented in Panels A, B and D of Figure 5.1 reproduced here for clarity. The network diagrams show the comparisons between the interventions that were made by individual studies and the number of comparisons in each network. All studies compared 2 intervention strategies, except Babul *et al.* (2007) (Babul *et al.*, 2007) which compared 3 strategies. Data on each outcome was not available for all interventions; i.e. for the storage of medicines and other household products outcomes, interventions ‘education plus home safety inspection’ and ‘education plus home visit’ were not investigated in any of the included studies, and for possession of a PCC number interventions, ‘education, provision of free/low cost equipment and fitting of equipment’ and provision of free/low cost equipment alone were not available.

Table 7.1: Subset of the poison prevention data displayed in Table 5.1

Comparison	First author and year of publication	IPD	Outcome information (no. of events/no. of households in control versus (vs.) treatment arm)		
			Safe storage of medicines	Safe storage of other household products	Possession of a PCC number
Usual care (1) *vs.* Education (2)	Gielen 2007	Yes	178/271 vs. 188/249	44/62 vs. 57/73	
	Nansel 2002	Yes	83/89 vs. 79/85	65/89 vs. 66/85	59/89 vs. 63/85
	Nansel 2008	Yes	72/74 vs. 140/144[†]	59/73 vs. 117/144[†]	50/59 vs. 90/119[†]
	Kelly B 1987	No	54/54 vs. 55/55	43/54 vs. 49/55	
	McDonald 2005	No	4/57 vs. 6/60	3/57 vs. 6/61	
	Kelly N 2003	No			45.56/136.68 vs. 112.95/137.63[*]
Usual care (1) *vs.* Education + free/low cost safety equipment (3)	Clamp 1998	Yes	68/82 vs. 79/83	49/82 vs. 59/83	
	Woolf 1987	No			29/143 vs. 47/119
	Woolf 1992	No		60/151 vs. 89/150	59/151 vs. 117/150
Usual care (1) *vs.* Education + equipment (3) *vs.* Education + equipment + home safety inspection (4)	Babul 2007	Yes	147/149 vs. 171/173 vs. 160/163		
Usual care (1) *vs.* Education + equipment + home safety inspection (4)	Hendrickson 2002	Yes		14/40 vs. 34/38	8/40 vs. 34/38
	Swart 2008	No	70.26/79.58 vs. 74.07/80[*]	46.86/57.96 vs. 50.87/58.27[*]	
	Kendrick 1999	Yes		317/367 vs. 322/363	
Usual care (1) *vs.* Education + equipment + fitting (5)	Watson 2005	Yes	683/738 vs. 712/762	327/669 vs. 368/693	
Usual care (1) *vs.* Education + home safety inspection (6)	Petridou 1997	No			67.26/100.12 vs. 71.08/97.83[*]
Usual care (1) *vs.* Education + equipment + home safety inspection + fitting (7)	Schwarz D 1993	No	88.42/248.37 vs. 128.16/248.37[*]		
	Phelan 2011	No			16/138 vs. 71/139
Usual care (1) vs. Home visit (8)	Johnson 2006	No			82/91 vs. 222/232[†]
Education (2) *vs.* Education + equipment (3)	Posner 2004	Yes	14/47 vs. 19/49	22/47 vs. 34/49	27/47 vs. 35/49
Education (2) vs. Education + equipment + fitting (5)	Sznajder 2003	Yes	44/49 vs. 43/45	32/41 vs. 40/48	
Education + equipment + home safety inspection (4) vs. Education + equipment + home safety inspection + fitting (7)	King J 2001	No		261/469 vs. 273/482	
Education + equipment (3) vs. Equipment (9)	Dershewitz 1979	No	22/101 vs. 20/104	1/101 vs. 0/104	

Treatment abbreviation and codes
Usual care = UC (1)
Education = E(2)
Education + free/low cost equipment = E + FE (3)
Education + equipment + home safety inspection = E + FE + HSI (4)
Education + equipment + fitting = E + FE + F (5)
Education + home safety inspection = E + HSI (6)
Education + equipment + home safety inspection + fitting = E + FE + HSI + F (7)
Education + home visit = E + HV (8)
Free/low cost equipment = FE(9)

*Effective sample size reported for cluster randomised studies after adjusting clustering, hence not whole numbers (details given in Kendrick *et al.* 2012[31])

The IPD for Gielen 2007 shows information on safe storage of medicines and safe storage of other household products was collected from different sets of households in this study (i.e. all the households that provided information for storage of medicines had missing data for safe storage of other household products and vice versa). Hence the Gielen 2007 IPD was not used to estimate the within-study correlations. †The intervention arms of Nansel 2008 and Johnson 2006 [32] comprises two groups that received different versions of a home safety intervention. The two versions were considered to be similar, hence combined into one intervention group for the analysis reported here.

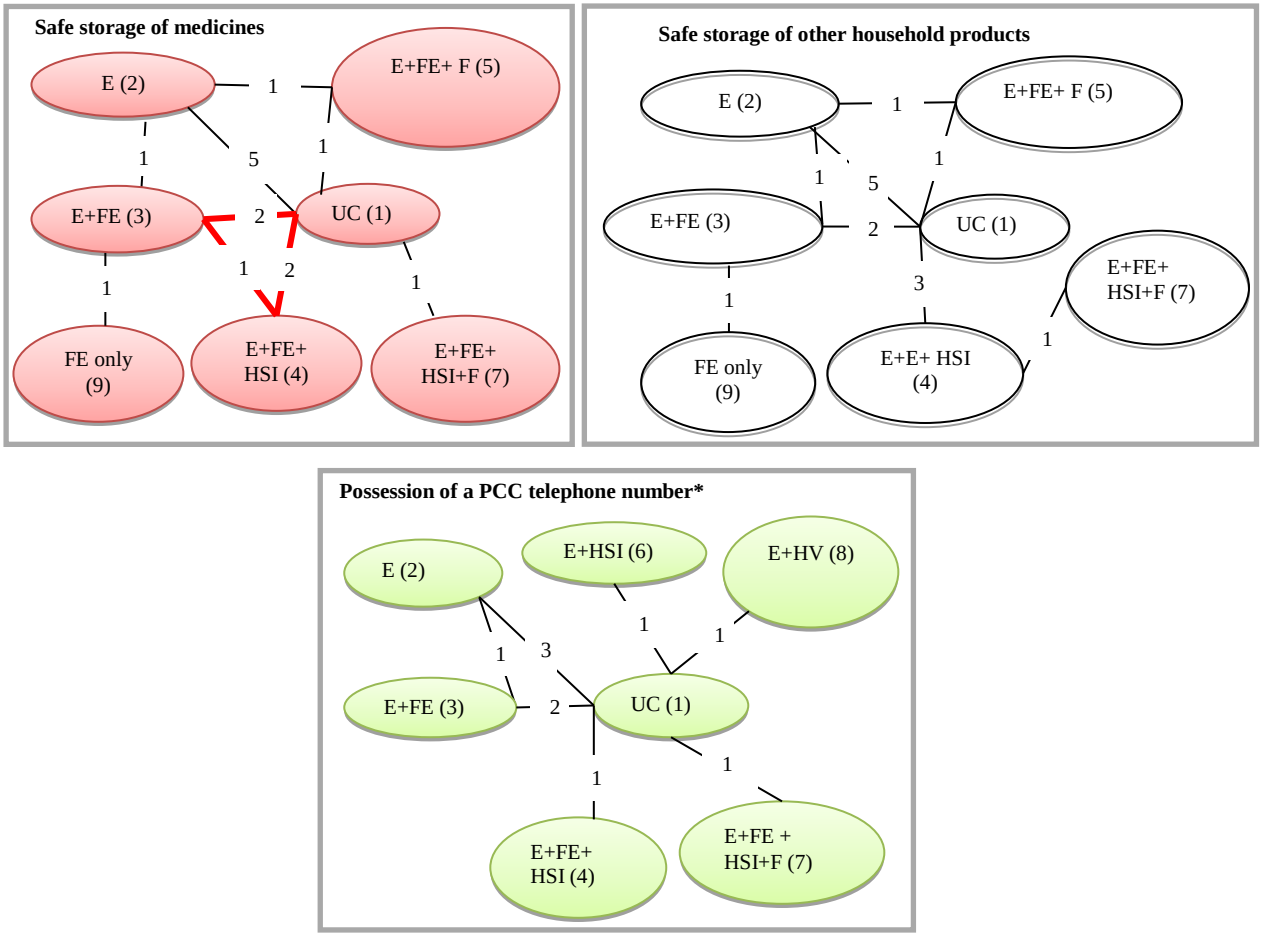


Figure 7.1: Intervention networks for the poisoning prevention outcome. These diagrams are same as Panels A, B and D of Figure 5.1.

7.4 Methods

In this section, a recap of the standard NMA model introduced in Chapter 2 Section 2.3.1 is first presented and then extended to the multiple outcome setting. Throughout the chapter, the single and multiple outcome models are referred to as univariate and multivariate NMAs respectively. Where studies report multiple outcomes, these will not be independent as each household provides information on the different outcome measures within intervention arms. The multivariate model takes this correlation structure into account by allowing the intervention effects measured by one outcome to be correlated with the intervention effects measured by other outcomes.

7.4.1 Model 1: Univariate NMA

Given arm-level binary data of the form presented in Table 7.1, the random effects NMA model of Lu and Ades (Lu and Ades, 2004) previously specified in Chapter 2 Section 2.3.1 is reproduced below for clarity. It is assumed that the occurrence of r_{ik} events out of a total of n_{ik} households in the *k*th-arm (*k*=*A*,*B*,*C*,….,) of the *i*th-study follow a binomial distribution with underlying event probability p_{ik} :

$$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ik}) = \begin{cases} \mu_{ib}, & if \quad k = b \\ \mu_{ib} + \delta_{i(bk)}, & if \quad k > b \end{cases} \qquad \text{for } b = A, B, C, \cdots, \tag{7.1}$$

$$\delta_{i(bk)} \sim \text{Normal}\left(d_{(bk)} = d_{(Ak} - d_{(Ab)}, \sigma^2_{(bk)}\right) \tag{7.2}$$

where $d_{AA} = 0$, μ_{ib} is a study-specific baseline effect (i.e. the log-odds for the control group in study *i* with baseline treatment *b*), $\delta_{i(bk)}$ is a study-specific log-odds ratio, $d_{(bk)}$ is the pooled effect of treatment *k* relative to treatment *b* (a quantity usually of interest in a meta-analysis) and $\sigma^2_{(bk)}$ is the between-study variance or heterogeneity parameter. As explained in section Chapter 2 Section 2.3.1, the homogeneous variance assumption allows for the distribution of effects (in a study with an arbitrary number of arms) to be expressed as a univariate marginal distribution and a series of univariate conditional distributions.

Specifically, for the *i*th-study with *p+1* arms and *p* treatment effect estimates relative to the reference treatment, if

$$\begin{pmatrix} \delta_{i(bk_1)} \\ \delta_{i(bk_2)} \\ \vdots \\ \delta_{i(bk_p)} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk_1)} \\ d_{(bk_2)} \\ \vdots \\ d_{(bk_p)} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{2} & \cdots & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \sigma^2 & \cdots & \frac{\sigma^2}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma^2}{2} & \frac{\sigma^2}{2} & \cdots & \sigma^2 \end{pmatrix} \right) \tag{7.3}$$

then the marginal and conditional univariate distributions for arm *j*, given the previous $1,2,\cdots,(j-1)$ arms are:

$$\delta_{i(bk_1)} \sim \text{Normal}\left(d_{(bk_1)}\ ,\ \sigma^2\right) \text{ for } j=1$$

$$\delta_{i(bk_j)} \mid \begin{pmatrix} \delta_{i(bk_1)} \\ \vdots \\ \delta_{i(bk_{j-1})} \end{pmatrix} \sim \text{Normal}\left(d_{(bk_j)} + \frac{1}{j}\sum_{t=1}^{j-1}\left(\delta_{i(bk_t)} - d_{(bk_t)}\right), \frac{(j+1)}{2j}\sigma^2 \right) \tag{7.4}$$

for $j = 2,\ldots,p$

Again, all other parts of the model including specification of prior distributions and estimation of parameters are as described in Section 2.3 of Chapter 2. Accordingly, minimally informative prior distributions were specified corresponding to a $\text{Normal}(0,10^3)$ prior distribution $d_{(bk)}$, and μ_{ib} and a $\text{Uniform}(0,2)$ prior distribution for the between-study standard deviation log odds ratio scale σ (Dias *et al.*, 2011a).

7.4.2 Model 2: Multivariate NMA

The univariate NMA model defined above is extended to the multiple outcomes settings in order to account for correlations between intervention effects on different outcomes. The method presented here is developed from the NMA with competing risks model (Ades *et al.*, 2010) where only the within-study correlations were taken into account. Their method is extended to account for the between-study correlation as well.

In Ades *et al.* (2010), a multinomial likelihood was appropriate as the three binary outcomes (relapse during treatment for Schizophrenia, discontinuation because of intolerable side effects, and discontinuation for other reasons) are mutually exclusive and event probabilities sum to 1 across outcomes. Here, however, a multinomial likelihood will not be appropriate for the example dataset because each household can have one, two or all three outcome events simultaneously, so that the event probabilities do not sum to 1 across outcomes. Instead, it was assumed that in each study *i* and for each *k*th arm, the estimates $\widehat{\theta}_{ikm}$ of the observed log-odds of an event on the *m*th outcome $(1,2,\cdots,M)$ jointly follow a multivariate normal distribution:

$$\begin{pmatrix} \hat{\theta}_{ik1} \\ \vdots \\ \hat{\theta}_{ikM} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \theta_{ik1} \\ \vdots \\ \theta_{ikM} \end{pmatrix}, \mathbf{S}_{ik} = \begin{pmatrix} s_{ik1}^2 & \cdots & r_{ik}^{1M} s_{ik1} s_{ikM} \\ & \ddots & \vdots \\ & & s_{ikM}^2 \end{pmatrix} \right)$$

$$\begin{pmatrix} \theta_{ik1} \\ \vdots \\ \theta_{ikM} \end{pmatrix} = \begin{cases} \begin{pmatrix} \mu_{ib1} \\ \vdots \\ \mu_{ibM} \end{pmatrix} & \text{if } k = b \\ \begin{pmatrix} \mu_{ib1} + \delta_{i(bk)1} \\ \vdots \\ \mu_{ibM} + \delta_{i(bk)M} \end{pmatrix} & \text{if } k > b \end{cases} \tag{7.5}$$

$$\text{for } b = A, B, C, \ldots$$

where $A, B, C, \cdots K$ are as defined in equation (7.1), (μ_{ib1}, μ_{ib2}, , μ_{ibM}) and ($\delta_{i(bk)1}$, $\delta_{i(bk)2}$, , $\delta_{i(bk)M}$) represent vectors of *'true'* baseline and study-specific effects in study *i* with baseline treatment *b* respectively. The quantities $\left(\hat{\theta}_{ik1}, \hat{\theta}_{ik2}, \cdots, \theta_{ikM}\right)$ and (θ_{ik1}, θ_{ik2}, , θ_{ikM}) represent vectors of observed and *'true'* log-odds of response in arm *k* of study *i* and $\mathbf{S}_{ik}$ is the associated within-study covariance matrix usually assumed known but estimated in practice from the data here as well (van Houwelingen *et al.*, 2002).

Elements of the vector $\left(\hat{\theta}_{ik1}, \hat{\theta}_{ik2}, \cdots, \theta_{ikM}\right)$ and the diagonal elements of $\mathbf{S}_{ik}$ were calculated using standard formulae for log-odds and variance of the log-odds (Sutton *et al.*, 2000). Continuity corrections were applied by adding 0.5 to the numerators and 1 to the denominators of studies with 0% or 100% event rate in one of the treatment arms (Dershewitz and Williamson, 1977; Kelly *et al.*, 1987). The off-diagonal elements of $\mathbf{S}_{ik}$ were

calculated from estimates of within-study correlations r_{ik}^{mn} between outcomes *m* and *n* (*m*≠*n*) in arm *k* of study *i* obtained from studies with IPD (see Box 7.1 and Table 7.3 below). The method used to estimate the correlations from the IPD is described in the implementation section below.

When summarising evidence across multiple endpoints, it is common to encounter instances where some studies do not report information for all outcomes of interest leading to incomplete vectors with missing study-specific effects for the outcomes not reported (Jackson *et al.*, 2011; Bujkiewicz *et al.*, 2013). Such studies can be included in the model under the assumption that the effects for outcomes not reported are missing at random. When implemented using the WinBUGS software, the missing study effects and standard errors are coded as NA in the data, a strategy previously outlined in Bujkiewicz *et al.* (Bujkiewicz *et al.*, 2013) and Dakin *et al.* (Dakin *et al.*, 2011). This enables WinBUGS to automatically 'impute' values with predicted distributions for the missing information under the missing at random assumption

Equation (7.5) is referred to as the within-study model and the model describing the distribution of the '*true*' effects across studies (presented below) as the between-study model following standard terminology in multivariate meta-analysis (Nam *et al.*, 2003; Riley *et al.*, 2007a; Jackson *et al.*, 2011; Mavridis and Salanti, 2012; Bujkiewicz *et al.*, 2013; Wei and Higgins, 2013a). For a network of two-arm trials, the between-study model for the *i*th study is thus given by:

$$\begin{pmatrix} \delta_{i(bk)1} \\ \vdots \\ \delta_{i(bk)M} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk)1} = d_{(Ak)1} - d_{(Ab)1} \\ \vdots \\ d_{(bk)M} = d_{(Ak)M} - d_{(Ab)M} \end{pmatrix}, \ \Sigma_{(bk)} \right)$$
$$\Sigma_{(bk)} = \begin{pmatrix} \sigma^2_{(bk)1} & \cdots & \rho^{1M}_{bk}\sigma_{(bk)1}\sigma_{(bk)M} \\ & \ddots & \vdots \\ & & \sigma^2_{(bk)M} \end{pmatrix} \tag{7.6}$$

where the *'true'* effects $\delta_{i(bk)m}$ ($m = 1, 2, \cdots, M$) jointly follow a Normal distribution with mean effects $d_{(bk)m}$. The parameters in equation (7.6) have the same interpretation as in equation (7.2) except that they are now specific to each outcome. The covariance matrix $\boldsymbol{\Sigma}_{(bk)}$ contains terms for the between-study variances, $\sigma^2_{(bk)m}$ for each outcome *m* and the between-study correlations ρ^{mn}_{bk} between effects measured by outcome *m* and *n* ($m \neq n$) specific to each *k* versus *b* comparison. Fitting the full model would thus require a large number of possibly multi-arm studies in order to make $\boldsymbol{\Sigma}_{(bk)}$ identifiable (Ades *et al.*, 2010; Jackson *et al.*, 2011). The number of parameters in $\boldsymbol{\Sigma}_{(bk)}$, can however be reduced if reasonable assumptions can be made about the covariance structure. In particular, most practical applications of NMA methods involve the assumption of a common between-study variance across treatment arms, often referred to as a homogenous variance assumption (Caldwell *et al.*, 2005; Cooper *et al.*, 2006; Carter *et al.*, 2014). Therefore, to simplify $\boldsymbol{\Sigma}_{(bk)}$ the additional assumption in this context of a common between-study correlation $\left(\rho^{mn}_{bk} = \rho^{mn}\right)$ was made leading to the following simplified between-study covariance structure for two-arm studies:

$$\begin{pmatrix} \delta_{i(bk)1} \\ \vdots \\ \delta_{i(bk)M} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk)1} = d_{(Ak)1} - d_{(Ab)1} \\ \vdots \\ d_{(bk)M} = d_{(Ak)M} - d_{(Ab)M} \end{pmatrix}, \Sigma_{(M \times M)} \right)$$
$$\Sigma_{(M \times M)} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\ \sigma_1 \sigma_M \\ & \ddots & \vdots \\ & & \sigma_M^2 \end{pmatrix} \tag{7.7}$$

where, as in the univariate case, σ_m represents the common between-study standard deviation or heterogeneity parameter specific to outcome *m*. Multi-arm studies are included in the analysis by extending equations (7.3) and (7.4) to the multiple outcome settings as shown in Appendix III. To complete model 2, μ_{ibm} and $d_{(Ak)m}$ are given minimally informative prior distributions:

$$\mu_{ibm}, d_{(Ak)m} \sim \text{Normal}\left(0, 10^3\right)$$

Prior distributions also need to be specified for $\mathbf{\Sigma}_{(M\times M)}$ which, in general, is non-trivial because of the positive definite constraint. Initially an Inverse-Wishart distribution (Spiegelhalter *et al.*, 2007) was specified:

$$\mathbf{\Sigma}_{(M\times M)} \sim \text{Inverse– Wishart}(\mathbf{K}, M)$$

where **K** is $M \times M$ scale matrix and M is the total number of outcomes. Specifying minimally informative Inverse-Wishart prior distributions is, however, problematic, especially when the amount of data is small relative to the dimensions of $\mathbf{\Sigma}_{(M\times M)}$ as is the case for the example data. Therefore, to allow for flexibility in formulating a prior distribution for $\mathbf{\Sigma}_{(M\times M)}$, the strategy outlined by (Lu and Ades, 2009) and more recently by Wei and Higgins (Wei and Higgins, 2013a) was followed to express $\mathbf{\Sigma}_{(M\times M)}$ in terms of a diagonal matrix of standard deviations $\mathbf{V}^{1/2}$ and squared positive semi-definite matrix of correlations **R** based on a separation strategy Barnard *et al.* (Barnard *et al.*, 2000)):

$$\mathbf{\Sigma} = \mathbf{V}^{1/2}\mathbf{R}\mathbf{V}^{1/2}$$

where the off-diagonal elements of **R** contain correlation terms and diagonal elements equal 1. Lu and Ades (Lu and Ades, 2009) and also Wei and Higgins (Wei and Higgins, 2013a) showed that **R** can be written as $\mathbf{R} = \mathbf{L}^{\mathbf{T}}\mathbf{L}$ using Cholesky decomposition where **L** is an upper triangular matrix. The spherical parameterization technique (Lu and Ades, 2009; Wei and Higgins, 2013a) can be used to express **R** in terms of sine and cosine functions of the elements in **L**. Using this latter technique, Uniform $(0,\pi)$ prior distributions were specified for the spherical coordinate ϕ_{mn} in the model to ensure that elements of the correlation matrix **R** lie in the interval $(1,-1)$. Finally, the elements of $\mathbf{V}^{1/2}$ correspond to the between-study standard deviation terms in $\mathbf{\Sigma}_{(M\times M)}$ and are given independent Uniform $(0,2)$ prior distributions as in the univariate case (model 1).

7.4.3 Model 3: Borrowing information outcomes

From Table 7.1, it can be seen that none of the studies had considered the interventions E+HSI and E+HV for storage of medicines and other household products. Similarly,

interventions E+FE+F and FE were not trialled by any of the included studies on possession of a PCC number. To estimate the full set of 24 basic intervention effects relative to usual care from 9 interventions on 3 outcomes, ideas originally proposed by DuMouchel and Harris (DuMouchel and Harris, 1983) and revisited by DuMouchel and Groer (DuMouchel and Groer, 1989) and Jones *et al.* (Jones *et al.*, 2009) were applied to the example problem. In doing so, it was assumed that the pooled effects of treatment *k* relative to usual care intervention $d_{(Ak)m}$, can be expressed as a sum of a treatment-specific effect α_k and an outcome-specific effect γ_m. This assumption replaces the minimally informative prior distribution $\text{Normal}(0,10^3)$ specified for $d_{(Ak)m}$ in model 2 with:

$$d_{(Ak)m} \sim \text{Normal}\left(\alpha_k + \gamma_m, \tau^2\right),\ k = B, C, \cdots, K;\ m = 1,2,\cdots,M \tag{7.8}$$

where *K* is the total number of treatments being evaluated across *M* outcomes, and for *k*=*A* (i.e. reference treatment *A*), $d_{(Ak)m}$ equal to zero. Note that on the logarithmic scale, this would imply that the ratio of any intervention effects is constant across outcomes as the γ_m cancel:

$$d_{(bk)m} = \left(d_{(Ak)m} - d_{(Ab)m}\right) \sim \text{Normal}\left(\alpha_k - \alpha_b, 2\tau^2\right) \tag{7.9}$$

Equation (7.8) thus embodies an assumption of equal or constant relative potency of treatments across outcomes which imply exchangeability of the relative effects between the non-reference/baseline interventions indicated by equation (7.9). For the example dataset, this implies that missing intervention effects for comparisons with the usual care can be predicted directly from equation (7.8) as a linear combination of α_k and γ_m assuming that each treatment effect relative to usual care is reported on at least one outcome. The missing intervention effects between non- reference/baseline treatments if required can similarly be predicted directly from the model as linear combinations of the intervention effects relative to usual care. The parameter τ controls the accuracy of the constant relative potency assumption. Values of τ close to zero would thus indicate a high degree of confidence (and support from the data) in the parallelism of effect profiles across outcomes and the constant relative potency assumption. Conversely, larger values of τ would correspond to the possibility of substantial deviation from parallelism of effect profiles across outcomes.

Multi-arm studies are included in model 3 based on the strategy outlined in Appendix II in the same way as in model 2. To complete model 3, the parameters α_k, γ_m and τ are given minimally informative prior distributions. For the mean effects, this is a normal distribution with zero mean and large variance:

$$\alpha_k, \gamma_m \sim \text{Normal}\left(0, 10^3\right).$$

The parameter τ was given a Uniform (0, 2) prior distribution, which is considered to be minimally informative on the log-odds ratio scale for most meta-analyses of effect sizes in medical applications (Dias *et al.*, 2011a). Sensitivity analyses were conducted to assess the impact of specifying alternative prior distributions for τ that are also considered minimally informative (Lambert *et al.*, 2005):

i) Normal prior distribution centred on 0 with large variance and constrained to be positive, $\tau \sim \text{Normal}\left(0, 10^2\right), \ \tau \geq 0$.

ii) Gamma prior distribution placed on the precision: $\tau^2 \sim \text{Gamma}\left(0.001, 0.001\right)$.

7.4.4 Some limitations when fitting model 3

There is a limitation to the amount of data (i.e. intervention effects relative to the usual care) on outcomes allowed to be missing for the model hyper-parameters to be identifiable. For *K* interventions and *M* outcomes, there will be $(K-1)\times M$ equations of the form in equation (7.8) that are used to estimate a total of $(K-1)+M$ hyper-parameters (i.e. $(K-1)$ of α_k and *M* of γ_m hyper-parameters). Therefore no more than $\left((K-1)\times M\right)-\left((K-1)+M\right)$ missing values in total are allowed. For example, for $K=3$ treatments and *M* = *2* outcomes, data has to be available on both outcomes for both treatment comparisons with the baseline when the prior distributions are non-informative. When a large number of data on outcomes is missing, placing informative prior distributions on the hyper-parameters can make them identifiable and improve convergence of the model.

7.5 Implementation

7.5.1 Estimating within-study correlations from IPD

A total of four models were fitted, models 1 and 3 as described above and two versions of model 2. In model 2a, an inverse-Wishart prior distribution was specified for the between-study covariance matrix $\mathbf{\Sigma}_{(M\times M)}$, whilst in model 2b, independent prior distributions were specified for the elements of $\mathbf{\Sigma}_{(M\times M)}$ based on the separation strategy in the previous section. All four models allowed for multi-arm trials to be included in the analysis. To fit the multivariate NMA models, the quantities $\left(\hat{\theta}_{ik1} \quad \hat{\theta}_{ik2} \quad \hat{\theta}_{ik3}\right)$ and the diagonals of $\mathbf{S}_{ik}$ were estimated using standard *2x2-table* formulae (Sutton *et al.*, 2000). Next, estimates of the within-study correlations were obtained from the IPD studies using the following three methods: i) Pearson correlation coefficient between the observed outcome events ii) Bootstrapping as described in Daniel and Hughes (1998), and iii) fitting a Generalised Estimating Equations model to each IPD study. The code used to fit these three estimation methods is presented Box 7.1 whilst the estimated correlations from the IPD data are presented in Table 7.2.

Box 7.1: Stata and R codes to estimate correlations between two log(odds) from IPD
Note: y1 and y2 refer to outcome 1 and outcome 2 respectively and take value 1 if a household has the desired outcome and 0 otherwise , t is a treatment group indicator taken values 0 for control group and 1 for intervention group

```
**Pearson correlation***
   pwcorr y1 y2

** GEE model **
** reshape from wide to long and fit GEE  ***
   qui reshape long y, i(id) j(outcome)

** create indicator variables, interaction terms and xtset data **
    tab outcome, gen(s)
    forvalues i=1/2 {
         gen  x1_`i' = s`i'*t
    }

** fit GEE model and estimate correlations between  pairs of log-odds **
  xtset id outcome
    xtgee y s1 s2, nocons i(id) link(logit) family(binom) robust
    estat vce, cor

** fit GEE model and estimate correlations between  pairs of log-odd ratios **
    xtgee y s1 s2 x1_1 x1_2, nocons i(id) link(logit) family(binom) corr(uns) robust
    estat vce, cor

# Correlation between pair of log-odds using bootstrap in R  (Daniel and Hughes 1998)
  Nb<-10000
  lodds1<-lodds2<-array(0,dim=Nb)
  p1<-p2<-c(rep(0,Nb))
  s<-seq(1:n)
  for (i in 1:Nb){
       sam<-sample(s, replace=T)
       new_out1<-y[sam,1]
       new_out2<-y[sam,2]
       p1[i]<-mean(new_out1)
       p2[i]<-mean(new_out2)
       lodds1[i]<-log(p1[i]/(1-p1[i]))
       lodds2[i]<-log(p2[i]/(1-p2[i]))
 }
 # pairwise correlations
  cor(lodds1,lodds2)
```

As can be seen in Table 7.2, all three methods produced broadly identical estimates of the correlations between pairs of outcome specific log-odds when fitted to each IPD study.

Table 7.2: Estimates of within-study correlations between pairs of log-odds ratios obtained from studies with IPD

IPD Study	Pearson correlation coefficient†			GEE Model‡			Bootstrap+		
	r^{12}	r^{13}	r^{23}	r^{12}	r^{13}	r^{23}	r^{12}	r^{13}	r^{23}

Nansel 2002	0.212	-0.115	0.072	0.212	-0.115	0.072	0.204	-0.000	0.104
Nansel 2008	0.235	0.013	-0.016	0.235	0.013	-0.016	0.234	0.010	-0.013
Clamp 1998	0.114			0.114			0.120		
Hendrickson 2002						0.458			
Watson 2005	0.235			0.235			0.197		
Posner 2004	0.324	-0.055	0.096	0.324	-0.055	0.096	0.396	-0.055	0.089
Sznajder 2003	-0.014			-0.014			-0.0004		
Mean (SE)	**0.184 (0.118)**	**-0.05 (0.064)**	**0.051 (0.059)**	**0.184 (0.118)**	**0.052 (0.064)**	**0.153 (0.209)**	**0.192 (0.131)**	**-0.015 (0.035)**	**0.060 (0.064)**

†Pearson correlation = correlation between observed outcome events obtained using **pwcorr** command in Stata.
‡GEE Model with unstructured correlation structure fitted in Stata using **xtgee** command (code given below).
+Bootstrap code given below.
Products, hence unable to estimate correlation from this IPD.
Outcome1 = Safe storage of medicines
Outcome2 = Safe storage of other household products
Outcome3 = Possession of a poison centre telephone number
r^{mn} = correlation between logs(odds) for outcome *m* and log(odds) for outcome *n* *(m≠n)*

7.5.2 Formulating prior distributions for the within-study correlations

The Pearson estimates were used to construct informative prior distributions for the correlation terms in $\mathbf{S}_{ik}$ of equation (7.5) by transforming the mean and standard error of each correlation term into parameters of the Uniform distribution (Lunn *et al.*, 2012):

$$r_{ik}^{mn} \sim \text{Uniform}\left(a^{mn}, b^{mn}\right)$$

where the $a^{mn} = \bar{r}^{mn} - \left(\frac{\sqrt{12 \times \text{var}\left(r^{mn}\right)}}{2}\right)$, $b^{mn} = \bar{r}^{mn} + \left(\frac{\sqrt{12 \times \text{var}\left(r^{mn}\right)}}{2}\right)$, r_{ik}^{mn} is the within-study correlation between the outcomes *m* and *n* effects measured on the log-odds scale in arm *k* of study *i*, and $\overline{r^{mn}}$ and *var* (r^{mn}) are the mean and variance of the within-study correlation between outcomes *m* and *n* effects measured by IPD respectively.

7.5.3 Assessing inconsistency in NMA with multiple outcomes

Currently, there are no known methods for assessing the inconsistency in the NMA with multiple outcomes. So the consistency of the evidence was assessed separately for each

outcome network using the node-splitting method (Dias *et al.*, 2010) which was introduced in Chapter 2 Section 2.3.3l. The results of these consistency assessments are displayed in Table 5.10. As stated in Section 5.5.3 of Chapter 5, these assessments found no evidence of conflict between the direct and indirect sources of evidence in all three outcome networks.

7.5.4 Goodness-of-fit and model selection

As explained in Chapter 2 Section 2.3.6, for analysis conducted within the Bayesian framework, the posterior mean residual deviance (McCullagh and Nelder, 1989) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) are used to assess the goodness-of-fit of models to the data and compare the fit of alternative models. These model fit assessments were however, not carried out in the analyses presented in this chapter. This is because, the multivariate models specified above contain missing information (from studies that do not report effects for all outcomes) which prevents automatic estimation and reporting of the DIC in WinBUGS. This was investigated further and the following albeit unsatisfactory explanation from http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/#q13 (accessed 5th June 2014) was found: "Why is DIC greyed out? DIC is currently greyed out in WinBUGS when one of the stochastic parents is a discrete node. The formal basis for DIC relies on approximate posterior normality for the parameter estimates and requires a plug-in estimate of each stochastic parent - for discrete nodes it is not clear which estimate to use". It was not clear how to calculate the DIC manually for a multivariate likelihood with missing data. A possible solution suggested by one reviewer (Jochem König) when the methods described in this Chapter were submitted for publication (Achana *et al.*, 2014a) is to assess model fit by calculating the residuals (i.e. differences between $d_{(Ak)m}$ of model 2 and model 3) and comparing these to the degrees of freedom available for estimating the extra parameter τ in model 3. However, because of limitations to the length of thesis, it was not possible to fully explore this further in the thesis, and will be suggested as possible opportunity for future work in the concluding chapter.

7.5.5 Model evaluation

All models described above were fitted in WinBUGS (Lunn *et al.*, 2000) using Markov Chain Monte Carlo (MCMC) simulations. The univariate models were fitted separately for each outcome using WinBUGS code available from Dias *et al.* (Dias *et al.*, 2011a). The WinBUGS code for the multivariate models is provided in Appendix IV. Convergence was assessed by examination of the trace and autocorrelation plots and the Rubin-Gelman statistic after running 400 000 simulations and discarding the first 200 000 samples as 'burn in samples'.

7.6 Results

7.6.1 Univariate and multivariate analyses

Parameters of interest were the posterior median estimate (and 95% credible intervals) of the pooled intervention effects relative to the usual care intervention, and the posterior median estimate (and 95% credible intervals) of the between-study standard deviation and correlation terms. Summary forest plots displaying effectiveness estimates relative to usual care on the odds ratio (OR) scale are presented in Figure 7.2. It can be seen that, all 4 models produced broadly similar estimates when the treatment effect is not extreme compared to the other effect estimates for the same outcome. Compared to the univariate analysis, the multivariate models produced noticeably less extreme estimates of intervention effects. This can be seen in the effect of 'education plus low cost/free equipment' on possession of PCC number being shifted towards the line of no effect from an OR of 39.35 (95% CrI 2.37 to 732.30) in model 1 to 23.55 (95% CrI 1.39, to 456.80) in model 2a, 20.37 (95% CrI 0.72, to 706.00) in model 2b and 4.20 (95% CrI 1.59 to 13.16) in model 3. Similarly, the OR for 'provision of low cost/free equipment alone' on safe storage of other household products shifted from 0.37 (95% CrI 0.00 to 15.10) in model 1 to 1.81 (95% CrI 0.63, to 5.37) in model 3.

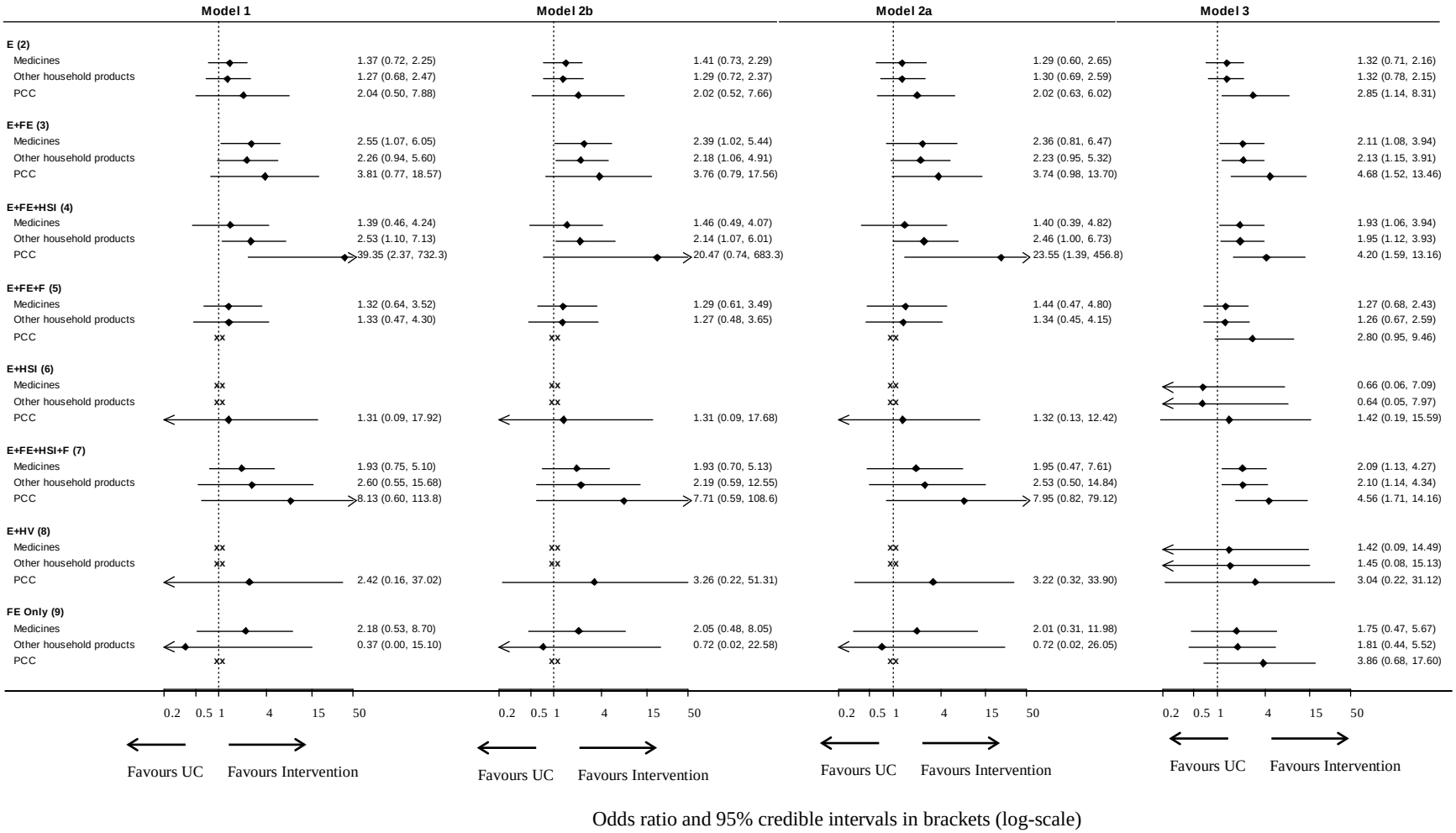


Figure 7.2: Summary forest plot of intervention effects relative to usual. Outcomes are safe storage of medicines, safe storage of other household chemicals and possession of a PCC telephone number

Mode1 1: Univariate NMA. Model 2a: Multivariate NMA (Wishart prior distribution). Model 2b: Multivariate NMA (separation strategy). Model 3: Multivariate NMA allowing for the relative effects between non-usual interventions to be exchangeable across outcomes. Effect estimate for which direct study data was not available are marked are indicated by xx on the forest plot. Intervention components: E = Education, FE=low cost/free equipment, HSI = Home safety inspection, HV = Home visit and F= Fitting of equipment

Posterior median and 95% credible intervals of the between-study standard deviations and correlations are presented in Table 7.3. The posterior medians of the between-study correlations from the multivariate models were small and estimated with considerable uncertainty (i.e. all had large variances). Estimates of the between-study standard deviations were broadly similar for the univariate NMA (model 1) and the multivariate NMA using the separation strategy (model 2b), and relatively high for multivariate NMA using the inverse-Wishart prior distribution (model 2a).

7.6.2 Model 3: Borrowing strength across outcomes

It can be seen from Figure 7.2 that the effect of 'education plus home safety inspection' and 'education plus home visit' relative to usual care intervention on safe storage of medicines and safe storage of other household products, and 'education plus low cost/free equipment plus fittings' and 'provision of low cost/free equipment alone' on possession of a PCC telephone number were only estimated in model 3 as none of the studies had trialled these interventions on the respective outcomes. In this model, estimates of relative effects between non- reference/baseline treatments were assumed to be exchangeable across outcomes, which enabled estimates to be obtained for all outcomes by predicting effects where the interventions have not been considered for the particular outcome of interest. For the intervention/outcome pair where data from trials were available, the extrapolation step had the additional effect of producing more precise estimates of the treatment effect in comparison to the models that do not assume exchangeability effects across outcomes.

Table 7.3: Posterior median and 95% credible intervals of the between-study standard deviation and correlation parameters

Parameter	Description/Prior distribution	Model 1 Univariate	Model 2a Multivariate using inverse-Wishart prior distribution for $\Sigma_{(M \times M)}$	Model 2b Multivariate using a separation strategy to specify priors for elements of $\Sigma_{(M \times M)}$	Model 3 Multivariate with extrapolation of effects across outcomes
σ_1	Between-study standard deviation: safe storage of medicines	0.26 (0.03, 1.02)	0.58 (0.33, 1.18)	0.27 (0.01, 1.08)	0.23 (0.01, 0.80)
σ_2	Between-study standard deviation: safe storage of other household products	0.56 (0.13, 1.27)	0.62 (0.35, 1.15)	0.47 (0.04, 1.18)	0.31 (0.01, 0.81)
σ_3	Between-study standard deviation: PCC	1.16 (0.57, 1.93)	0.94 (0.53, 1.99)	1.18 (0.57, 1.93)	1.08 (0.58, 1.85)
τ	Primary analysis: $\tau \sim$Uniform *(0, 2)*				0.10 (0.01, 0.53)
τ	Sensitivity analysis: $\tau \sim$ Normal(0, 10^2), $\tau \geq 0$				0.11 (0.00, 0.56)
τ	Sensitivity analysis: $\tau^2 \sim$ Inverse − Gamma (0.001, 0.001)				0.08 (0.02, 0.36)
ρ^{12}	Between-study correlation[medicines, other household products]		0.03 (−0.73, 0.76)	0.05 (−1.00, 1.00)	0.45 (−0.99, 1.00)
ρ^{13}	Between-study correlation[medicines, PCC]		0.06 (−0.80, 0.81)	0.20 (−1.00, 1.00)	0.50 (−0.98, 1.00)
ρ^{23}	Between-study correlation[Other household products , PCC]		0.08 (−0.81, 0.83)	0.13 (−0.97, 0.98)	0.60 (−0.87, 0.99)

7.6.3 Sensitivity analysis

The results of the sensitivity analyses are presented in Figure 7.3. The posterior median and 95% credible intervals of intervention effects relative to usual care were unaffected by placing alternative minimally informative prior distributions on τ. The posterior median and credible intervals for τ (Table 7.3) were similarly not sensitive to the choice of prior distribution placed on τ in the primary and sensitivity analyses. The posterior median estimates were all close to zero, which suggest that assumptions about the parallelism of effect profiles across outcomes is supported by the data.

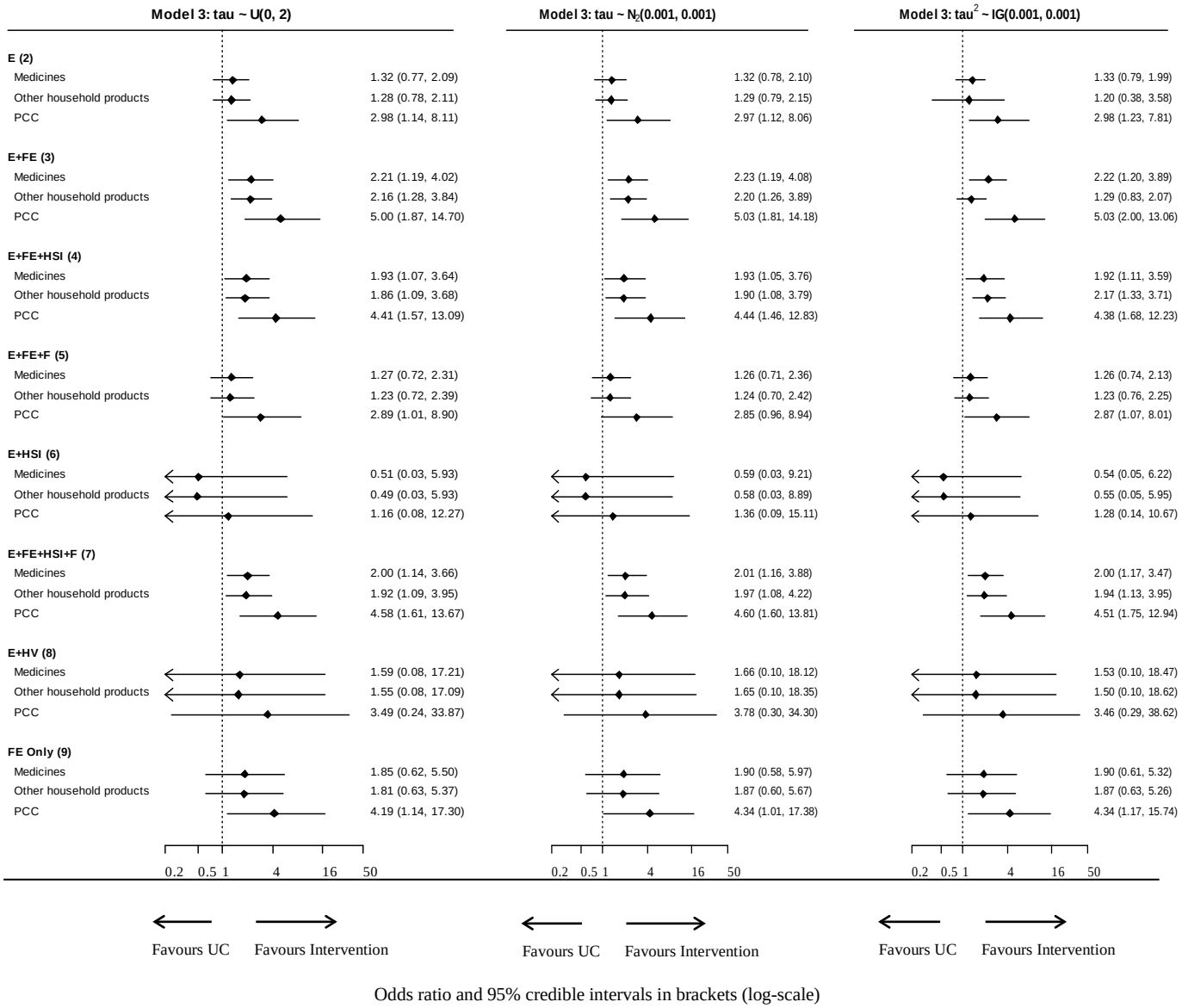


Figure 7.3: Results of sensitivity analysis to different specifications of prior distributions for τ in model 3. Intervention components: E = Education, FE=Free equipment, HSI = Home safety inspection, HV = Home visit and F= Fitting of equipment. IG = Inverse-Gamma distribution, N=Normal distribution and U = Uniform distribution

7.7 Discussion

In this chapter, methods were developed for simultaneous comparison of multiple treatments across multiple outcome measures while preserving the internal randomisation of individual studies. The methods may be viewed as an extension of Ades *et al.'s* (2010) NMA with competing risks paper (Ades *et al.*, 2010) wherein only the within-study correlation is taken into account. Their method was extended to account for the dependency between outcome effects across studies as well as within-studies. In this particular application of the multivariate approach to the example dataset, accounting for the correlation between outcomes alone (models 2a and 2b) did not reduce the uncertainty around estimates of intervention effects compared to analysing each outcome separately (model 1). Assuming that intervention effects are exchangeable across outcome did however lead to a modest reduction in uncertainty around effectiveness estimates (model 3).

The between-study correlations were estimated with considerable uncertainty (Table 7.3) and appear to have little impact on overall effect estimates. This may be because the between-study correlation arises due to, among other things, differences in study-level characteristics that also give rise to between-study heterogeneity in a meta-analysis. Based on a criterion outlined in Spiegelhalter *et al.* (Spiegelhalter *et al.*, 2004) the posterior median estimates of the between-study standard deviations, σ_1 and σ_2 on the log odds ratio scale (Table 7.3) could be interpreted as indicating evidence of low to moderate heterogeneity for storage of medicines and storage of other household products outcomes. Only the estimates for possession of poison control centre number exhibited a considerable degree of heterogeneity. Correspondingly, the posterior medians of the between-study correlations were small. There was therefore very little gain (in terms of increasing the precision of estimates) from formulating the between-study covariance structure described for the analysis presented here. Accounting for the between-study correlation is likely to be beneficial in situations where the between-study variance (heterogeneity) is large relative to within-study variances.

The within-study correlations were incorporated through the arm-specific effects (log-odds) rather than the study-specific treatment difference (log-odds ratio) as is often done in multivariate meta-analysis (Berkey *et al.*, 1998; Arends *et al.*, 2003; Riley, 2009; Mavridis

and Salanti, 2012). This approach greatly simplifies the likelihood for multi-arm studies because treatment arms can be considered independent as a consequence of randomisation. Hence, there is no requirement to account for the additional correlations between effect estimates which share a common comparator treatment in the model likelihood (Franchini *et al.*, 2012). The arm-based approach is also likely to be useful when (as is typical with many practical application of multivariate meta-analysis) the within-study correlations are not available (Riley, 2009; Kirkham *et al.*, 2012; Bujkiewicz *et al.*, 2013; Wei and Higgins, 2013b) and have to be obtained from an external source such as expert opinions (Efthimiou *et al.*, 2014). In such situations, formulating questions about correlations between outcome-specific event probabilities (which can be used directly in an arm-based approach) is more likely to be intuitive and easily understood by non-statistician healthcare experts than questions about correlations between intervention effects. The correlations between the intervention effects if required can easily be obtained from the correlations between the outcomes (Wei and Higgins, 2013b; Efthimiou *et al.*, 2014).

At the between-study level, a common correlation structure was assumed in equation (7.7) across treatments in addition to the common variance assumption underlying most practical application of NMA methods. The common correlation assumption implies that if several separate multivariate meta-analyses were conducted with the same outcomes, each with a different set of *k* versus *b* comparison, the assumption is that the between-study correlations would be the same across the different sets of *bk* comparisons. This structure was suggested to simplify the covariance structure and reduce the number of parameters in the model.

Initially an inverse-Wishart prior distribution was specified for the between-study covariance matrix $\mathbf{\Sigma}_{(M\times M)}$. However, this prior distribution is believed to be influential due to the small number of studies in the example dataset relative to the number of outcomes. Under these conditions, the inverse-Wishart prior distribution produced upwardly-biased estimates of σ_1 and σ_2 and downward bias in the estimate for σ_3 when compared to the corresponding estimates obtained from the univariate model (Table 7.3). These findings are consistent with observations in the univariate case where the use of a Gamma prior distribution (which is the univariate analogue of the Inverse-Wishart prior distribution) can lead to an overestimation of the heterogeneity parameter when the true value is close to 0 (Lambert *et al.*, 2005; Gelman,

2006). As an alternative to an inverse-Wishart prior distribution, the spherical decomposition technique suggested by Lu and Ades (Lu and Ades, 2009) was followed to specify prior distributions for the correlation and standard deviations terms in $\mathbf{\Sigma}_{(M\times M)}$. This parameterization offered greater flexibility in formulating independent prior distributions for the standard deviation and correlation terms in $\mathbf{\Sigma}_{(M\times M)}$.

An obvious limitation to implementation of the multivariate models presented in this chapter is the limited availability of data including i) the problem of missing within-study correlations and ii) the requirement for a relatively large number of studies to estimate all model parameters. The problem of missing within-study correlations has traditionally hampered the widespread application of multivariate meta-analysis (Riley *et al.*, 2007b; Riley, 2009; Bujkiewicz *et al.*, 2013). In the example data used in this chapter, IPD was available from a proportion of the included studies, allowing the within-study correlations to be estimated. Alternative approaches to dealing with missing within-study correlations when IPD is not available include: i) using the observed correlation from the summary study-specific effects (Kirkham *et al.*, 2012), ii) eliciting information about the correlations from external sources such as clinical experts (Efthimiou *et al.*, 2014) and iii) specifying 'vague' prior distributions for analysis conducted within a Bayesian framework (Nam *et al.*, 2003).

The second data issue concerns the number of studies needed to estimate the full unstructured between-study covariance matrix presented in equation (7.6). It is anticipated that a large number of multi-arm studies reporting across the three outcomes will be needed to identify $\mathbf{\Sigma}_{(bk)}$ and estimate all model parameters. This can be problematic considering the fact that most applications of network meta-analysis typically include mostly two-arm studies with very small numbers of multi-arm studies. Even with the simplification of the between-study covariance matrix given in equation (7.5), a relatively large number of studies in comparison to the total number of outcomes being considered may still be needed. It is, however, not clear how many studies should be considered large enough for a NMA with multiple outcomes. As a guide, Wei and Higgins (Wei and Higgins, 2013a) recently estimated 15, 27 and 42 studies as a minimum for multivariate pairwise meta-analysis with two, three and four-outcomes respectively. Hence, an even larger number of studies may be required for the NMA with multiple outcomes.

Another limitation of the multivariate models presented here is that they rely on the normal approximation to binomial distribution to incorporate the within-study correlations in the model. The normal approximation frequently fails and may not provide adequate fit to the data in the presence of studies with zero or a small number of events, necessitating use of continuity corrections. An exact binomial distribution was not used to model the within-study likelihood specified in section 7.4.2 because the primary interest was to develop models for summary binary data where outcomes are not mutually exclusive, and where it is not reasonable to assume that within-study correlations are zero so that the likelihood factorises easily as in Arends *et al.* (Arends *et al.*, 2003). Further methodological investigations into modelling multivariate summary data that is not normally distributed will therefore be useful. An example is provided in Chu *et al.* (Chu *et al.*, 2009) where parameterization of the within-study model enabled the special case of diagnostic sensitivity and specificity to be jointly modelled with disease prevalence using a trivariate binomial likelihood. In the interim, an alternative formulation which bypasses the need for approximating normal distributions is to model the IPD directly where this is available. This will require extending Saramago *et al.'s* (Saramago *et al.*, 2012) NMA model with aggregate and individual participant level data from single outcome to multiple outcome settings.

The consistency of each outcome network was assessed separately using the method of node splitting (Dias *et al.*, 2010) and found no evidence of conflict between the direct and indirect sources on the pairwise contrasts that have both sources of evidence. The consistency of the multivariate estimates partly was however not assessed because the current methods may not easily generalise to the multivariate case. Extensions of the node-split method to multiple outcome networks and the effect of jointly synthesising evidence across multiple endpoints on evidence consistency are currently being investigated in a simulation study.

The initial motivation for a multiple outcome NMA was to estimate intervention effects for all the outcomes, including effects of interventions on outcomes not considered by any of the studies included in the analysis. This requires the correlation structure between effects on multiple outcomes to be appropriately modelled and also implementing a mechanism of "borrowing strength" across outcomes through the assumption of exchangeability of the random effect across outcomes. This implies a priori assumption that outcomes are related but different and that there is no way of knowing the order of magnitude of effects on outcomes. If this assumption does not hold, it may potentially lead to worse or more biased

effectiveness estimates. In the example, the outcomes are similar and measured on the same scale. It would be clearly inappropriate to assume that intervention effects are exchangeable across outcomes that are different in some important respects such as being measured on different scales (e.g. where one outcome reports a weighted mean difference and another outcome reports a log-odds ratio) as such estimates will differ in terms of the precision with which they are estimated.

7.8 Chapter summary

Methods for simultaneous comparison of multiple treatments across multiple outcome measures while preserving the internal randomisation of individual studies were presented in this chapter. Application of the method to the poison prevention data yielded similar point estimates of treatment effect to those obtained from a univariate NMA but the uncertainty around the multivariate estimates increased or decreased depending on the prior distributions specified for the between-study covariance structure. Application of the results from both methods in the economic evaluation of PH interventions will be demonstrated through sensitivity analysis in Chapter 8.

8. MEDICINAL POISONS DECISION MODEL

8.1 Chapter overview

This chapter describes a probabilistic decision model developed to evaluate the cost-effectiveness of home safety interventions to prevent accidental poisonings in pre-school children introduced in Chapter 1 Section 1.3. Results of the network meta-analysis conducted in Chapter 5 are used to inform the effectiveness estimates for the base case analysis. Sensitivity analyses are conducted using results from fitting the methods developed in Chapter 6 (Achana *et al.*, 2013) and Chapter 7 (Achana *et al.*, 2014a) to the example data to investigate the effect of adjusting for the control group event rate and accounting for the correlations between interventions effects across multiple outcomes on the cost-effectiveness estimates.

8.2 Base case analysis

The base case analysis is developed for the prevention of unintentional poisons resulting from exposure to medicinal substances. The cost-effectiveness of interventions to prevent non-medicinal poisonings is investigated through sensitivity analysis. Table 8.1 summarises the important features of the base analysis. The population of interest is pre-school children (i.e. children aged 0-4 years old), the exposure variable is safe storage of medicines in the home and the reference intervention for the purpose of estimating cost-effectiveness is usual care, which is defined to include usual or no safety education. Safe storage is defined as storage above adult eye level or in locked cabinets and/or drawers so that they are out of reach of children (Kendrick *et al.*, 2008). The outcome variable is medically attended for unintentional ingestion of medicinal substance.

Table 8.1: Base case analysis for safe storage of medicines

Parameter	Description
Type of economic evaluation	Cost-effectiveness analysis
Modelled population	Preschool children (0-4 years of age)
Exposure variable	Safe storage of medicines
Outcome event	Accidental ingestion of medicinal substance
Unit of analysis	Household with at least one child
Perspective on costs	UK NHS and Personal and Social Services (PSS)
Health outcomes (Utilities)	Quality Adjusted-Life Year (QALY)
Base year for calculating costs/prices	2012
Currency unit	British pound (£)
Hypothetical cohort size	100,000 households
Effectiveness evidence (Chapter 5)	13 studies
Comparator or reference intervention	Usual care[1] intervention
Number of intervention strategies (Chapter 5)	7
Number of health states (Markov model)	6
Cycle length for Markov model	1 year
Half-cycle correction	No
Time horizon	100
Discount rate for costs	3.5%
Discount rate for utilities	3.5%

1Usual care (UC) intervention is defined to include usual safety education or no safety education.

The unit of analysis is the household when estimating the relative effectiveness of interventions but the individual when modelling cost-effectiveness. Households were chosen as the unit of analysis because households were the primary unit of randomisation in the effectiveness evidence described in Chapter 5 Section 5.2. It is assumed that interventions act to increase the proportion of households with safe storage (at a rate determined by the relative effectiveness estimates) above and beyond the baseline prevalence of safety practices in the UK. Households with safe storage are assumed to present a lower risk of accidental ingestion compared to households without safe storage. The aim of the decision analysis then is to estimate the likelihood of an unintentional poisoning event in pre-school children for a given intervention and use this to estimate NHS costs and consequences/benefits associated with treating such events over the life time of the individual.

The model has a cycle length of 1 year and a life–time horizon of 100 years. These were chosen to ensure that individuals remain in health-states long enough for the costs and consequences associated with poisoning injury to be realised and to ensure that long-term costs and consequences associated with injury during the pre-school years are captured. Health outcomes are expressed using quality-adjusted life years (QALYs). Costs are considered from the perspective of UK NHS and personal social services (PSS). Costs and QALYs are discounted over the time horizon at 3.5% per annum in line with the NICE methods guidance for public health evaluations (NICE, 2012). Therefore to costs incurred by parents for lost time from work to care for child and other similar non-medical or health sector costs are not included in the analysis. The base year is set to 2012 for purpose of estimating costs and inflating prices incurred in the past.

8.3 Model structure

A cohort simulation model is developed to estimate the life-time costs and QALYs associated with home safety intervention compared to usual care intervention. The structure comprises of both decision tree and Markov models which were described in Chapter 3 and is based on two previous decision analytic models developed to evaluate the cost-effectiveness of smoke alarm give-away schemes on health outcomes in children (Pitt *et al.*, 2009; Saramago *et al.*, 2014). Figure 8.1 shows the structure of the model. It comprises three distinct but interlinked sub-models:

i) First stage decision tree model referred to as the ‘intervention model’
ii) Second stage Markov state transition model referred to as the ‘preschool model’ and
iii) A third stage Markov state transition model referred to as the ‘long-term model’.

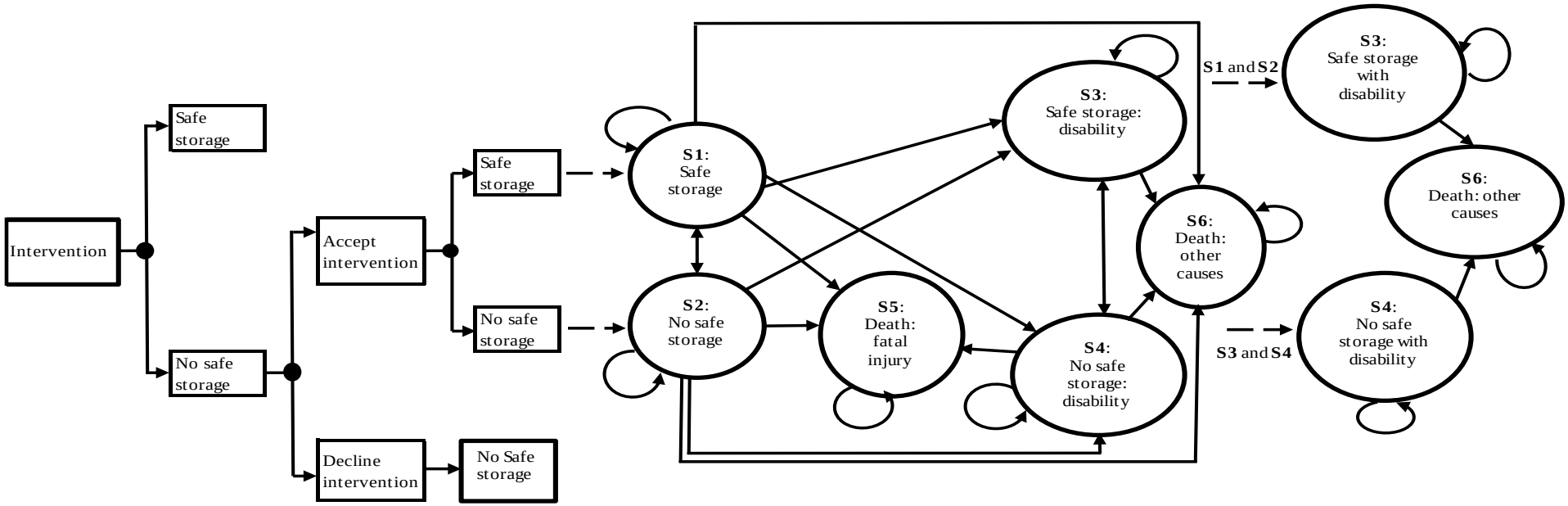


Figure 8.1: Decision model structure. Arrow heads indicate direction of movement of households/individuals through the model

8.3.1 Intervention model

The first part of the decision model uses a decision tree introduced in Chapter 3 Section 3.3.2 to estimate costs and outcomes associated with the interventions being evaluated. The costing element concerns mainly the cost of providing the intervention strategies while the outcomes are the proportion of households in each intervention cohort with or without safe storage. Costs and outcomes were estimated by taking account of the baseline prevalence of safety practices in the general population, the acceptance rate of interventions in the modelled population and the relative effectiveness of interventions. In the model, interventions act to increase the proportion of households with safe storage (at a rate determined by the relative effectiveness estimates) above and beyond the baseline prevalence of safe storage practices in the UK. The outcomes from the intervention model serve as inputs for the second stage 'preschool' model which is described below.

8.3.2 Stage 2: Preschool model

The preschool model utilizes the Markov model structure introduced in Chapter 3 Section 3.3.3 to estimate the costs and QALYs associated with each intervention strategy being evaluated in the first 5 years of life (ages 0-4years). There are 6 distinct health states as shown in Figure 8.1: safe storage (S1), no safe storage (S2), safe storage with disability (S3), no safe storage with disability (S4), death from poisoning injury (S5) and death from causes unrelated to poisoning injury (S6). As stated above, the intervention model provides the input parameters for the pre-school model. This implies that households initially enter the preschool model through one of two states (S1 - Safe storage or S2 - No safe storage) based on the outcomes from the intervention model. Subsequent movement of individuals between health states permitted in the model are indicated by lines joining the respective health states with the arrows pointing in the direction of movement (Figure 8.1). Individuals move between health states in yearly cycles or remain in their respective states. The transition probabilities governing these movements are estimated based on the available evidence assembled from the literature and described in section 8.4. It is assumed that any disability or chronic health condition caused by a poisoning injury is permanent and persists for the rest of the individual's life (hence individuals in S3 and S4 cannot return to S1 or S2). The two death

states are S5 and S6 and represent absorbing states so that individuals cannot leave once in these states.

The model does not account for multiple events in households with two or more pre-school children in the same year. In other words, although it is possible for households to have more than one child, it is assumed that only one child can experience an accidental poisoning event in any one single year. Observed or suspected cases of accidental ingestion are assumed to present to the emergency department for assessment and or treatment. No attempt is made to account for seeking outside the standard emergency response system (defined here as ambulance to emergency department).

On arrival at the emergency department, cases are assumed to be triaged on initial assessment as either serious/toxic or not serious/not toxic. This categorisation was arrived at with input from specialists in emergency department medicine (Personal communication Dr Philip Miller). There was no information available from reviewing the literature to classify poisoning cases as minor, moderate and major/severe. There was, however, information from Hospital Episode Statistics (Health & Social Care Information Centre, 2013b) on the proportion of cases admitted for an in-patient stay following a period of assessment at emergency department. Based on this information, it was assumed that only cases triaged as serious/toxic are admitted for a period of inpatient stay and treatment. Consequently, minor cases are assumed to be treated and discharged from the emergency department. Finally, the preschool model allows for the possibility of serious poisoning cases to be fatal, chronic (i.e. lead to long-term health condition) or completely resolved (with a return to full health).

8.3.3 Stage 3: Long-term model

This part of the model applies to individuals aged from 5 years and accounts for the lifetime costs and QALYs associated with chronic injury in the first 5 years of life and assumed to last the life time of the individual. A Markov structure with 3 health states (S3, S4 and S6 defined above) as shown in Figure 8.1 is used to model the long-term costs and consequences associated with injuries that occur during the preschool years. No cases of accidental ingestion are assumed to occur during this period (i.e. at ages above 5 years), or if they do

occur they are not taken into account in the model. Therefore, the only possible transitions in this part of the model are movements of individuals from disability states to death from causes unrelated to the poisoning injury. This also implies that deaths from a poisoning related injury are assumed to occur before and not after age 5 in the model.

8.4 Evidence for base case analysis

This section describes the sources of evidence used for the intervention and preschool models described above. Table 8.2 summarises the parameters of the model and the underlying evidence base informing each parameter.

Table 8.2: Probabilities used in the base case (medicinal poisons) decision analysis

Parameter	Description	Source	Value	Distribution
pSafe$^{(1)}$	Baseline prevalence of safe storage of medicines?	Prevalence rate among community controls from study A	1527/ 2033 = 75%	Beta
pAccept$^{(k)}$	Probability of accepting the intervention *k* (*k = 2, …., 7*)	Assumption based on value in Functional smoke alarm model(Saramago *et al.*, 2014). Set the same for all interventions.	90%	Fixed
pEff$^{((k)}$	Probability intervention *k* is effective, *k* (*k= 1,2, …., 7*)	NMA analysis (Achana et al 2014) : (1) Usual care (2) Education (3) Education + provision of low cost/free equipment (4) Education + provision of low cost/free equipment + home safety inspection (5) Education + provision of low cost/free equipment + fitting (6) Education + provision of low cost/free equipment + home safety inspection + fitting (7) Provision of low cost/free low cost/free equipment	Mean (95% CrI) =0.90 (0.84, 0.94) =0.87 (0.83, 0.91) =0.95 (0.89, 0.98) =0.90 (0.76, 0.96) =0.90 (0.81, 0.96) =0.93 (0.83, 0.97) =0.94 (0.78, 0.98)	
pIngest$^{(1)}$	Probability of accidental exposure/ingestion	Poisoning cases in pre-school children (n=10837), UK pre-school population in 2005-2009 period (n=3599180) from (Orton et al 2014, unpublished). The numerator (n=10837*0.6 = 6502) was derived based on information (Tyrrell *et al.*, 2012) suggesting that 1316 (60%) of the 2193 medically reported poisonings identified in the THIN database were due to ingestion of a medicinal substance	6502/3599180 = 0.181%	Beta
orIngest	Relative risk of exposure to a medicinal substance comparing children with a poisoning to community controls	Study A: Community controlled adjusted analysis OR=1.67 (95% CI 1.23 to 2.27)	-0.513 (0.155)*	Normal
pAmb	Probability of using emergency ambulance	Hospital Episode Statistics (2012b): 24.2% of all cases arrived by emergency transfer (ambulance/helicopter).	0.242	Fixed
pAdmit	Probability of in-patient admission following a medicinal poisoning injury (ICD-10: X40-X44)	Hospital Episode Statistics, 2012-2013)(Health & Social Care Information Centre, 2013b): Number of poisoning cases (X40-X44) admitted in 0-4yr olds (period 2012-2013) in England= 3909. Scaled up by a factor of 1.163 (i.e. 3909*1.163= 4546 cases for whole of UK) based on mid-2012 population estimates for UK and England, ONS 2012a(Office for National Statistics, 2013).	4546/ 6502 = 69.92%	Beta
pSevere	Probability of severe injury	NPDS 2012 report (Mowry *et al.*, 2013), Table 13, page 968) 1.91% of major poisoning cases (across all age groups) resulted in a permanent health condition . Numerator = 0.019* 4546 =87.	87/ 4546= 1.91%	Beta
pFatal	Probability of fatal injury	UK mortality statistics (Office for National Statistics). 1 fatality from medicinal poisonings in 0-4 years old (assumed fatality occur after a long inpatient stay).	1/86 = 1.16%	Beta
pDead	UK mortality statistics	UK mortality statistics(Office for National Statistics, 2010)		Normal

*log odds ratio and standard error in brackets

8.4.1 Intervention strategies

The evidence on intervention effectiveness was derived primarily from the Cochrane home safety systematic review (Kendrick *et al.*, 2012b) supplemented with evidence from the overview of reviews (Young *et al.*, 2013). Chapter 5 describes the network meta-analyses that were used to summarise the evidence on 9 interventions across 5 poison prevention outcomes. Seven of the 9 interventions were included in the NMA for safe storage of medicines (which is the exposure variable for the base case analysis reported here):

- Usual care
- Education
- Education + provision of low cost/free equipment
- Education + provision of low cost/free equipment + home safety inspection
- Education + provision of low cost/free equipment + fitting
- Education + provision of low cost/free equipment + home safety inspection + fitting
- Provision of low cost/free low cost/free equipment

The effectiveness estimates from the NMA were reported as odds ratios in the analysis presented in Chapter 5. The probability of having safe storage of medicines for each intervention is obtained by applying the odds ratios to baseline risk in the population of interest. This is achieved automatically in the decision analytic model when implemented within the comprehensive decision modelling framework (Cooper *et al.*, 2004). Estimates of the probabilities of safe storage for the intervention strategies obtained this way are summarised in Table 8.2.

Note that categorising the interventions into seven distinct intervention packages above does not completely remove all heterogeneity in intervention definition across studies. For example, education may be something as simple as home safety information leaflet, or face-to-face interview with a trained professional, or a computer programme designed to deliver tailored home safety advice. Similarly, low cost or free equipment may be related to poison prevention (for example, cupboard locks and latches) or unrelated equipment such as smoke alarms, for example. Equipment may be provided as a stand-alone scheme or as part of an equipment package scheme that includes non-poison prevention related equipment such as

smoke alarms, stair gates and window locks. This type of heterogeneity in the intervention classification will most likely have implications for estimating relative effectiveness and costs for each intervention. However, it was not possible to classify the intervention strategies further into more tightly defined intervention categories while maintaining a connected and consistent network of evidence required in order to fit a network meta-analysis model. While it is not possible to disentangle the effectiveness between varying degrees of intervention strategies without reclassification of intervention strategies, it would be at least possible to assign costs based on alternative definitions of strategies being evaluated. Therefore, sensitivity analysis will be used (see Section 8.8.2) to investigate the influence of assigning costs based on alternative more refined definition of interventions on results of the analysis.

8.4.2 Baseline prevalence of safety practices

The probability of moving from a 'no safe storage state' to a 'safe storage state' in cycles 1 to 5 for households in the usual care intervention cohort, denoted by $pSafe^{(1)}$ was assumed to be equal to the baseline prevalence of safety practices in the general population. The parameter was informed by evidence from a case control study of poison risk factors (Majsak-Newman 2014, unpublished, published protocol (Kendrick *et al.*, 2012a)) suggesting that 1527 (75%) out of a total of 2033 households among community controlled cohort have safe storage of medicines (Table 8.2). The parameter $pSafe^{(1)}$ was assumed to be drawn from a beta distribution with parameters based on the evidence from the published study to allow for parameter uncertainty:

$$pSafe^{(1)} \sim \text{Beta}\left(a = 1527, b = 2033 - 1527\right) \tag{8.1}$$

Next, the probability of moving from 'no safe storage state' to 'safe storage state' for households in each of the active intervention cohorts, denoted by $pSafe^{(k)}$, is derived by combining the effectiveness probability $p_{eff}^{(k)}$ for intervention *k* and the baseline probability of safe storage as follows:

$$pSafe^{(k)} = pSafe^{(1)} + \left(1 - pSafe^{(1)}\right) \times pEff^{(k)} \qquad k = 2,3,\cdots K \tag{8.2}$$

where $K=7$ is the total number of interventions being evaluated for safe storage of medicines.

8.4.3 Probability of accidental ingestion

The probability of accidental ingestion (of a medicinal substance) given 'no safe storage' was derived using estimates of the annual number of poisoning cases amongst preschool children presenting at emergency departments in the UK, the proportion of cases caused by exposure to a medicinal substance and the at risk population (number of preschool children in the UK). Evidence on the annual number of poisoning cases presenting at emergency departments was obtained from Orton et al (2014, unpublished) based on analysis of injury rates reported in the Health Improvement Network (THIN), a nationally representative dataset of the UK population. Orton et al. estimated that on average, approximately 10387 first episode cases of unintentional poisonings present annually at the emergency department among a UK preschool population of 3599180 children in the period 2005-2009. To obtain the number of cases attributed to ingestion of a medicinal substance given no safe storage, the 10387 estimated cases were multiplied by 0.6 based on evidence from Orton et al 2012 which shows that approximately 60% of the unintentional poisonings among preschool children recorded in the THIN database are caused by accidental ingestion of a medicinal substance. The required probability (i.e. of accidental ingestion of a medicinal substance given no safe storage for the pre-school age group) , denoted by $pIngest^{(1)}$, was then estimated to be 0.018% based on the above data as follows:

$$pIngest^{(1)} = \frac{10837 \times 0.6}{3599180} = \frac{6502}{3599180} = 0.00181 \tag{8.3}$$

The uncertainty around this estimate was taken into account by expressing $pIngest^{(1)}$ as a beta distribution with parameters ($a= 6502$ and $b=3599180-6502$) as in equation (8.1). Next, the probability of unintentional ingestion (of a medicinal substance) given safe storage, denoted by $pIngest^{(2)}$ was derived by applying the relative risk ratio, $orIngest$ comparing

the risk of unintentional ingestion given safe storage to the risk of corresponding risk given no safe storage as follows:

$$pIngest^{(2)} = pIngest^{(1)} \times orIngest \tag{8.4}$$

Evidence on the effectiveness of safe storage to prevent accidental ingestion of a medicinal substance was obtained from the case control study of the risk and protective factors for poisoning injury in childhood (Majsak-Newman et al., 2014). The adjusted community controlled analysis from this study indicated that compared with no safe storage, safe storage reduced the risk of accidental ingestion of a medicinal substance by about 40% (OR 0.60, 95% CI 0.44 to 0.81). To incorporate the uncertainty around this estimate, the odds ratios were assumed to be normally distributed on the logarithmic scale with mean *d* and variance σ_d^2:

$$\log(orIngest) \sim \text{Normal}(d, \sigma_d^2) \tag{8.5}$$

where $d = \log(0.6) = -0.511$ and $\sigma_d^2 = \frac{\log(0.81) - \log(0.44)}{3.92} = 1.557$. In the model odds ratios were assumed to be equivalent to relative risk and used to derived $pIngest^{(2)}$ using equation (8.4) which has a mean value of 0.011%.

8.4.4 Probability of inpatient admission following a poisoning injury

The probability of inpatient admission following assessment at the emergency department was based on evidence from the Hospital Episode Statistics (Health & Social Care Information Centre, 2013b) for England which indicated that approximately 3909 cases of medicinal poisoning (ICD-10 codes X40-X44) in pre-school age group were admitted in the year 2012. Because the estimates reported in Orton et al (2014) were for the whole UK population, the number of admitted cases in England was scaled up by 1.16 to give 4534 as the estimated annual number of medicinal poisoning cases admitted for the whole of UK. The scaling factor was based on the assumption that approximately 84% of the UK population live in England (i.e. 16% live in the rest of the UK). The probability of inpatient admission was then calculated by dividing 4543 by 6502 to obtain 70% as the proportion of cases admitted. As was done in equation (8.1) this information was encoded as a beta distribution

with parameters *a=6502 and b= (3599180-6502)* to account for uncertainty in the estimate. The probability of a minor injury was obtained by subtracting the probability of inpatient admission from 1 based on the assumptions that only serious cases were admitted as outlined previously.

8.4.5 Probability of moderate, severe and fatal poisoning injury

In the preschool model described in section 8.3.2, admitted cases were classed as either moderate, if the individual makes a complete recovery without any long term health problems on discharge, or severe if the injury is fatal or the individual develops a chronic/long term health condition. No evidence was found from a review of the literature to support this severity based classification of poisoning injury. The only (next best) evidence available was obtained from the 2012 report of the American Association of Poison Control Centres (Mowry *et al.*, 2013) which indicated that approximately 1.91% of major poisoning cases (in all age groups) resulted in a chronic/permanent injury. This figure was therefore taken as the probability of a severe injury among the cases admitted for a period of inpatient stay. This also suggests that the majority (about 98%) of cases admitted for an inpatient stay make a complete recovery in the model. Applying the 1.19% (i.e. the probability of a major poisoning) to the number of admitted cases estimated in section 8.4.4 resulted in 88 (out of the estimated 4543 poisoning cases admitted) cases classed as severe, of which 1 would be fatal on average based on data from the 2012 UK mortality statistics (Office for National Statistics, 2012).

In summary, based on the above evidence, the following probabilities were used in the model:
Probability of a severe injury among the admitted cases, *pSevere= 0.0191*.
Probability of moderate injury among the admitted cases, *pModerate = 1-pSevere= 0.9810*.
Probability of a fatal injury among the severe cases, *pFatal = 1/88 = 0.0114*.
These probabilities were inputted in the model by expressing them as beta distributions to account for uncertainty in the estimation, based on equation (8.1).

8.5 Costs

This section describes the procedure used to estimate costs and resource use and also reports the source of evidence informing each estimate. Costs are considered from the NHS/PSS perspective; therefore only the cost of providing the interventions and the NHS costs of treating unintentional poisoning related injuries are considered. Costs incurred or reported prior to 2012 are converted to 2012 prices using the Bank of England inflation calculator to take account of inflation (as described in Chapter 3 Section 3.2.4).

8.5.1 Intervention costs

It can be seen from the effectiveness evidence presented in section 8.4.1 that each of the intervention strategies consisted of a number of components either alone or in combination with each other. Therefore, the cost of providing each intervention was estimated by summing up the costs of the constituent components, and adding to this, a fixed cost associated with setting up the intervention scheme to cover things like administrative, transport and telephone costs (Table 8.3). The fixed cost of setting up an intervention scheme was obtained from the functional smoke alarm decision analysis (Saramago *et al.*, 2014) which provided an estimate of £79,529 as the cost of setting up a smoke alarm give way scheme in the year 2012 for a cohort of 100,000 households. The cost of home visit or home safety inspection was estimated to be £22.67, based on the mean hourly rate for a local authority (LA) funded home care worker of £34 (Curtis, 2012), and assuming a typical home safety inspection took 40 minutes to complete.

Also as noted in Section 8.4.1, there was still heterogeneity in the way the interventions were defined even after being categorised into more homogenous groups. This has implications for estimating the costs of the interventions. For example, home safety education may involve a face-to-face contact with trained professionals and is hence likely to be more expensive than providing a home safety information leaflet.

Table 8.3: Intervention costs used in base case model

Parameter	Description	Source	Value	Distribution
cFixed	Fixed cost of setting up an intervention scheme for 100 000 households	Functional smoke alarm model (Saramago *et al.*, 2014)	£79,529	Fixed
cEdu	Cost of home safety education (based on 20 minutes of local authority working time)	PSSRU 2012(Curtis, 2012)	£6.66	Fixed
cAccept	Cost of accepting intervention	Functional smoke alarm model (Saramago *et al.*, 2014)	£0.46	Fixed
cHSI	Cost of 40 minutes home safety inspection	PSSRU (Curtis, 2012) – the mean hourly cost of local authority funded homecare was £34, hence 40 minutes = 0.67*£34 = £22.67	£22.67	Fixed
cEquip	Cost of safety equipment (cupboard locks x2) updated to 2012 prices	Locks for kitchen cupboards (£3, range £2-6 per lock in 2009) reported in NICE PH30 costing template (NICE, 2010a).	£6.8 (range £4.54 – £13.62	Fixed
cInstall	Cost of installing cupboard locks (x2)	Functional smoke alarm model (Saramago *et al.*, 2014)	£11.83	Fixed

In the base case analysis, it was assumed that education was a face-to-face activity delivered by a trained professional costing approximately 20 minutes of a local authority workers time. Similarly, free or low cost equipment was assumed to consist of two sets of cupboard locks provided at a cost of £6.80 (range £4.54 to £13.62) per set as reported in the costing template for NICE PH30 (NICE, 2010a). Sensitivity analyses were conducted to investigate the impact of providing: i) a home safety information pack costing £0.56 (including VAT) per family (Errington et al. 2011) instead of a face-to-face educational activity, and ii) equipment schemes as part of the intervention package. The cost of installing the 2 sets of cupboard locks was assumed to be the same as the cost of installing a smoke alarm reported in DiGuiseppi *et al.* (DiGuiseppi and Higgins, 2001) and estimated to be £11.83 updated to 2012 UK pound sterling.

8.5.2 Healthcare costs of treatment

The cost of treating unintentional poisoning injury was estimated based on NHS reference costs for hospital services obtained from PSSRU Unit Costs of Health and Social Care 2012 (Curtis, 2012). The NHS reference costs in Curtis (2012) were reported as national averages together with the interquartile range. The standard error associated with each cost element was therefore derived from the interquartile statistics under assumptions of normality:

$$\sigma_X = \frac{UQ_X - LQ_X}{2 \times 0.675} \tag{8.6}$$

where UQ_X and LQ_X represent the upper and lower quartiles and σ_X is the required standard error for cost element *X*.

Estimates of the health sector costs used in the base case analysis are presented in Table 8.4. The mean cost (standard error) was £263 (£21.48) for emergency transfers, £112 (£27.41) for cases treated and discharged from the emergency department and £146 (£42.22) for cases admitted for inpatient stay following emergency department assessment and or treatment. The mean (standard error) cost of hospital inpatient admission was £586 (£223.70) for non-elective short-stay (<2 days) and £2,461 (£810.37) for non-elective long inpatient admission (≥2days).

Table 8.4: Health sector costs (2012 prices)

Parameter	Description	Source	Value (SE)	Distribution
cAmb	Cost of emergency transfers	PSSRU (Curtis, 2012)	£263 (£21.48)	Gamma
cED1	Cost of emergency department treatment of cases not leading to hospital inpatient stay (minor injury)	PSSRU (Curtis, 2012)	£112 (£27.41)	Gamma
cED2	Cost of emergency department treatment for cases leading to hospital inpatient stay (major injury)	PSSRU (Curtis, 2012)	£146 (£42.22)	Gamma
cAdmit1	Cost of a non-elective short (<2 days) inpatient admission	PSSRU (Curtis, 2012)	£586 (£223.70)	Gamma
cAdmit2	Cost of a non-elective long (≥2days) inpatient admission	PSSRU (Curtis, 2012)	£2461 (£810.37)	Gamma
cChro	Annual cost of chronic ill-health	HALO study (Nicholls 2009)	£386.42 (£96.72)	Gamma
cFatal	Cost of fatal injury	Functional smoke alarm model (Saramago *et al.*, 2014)	£205.50	Fixed
cGP	Cost of 11.7 minutes GP consultation	PSSRU (Curtis, 2012)	£43	Fixed

cHV	Cost of a Health visitor lasting 40 minutes	PSSRU (Curtis, 2012)	£44 (£15.56)	Gamma

Uncertainty around the cost estimates was taken into account using a gamma distribution:

$$C_X \sim \text{Gamma}(aC_X, bC_X) \tag{8.7}$$

where $aC_X = \frac{\mu_X^2}{\sigma_X^2}$, $bC_X = \frac{\mu_X}{\sigma_X^2}$, μ_X is the mean of cost of element *X* and σ_X is the associated standard error. For example, the cost of emergency transfer with mean $\mu_X = £263$ and standard error $\sigma_X = £21.48$ will be included in the analysis as $\text{Gamma}(149.89, 0.570)$ distribution. Other health sector costs considered in the analysis (Table 8.4) are the additional costs of a poisoning related fatality (i.e. coroners, autopsy), follow-up GP consultation lasting 11 minutes, health visitors' time lasting 40 minutes and the annual costs of chronic ill-health.

Evidence from the Hospital Episode Statistics indicates that approximately 24% of emergency department attendances arrived by emergency ambulance in the 2011-12 reporting period (Health & Social Care Information Centre, 2013a). Therefore the cost of emergency transfer was weighted by 0.242 to account for the fact that not all cases of accidental poisoning would incur emergency transfer costs. Also as stated in the preschool model (section 8.3.2), cases were assumed to be admitted for short-inpatient stay if the injury was moderate leading to complete recovery, and to long-inpatient stay if the injury was severe leading to a chronic health condition or fatality . It was assumed that only admitted cases were referred for follow-up GP consultation and only severe cases were referred for follow-up home visitation in addition to GP consultation. The additional cost of a fatality, GP follow-up consultation and home visitor follow-up were included in the analysis as fixed costs as there was no measure of uncertainty available for these two costs. The NHS costs of treating poisoning injuries with varying degree of severity estimated based on the above cost information and analysis are summarised in Table 8.5.

Table 8.5: Estimates of the NHS cost of treating poisoning injury by severity category

Injury Category	Cost components	Mean cost (95% CrI)
Minor	$cMinor = (0.242 \times cAmb) + cED_1$	£175.70 (£128.10, £235.30)
Moderate	$cModerate = (0.242 \times cAmb) + cED_2 + cShort + cGP$	£842.70 (£478.00, £1368.00)
Severe	$cSevere = (0.242 \times cAmb) + cED_2 + cLong + cGP + cHV$	£2758.00 (£2663.00, £4591.00)
Fatal	$ctFatal = (0.242 \times cAmb) + cED_2 + cLong + cGP + cFatal$	£2882.00 (£2788.00, £4724.00)

cAmb = cost per case of emergency transfer
cED_1 = cost per case of emergency department treatment (cases not admitted)
cED_2 = cost per case of emergency department treatment (cases admitted)
cShort = cost per case of non-elective short inpatient stay
cLong = cost per case of non-elective inpatient stay
cFatal = cost per case fatality
cGP = cost of 11.3 minutes follow-up GP consultation
cHV = cost of a home visit

8.6 Utilities

The primary unit of health benefit (utility) in the analysis is the quality adjusted life-year (QALY). These were estimated at each cycle of the model and summed across all cycles to obtain the total QALYs associated with each intervention group. Because of the way the decision model is structured, utilities are not directly attached to health-states within the Markov state-transition structure (i.e. the preschool and long-term models) described in Section 8.3. Instead, the total utility for each health-state at the end of each cycle was calculated as a weighted sum of utilities associated with the health outcomes and weighted by the product of the respective pathway probabilities. This requires baseline data on age-specific utilities for individuals with no poison-related injury and data on the utility decrement associated with poison-related injuries of varying severity. Utility for non-injured individuals (i.e. those with no poison-related injury) was obtained from general UK population utility norms (Kind *et al.*, 1998) for people aged 18 years and above (Table 8.6).

Table 8.6: Utilities used in base case analysis

Parameter	Description	Source	Value (SE)	Distribution
uPop	UK non-injured population utilities	(Kind *et al.*, 1998)	<25yrs 0.94 (SE=0.12) 25-34yrs 0.93 (SE=0.15) 35-44yrs 0.91 (SE=0.16) 45-54yrs 0.85 (SE=0.25) 55-54yrs 0.80 (SE=0.26) 65-74yrs 0.78(SE=0.26) ≥75yrs 0.73 (SE=0.27)	Normal
uMinor	Utility deficit for minor injury	Miller 2000 (Miller *et al.*, 2000). Assumed standard error is 10% of mean(Anokye *et al.*, 2011; Pavey *et al.*, 2011)	0.03 (SE=0.003)	Beta
uModerate	Utility deficit for moderate injury	Utility decrement 0.046 for poisoning injury (Miller *et al.*, 2012)	0.046 (SE=0.0046)	Beta
uSevere	Utility deficit for severe injury	Utility decrement 0.046 for poisoning injury (Miller *et al.*, 2012) and decrement associated with disability of 0.1 from the HALO study	0.146 (SE=0.0146)	Beta
uChronic	Utility deficit associated with disability per year	HALO Study (Nichol et al 2009)	0.10 (SE=0.025)	Beta

There was no baseline utility data for the non-injured individuals younger than 18 years in the UK and a paucity of evidence on decrement associated with poison related injuries in children under 5 years old. As a result, assumptions were made in order to allow the cost-effectiveness analysis to be conducted using the available evidence. It was assumed that non-injured individuals under 18 years of age have the same quality of life value as individuals between the age 18 to 25 year olds (Kind *et al.*, 1998).

Evidence on quality of life associated with unintentional poisoning injury in childhood came from two articles by the same author (Miller *et al.*, 2000; Miller *et al.*, 2012). Miller *et al.* (2000) reported a QALY loss of 0.046 while Miller *et al.* (2012) reported a QALY loss of 0.08 for childhood poisoning injury (Table 8.6). These two figures are estimates of the utility for any poison-related injury irrespective of the severity of the injury. It was therefore assumed that minor poisoning injury were associated with a QALY decrement of 0.046 (the lower of the two estimates) and moderate injury was associated with QALY decrement of 0.046 (the upper estimate). The utility decrement for severe injury was obtained by adding the upper estimate of 0.046 to the QALY decrement associated with chronic injury reported in the HALO study (Nichol *et al.* 2009). It is not clear whether this is a reasonable assumption to make. However, given paucity of data on poison related utilities and the

definition of severe injury as any poison injury that causes a permanent or chronic health condition in the preschool model, it was felt this is the best way to make use of the available evidence. Uncertainty was incorporated in estimating the utilities by using a strategy reported in Anokye et al. (Anokye *et al.*, 2011) where it was assumed that the standard error of each utility decrement equals 10% of the mean value. Sensitivity analyses are conducted (the results of which are displayed in Table 8.9 and reported in Section 8.8.2 to investigate the extent to which the cost-effectiveness estimates are affected by the estimates of QALY decrements used in the analysis.

8.7 Model evaluation

8.7.1 Outcomes of economic evaluation

As stated in Chapter 3 Section 3.4, the ultimate objective in a cost-effectiveness analysis is to estimate the expected costs and QALYs associated with the interventions being evaluated. From these two statistics, an incremental cost-effectiveness ratio (ICER) can be calculated and used to compare the costs and effects of one intervention relative to another (equation (3.1) in Section 3.4.1). The probability that each intervention is the most cost-effective can then be calculated for a range of willingness-to-pay ratios (i.e. the price that a decision maker is willing to pay for a unit of effectiveness, usually the QALY) and plotted on a cost-effectiveness acceptability curve (Section 3.4.2). In the sections that follow from this point, the methods introduced in Chapter 3 will be applied to each part of the decision analytic model outlined in Figure 8.1 and used to estimate ICERs for intervention k $(k = 2,3,\cdots,K)$ relative to 'usual care' and the probability that each intervention is the most cost-effective at various willingness-to-pay ratios.

Figure 8.2 shows a decision tree diagram showing the pathways through the intervention model and the respective pathway probabilities at each decision node. The intervention model is evaluated at the beginning of the decision model (i.e. at time, $c = 1$). It is assumed that all individuals initially enter the model in a state of good health; hence only costs associated with providing the intervention are incurred and no utility decrement is applied at this stage of the model.

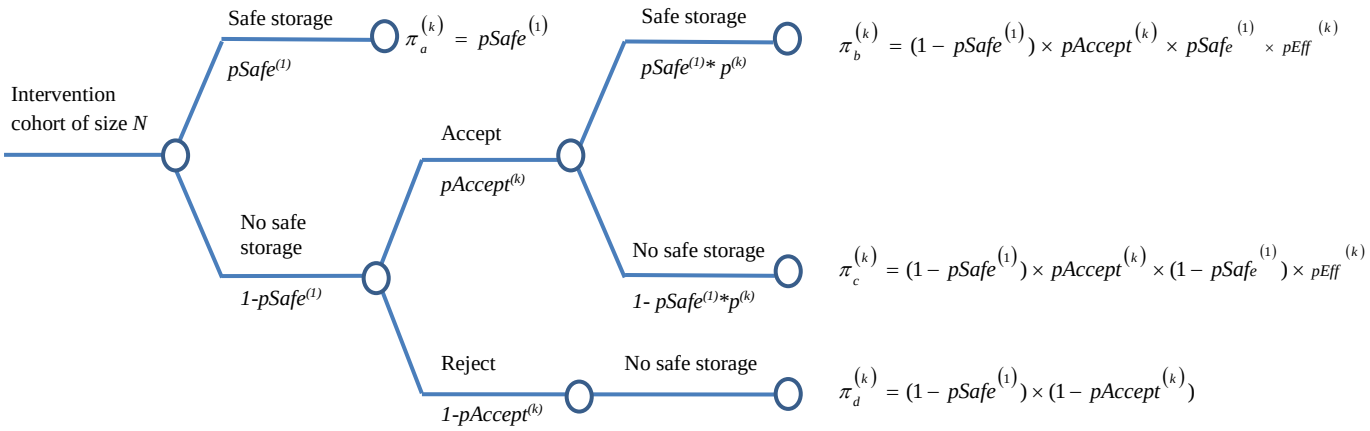


Figure 8.2: Decision tree diagram of the intervention model
Probabilities attached to each decision tree node are defined in Table 8.2

Therefore, the total cost incurred, $Cost_1^{(k)}$ and total utility, $Utility_1^{(k)}$ at time *c=1* for intervention *k* are given by:

$$Cost_1^{(k)} = cFixed + \left(\left(cInterv^{(k)} + cAccept^{(k)}\right) \times \left(\pi_b^{(k)} + \pi_c^{(k)}\right) \times N\right) \tag{8.8}$$

$$Utility_1^{(k)} = N \times \mu_1 \tag{8.9}$$

where $cInterv^{(k)}$ is the cost of intervention *k*, estimated as a sum of the costs of the constituent components, *cFixed* is the fixed cost of setting up an intervention scheme, $cAccept^{(k)}$ is the cost associated with the intervention being accepted, u_1 is the age-specific baseline utility corresponding to time c=1, *N* is the cohort size and the probabilities $\pi_b^{(k)}$ and $\pi_c^{(k)}$ are as defined in Figure 8.2 (i.e. the probabilities of safe storage and non-safe storage given that the intervention is accepted).

The outcome of the intervention model determines the number of households that enter the Markov model at time *c=2* in the safe storage state with probability $\pi_1^{(k)} = \pi_a^{(k)} + \pi_b^{(k)}$ or in the no safe storage state with probability $\pi_2^{(k)} = \pi_c^{(k)} + \pi_c^{(k)}$ as shown Figure 8.2 for the *k*th intervention cohort. The Markov part of the model was evaluated by simulating the movement of individuals in a hypothetical cohort of size *N* between health states defined in Figure 8.1 and estimating the costs and QALYs associated with each intervention over a life time horizon (i.e. 100 years). Figure 8.3 displays the possible pathways that individuals are able to take in moving between health states in any one cycle. Let $\lambda_{ss'}^{(k)}$ be the transition probability from state *s* to s' and $m_{cs'}^{(k)}$ be the number of individuals in state s' at time cycle *c* specific to *k*th intervention cohort.

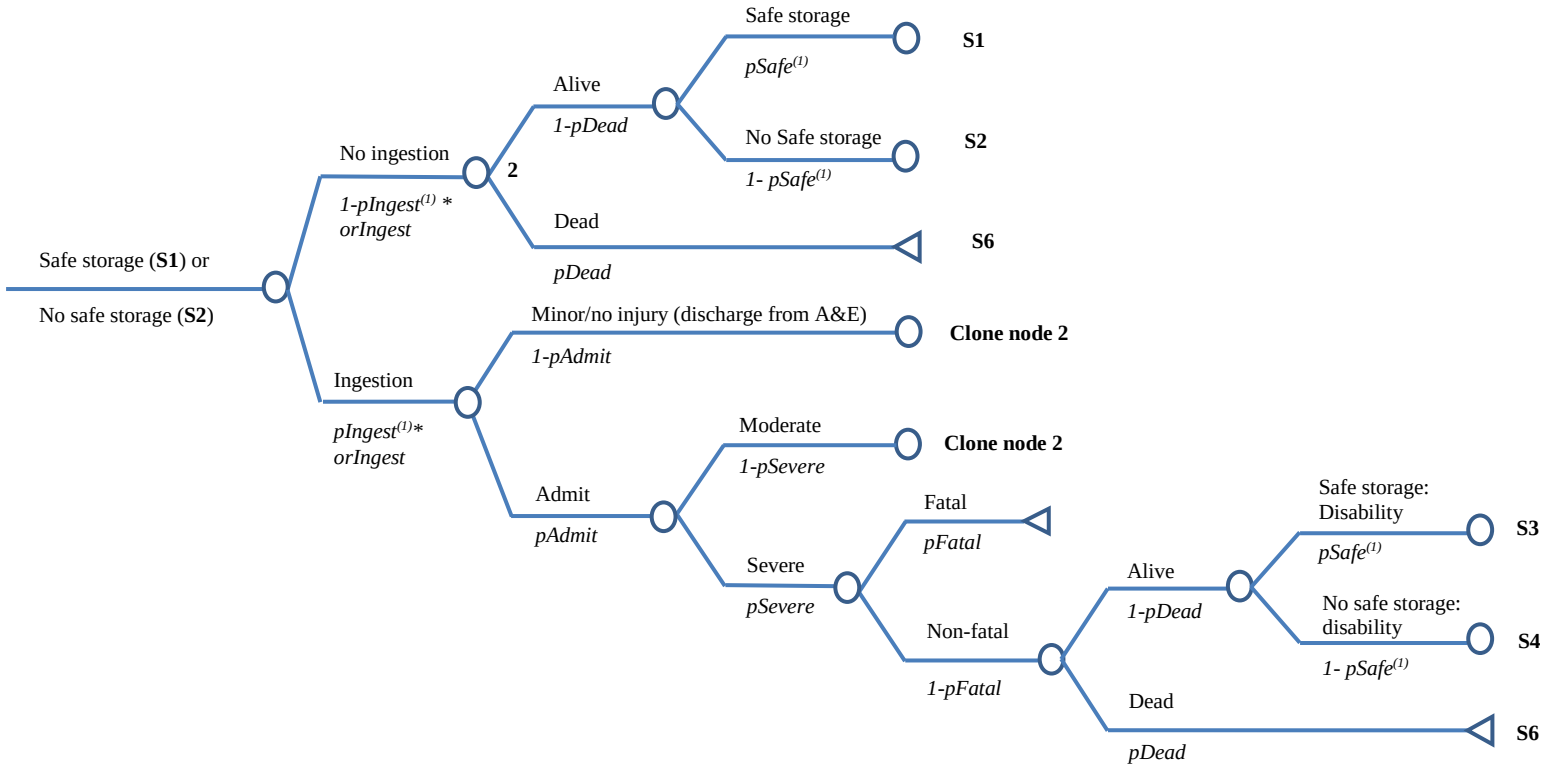


Figure 8.3: Movement of individuals between health-states in preschool model
Probabilities defined in Table 2 (*orIngest*[1] = *1* if starting in no safe storage state). Health-states are labelled S1 to S6.

Then the distribution of individuals among the *S* states at time cycle (*c+1*) is given by:

$$m_{(c+1)s}^{(k)} = m_{c1}^{(k)}\lambda_{1s}^{(k)} + m_{c2}^{(k)}\lambda_{2s}^{(k)} + \cdots + m_{cS}^{(k)}\lambda_{1S}^{(k)}$$

$$= \sum_{s'=1}^{S} m_{cs'}^{(k)}\lambda_{s's}^{(k)} \qquad (8.10)$$

where the vector $\mathbf{m}_1^{(k)} = \left(\left(\pi^{(k)} \times N\right),\ \left(1-\pi^{(k)}\right)N,\ 0,\ 0,\ 0,\ 0\right)$ represent the distribution of households at c=*1*. Equation (8.10) is an inner product of the vector $\mathbf{m}_c^{(k)}$ representing the number of individuals in the various states at time cycle *c*, and the transition probability matrix $\boldsymbol{\lambda}^{(k)}$. Note that in this model, the matrix $\boldsymbol{\lambda}^{(k)}$ is time dependent as it incorporates age-specific mortality statistics which depends on the age of the individual.

Once the number of individuals is known, the total cost and total QALYs in the remaining cycles can be estimated. Firstly, the costs and QALYs associated with each state at time cycle *c >1* is calculated by summing up all the costs and utilities associated with the pathways that individuals take through the model to arrive at the given state weighted by the respective pathway probabilities as depicted in Figure 8.3. The total costs incurred across all states in each cycle for the intervention *k*, denoted as $Cost_c^{(k)}$ is then obtained by summing up the product of the weighted cost $c_{cs}^{(k)}$ and the number of individuals $m_{cs}^{(k)}$ in health-state, *s* and discounted at a rate, *r* = 3.5% per cycle as recommended by NICE (NICE, 2012):

$$Cost_c^{(k)} = \frac{m_{c1}^{(k)}c_{c1}^{(k)} + m_{c2}^{(k)}c_{c2}^{(k)} + \cdots + m_{cs}^{(k)}c_{cs}^{(k)}}{(1-r)^{(c-1)}}$$

$$= \frac{\sum_{s=1}^{S} m_{cs}^{(k)}c_{cs}^{(k)}}{(1-r)^{(c-1)}} \qquad (8.11)$$

The total utility gained at each cycle is obtained in a similar way at each cycle:

$$Utility_c^{(k)} = \frac{m_{c1}^{(k)}u_{c1}^{(k)} + m_{c2}^{(k)}u_{c2}^{(k)} + \cdots + m_{cS}^{(k)}u_{cS}^{(k)}}{(1-r)^{(c-1)}}$$

$$= \frac{\sum_{s=1}^{S} m_{cs}^{(k)}u_{cs}^{(k)}}{(1-r)^{(c-1)}} \qquad (8.12)$$

where $Utility_c^{(k)}$ is the total utility gained in cycle *c* for intervention *k* and $\mu_{cs}^{(k)}$ is the utility in health state *s*. The results of equations (8.11) and (8.12) are used to estimate the mean costs and mean utility by summing up the costs and utilities across all cycles divided by the number of individuals in the cohort:

$$MeanCost^{(k)} = \frac{1}{N}\sum_{c=1}^{T} Cost_c^{(k)}$$

(8.13)

$$MeanUtility^{(k)} = \frac{1}{N}\sum_{c=1}^{T} Utility_c^{(k)}$$

(8.14)

To assess the cost-effectiveness of each of the 6 active interventions compared to usual care in the base-case analysis, the incremental-cost-effectiveness ratio was calculated as a difference between the mean costs and the mean utility as:

$$ICER^{(k)} = \frac{MeanCost^{(k)} - MeanCost^{(1)}}{MeanUtility^{(k)} - MeanUtility^{(1)}} = \frac{\Delta_C^{(k)}}{\Delta_E^{(k)}}$$

(8.15)

The ICERs obtained from equation (8.15) can be used to estimate an incremental net monetary benefit and the probability that each intervention *k* is cost-effective for a range of ceiling or willingness-to-pay ratios per additional unit of health, α. To do this, the incremental net benefit $INB^{(k)}$ for intervention *k* compared to usual care is given by the equation (Welton *et al.*, 2012):

$$INB^{(k)} = \alpha\Delta_E^{(k)} - \Delta_C^{(k)} > 0 \qquad (8.16)$$

For a probabilistic decision model, the net benefit function in equation (8.16) can be used to estimate the probability that the intervention *k* is the most cost-effective compared to usual care at different values of α :

$$pCE_{\alpha}^{(k)} = \frac{\text{Number of simulations } INB_{\alpha}^{(k)} > 0}{\text{Total number of simulations}} \tag{8.17}$$

A plot of $pCE_{\alpha}^{(k)}$ against α gives the cost-effectiveness acceptability curve (CEAC) for intervention *k* relative to usual care intervention.

8.7.2 Implementing the analysis

The analysis was conducted within a Bayesian framework utilising the comprehensive decision modelling approach introduced earlier in Section 3.3.4. This allows the synthesis of evidence (i.e. network meta-analysis model) and the cost-effectiveness evaluation to be conducted in one analysis model. This approach has the advantage that correlations between model parameters are automatically incorporated and propagated through to model outputs together with uncertainty (Cooper *et al.*, 2004; Welton *et al.*, 2012). All data sources used to inform model parameters are presented in Tables 8.1 to 8.6. The parameters of the decision model itself are given informative prior distributions based on the available evidence; hence minimally informative prior distributions were specified only for the parameters of the network meta-analysis embedded within the comprehensive decision model. Accordingly the pooled mean effects relative to usual care intervention and the study-specific effects were given Normal (0, 10^3) prior distributions and the between-study standard deviation was given a Uniform(0,2) prior distribution which are considered to be minimally informative on the log-odds ratio scale (Dias *et al.*, 2012).

The model was fitted in WinBUGS with the parameters estimated by means of Markov Chain Monte Carlo (MCMC) simulations. Estimates were obtained after running 3 MCMC chains for 30 000 iterations using disparate starting values. The first 10 000 iterations from each chain were discarded as ‘burn-in’ samples to ensure that the starting values do not influence the samples on which inference is based (Spiegelhalter *et al.*, 2007). Convergence diagnostics

were assessed in the same way as reported in Chapter 5. The WinBUGS code used to implement the model is given in Appendix V.

8.7.3 Incorporating uncertainty

Uncertainty in the model was addressed through probabilistic modelling and deterministic sensitivity and scenario analyses. The uncertainty in model inputs was taken into account by expressing parameters in the model as probability distributions. In addition to the base-case analysis, a number of sensitivity analyses (SA) were also conducted to investigate: i) the impact of assumptions underlying the model on estimates of cost-effectiveness, ii) uncertainty associated with multiple sources of evidence for a single parameter where it was uncertain as to which was the most appropriate, iii) parameters included in the model as fixed values due to lack of data on the appropriate measure of variability and iii) what if scenarios and best-case versus worse-case scenarios. The list of sensitivity analyses that were conducted are displayed in **Box 8.1.**

Box 8.1: List of sensitivity analyses

SA1	Baseline probability of safe storage changed from 75% (KCS community controls) to 93% (Patel et al 2008)
SA2	Baseline probability of safe storage changed from 75% (KCS community controls) to 50% (Assumption)
SA3	Probability intervention is accepted changed from 90% to 50% (Assumption)
SA4	Proportion admitted changed from 70% (HSE, 2012) to 83.3% (Phil Miller, personal communication)
SA5	Probability of permanent injury among admitted cases changed from 1.9% (NSPD 2012 report) to 4.2% (Assumption based on HASS 2002)
SA6	Cost of education changed from £11.33 (based on 20 minutes of a local authority workers time) to £0.56 (cost of home safety information pack per family reported in the Safe At Home Project report, 2011).
SA7	Reduce the number of cupboard locks from two locks (costing £6.80) to one lock costing £3.40.
SA8	Increase the number of children in a household from 1 to 1.8.
SA9	Increase the uncertainty associated with the utility decrements for poisoning injuries changed from 10% of the utility decrement value to 20% (i.e. utility decrement entered in the model as fixed values without uncertainty).
SA10	The uncertainty associated with the utility decrements for poisoning injuries changed from 10% of the utility decrement value to zero (i.e. utility decrement entered in the model as fixed values without uncertainty).

8.8 Results

8.8.1 Base case analysis

The results of the base case cost-effectiveness evaluation are presented in Table 8.7. Estimates of the mean and incremental costs and QALYs are expressed per 1000 households because the gain in utility per household was very small for the intervention groups compared to usual care. The expected utility per 1000 households was about 25056 years (or a slightly more than 25 years per individual) of perfect health for all intervention groups. It can be seen that the usual care cohort has the lowest mean cost (about £4,582.484 (95% CrI £3,206.543 to £6,794.282) per 1000 households). The intervention with 'education, home inspection and provision and installation of equipment', which is also the most intensive strategy, has the highest mean cost of about £14,139.442 (95% CrI £12,575.590 to £16,392.710). This is to be expected as the low incidence of accidental ingestion implies that most households receiving the intervention with associated costs do not have unintentional poisoning injury and therefore do not have the substantial costs associated with the treatment pathways through the model. Compared to usual care, the interventions with the lowest cost and hence the lowest ICERs were education (£75,090.867, 95% CrI £44,454.684 to £175,345.172) per QALY gained and provision of low cost/free equipment (£74,073.852, 95% CrI £43,436.203 to £175,604.781) per QALY gained while the most intensive intervention comprising of 'education, low cost/free equipment, home safety inspection and fitting' has the highest ICER at £360,345.594 (95%CrI £220,222.875 to £807,064.000) per QALY gained.

Table 8.7: Base case analysis results

Intervention	Expected QALYs	Expected Costs (£s)	Incremental QALYs	Incremental Costs (£s)	ICER (£s per QALY)	Probability CE (£30,000)	Probability CE (£50,000)
UC	25,056.324 (25,039.085 to 25,073.410)	4,582.484 (3,206.543 to 6,794.282)				1	1
E	25,056.349 (25,039.112 to 25,073.439)	6,527.061 (5,183.626 to 8,672.029)	0.026 (0.012 to 0.042)	1,943.250 (1,758.051 to 2,108.687)	75,090.867 (44,454.684 to 175,345.172)	0	0
FE	25,056.349 (25,039.114 to 25,073.439)	6,553.776 (5,209.264 to 8,695.595)	0.026 (0.012 to 0.043)	1,968.446 (1,773.344 to 2,140.042)	74,073.852 (43,436.203 to 175,604.781)	0	0
E + FE	25,056.349 (25,039.114 to 25,073.439)	8,039.002 (6,670.992 to 10,190.743)	0.026 (0.012 to 0.043)	3,453.429 (3,191.833 to 3,716.915)	130,234.188 (78,037.828 to 299,593.656)	0	0
E + FE + HSI	25,056.347 (25,039.112 to 25,073.435)	11,491.245 (10,022.280 to 13,693.583)	0.026 (0.012 to 0.042)	6,897.82 (6,417.179 to 7,404.3260)	268,942.563 (163,016.766 to 608,722.000)	0	0
E + FE + F	25,056.347 (25,039.112 to 25,073.439)	10,703.370 (9,264.202 to 12,896.240)	0.026 (0.012 to 0.042)	6,114.845 (5,686.187 to 6,559.178)	238,080.594 (144,275.344 to 534,364.813)	0	0
E + FE + HSI + F	25,056.349 (25,039.114 to 25,073.441)	14,139.442 (12,575.590 to 16,392.710)	0.026 (0.012 to 0.043)	9,538.38 (8,868.862 to 10,244.129)	360,345.594 (220,222.875 to 807,064.000)	0	0

Data are expected QALY (95% credibility interval) and expected costs (95% credibility interval) per 1,000 households. UC = usual care; (2) E = education; (3) E + FE = education + low cost/free equipment; (4) E + FE+HSI = education + low cost/free equipment + home safety inspection; (5) E+ FE + F = education + low cost/free equipment + Fitting; (6) E + FE + HSI + F+ fitting = education + low cost/free equipment + home safety inspection + Fitting; (7) FE = low cost/free equipment . Probability CE = probability that intervention is cost effective at a £30,000/£50,000 threshold value. QALYs = quality-adjusted life years.

Figure 8.4 is a plot of simulated 4000 samples of the incremental costs versus incremental QALYs for each of the 6 interventions compared to usual care on a cost-effectiveness plane. Each of the simulated (4000 samples) results of probabilistic ICERs for all 6 interventions compared to usual care is presented on a cost-effectiveness plane. All the ICERs lie in the north-east quadrant of the plane, suggesting that each of the 6 intervention strategies is more costly but also more effective than usual care. Education and low cost/free equipment have identical ICERs compared to usual care intervention; hence the simulated results for these two ICERs almost completely overlap each other in Figure 8.4.

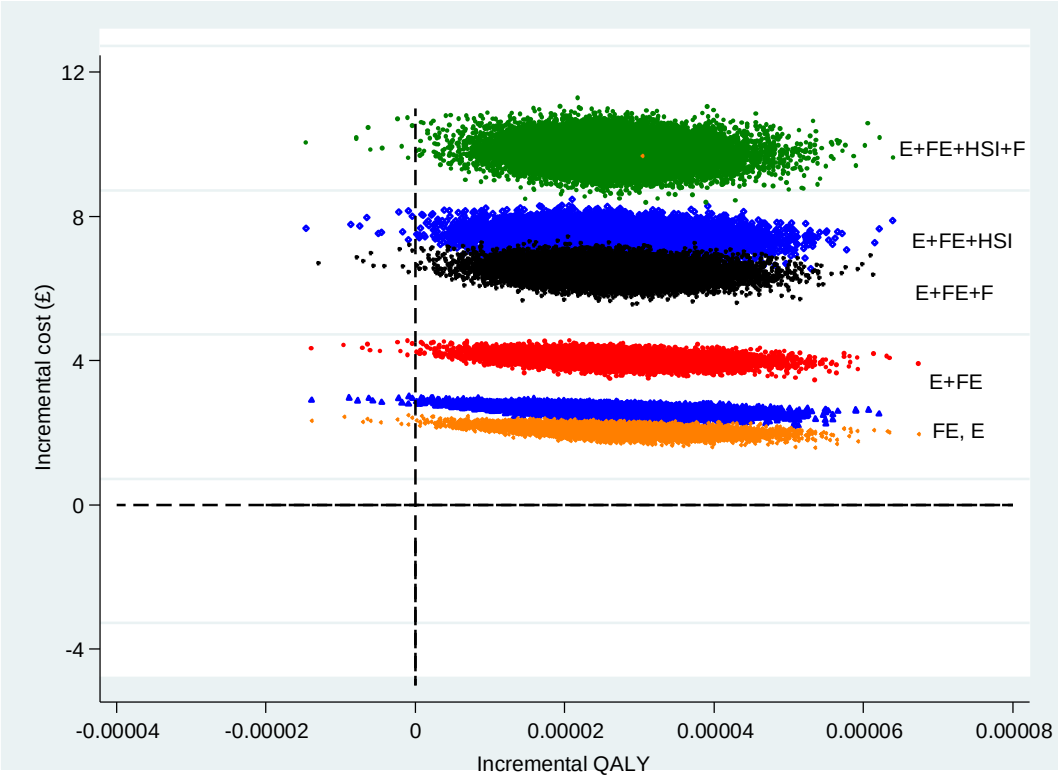


Figure 8.4: Cost-effectiveness of home safety interventions to reduce unintentional poisoning in children under 5years old from the base case analysis
Abbreviations (E=Education, FE= Provision of free/low cost equipment, HSI= Home safety inspection, F=Fitting of equipment). NB: education and low cost/free equipment have similar ICERs compared to usual care, hence the simulated ICERs representing these two interventions overlap each other

Also notice that simulated ICERs are almost all parallel to the x-axis in Figure 8.4. This suggests all the interventions produced a broadly similar gain in QALYs but differed only in

the incremental cost which increased with increasing intensity of intervention. Hence the cost of providing the intervention was the main driver of cost effectiveness in the base case analysis. The probabilistic ICERs for education and low cost/free equipment as stand-alone interventions therefore overlap each other on the cost-effectiveness plane while the ICERs for 'education, low cost/free equipment, home safety inspection and fitting' lie at the top end of the plot as it is the intervention with the most incremental costs. At a willingness-to-pay ratio of £30,000 per QALY gained, usual care has a 100% probability of being the most cost-effective intervention (Table 8.7).

Figure 8.5 is a plot of the probability that each intervention is the most cost-effective at different willingness-to-pay ratios (α) for the base case analysis. Usual care, education and provision of free/low cost equipment are the only interventions that have non-zero probabilities of being the most cost-effective intervention between α values ranging from £0 per QALY to £100,000 per QALY gained. Hence only the cost-effectiveness acceptability curves for these three interventions are displayed on Figure 8.5. The plot shows that the usual care intervention has the greatest probability of being the most cost-effective intervention at α values below £75, 000 per QALY above which low cost/free equipment becomes the most cost-effectiveness intervention at higher values of α.

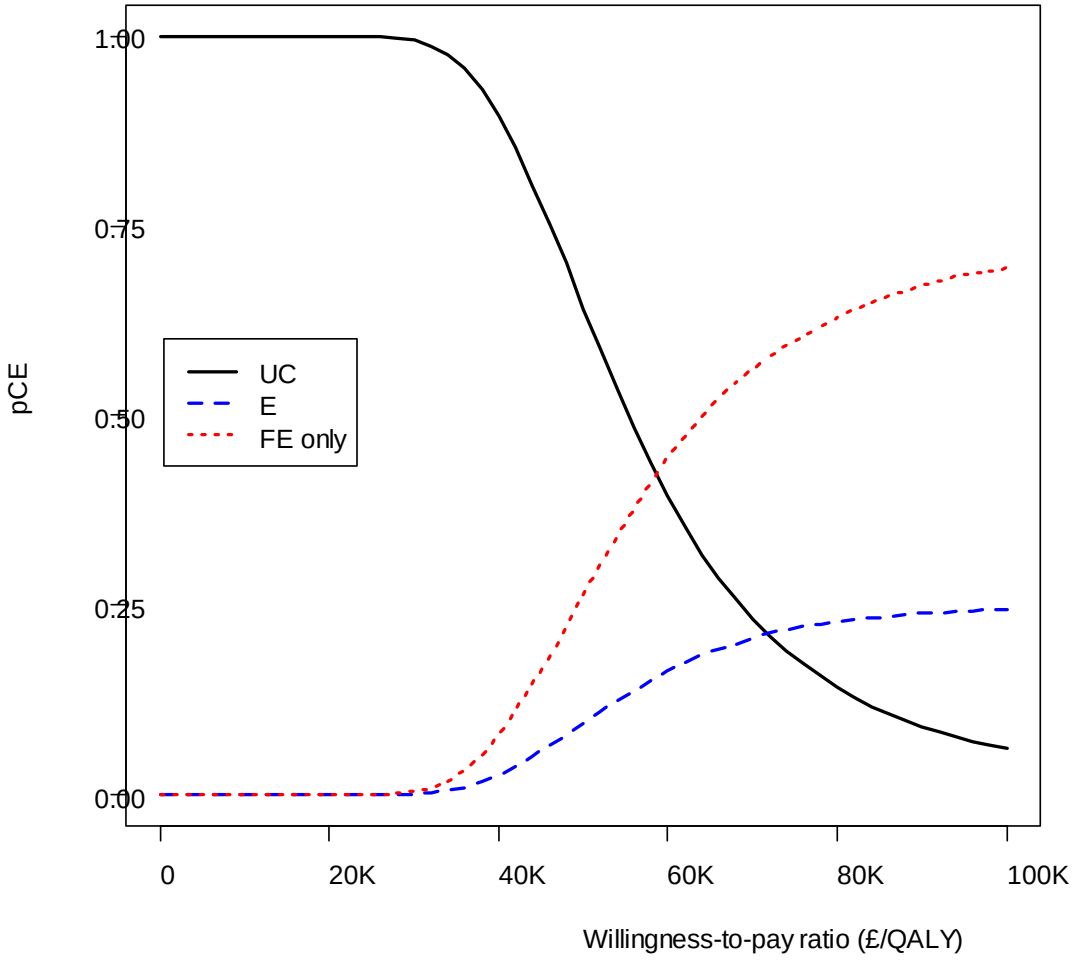


Figure 8.5: Cost-effectiveness acceptability curves for the base case analysis
Curves indicate the probability that each intervention is the most cos-effective for a range of willingness-to-pay ratios (20K stands for £20,000/QALY etc.)

8.8.2 Sensitivity analyses I results

Tables 8.8 and 8.9 and display the results of sensitivity analyses conducted to investigate the effect of varying particular assumptions and parameters of the model on the cost-effectiveness estimates. Only results for the three interventions (usual care, education and low cost/free equipment) with a non-zero probability of being the most cost-effective intervention at α values range from £0 - £100,000 per QALY are presented. These results show that cost-effectiveness estimates are robust to changes of assumptions and parameter estimates tested in most of the sensitivity analyses. At a willingness-to-pay ratio of £30,000/QALY, usual care continues to have the highest probability of being the most cost-effective strategy.

Table 8.8: Probability that an intervention is the most cost-effective at a willingness-to-pay ratio of £30,000 per QALY from base-case and a number of sensitivity analyses.

Analysis	Description	Probability intervention is the most cost-effective at £30,000 per QALY		
		Usual Care	Education	Equipment
BCA	Base case analysis	1	0.000	0.000
SA1	Baseline probability of safe storage changed from 75% (KCS community controls) to 93% (Patel et al 2008)	1	0.000	0.000
SA2	Baseline probability of safe storage changed from 75% (KCS community controls) to 50% (Assumption)	1	0.000	0.000
SA3	Probability intervention is accepted changed from 90% to 50% (Assumption)	1	0.000	0.000
SA4	Proportion admitted changed from 70% (HSE, 2012) to 83.3% (Phil Miller, personal communication)	0.998	0.0004	0.001
SA5	Probability of permanent injury among admitted cases changed from 1.9% (NSPD 2012 report) to 4.2% (Assumption based on HASS 2002)	0.881	0.037	0.082
SA6	Cost of education changed from £11.33 (based on 20 minutes of a local authority workers time) to £0.56 (cost of home safety information pack per family reported in the Safe At Home Project report, 2011).	0.224	0.776	0.000
SA7	Reduce the number of cupboard locks from two locks (costing £6.80) to one lock costing £3.40.	0.916	0.000	0.084
SA8	Increase the number of children in a household from 1 to 1.8.	0.854	0.040	0.106
SA9	Increase the uncertainty associated with the utility decrements for poisoning injuries changed from 10% of the utility decrement value to 20% (i.e. utility decrement entered in the model as fixed values without uncertainty).	1	0.000	0.000
SA10	The uncertainty associated with the utility decrements for poisoning injuries changed from 10% of the utility decrement value to zero (i.e. utility decrement entered in the model as fixed values without uncertainty).	1	0.000	0.000

Only changing the cost of the intervention and increasing the number of children per household from 1 to 1.8 appeared to influence the cost-effectiveness estimates. In sensitivity analysis 6, for example, the cost of home safety education was reduced from £11.33 (based on 20 minutes of a local authority workers time) to £0.56 (which is the cost of home safety information pack per family reported in the Safe At Home Project report (Errington *et al.*, 2011)). This makes education the most cost-effective intervention with a probability of 1 (at $\alpha = £30{,}000$ per QALY gained) and an ICER of £22,193.77 (95% CrI £11,091.88 to £58,994.16) per QALY gained compared to usual care (see SA6 of Table 8.9). Similarly, reducing the number of cupboard locks supplied as part of an intervention scheme from two locks (costing £6.80) to one lock (costing £3.40) reduced the ICER for low cost/free equipment to £41,152.13(95% CrI 24,131.23 to £97,558.10) per QALY gained compared with usual care (see SA7 of Table 8.9). Probability of being the most cost-effective intervention at $\alpha = £30{,}000$ per QALY gained in this sensitivity analysis was 0.342 for low cost/free equipment and 0.677 for usual care.

Table 8.9: Results of sensitivity analysis from the decision analysis carried out in Chapter 8
Only results for the three interventions (usual care, education and low cost/free equipment) are displayed

	Expected QALYs	Expected Costs (£s)	Incremental QALYs	Incremental Costs (£s)	ICER (£s per QALY)	Probability CE (£30,000)	Probability CE (£50,000)
SA1: Baseline probability of safe storage changed from 75% (KCS community controls) to 93% (Patel et al 2008)							
UC	25,056.379 (25,039.171 to 25,073.389)	5,408.819 (3,821.254 to 7,859.981)				1.000	1.000
E	25,056.406 (25,039.200 to 25,073.420)	8,736.729 (7,202.928 to 11,099.263)	0.030(0.014 to 0.051)	3,327.817 (3,124.751 to 3,510.758)	108,672.828 (63,916.242 to 252,348.609)	0.000	0.000
FE	25,056.406 (25,039.202 to 25,073.421)	8,794.313 (7,258.048 to 11,161.565)	0.031(0.014 to 0.052)	3,384.031 (3,173.847 to 3,571.45)	107648.438 (62768.465 to 253147.891)	0.000	0.000
SA2: Baseline probability of safe storage changed from 75% (KCS community controls) to 50% (Assumption)							
UC	25,056.425 (25,039.188 to 25,073.523)	3,365.399 (2,145.585 to 5,368.190)				1.000	1.000
E	25,056.429 (25,039.188 to 25,073.524)	4,603.245 (3,309.709 to 6,677.042)	0.002(0.000 to 0.004)	1,211.478(989.988 to 1,609.363)	764,751.625 (351,516.000 to 2,194,121.50)	0.000	0.000
FE	25,056.429 (25,039.188 to 25,073.524)	4,613.529 (3,317.301 to 6,688.691)	0.002(0.000 to 0.004)	1,220.869(994.35 to 1,627.865)	748,935.625 (345,134.312 to 2,173,635.750)	0.000	0.000
SA3: Probability intervention is accepted changed from 90% to 50% (Assumption)							
UC	25,056.324 (25,039.085 to 25,073.410)	4,582.488 (3,206.546 to 6,794.293)				1.000	1.000
E	25,056.337 (25,039.103 to 25,073.425)	6,017.573 (4,659.133 to 8,190.004)	0.014 (0.006 to 0.023)	1,433.045 (1,330.156 to 1,524.955)	99,734.570 (59,950.289 to 229,715.234)	0.000	0.000
FE	25,056.339 (25,039.103 to 25,073.425)	6,030.925 (4,674.281 to 8,201.463)	0.015 (0.007 to 0.024)	1,447.043 (1,338.653 to 1,542.374)	98,156.273 (58,349.586 to 229,039.672)	0.000	0.000
SA4: Proportion admitted changed from 70% (HSE, 2012) to 83.3% (Phil Miller, personal communication)							
UC	25,056.532 (25,039.431 to 25,073.469)	5,227.188 (3,556.498 to 7,888.824)				1.000	1.000
E	25,056.562 (25,039.471 to 25,073.498)	7,136.789 (5,514.080 to 9,733.726)	0.029 (0.013 to 0.047)	1,909.374 (1,702.26 to 2,089.349)	65,745.500 (38,683.371 to 156,058.344)	0.000	0.000

FE	25,056.566 (25,039.473 to 25,073.500)	7,160.833 (5,535.07 to 9,758.969)	0.030 (0.013 to 0.049)	1,934.151 (1,718.117 to 2,121.301)	64,822.867 (37,509.660 to 157,791.625)	0.000	0.000
SA5: Probability of permanent injury in admitted cases changed from 1.9% (NSPD 2012 report) to 4.2% (Assumption based on HASS 2002)							
UC	25,056.211 (25,039.001 to 25,073.324)	5,652.26 (4,052.068 to 7,963.166)				0.737	0.185
E	25,056.253 (25,039.028 to 25,073.355)	7,504.056 (5,956.074 to 9,775.334)	0.042 (0.018 to 0.070)	1,861.689 (1,652.01 to 2,056.304)	44,502.5703125 (24,908.361 to 111,361.195)	0.090	0.274
FE	25,056.255 (25,039.026 to 25,073.357)	7,542.185 (5,993.337 to 9,809.694)	0.043 (0.018 to 0.072)	1,897.872 (1,681.319 to 2,101.732)	44,093.344 (24,551.545 to 113,724.445)	0.173	0.541
SA6: Cost of education changed from £11.33 (based on 20 minutes of a local authority workers time) to £0.56 (cost of home safety information pack per family reported in the Safe At Home Project report, 2011).							
UC	25,056.324 (25,039.085 to 25,073.410)	4,582.488 (3,206.546 to 6,794.293)				0.000	0.000
E	25,056.349 (25,039.112 to 25,073.439)	5,163.133 (3,824.456 to 7,298.587)	0.026 (0.012 to 0.042)	579.268 (416.319 to 706.119)	22,193.7734375 (11,091.881 to 58,994.156)	1.000	1.000
FE	25,056.349 (25,039.114 to 25,073.439)	6,553.782 (5,209.266 to 8,695.605)	0.026 (0.012 to 0.043)	1,968.446 (1,773.343 to 2,140.041)	74,073.836 (43,436.219 to 175,604.594)	0.000	0.000
SA7: Reduce the number of cupboard locks from two locks (costing £6.80) to one lock costing £3.40.							
UC	25,056.324 (25,039.085 to 25,073.410)	4,582.488 (3,206.546 to 6,794.293)				0.677	0.155
E	25,056.349 (25,039.112 to 25,073.439)	6,527.066 (5,183.629 to 8,672.038)	0.026 (0.012 to 0.042)	1,943.249 (1,758.050 to 2,108.686)	75,090.891 (44,454.703 to 175,345.063)	0.000	0.000
FE	25,056.349 (25,039.114 to 25,073.439)	5,791.69 (4,455.652 to 7,929.708)	0.026(0.012 to 0.043)	1,208.62 (1,031.219 to 1,349.098)	45,325.625 (25,627.043 to 111,643.75)	0.323	0.844
SA8: Increase the number of children in a household from 1 to 1.8							
UC	25,056.324 (25,039.085 to 25,073.41)	4,582.488 (3,206.546 to 6,794.293)				0.534	0.090
E	25,056.349 (25,039.112 to 25,073.439)	6,527.066 (5,183.629 to 8,672.038)	0.026 (0.012 to 0.042)	1,943.249 (1,758.050 to 2,108.686)	41,717.160 (24,697.057 to 97,413.930)	0.123	0.239
FE	25,056.349 (25,039.114 to 25,073.439)	6,553.782 (5,209.266 to 8,695.605)	0.026 (0.012 to 0.043)	1,968.446 (1,773.343 to 2,140.041)	41,152.133 (24,131.232 to 97,558.102)	0.342	0.671

SA9: Increase the uncertainty associated with the utility decrements for poisoning injuries changed from 10% of the utility decrement value to 20%							
UC	25,056.496 (25,039.32 to 25,073.778)	4,598.429 (3,187.581 to 6,815.030)				1.000	1.000
E	25,056.524 (25,039.343 to 25,073.799)	6,544.405 (5,164.149 to 8,696.96)	0.026 (0.011 to 0.042)	1,942.735 (1,758.842 to 2,107.723)	74,677.891 (44,061.750 to 178,996.891)	0.000	0.000
FE	25,056.526 (25,039.343 to 25,073.801)	6,570.660 (5,193.505 to 8,723.189)	0.026 (0.012 to 0.043)	1,969.168 (1,778.752 to 2,139.504)	74,041.211 (43,100.152 to 181,167.438)	0.000	0.000
SA10: The uncertainty associated with the utility decrements for poisoning injuries changed from 10% of the utility decrement value to zero (i.e. utility decrement entered in the model as fixed values without uncertainty).							
UC	25,056.479 (25,039.433 to 25,074.003)	4,597.515 (3,202.75 to 6,749.747)				1.000	1.000
E	25,056.503 (25,039.463 to 25,074.028)	6,540.991 (5,191.194 to 8,643.058)	0.026 (0.012 to 0.041)	1,943.535 (1,759.033 to 2,107.683)	74,984.555 (44,704.977 to 176,456.95)	0.000	0.000
FE	25,056.503 (25,039.461 to 25,074.028)	6,564.879 (5,214.495 to 8,655.904)	0.027 (0.012 to 0.043)	1,967.813 (1,775.796 to 2,137.956	73,625.648 (43,590.629 to 177,627.906)	0.000	0.000

8.8.3 Sensitivity analysis II: Baseline risk adjusted

The effectiveness estimates from the NMA that informed the economic evaluation were adjusted for baseline risk based on the methodology developed in Chapter 6 (Achana *et al.*, 2013). Specifically, Model 1A of section 6.4.3 which assumed a common regression coefficient for all *treatment* × *covariate* interactions (i.e. $\beta_{Ak} = \beta$) with unconstrained baseline risk, μ_{iA} across studies was used and β and μ_{iA} given minimally informative prior distributions:

$$\beta, \mu_{iA} \sim \text{Normal}\left(0, 10^3\right)$$

This adjusted analysis was conducted first to demonstrate application of the new synthesis methodology to economic evaluation within a PH context and secondly to investigate whether or not differences in the distribution of the control group event rate (i.e. proportion of households with safe storage of medicines in the usual care intervention arm) have an effect on the cost-effectiveness estimates. The regression coefficient β was estimated to be -0.152 (95% CrI -0.359 to 0.115) indicating that there was no evidence to suggest the intervention effects were associated with baseline safety practices as estimated by the proportion of households with safe storage of medicines rates in the usual care arm of studies. Consequently using the adjusted estimates in the decision analysis had no effect on the cost-effectiveness estimates. These are displayed in Table 8.10 and are almost identical to the results for the base case analysis with usual care still being the most cost-effective intervention at λ values of £30,000 and £50,000 per QALY respectively.

Note that if required, the models specified in Chapter 6 can be used to predict the treatment effect parameters in a population with known baseline risk. For example, the predicted effect of treatment *k* relative to treatment *b* denoted by d_{ibk}^{pred} for a population *i* with baseline risk x_i follows:

$$d_{ibk}^{pred} = d_{bk} + \beta_{bk}\left(x_i - \bar{x}\right) \tag{8.18}$$

where d_{bk} and β_{bk} are as defined in equation 6.2 (namely d_{bk} is the mean effect of intervention *k* relative to intervention *b* adjusted at the centred baseline risk value $\bar{x}$, and β_{bk}

is the change in the log-odds ratio of an event per unit change in the baseline risk for intervention *k* relative to *b*). The predicted effectiveness estimates d_{ibk}^{pred}, can then be incorporated directly into the decision analytic model as described above.

Table 8.10: Sensitivity analysis results incorporating adjusted estimates based on the methods presented in Chapter 6.

Intervention	Expected QALYs	Expected Costs (£s)	Incremental QALYs	Incremental Costs (£s)	ICER (£s per QALY)	Probability CE (£30,000)	Probability CE (£50,000)
UC	25,056.49 (25,039.261 to 25,073.671)	4,589.209 (3,180.039 to 6,828.071)				1	1
E	25,056.519 (25,039.282 to 25,073.698)	6,537.452 (5,161.843 to 8,714.782)	0.027 (0.012 to 0.044)	1,943.498 (1,758.1630 to 2,106.994)	71,177.08 (42,410.88 to 165,196.80)	0	0
FE	25,056.519 (25,039.282 to 25,073.702)	6,562.115 (5,188.673 to 8,736.221)	0.028 (0.013 to 0.045)	1,967.684 (1,775.143 to 2,137.254)	69,956.44 (41,288.46 to 165,001.90)	0	0
E + FE	25,056.52 (25,039.282 to 25,073.70)	8,053.416 (6,643.763 to 10,234.573)	0.028 (0.013 to 0.045)	3,452.978 (3,191.665 to 3,707.736)	123,048.90 (74,170.0 to 283,082.0)	0	0
E + FE + HSI	25,056.519 (25,039.282 to 25,073.70)	11,490.844 (10,001.724 to 13,718.154)	0.028 (0.013 to 0.045)	6,890.954 (6,403.567 to 7,389.143)	246,780.80 (150,953.0 to 553,591.40)	0	0
E + FE + F	25,056.519 (25,039.282 to 25,073.698)	10,709.185 (9,243.674 to 12,920.988)	0.027 (0.012 to 0.044)	6,109.439 (5,680.384 to 6,549.414)	222,093.10 (135,688.1 to 495,035.0)	0	0
E + FE + HSI + F	25,056.519 (25,039.282 to 25,073.70)	14,150.879 (12,569.782 to 16,442.518)	0.027 (0.012 to 0.044)	9,545.723 (8,868.781 to 10,238.464)	349,557.40 (213,685.8 to 776,726.60)	0	0

Data are expected QALY (95% credibility interval) and expected costs (95% credibility interval) per 1,000 households. UC = usual care; E = education; E + FE = education + low cost/free equipment; E + FE+HSI = education + low cost/free equipment + home safety inspection; E+ FE + F = education + low cost/free equipment + Fitting; E + FE + HSI + F+ fitting = education + low cost/free equipment + home safety inspection + Fitting; FE = low cost/free equipment. Probability CE = probability that intervention is cost effective at a £30,000/£50,000 threshold value. QALYs = quality-adjusted life years.

8.8.4 Sensitivity analysis III: Using effectiveness estimates from the multivariate analysis

In this sensitivity analysis, the methodology developed in Chapter 7 (Achana *et al.*, 2014a) to synthesise effectiveness evidence across multiple outcomes (i.e. injury prevention domains) is applied to the comprehensive decision model developed this Chapter. As stated in the overview of Chapter 7, this type of analysis enable all available evidence relevant to a decision problem to be taken into account when summarising evidence to inform an economic evaluation. This can be useful in a PH evaluation context where the evidence base is often scarce or limited in one way or another.

The results of the first stage multivariate analysis (Achana *et al.*, 2014a) described in Chapter 7 (see section 7.4.2) are not used because these were broadly similar to estimates from the univariate NMA used in the base case decision model. Thus in this sensitivity analysis, only the results of the second stage multivariate analysis (i.e. Model 3 of Chapter 7 Section 7.4.3) are used to inform the cost-effectiveness analysis. In this model, intervention effects were assumed to be exchangeable across outcomes which enabled more precise effectiveness estimates to be obtained for all interventions including predicted estimates of effects where direct trial data is not available. For example the estimate for education versus usual care on safe storage of medicines changed from an OR of 1.39 (95% CrI 0.73 to 2.28) in standard NMA analysis to 1.32 (95% CrI 0.71 to 2.16) in the second stage multivariate model.

To fit this analysis within the comprehensive decision model, μ_{iA}, representing the log-odds of safe storage in the usual care arm of studies were assumed to be normally distributed with mean $\overline{\mu}$ and variance σ_{μ}^{2} in order to make missing μ_{iA} s from studies that do not have usual care-arm identifiable:

$$\mu_{iA} \sim \text{Normal}\left(\overline{\mu}, \sigma_{\mu}^{2}\right)$$

where $\overline{\mu} \sim \text{Normal}\left(0, 10^{3}\right)$ and $\sigma_{\mu} \sim \text{Uniform}\left(0, 2\right)$ prior distributions were specified for the respective parameters.

Table 8.11 displays the cost-effectiveness estimates for based for this sensitivity analysis. First, it can be seen that ICERs were obtained for all 8 interventions relative to usual care including estimates for 'education and home safety inspection' and 'education and home visit' which were not available from the base case analysis. The estimated ICERs in were lower in the multivariate analysis than those obtained from the base case model but UC remained the most cost-effective intervention at $\lambda = £30{,}000$ per QALY with probability 0.69 followed by provision of low cost/free equipment alone with probability 0.19 and education alone with probability 0.12 (Figure 8.6). All the other interventions have a zero probability of being the most cost-effective intervention at λ values up to £100,000 per QALY, so cost-effectiveness acceptability curves for these are not displayed in Figure 8.6.

Table 8.11: Sensitivity analysis using effectiveness estimates on safe storage of medicines from multivariate NMAs (model 3) presented in Chapter 7.

Intervention	Expected QALYs	Expected Costs (£s)	Incremental QALYs	Incremental Costs (£s)	ICER (£s per QALY)	Probability CE (£30,000)	Probability CE (£50,000)
UC	25,056.225 (25,039.469 to 25,073.578)	5,393.226 (3,971.242 to 7,563.625)				0.929	0.410
E	25,056.252 (25,039.492 to 25,073.605)	6,552.195 (5,173.962 to 8,663.757)	0.024 (0.011 to 0.039)	1,163.253 (987.578 to 1,323.656)	47,876.14 (27,001.615 to 119,470.5)	0.040	0.347
FE	25,056.252 (25,039.492 to 25,073.605)	6,585.009 (5,208.121 to 8,698.240)	0.024 (0.010 to 0.040)	1,194.933 (1,013.128 to 1,359.812)	49,366.21 (27,060.912 to 131,775.0)	0.031	0.243
E + FE	25,056.253 (25,039.494 to 25073.606)	8,067.562 (6,662.20 to 10,193.320)	0.026 (0.011 to 0.042)	2,671.604 (2,416.564 to 2,925.679)	103,187.70 (61,917.477 to 241,381.5)	0	0
E + FE + HSI	25,056.253 (25,039.494 to 25,073.605)	11,502.769 (10,021.366 to 13,685.529)	0.026 (0.011 to 0.042)	6,104.935 (5,623.281 to 6,593.228)	237,496.2 (144,462.047 to 544,826.9)	0	0
E + FE + F	25,056.252 (25,039.492 to 25,073.605)	10,729.288 (9,270.933 to 12,895.501)	0.024 (0.011 to 0.040)	5,331.774 (4,905.398 to 5,762.07)	218,508.9 (130,430.039 to 511,820.2)	0	0
E+HSI	25,056.248 (25,039.492 to 25,073.597)	10,029.7 (8,579.253 to 12,192.57)	0.019 (0.005 to 0.038)	4,635.095 (4,250.595 to 5,019.478)	237,447.9 (118,198.68 to 985,403.2)	0	0
E+HV	25,056.25 (25,039.494 to 25,073.606)	10,011.304 (8,563.152 to 12,172.209)	0.022 (0.005 to 0.040)	4,614.602 (4,230.198 to 5,002.835)	209,116.8 (112,401.719 to 895,005.9)	0	0
E + FE + HSI + F	25,056.253 (25,039.494 to 25,073.603)	14,157.283 (12,566.775 to 16,383.749)	0.026 (0.011 to 0.042)	8,754.287 (8,085.174 to 9,435.503)	338,497.6 (207,613.844 to 770,333.4)	0	0

Data are expected QALY (95% credibility interval) and expected costs (95% credibility interval) per 1,000 households. UC = usual care; E = education; E + FE = education + low cost/free equipment; E + FE+HSI = education + low cost/free equipment + home safety inspection; E+ FE + F = education + low cost/free equipment + Fitting; E + FE + HSI + F+ fitting= education + low cost/free equipment + home safety inspection + Fitting; FE = low cost/free equipment. Probability CE = probability that intervention is cost effective at a £30,000/£50,000 threshold value. QALYs = quality-adjusted life years.

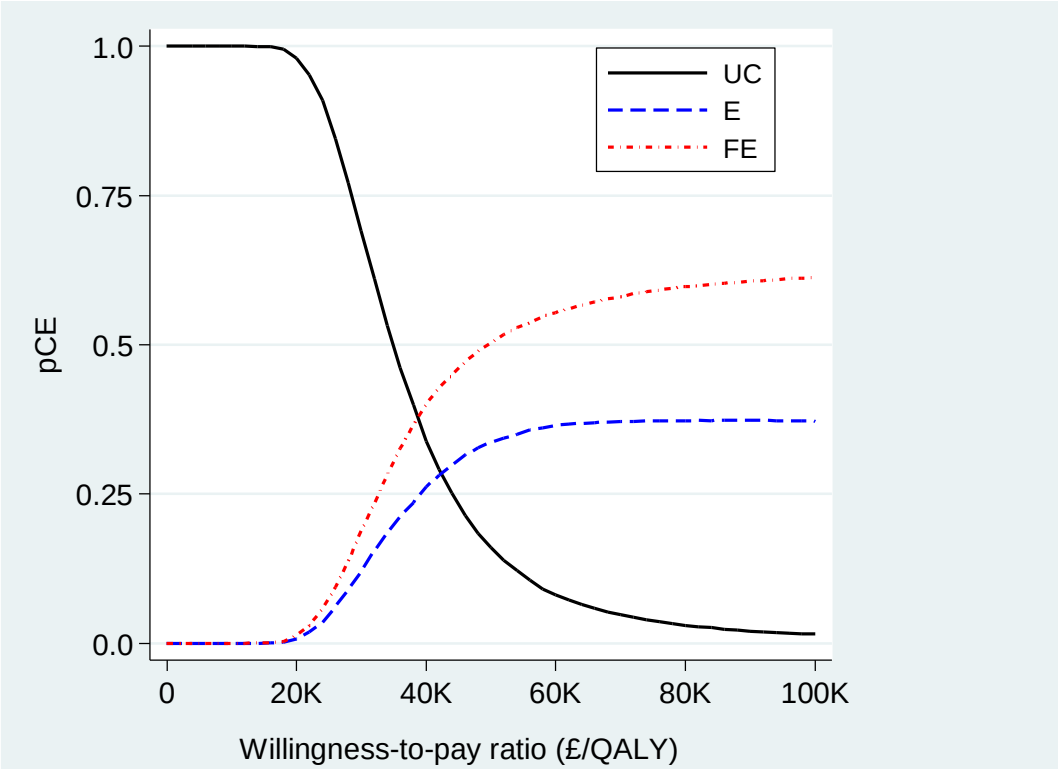


Figure 8.6: Cost-effectiveness acceptability curves using estimates from the multiple outcomes NMA (Model 3) for safes storage of medicines described in Chapter 7.

8.9 Discussion

8.9.1 Summary of findings

The analysis presented in this chapter is believed to be the first to evaluate the cost-effectiveness of interventions to prevent medicinal poisonings in children under 5 years of age. Interventions that have one safety element or component such as home safety education alone or providing low cost/free equipment alone were associated with the lowest ICERs compared to usual care, whereas multi-component interventions generally had higher ICERs. Compared to usual care intervention, education alone was estimated to provide an extra QALY at £75,090 (95% CrI £44,454 to £175,345) where as providing low cost or free equipment alone provided an extra QALY at £56,346.99 (95% CrI 34,364.35 to 133,447.67). All other interventions had much higher ICERs when compared to usual care. The analysis above has been conducted from the perspective of the UK health service and Personal Social Services, and, as stated in Chapter 3 Section 3.4.2, NICE (NICE, 2013) would usually

consider ICERs below £30,000 per QALY gained as cost-effective use of NHS resources when appraising health technologies. Applying this criterion to the base case analysis results would thus suggest that home safety education and/or provision of safety equipment interventions are unlikely to be cost-effective compared to usual care unless policy makers are willing to pay more than £75,000 for a unit of health (QALY).

There was considerable uncertainty in these estimates and a number of sensitivity analyses were carried out to assess the robustness of the findings. These extra analyses indicate the base case results were largely robust to alternative model specifications and to most model inputs including adjusting the effectiveness estimates from the NMA for baseline risk differences across studies (Achana *et al.*, 2013). However, the results were quite sensitive to the cost of home safety education and provision of low cost/free equipment. For example, reducing the cost of home safety education from £11.33 (based on 20 minutes of a local authority workers time) to £0.56 (based on the cost of home safety information pack per family reported in the Safe At Home Project report (Errington *et al.*, 2011) increased the probability that education is the most cost-effective intervention from 0 in the base case analysis to 0.88 at a willingness-to-pay ratio of £30,000 per QALY gained. Similarly, the probability of low cost/free equipment provision being the most cost-effective intervention changed from 0 in the base case analysis to 0.32 at willingness-to-pay ratio of £30,000 per QALY gained when the number of cupboard locks was reduced from two (costing £6.80) to one lock costing £3.40. These findings would seem to suggest that the key driver of cost-effectiveness in the analysis presented here is the cost of providing the intervention strategy with the more expensive multi-component interventions having the highest ICERs compared to usual care.

The results were also sensitive when the effectiveness estimates from the multivariate analysis (Model 3), where information was borrowed across different outcome networks (Achana *et al.*, 2014a), were used to inform the cost-effectiveness analysis. In this particular case, the ICERs for education and provision of low cost/free equipment were £47,876.14 (27,001.62 to £119,470.5) and £49,366.21 (95% CrI £27,060.91 to £131,775.0) per QALY gained compared with usual care. Hence usual care continues to be the most cost-effective strategy at a willingness-to-pay ratio of £30,000 per QALY with a probability of 0.929.

8.9.2 Strengths and limitations

The cost-effectiveness evaluations in this chapter have a number of strengths and limitations. The evaluations are fully probabilistic, allowing for parameter uncertainty to be taken into account in the cost-effectiveness estimates. Where uncertainty remains, for example because of model assumptions or uncertainty about which piece of evidence to use when multiple sources are available for the same parameter, these were investigated through scenario and deterministic sensitivity analyses. These sensitivity analyses indicated that the cost-effectiveness estimates were largely robust to most changes in parameter values with the exception of the costs associated with providing the intervention.

The approach to evidence synthesis adopted throughout this thesis, and which have been used to inform the decision analysis, strengthens the analysis presented here in that using the network meta-analysis (NMA) has enabled the simultaneous evaluation of the effectiveness and cost-effectiveness of multiple interventions in a single coherent analysis. This ensures that correlations between the effectiveness estimates are taken into account and propagated through the model to produce final estimates of the cost-effectiveness together with uncertainty (Cooper *et al.*, 2004; Welton *et al.*, 2012).

The main limitation of the analysis presented in this chapter is the lack of data to inform several parameters of the model. Data on clinical history and prognosis of the childhood poisoning injuries, background utilities, and quality of life data associated with childhood poison injuries were either lacking or when available, of poor quality and in a form unsuitable for use in the cost-effectiveness analysis. Because of this, a number of assumptions were made in order to simplify the model structure and make use of the available data. The assumptions are a potential source of uncertainty in the model results. As much as possible, robustness of the results to such modelling assumptions was assessed through sensitivity analyses.

8.9.3 Comparisons with previous work

The only other known economic evaluation to have considered the cost-effectiveness of strategies for preventing accidental injuries in children is the analysis undertaken to inform the development of NICE PH30 [Preventing unintentional injuries among under-15s in the home] (Pitt *et al.*, 2009). Pitt *et al.* evaluated a generic home safety intervention versus no intervention for the prevention of any injury in the home irrespective of injury mechanism (i.e. injury included falls, scalds, poisonings, etc.). The generic intervention in Pitt et al. was described as "General home safety programme includes measures such as: home safety consultation visits, provision of educational materials and advice, as well as the free supply and installation of a range of home safety equipment (including smoke alarms, stair gates, cupboard and window locks, etc.). The analysis presented in this Chapter builds on Pitt *et al.'s* by evaluating the cost-effectiveness of several interventions (all of which are more homogenously defined than the strategy in Pitt *et al.*) to prevent medicinal poisons in pre-school children.

8.10 Chapter summary

In this Chapter, a probabilistic decision-analytic model was developed to evaluate the cost-effectiveness of home safety interventions compared with usual care intervention in preventing unintentional medicinal poisonings in children under 5 years old. In the recently published systematic review of synthesis methods in PH evaluations (Achana *et al.*, 2014b), a number of issues were identified as barriers to quantitative synthesis of PH evidence including heterogeneity of methods, outcomes and intervention across studies. It was suggested and demonstrated in that review that more complex synthesis methodology can be employed to overcome some of the issues identified as barriers to quantitative synthesis in PH evaluations. The analyses presented in this chapter demonstrated how the recommendations in Achana *et al.*, to make PH evaluations more quantitative, can be implemented within a decision-analytic modelling framework when evaluating the cost-effectiveness of PH interventions. The methodology employed here involved using network meta-analysis and various extensions of it that were developed in this thesis (Achana *et al.*, 2013; Achana *et al.*, 2014a). These methods allow for heterogeneity in intervention definitions to be taken into account through extending the network to include multiple interventions and outcomes (Achana *et al.*, 2014a) and to adjust for the baseline risk (Achana

et al., 2013) as a proxy for measured or unmeasured modifiers of the treatment effect and hence a potential source of heterogeneity in the meta-analysis.

9. DISCUSSION

9.1 Chapter summary

In this concluding chapter, a summary of the main findings of the work undertaken in the thesis is presented. The strengths and limitations of the work are also discussed. The Chapter concludes by looking at opportunities for further work.

9.2 Thesis summary

The thesis considered the challenges present when summarising evidence to inform an economic evaluation of public health (PH) interventions by:

a) Reviewing the evidence to identify the barriers to quantitative synthesis in PH evaluations of interventions,
b) Developing and applying new synthesis methodology to overcome the challenges of evidence synthesis identified in (a) above, and
c) Applying the methods developed in (b) above to a substantive decision analytic model in order to assess the effectiveness and cost-effectiveness of poison prevention strategies for preschool children at home.

In the account below, a summary of the key features and principal findings of the 8 chapters preceding this discussion chapter is presented. The first three chapters provided a background introduction to the challenges of evidence in PH, the example problem and an overview of the methods to be used in the thesis.

Chapter 4 then presented a systematic review (Achana *et al.*, 2014b) carried out to determine how evidence is currently being synthesised in the NICE PH appraisal process and to determine barriers to quantitative synthesis of evidence in PH evaluations. The review identified that the main barrier to the use of quantitative synthesis methods in PH systematic

reviews is the heterogeneous nature of PH evidence. Specific issues identified include heterogeneity in i) many aspects of study designs, ii) definition of intervention strategies and outcome measures, and iii) the wider scope of many PH research questions. This last point makes it increasingly impractical to define interventions and outcomes clearly. Even where an attempt was made to summarise the evidence from the systematic review quantitatively, the synthesis tended to use the simplest methods such as fixed or random effects pairwise meta-analysis. Although carrying out pairwise meta-analysis may be preferred than carrying out a less focused narrative summary of the evidence, only comparison between any two interventions is possible at any one time which potentially limits the scope of the analysis and the utility of the findings (Achana *et al.*, 2014b).

Exploration of quantitative synthesis methods carried out in response to the systematic review findings indicated that more advanced synthesis methodology can be employed to overcome some of the issues identified as barriers to widespread use of meta-analysis in PH evaluations (Achana *et al.*, 2014b). Such methods have the potential to model the data more realistically and to answer policy relevant questions directly. Examples of the methods being advocated in Achana *et al.*(2014b) include extending the standard pairwise meta-analysis to: i) incorporate individual participant data (where available) and adjust for both summary and individual-level covariates (Sutton *et al.*, 2008), and ii) network meta-analysis and meta-regression (Cooper *et al.*, 2009) of summary (Caldwell *et al.*, 2005) and individual-level (Saramago *et al.*, 2012) data where a large number of interventions can be compared with one another in a coherent analysis (Caldwell *et al.*, 2005). The second of these two points was implemented in Chapter 5 where NMAs were used to compare the effectiveness of different interventions in promoting poison prevention behaviours in households with children (Achana *et al.* 2014, paper under review). The results of the NMAs suggested that more intensive interventions were more effective than education alone in promoting uptake of poison prevention practices in the home. These findings could not have been established using pairwise meta-analysis which would have required interventions of differing intensity to be lumped together in order to facilitate quantitative summary of the evidence, as was done in the Cochrane home safety systematic review (Kendrick *et al.*, 2012b).

Further modelling extensions of the standard network meta-analysis model to include a covariate for the baseline risk (Achana *et al.*, 2013) and to compare multiple interventions across multiple outcomes (Achana *et al.*, 2014a) were developed in Chapters 6 and 7 respectively. The methods developed in Chapter 6 specifically allowed for baseline imbalances in the control group event rate across studies (often referred to as baseline risk) to be taken into account in the network meta-analysis. This is important in a decision making context as heterogeneity in the baseline risk (i.e. in the control group event rate) may have implications when deciding which patient groups are most likely to benefit from the intervention. Application of the baseline risk NMA model to the accidents data found no evidence to suggest that the effectiveness estimates were related to the baseline rate of safety practices in households with children. In Chapter 7, a multivariate NMA model was developed to allow for multiple interventions to be compared with one another across multiple outcome measures while accounting for the correlation structure between outcomes (Achana *et al.*, 2014a). Extensions of the model allowed for extrapolation of evidence across a series of evidence networks. This enabled information sharing on the effectiveness of interventions across a wide range of different poison prevention measures. This type of analysis can lead to more precise estimates of the treatment effects by making use of all available information relevant to the decision problem (including information from closely related evidence networks). This can have the added benefit of reducing decision uncertainty when the analysis is used to inform an economic evaluation. Such an analysis would be useful in situations where the evidence base is either sparse or limited in other respects as is often the case in PH evaluations.

Finally in Chapter 8, a probabilistic decision analytic model was developed to assess the cost-effectiveness of poison prevention strategies for pre-school children in the home. This economic evaluation is believed to be the first to look at the cost-effectiveness of strategies to prevent accidental poisons in children under 5 years of age. The methods and analysis reported in Chapters 5, 6 and 7 were employed to synthesise the evidence on intervention effectiveness and used to inform the cost-effectiveness analysis. The base case analysis suggested that home safety interventions were unlikely to be cost-effective compared to usual care (i.e. do nothing approach) for the UK NHS unless policy makers are willing to pay upwards of £75,000 for every QALY gained. There was however considerable uncertainty in

these estimates due mainly to limitations of the data but also because of uncertainties about model structure and assumptions. The sensitivity analyses indicated that the base case results were largely robust to those changes in the parameter values and assumption of the model that were implemented. However, the results were quite sensitive to the cost of home safety education and/or provision of low cost/free equipment, lending support to the conclusion that cost-effectiveness of poison prevention strategies were largely driven by the cost of providing the intervention.

9.3 Strengths and limitations

This section takes a general overview of the strengths and limitations of the more quantitative approach to PH evaluations advocated and implemented in this thesis. Detailed discussions of the advantages and limitations of the specific analysis or methodology developed have been presented in the concluding sections of the respective chapters. In the account below, the strengths and limitations of the evidence synthesis models are presented first, followed by those relating to the decision analytic models.

- **Evidence synthesis**

This section discusses the synthesis of PH evidence and is based on the issues identified in the systematic review by Achana *et al.* (Achana *et al.*, 2014b). Many of the challenges of evidence synthesis identified in that review (often cited as reasons for not pooling the data in PH systematic reviews) are related to the heterogeneous nature of PH evidence and include variations in many aspects of study design, interventions and outcome measures. The quantitative synthesis of PH evidence suggested and applied throughout this thesis demonstrated that more advanced evidence synthesis methodology can be employed to overcome the specific issues of heterogeneity. These methods enable a more realistic modelling of the type of data commonly available in PH evaluations. They are potentially more useful when summarising evidence to inform an economic evaluation than carrying out simple pairwise meta-analysis or less focused and detailed reviews of the literature. Underlying this desire for public health evaluations to become more quantitative, in the face of the challenges encountered, is a firm belief that a structured and transparent description

and analysis of the decision question is desirable. This belief is based on the premise that while concluding the evidence base to be "too heterogeneous for meta-analysis", may be better than carrying out a naively simple meta-analysis, not being able to present a quantitative analysis severely restricts the utility of the review; particularly for decision making (Achana *et al.*, 2014b).

The more quantitative approach to evidence synthesis in PH evaluations advocated in this thesis thus far is not without limitations. Firstly, the systematic review (Achana *et al.*, 2014b) carried out to identify the barriers to meta-analysis of PH evidence only considered evaluations carried out by NICE. The review findings and conclusions may therefore not directly apply to PH evaluations in other contexts. Secondly, as exemplified by the poison prevention data presented in Chapter 5 and articulated earlier on in this thesis (Chapter 1 Section 1.2), many PH systematic reviews include observational evidence and non-randomised study designs which are more prone to high risk of bias than the traditional RCT. Despite this, beyond carrying out a risk of bias assessment and sensitivity analyses excluding non-RCT studies, none of the methods and analyses outlined directly considered the influence of the study quality/validity in the effectiveness and cost-effectiveness evaluations in this thesis. This is a potential limitation of the evidence synthesis undertaken in this thesis. An alternative, potentially more useful and sophisticated approach than a 'leave one study out' sensitivity analysis, is to model the various sources of biases in each study more directly when including observational evidence (Turner *et al.*, 2009; Thompson *et al.*, 2010). Finally, regarding the specific injury prevention context, for the analyses presented in Chapter 5, even when categorising the interventions into seven distinct groups, there is still residual heterogeneity in intervention definition; for example, education may be a leaflet designed for the prevention of an injury in the home, but it may also include a face-to-face interview, or a computer questionnaire producing tailored advice based on the user answers.

- **Decision analysis**

The analysis presented in Chapter 8 represents a transparent attempt to assess the cost-effectiveness of several home safety interventions compared with usual care to prevent unintentional medicinal poisons in preschool at home. The only other economic analysis in the area of childhood accidents prevention (Pitt *et al.*, 2009) evaluated the cost-effectiveness

of a generic home safety intervention versus no intervention for all home accidents irrespective of the mechanism or cause of injury (i.e. injury included falls, scalds, poisonings, etc). In contrast, this analysis considered interventions that are more clearly defined than the generic intervention in Pitt *et al*. (2009) and aimed at the prevention of medicinal poisons in pre-school children. This analysis therefore directly answers the question – 'What is the most cost-effective strategy for the prevention of accidental poisonings in the preschool age group who are most at risk of accidental injury?'.

The cost-effectiveness evaluation is also fully probabilistic, allowing for parameter uncertainty to be taken into account in the analysis. Where there remains uncertainty in model assumptions and structure, or uncertainty regarding multiple sources of evidence to use, these were investigated through scenario and deterministic sensitivity analyses. These sensitivity analyses indicated that the cost-effectiveness estimates were largely robust to changes in those model inputs that were investigated. Nevertheless, there are a number of limitations, the most obvious of which is the paucity of data that was available to inform the decision model. Data on the clinical history and outcomes of childhood poison injury was not always available in a form suitable for the analysis, necessitating assumptions in order to carry out the analysis. For example, data on the number of unintentional poison cases being admitted was only available for England rather than for the whole UK population as required in the decision model. In another example, the proportion of poison cases that developed into permanent injuries was taken from an American study (Mowry *et al.*, 2013) that may not be very relevant to the UK population. These data limitation issues and the resulting assumptions all represent sources of uncertainty that are difficult to quantify without availability of good data.

Utility data were also poor and almost not available in a form suitable for the cost-effectiveness evaluations. Data on the general background utility for the non-injured UK population under 18 years of age was not available. Hence utility norms for children in the model were assumed equal to the utility norm for the 18-25 year group in the UK population utility norms report (Kind *et al.*, 1998). There was also paucity of data on the utility decrement for poison injuries of varying severity. Only two sources (Miller *et al.*, 2000;

Miller *et al.*, 2012) of utility decrement in the form of QALY loss for poison related injury in children were found. Both sources were American studies, which hence may not adequately quantify the impact of poison injury in a UK population. Sensitivity analyses were thus conducted in which the uncertainty associated with the utility decrements used in the model was increased from 10% to 20% of the respective mean value. These analyses suggested the cost-effectiveness estimates were largely robust, but it must be stressed that the impact of such sensitivity analysis (which were based on the analyst's personal opinion rather than supported evidence) is itself uncertain and highly speculative.

9.4 Future work

This section outlines potential further extensions (both applied and methodological) of the work presented in this thesis. A number of these investigations concern issues that came to light during the conduct of the analysis in this thesis but were not adequately addressed for reasons such as lack of data and methods to conduct an appropriate analysis and constraints imposed by time resources and the permitted length of the thesis. First the opportunities for further work relating to the evidence synthesis models are discussed followed by those relating to the decision analysis.

- **Evidence Synthesis**

The network meta-analyses carried out in Chapter 5, results of which informed the decision analyses in Chapter 8, could be extended to include both summary data and individual patient data (IPD) when available. Methods already exist to allow network meta-analysis to be conducted using IPD (Donegan *et al.*, 2012; Saramago *et al.*, 2012). This would enable heterogeneity in intervention effects and patient-level treatment-by-covariate interactions to be fully explored and used to assess the appropriateness of the underlying assumptions of the network meta-analysis. Further methodological work in this area can also be undertaken including extending the IPD NMA model to the multiple outcomes setting, building on the work of Achana *et al.* (Achana *et al.*, 2014a). Such modelling extensions will potentially allow for treatment-by-covariate effects to be explored while accounting for the correlation structure within the data.

As a starting point and for IPD with a binary outcome variable, a Bernoulli distribution could be used to model the within-study covariance structure in a hierarchical modelling framework. This would avoid the need to introduce artificial cell corrections when studies report a 0 or 100% event rate in one of the intervention arms as in Achana et al. (Achana *et al.*, 2014a). The IPD network meta-regression models could then be integrated within a decision analytic model to carry out a cost-effectiveness analysis (Saramago Goncalves, 2012). This would potentially increase the statistical power of the modelling, allowing differential effects over population subgroups to be explored in order to customise the cost-effectiveness analysis to specific patient groups (for example, to evaluate the cost-effectiveness in populations with different levels of deprivation).

- **Decision analysis**

The decision analyses presented in this thesis were conducted from the perspective of the NHS and Personal Social Services. Thus only heath sector costs and benefits were considered in the cost-effectiveness evaluations. Further modelling extensions could be carried out to encompass alternative perspectives, for example, a societal perspective which can be argued to be a more appropriate in economic evaluation of PH interventions because of the emphasis on the community and society rather than the individual in PH (Weatherly *et al.*, 2009). This is because the benefits of PH interventions could extend beyond health benefits to include non-health-related benefits. Therefore, it may be argued that only analyses conducted from the perspective of the public sector could incorporate the full extent of the benefits and consequences associated with PH health intervention.

As stated above, a number of assumptions were made during the construction of the decision model, in some cases to simplify the model structure and in others simply to facilitate the conduct of the analysis given the available data. These assumptions introduce a degree of uncertainty in the decision model results. It would be useful to investigate the impact or otherwise of such modelling assumptions when feasible to do so. As an example, to simplify the model, the incidence of accidental poisonings as a result of exposure to a medicinal substance was assumed to be constant in the first five years of a child's life. However, there is evidence to suggest that the rate of accidental poisonings in children increases with age in the first years of life to a peak in about year 2 and then decreases with increasing age between

years 2 and 5 (Tyrrell *et al.*, 2012). This can easily be incorporated in the analysis simply by allowing the rate of accidental poisonings to depend on age.

As pointed out in Section 9.3 above, the main limitation of the cost-effectiveness evaluation was the paucity of data to populate the decision analytic models. Further primary research is therefore required to, for example, map out the clinical history and prognosis of childhood poisonings and estimate health outcomes and quality of life information (QALYs) for childhood poisoning injury. Given the paucity of data to conduct even the basic cost-effectiveness evaluation, no attempt was made to investigate and quantify the value of further research in reducing decision uncertainty associated with parameters of the model where there was a high degree of uncertainty (Chapter 12 of Welton et al. 2012 - "Expected value of information for research prioritisation and study design"). This type of value of information analysis has already been used to assess the value of conducting further research on decision uncertainty associated with whether or not to adopt a smoke alarm give away scheme in households with children (Saramago *et al.*, 2014). Similar analyses can be carried out if and when data becomes available, to quantify expected costs and consequences associated with decision uncertainty about whether or not to adopt a particular poison prevention strategy, given the available information.

9.5 Conclusion

In conclusion, this thesis has demonstrated through application to an active area of PH research, that, when summarising evidence to inform an economic evaluation within PH, more advanced synthesis methods can be employed to handle issues of heterogeneity. This will allow for a more realistic modelling of data and answer the policy relevant questions in PH evaluations than is often possible with pairwise meta-analysis or narrative summary of the evidence. Researchers working on PH evaluations should therefore consider expanding their toolbox and skills in the use of more sophisticated synthesis methods. Regarding the specific PH example (namely the prevention of unintentional poisonings in preschool children at home) used in this thesis, although the approach to synthesis of the effectiveness enabled more realistic analysis of the data, the cost-effectiveness evaluations were limited by

the lack of good data to inform many parameters of the model. The findings of the cost-effectiveness analyses should therefore be interpreted in the light of these limitations. Nevertheless, the cost-effectiveness analysis carried out in this thesis represent a more coherent and transparent process of integrating multiple and diverse sources of evidence, and is an example of approaches which could be more widely adopted in PH contexts. This helps answer policy relevant questions such as "what the most cost-effective strategy(s) that the NHS should adopt for preventing unintentional poisoning in pre-school children".

REFERENCE

Achana, F., *et al.*, (2014a). "Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes." BMC Medical Research Methodology **14**(1): 92.

Achana, F., *et al.*, (2014b). "An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial." Journal of Clinical Epidemiology **67**(4): 376-390.

Achana, F. A., *et al.*, (2013). "Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis." Statistics in Medicine **32**(5): 752-771.

Ades, A. E., *et al.*, (2005). "The Interpretation of Random-Effects Meta-Analysis in Decision Models." Medical Decision Making **25**(6): 646-654.

Ades, A. E., *et al.*, (2010). "Network meta-analysis with competing risk outcomes." Value in Health **13**(8): 976-983.

Anokye, N. K., *et al.*, (2011). "The cost-effectiveness of exercise referral schemes." BMC Public Health **11**.

Arends, L. R., *et al.*, (2000). "Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses." Statistics in Medicine **19**(24): 3497-3518.

Arends, L. R., *et al.*, (2003). "Combining multiple outcome measures in a meta-analysis: an application." Statistics in Medicine **22**(8): 1335-1353.

Armstrong, R., *et al.*, (2008). Chapter 21: Reviews in health promotion and public health. In Higgins JPT, Green S (editors). Cochrane Hanbook for Systematic reviews of Interventions. Chichester (UK), John Wiley & Sons.

Babul, S., *et al.*, (2007). "A randomized trial to assess the effectiveness of an infant home safety programme." International Journal of Injury Control and Safety Promotion **14**(2): 109-117.

Baio, G. (2013). Bayesian Methods in Health Economics. Boca Raton, Chapman & Hill.

Bank of England (2014). "Inflation Calculator. Available from http://www.bankofengland.co.uk/education/Pages/resources/inflationtools/calculator/index1.aspx. Accessed 15April 2014."

Barnard, J., *et al.*, (2000). "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage." Statistica Sinica **10**(4): 1281-1311.

Benson, B. E., *et al.*, (2013). "Position paper update: gastric lavage for gastrointestinal decontamination." Clinical Toxicology **51**(3): 140-146.

Berkey, C. S., *et al.*, (1998). "Meta-analysis of multiple outcomes by regression with random effects." Statistics in Medicine **17**(22): 2537-2550.

Bojke, L., *et al.*, (2009). "Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods." Value in Health **12**(5): 739-749.

Borenstein, M., *et al.*, (2009). Introduction to Meta-Analysis, Wiley 2009.

Brand, R. and H. Kragt (1992). "Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials." Statistics in Medicine **11**(16): 2077-2082.

Briggs, A. (2000). "Handling Uncertainty in Cost-Effectiveness Models." Pharmacoeconomics **17**(5): 479-500.

Briggs, A., *et al.*, (2006). Decision Modelling for Health Economic Evaluation. Oxford, Oxford University Press.

Briggs, A. and M. Sculpher (1998). "An introduction to Markov modelling for economic evaluation." Pharmacoeconomics **13**(4): 397-409.

Brooks, S. and A. Gelman (1998). "Alternative methods for monitoring convergence of iterative simulations." Journal of Computational and Graphical Statistics (7): 434-455.

Bujkiewicz, S., *et al.*, (2013). "Multivariate meta-analysis of mixed outcomes: a Bayesian approach." Statistics in Medicine **32**(22): 3926-3943.

Caldwell, D. M., *et al.*, (2005). "Simultaneous comparison of multiple treatments: combining direct and indirect evidence." British Medical Journal **331**(7521): 897-900.

Caldwell, D. M., *et al.*, (2010). "Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency." Journal of Clinical Epidemiology **63**(8): 875-882.

Carter, P., *et al.*, (2014). "A Mediterranean diet improves HbA1c but not fasting blood glucose compared to alternative dietary strategies: a network meta-analysis." Journal of Human Nutrition and Dietetics **27**(3): 280-297.

Chalkidou, K., *et al.*, (2008). "Cost-effective public health guidance: Asking questions from the decision-maker's viewpoint." Health Economics **17**(3): 441-448.

Children Act (2004). Children Act 2004. UK.

Chu, H. T., *et al.*, (2009). "Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection." Statistics in Medicine **28**(18): 2384-2399.

Cooper, N. J., *et al.*, (2011a). "Network Meta-analysis to Evaluate the Effectiveness of Interventions to Increase the Uptake of Smoke Alarms." Epidemiologic Reviews.

Cooper, N. J., *et al.*, (2011b). "How Valuable are Multiple Treatment Comparison Methods in Evidence-Based Health-Care Evaluation?" Value in Health **14**(2): 371-380.

Cooper, N. J., *et al.*, (2004). "Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach." Health Economics **13**(3): 203-226.

Cooper, N. J., *et al.*, (2006). "Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation." Archives of Internal Medicine **166**(12): 1269-1275.

Cooper, N. J., *et al.*, (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.

Curtis, L. (2012). Unit Costs of Health & Social Care. In.; 2012. Avaialble from http://www.pssru.ac.uk/project-pages/unit-costs/2012/. Accessed 6 May 2014.

Dakin, H. A., *et al.*, (2011). "Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension." Statistics in Medicine **30**(20): 2511-2535.

Davis, S., *et al.*, (2014). NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation. Available from http://www.nicedsu.org.uk.

Department of Health (1999). Saving Lives: Our Healthier Nation. London, The Stationery Office.

Department of Health (2002). Preventing accidental injury– priorities for action. Report to the Chief Medical Officer from the Accidental Injury Task Force. . London.

Department of Health (2005). Choosing Health: Making healthy choices easier. London.

Dershewitz, R. A. and J. W. Williamson (1977). "Prevention of Childhood Household Injuries - Controlled Clinical-Trial." American Journal of Public Health **67**(12): 1148-1153.

Dias, S., *et al.*, (2012). "A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials." Medical Decision Making.

Dias, S., *et al.*, (2010). "Checking consistency in mixed treatment comparison meta-analysis." Statistics in Medicine **29**(7-8): 932-944.

Dias, S., *et al.*, (2011a). "NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-analysis of Randomised Controlled Trials: Available from http://www.nicedsu.org.uk."

Dias, S., *et al.*, (2011b). NICE Technical Support Document 4: Inconsistency in networks of evidence based on randomised trials: Available from http://www.nicedsu.org.uk.

Dias, S., *et al.*, (2013). "Evidence Synthesis for Decision Making 4: Inconsistency in Networks of Evidence Based on Randomized Controlled Trials." Medical Decision Making **33**(5): 641-656.

Dickersin, K., *et al.*, (1987). "Publication Bias and Clinical-Trials." Controlled Clinical Trials **8**(4): 343-353.

DiGuiseppi, C., *et al.*, (2001). "Interventions for promoting smoke alarm ownership and function." Cochrane Database of Systematic Reviews(2): 1-77.

DiGuiseppi, C. and J. P. Higgins (2001). "Interventions for promoting smoke alarm ownership and function." Cochrane Database of Systematic Reviews(2): CD002246.

DiGuiseppi, C. and J. P. T. Higgins (2000). "Systematic review of controlled trials of interventions to promote smoke alarms." ARCHIVES OF DISEASE IN CHILDHOOD **82**(5): 341-348.

Donegan, S., *et al.*, (2012). "Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates." Statistics in Medicine **31**(29): 3840-3857.

Draper, N. R. and H. Smith (1998). Applied Regression Analysis. London.

Drummond, M. F., *et al.*, (2005). Methods for the Economic Evaluation of Health Care Programmes Oxford, Oxford University Press

DuMouchel, W. and P. G. Groer (1989). "A Bayesian Methodology for Scaling Radiation Studies from Animals to Man." Health Physics **57**: 411-418.

DuMouchel, W. H. and J. E. Harris (1983). "Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species." Journal of the American Statistical Association **78**(382): 293-308.

Duval, S. and R. Tweedie (2000). "A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis." Journal of the American Statistical Association **95**(449): 89-98.

Eden, J., *et al.*, (2012). Finding What Works in Health Care: Standards for Systematic Reviews. Available from http://www.nap.edu/catalog.php?record_id=13059.

Efthimiou, O., *et al.*, (2014). "An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios." Statistics in Medicine **33**(13): 2275-2287.

Egger, M., *et al.*, (1997). "Bias in meta-analysis detected by a simple, graphical test." British Medical Journal **315**(7109): 629-634.

Egger, M., *et al.*, (2000). "Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature." Journal of Clinical Epidemiology **53**(11): 1119-1129.

Errington, G., *et al.*, (2011). Evaluation of the National Safe At Home Scheme. Final report for the Royal Society for the Prevention of Accidents. Avialable online from http://www.rospa.com/homesafety/safeathome/final-evaluation-report.pdf. Accessed 10 May 2014.

Franchini, A. J., *et al.*, (2012). "Accounting for correlation in network meta-analysis with multi-arm trials." Research Synthesis Methods **3**(2): 142-160.

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models(Comment on an Article by Browne and Draper)." Bayesian Analysis **1**(3): 515-533.

Ghidey, W., *et al.*, (2007). "Semi-parametric modelling of the distribution of the baseline risk in meta-analysis." Statistics in Medicine **26**(30): 5434-5444.

Ghidey, W., *et al.*, (2011). "Modelling the effect of baseline risk in meta-analysis: A review from the perspective of errors-in-variables regression." Statistical Methods in Medical Research.

Glass, G. V. (1976). "Primary, secondary and meta-analysis of research." Educational Researcher **5**: 3-8.

Gray, A. M., *et al.*, (2011). Applied Methods of Cost-effectiveness Analysis in Health Care. Oxford, Oxford University Press.

Harbord, R. M., *et al.*, (2006). "A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints." Statistics in Medicine **25**(20): 3443-3457.

Health & Social Care Information Centre (2013a). "Hospital Episode Statistics for England, Accident and Emergency (A&E) statistics, 2011-12. Avialable http://www.hscic.gov.uk/catalogue/PUB13464 ".

Health & Social Care Information Centre (2013b). "Hospital Episode Statistics for England. Inpatient statistics, 2012-13. Available from http://www.hscic.gov.uk/hes. Accessed 5 May 2014."

Health, D. o. (2007). National Service Framework for Children,Young People and Maternity Services: Executive Summary. London.

Higgins, J. P. T. and S. E. Green (2011). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0., The Cochrane Collaboration 2011.

Higgins, J. P. T. and S. G. Thompson (2002). "Quantifying heterogeneity in a meta-analysis." Statistics in Medicine **21**(11): 1539-1558.

Higgins, J. P. T., *et al.*, (2009). "A re-evaluation of random-effects meta-analysis." Journal of the Royal Statistical Society Series a-Statistics in Society **172**: 137-159.

Hojer, J., *et al.*, (2013). "Position paper update: ipecac syrup for gastrointestinal decontamination." Clinical Toxicology **51**(3): 134-139.

Hong, H., *et al.*, (2013). "Methods Research Report: A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons. Avaialable online from http://www.ncbi.nlm.nih.gov/books/NBK116689/pdf/TOC.pdf."

Hubbard, S., *et al.*, (2014). "Network meta-analysis to evaluate the effectiveness of interventions to prevent falls in children under age 5 years." Injury Prevention.

Jackson, C. H., *et al.*, (2009). "Accounting for uncertainty in health economic decision models by using model averaging." Journal of the Royal Statistical Society: Series A (Statistics in Society) **172**(2): 383-404.

Jackson, D., *et al.*, (2011). "Multivariate meta-analysis: Potential and promise." Statistics in Medicine **30**(20): 2481-2498.

Johnston, B. D., *et al.*, (2006). "Healthy steps in an integrated delivery system - Child and parent outcomes at 30 months." Archives of Pediatrics & Adolescent Medicine **160**(8): 793-800.

Jones, D. R., *et al.*, (2009). "Interspecies extrapolation in environmental exposure standard setting: A Bayesian synthesis approach." Regulatory Toxicology and Pharmacology **53**(3): 217-225.

Kelly, B., *et al.*, (1987). "Safety education in a pediatric primary care setting." Pediatrics **79**(5): 818-824.

Kendrick, D., *et al.*, (2007). "Home safety education and provision of safety equipment for injury prevention." Cochrane Database of Systematic Reviews(1): CD005014.

Kendrick, D., *et al.*, (2012a). "Keeping children safe at home: protocol for three matched case–control studies of modifiable risk factors for falls." Injury Prevention **18**(3): e3.

Kendrick, D., *et al.*, (2008). "Effect of education and safety equipment on poisoning-prevention practices and poisoning: systematic review, meta-analysis and meta-regression." Archives of Disease in Childhood **93**(7): 599-608.

Kendrick, D., *et al.*, (2012b). "Home safety education and provision of safety equipment for injury prevention." Cochrane Database of Systematic Reviews(9).

Kendrick, D., *et al.*, (2012c). "Home safety education and provision of safety equipment for injury prevention." Cochrane Database of Systematic Reviews. The Cochrane Library **9**.

Kind, P., *et al.*, (1998). "Variations in population health status: results from a United Kingdom national questionnaire survey." British Medical Journal **316**(7133): 736 - 741.

Kirkham, J. J., *et al.*, (2012). "A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews." Statistics in Medicine **31**(20): 2179-2195.

Lambert, P. C., *et al.*, (2005). "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS." Statistics in Medicine **24**(15): 2401-2428.

Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." Statistics in Medicine **23**(20): 3105-3124.

Lu, G. and A. E. Ades (2006). "Assessing Evidence Inconsistency in Mixed Treatment Comparisons." Journal of the American Statistical Association **101**(474): 447-459.

Lu, G., *et al.*, (2007). "Meta-analysis of mixed treatment comparisons at multiple follow-up times." Statistics in Medicine **26**(20): 3681-3699.

Lu, G. B. and A. Ades (2009). "Modeling between-trial variance structure in mixed treatment comparisons." Biostatistics **10**(4): 792-805.

Lumley, T. (2002). "Network meta-analysis for indirect treatment comparisons." Statistics in Medicine **21**(16): 2313-2324.

Lunn, D., *et al.*, (2012). The BUGS Book - A Practical Introduction to Bayesian Analysis.

Lunn, D. J., *et al.*, (2000). "WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility." Statistics and Computing **10**(4): 325-337.

Mavridis, D. and G. Salanti (2012). "A practical introduction to multivariate meta-analysis." Statistical Methods in Medical Research.

McCullagh, P. and J. Nelder (1989). Generalized Linear Models. London, Chapman and Hall.

McDaid, C., *et al.*, (2010). "Paracetamol and selective and non-selective nonsteroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery:a systematic review." Health Technology Assessment **14**(17): 1-+.

McIntosh, M. W. (1996). "The population risk as an explanatory variable in research synthesis of clinical trials." Statistics in Medicine **15**(16): 1713-1728.

Miller, T. R., *et al.*, (2012). The Cost of Child and Adolescent Injuries and the Savings from Prevention. Injury Prevention for Children and Adolescents: Research, Practice and Advocacy. K. D. Liller. Washington, American Public Health Association**:** 21-82.

Miller, T. R., *et al.*, (2000). "The Cost of Childhood Unintentional Injuries and the Value of Prevention. Available from http://futureofchildren.org/futureofchildren/publications/docs/10_01_05.pdf. Accessed 05 May 2014."

Mills, E. J., *et al.*, (2011). "Multiple treatment comparison meta-analyses: a step forward into complexity." Clin Epidemiol **3**: 193-202.

Minkovitz, C. S., *et al.*, (2003). "A practice-based intervention to enhance quality of care in the first 3 years of life the - Healthy steps for young children program." Journal of the American Medical Association **290**(23): 3081-3091.

Moreno, S. G., *et al.*, (2009). "Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications." British Medical Journal **339**.

Mowry, J. B., *et al.*, (2013). "2012 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 30th Annual Report." Clinical Toxicology **51**(10): 949-1229.

Nam, I. S., *et al.*, (2003). "Multivariate meta-analysis." Statistics in Medicine **22**(14): 2309-2333.

Nansel, T. R., *et al.*, (2008). "Preventing unintentional pediatric injuries: a tailored intervention for parents and providers." Health Educ Res **23**(4): 656-669.

NICE (1999). The National Institute for Clinical Excellence (Establishment and Constitution) Order 1999 Avaialable from

http://www.legislation.gov.uk/uksi/1999/220/pdfs/uksi_19990220_en.pdf.

NICE (2006). Methods for the development of NICE public health guidance. Avialable from http://www.nice.org.uk/media/FB9/59/PHMethodsManual2006.pdf.

NICE (2008). "Updated guide to the methods of technology appraisal - June 2008. Available from http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf."

NICE (2009). Methods for the development of NICE public health guidance (second edition). Available from http://www.nice.org.uk/media/CE1/F7/CPHE_Methods_manual_LR.pdf.

NICE (2010a). "PH30 Preventing unintentional injuries among under-15s in the home: costing template. Available from http://guidance.nice.org.uk/PH30/CostingTemplate/xls/English. Accessed 5 May 2014."

NICE (2010b). "Preventing unintentional injuries among the under-15s in the home. Available from http://www.nice.org.uk/guidance/ph30. Accessed 5 September 2014."

NICE (2012). Methods for the development of NICE public health guidance (third edition). Available from http://publications.nice.org.uk/methods-for-the-development-of-nice-public-health-guidance-third-edition-pmg4.

NICE (2013). Guide to the methods of technology appraisal 2013. Avaialable from http://publications.nice.org.uk/pmg9.

Novielli, N., *et al.*, (2010). "How Is Evidence on Test Performance Synthesized for Economic Decision Models of Diagnostic Tests? A Systematic Appraisal of Health Technology Assessments in the UK Since 1997." Value in Health **13**(8): 952-957.

Ntzoufras, I. (2009). Bayesian Modelling Using WinBUGS. New Jersey, John Wiley & Sons.

Office for National Statistics (2010). Interim Life Tables, England & Wales http://www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/interim-life-tables/index.html.

Office for National Statistics (2012) "Death Registration Summary Tables - England and Wales, 2012. Available from http://www.ons.gov.uk/ons. Accessed 5 May 2014."

Office for National Statistics (2013) "Statistical Bulletin: Annual Mid-year Population Estimates, 2011 and 2012. Available from http://www.ons.gov.uk/ons/dcp171778_320900.pdf. Accessed 05 May 2014."

Pavey, T. G., *et al.*, (2011). "The clinical effectiveness and cost-effectiveness of exercise referral schemes: a systematic review and economic evaluation." Health Technology Assessment **15**(44): 1-+.

Pearson, M., *et al.*, (2009). PREVENTING UNINTENTIONAL INJURIES AMONG UNDER-15s IN THE HOME. Report 1:Systematic reviews of effectiveness and cost-effectiveness of home safety equipment and risk assessment schemes. Available from http://www.nice.org.uk/guidance/ph30/resources/preventing-unintentional-injuries-among-under-15s-in-the-home-review-of-effectiveness-and-cost-effectiveness2. Accessed 5 September 2014.

Peters, J. L., *et al.*, (2008). "Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry." Journal of Clinical Epidemiology **61**(10): 991-996.

Peters, J. L., *et al.*, (2010). "Assessing publication bias in meta-analyses in the presence of between-study heterogeneity." Journal of the Royal Statistical Society Series a-Statistics in Society **173**: 575-591.

Petticrew, M. and M. Egan, Eds. (2006). Relevance, rigour and systematic reviews. Moving beyond effectiveness in evidence synthesis Methodological issues in the synthesis of diverse sources of evidence http://www.nice.org.uk/niceMedia/docs/Moving_beyond_effectiveness_in_evidence_synthesis2.pdf (accessed 06 June 2012).

Philips, Z., *et al.*, (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., *et al.*, (2004). "Review of guidelines for good practice in decision-analytic modelling in health technology assessment." Health Technology Assessment **8**(36): 1-+.

Phillips, C. J. (2005). Health Economics: An Introduction for Health Professionals, Blackwell.

Pitt, M., *et al.*, (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. Available from http://guidance.nice.org.uk/PH30/SupportingEvidence. Accessed 28 May 2013.

Reich, S. M., *et al.*, (2011). "Using Baby Books to Increase New Mothers' Safety Practices." Academic Pediatrics **11**(1): 34-43.

Riley, R. D. (2009). "Multivariate meta-analysis: the effect of ignoring within-study correlation." Journal of the Royal Statistical Society Series a-Statistics in Society **172**: 789-811.

Riley, R. D., *et al.*, (2007a). "An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes." Statistics in Medicine **26**(1): 78-97.

Riley, R. D., *et al.*, (2007b). "Bivariate random-effects meta-analysis and the estimation of between-study correlation." BMC Medical Research Methodology **7**.

Riley, R. D., *et al.*, (2007c). "Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods." Journal of Clinical Epidemiology **60**(5): 431-439.

Riley, R. D., *et al.*, (2008). "An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown." Biostatistics **9**(1): 172-186.

Rodgers, M., *et al.*, (2009). "Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews." Evaluation **15**(1): 49-73.

Rosen, L., *et al.*, (2006). "In defense of the randomized controlled trial for health promotion research." American Journal of Public Health **96**(7): 1181-1186.

Rucker, G., *et al.*, (2008). "Arcsine test for publication bias in meta-analyses with binary outcomes." Statistics in Medicine **27**(5): 746-763.

Salanti, G., *et al.*, (2008). "Evaluation of networks of randomized trials." Statiscal Methods Medical Research **17**(3): 279-301.

Salanti, G., *et al.*, (2009). "A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered." Journal of Clinical Epidemiology **62**(8): 857-864.

Saramago Goncalves, P. (2012). Methodological issues in the analysis of individual- and aggregate-participant level data for cost effectiveness analysis. PhD thesis, University of York. .

Saramago, P., *et al.*, (2014). "Cost-effectiveness of interventions for increasing the possession of functioning smoke alarms in households with pre-school children: a modelling study." BMC Public Health **14**(1): 459.

Saramago, P., *et al.*, (2012). "Mixed treatment comparisons using aggregate and individual participant level data." Statistics in Medicine **31**(28): 3516-3536.

Sharp, S. J. and S. G. Thompson (2000). "Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches." Statistics in Medicine **19**(23): 3251-3274.

Sharp, S. J., *et al.*, (1996). "The relation between treatment benefit and underlying risk in meta-analysis." British Medical Journal **313**(7059): 735-738.

Simmonds, M. C., *et al.*, (2005). "Meta-analysis of individual patient data from randomized trials: a review of methods used in practice." Clinical Trials **2**(3): 209-217.

Song, F., *et al.*, (2003). "Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses." British Medical Journal **326**(7387): 472-475.

Song, F. J., *et al.*, (2011). "Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study." British Medical Journal **343**.

Spiegelhalter, D., *et al.*, (2000). "Bayesian methods in health technology assessment: a review." Health Technology Assessment **4**(38): 130.

Spiegelhalter, D., *et al.*, (2007). " WinBUGS User Manual Version 1.4 January 2003. Upgraded to Version 1.4.3.".

Spiegelhalter, D. J., *et al.*, (2004). Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester, John Wiley & Sons.

Spiegelhalter, D. J., *et al.*, (2002). "Bayesian measures of model complexity and fit." Journal of the Royal Statistical Society (B) **64**: 1-34.

Spiegelhalter, D. J., *et al.*, (2014). "The deviance information criterion: 12 years on." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **76**(3): 485-493.

Stata Corporation (2012). Stata Statistical Software: Release 12. College Station, TX, StataCorp LP.

Sterne, J. A. C. and M. Egger (2001). "Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis." Journal of Clinical Epidemiology **54**(10): 1046-1055.

Sterne, J. A. C., *et al.*, (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials.

Stijnen, T., *et al.*, (2010). "Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data." Statistics in Medicine **29**(29): 3046-3067.

Sun, X. and T. Faunce (2008). "Decision-analytical modelling in health-care economic evaluations." The European Journal of Health Economics **9**(4): 313-323.

Sutton, A., *et al.*, (2000). Methods for Meta-analysis in Medical Research. Chichester, Wiley.

Sutton, A. J., *et al.*, (2009). "An encouraging assessment of methods to inform priorities for updating systematic reviews." Journal of Clinical Epidemiology **62**(3): 241-251.

Sutton, A. J., *et al.*, (2008). "Meta-analysis of individual- and aggregate-level data." Statistics in Medicine **27**(5): 651-669.

The Audit Commission (2007). Better safe than sorry. London.

Thompson, S., *et al.*, (2010). "A proposed method of bias adjustment for meta-analyses of published observational studies." International Journal of Epidemiology.

Thompson, S. G., *et al.*, (1997). "Investigating underlying risk as a source of heterogeneity in meta-analysis." Statitics in Medicine **16**(23): 2741-2758.

Torgerson, D. and J. Raftery (1999). "Economics notes - Measuring outcomes in economic evaluations." British Medical Journal **318**(7195): 1413-1413.

Turner, R. M., *et al.*, (2009). "Bias modelling in evidence synthesis." Journal of the Royal Statistical Society Series a-Statistics in Society **172**: 21-47.

Tyrrell, E. G., *et al.*, (2012). "Children at risk of medicinal and non-medicinal poisoning: a population-based case-control study in general practice." British Journal of General Practice **62**(605).

van Houwelingen, H. and S. Senn (1999). "Investigating underlying risk as a source of heterogeneity in meta-analysis." Statistics in Medicine **18**(1): 110-115.

van Houwelingen, H. C., *et al.*, (2002). "Advanced methods in meta-analysis: multivariate approach and meta-regression." Statistics in Medicine **21**(4): 589-624.

Walter, S. D. (1997). "Variation in baseline risk as an explanation of heterogeneity in meta-analysis." Statistics in Medicine **16**(24): 2883-2900.

Wanless, D. (2004). Securing good health for the whole population: Final report - February 2004 (Chapter 6). http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4074426 (accessed 06 June 2012).

Weatherly, H., *et al.*, (2009). "Methods for assessing the cost-effectiveness of public health interventions: Key challenges and recommendations." Health Policy **93**(2-3): 85-92.

Wei, Y. and J. P. T. Higgins (2013a). "Bayesian multivariate meta-analysis with multiple outcomes." Statistics in Medicine **32**(17): 2911-2934.

Wei, Y. and J. P. T. Higgins (2013b). "Estimating within-study covariances in multivariate meta-analysis with multiple outcomes." Statistics in Medicine **32**(7): 1191-1205.

Welton, N. J., *et al.*, (2008). "Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B." Statistics in Medicine **27**(27): 5620-5639.

Welton, N. J., *et al.*, (2012). Evidence Synthesis for Decision Making in Healthcare, John Wiley & Sons, Ltd.

World Health Organisation (2014). "Metrics: Disability-Adjusted Life Year (DALY). Avaialable from http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/. Accessed 10April 2014."

Young, B., *et al.*, (2013). "Preventing childhood falls within the home: Overview of systematic reviews and a systematic review of primary studies." Accident Analysis & Prevention **60**(0): 158-171.

APPENDICES

Appendix I: Extra table of results for Chapter 4

Reported reasons why meta-analysis was not conducted in systematic review of the effectiveness and or cost-effectiveness evidence for NICE Public Appraisals

Appraisal title (Reference number)	Systematic review title	Reported reasons for not doing a meta-analysis (page number)	Source of heterogeneity[+]
Brief interventions and referral for smoking cessation (PH1)	Review 25 January 2006	Overview of reviews. Summarised review level evidence	nr
Four commonly used methods to increase physical activity (PH2)	Brief interventions review 25 Jan 2006	Not reported	nr
	Pedometers Review 25 Jan 2006	Not reported	
	Exercise referral Review May 2006	Not reported	
	Walk Cycling Review 25 Jan 2006	Not reported	
Interventions to reduce substance misuse among vulnerable young people (PH4)	Substance misuse: effectiveness review - main report (PHIAC 5.3a revised)	A narrative approach to synthesis was undertaken since despite common primary outcome data (i.e. substance use), there was great variation across primary studies in how this was collected and reported (page 73).	o
Workplace interventions to promote smoking cessation (PH5)	PH5 Workplace interventions to promote smoking cessation: effectiveness review	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 16).	m
Behaviour change (PH6)	Behaviour change: Review 1 - Effectiveness review	No formal synthesis (such as meta-analysis) was undertaken, as a narrative summary of the results was more appropriate for a review of reviews (page 8).	nr
	Behaviour change: Review 2 - Road safety	No formal synthesis was undertaken as this was a review of reviews; instead, a narrative summary of the results was provided (page 19).	
Physical activity and the environment (PH8)	Building design evidence review	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 22).	i,m,o
	Environmental correlates of physical activity review	As there are different ways that exposure and outcome factors have been measured, a formal synthesis and quantitative comparison is not possible (page 4).	
	Natural environment evidence review	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 21).	
	Policy evidence review	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 18).	
	Transport evidence review	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 22).	
	Urban planning & design evidence review	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 22).	

Community engagement (PH9)	Community engagement: review 2 - approaches and methods	As this work covered differing methodologies, several community engagement methods/approaches, numerous intervention strategies, health behaviours and outcome measures it was considered inappropriate to synthesise the data using meta-analysis (page 31).	i,m,o,p
Maternal and child nutrition (PH11)	PH11 Maternal and child nutrition: review 1 preconception	Not stated	m
	PH11 Maternal and child nutrition: review 2 - nutrition	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 22).	
	PH11 Maternal and child nutrition: review 3 post-partum	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 27).	
	PH11 Maternal and child nutrition: review 4 - milk feeding	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 28).	
	PH11 Maternal and child nutrition: review 5 - 6-24 months	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 24).	
	PH11 Maternal and child nutrition: review 6 - 2-5 years	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 24).	
	PH11 Maternal and child nutrition: review 7 - vitamin D	Not stated.	
Promoting physical activity in the workplace (PH13)	Promoting physical activity in the workplace: final evidence review	Evidence is provided using a narrative synthesis, supported by evidence tables, drawing out the key features of each study (page 5).	nr
Preventing the uptake of smoking by children and young people (PH14)	Preventing the uptake of smoking by children and young people: review of effectiveness	Overall, due to heterogeneity of design among the studies, a narrative synthesis was conducted (page 29).	m,o
	Services in disadvantaged areas: Smoking review	Some studies did not include quantifiable outcomes. As a result, it was not possible to conduct data synthesis in the traditional way by, for example, pooling intervention effects between studies and generating forest plots to illustrate effects (page 17).	
Identifying and supporting people most at risk of dying prematurely (PH15)	Services in disadvantaged areas: Statins report	There was a large degree of heterogeneity in terms of interventions, settings, and populations so a narrative synthesis of the results was carried out (page 8).	i,m,p
Mental wellbeing and older people (PH16)	PH16 Mental wellbeing and older people: effectiveness and cost effectiveness review	Often the diversity of the interventions, the settings in which they were delivered, and the outcomes measured means that pooled estimates of effect are not appropriate (page 41).	i,m,o
Promoting physical activity for children and young people (PH17)	Promoting physical activity for children: review 2 - quantitative correlates	Overview of reviews. Summarised review level evidence	i,m,o
	Promoting physical activity for children: review 4 - interventions for under eights (revised July 2008)	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 20).	
	Promoting physical activity for children: review 5 - active travel interventions (revised July 2008	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page26).	
	Promoting physical activity for children: review 6 - interventions for adolescent girls	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 24).	
	Promoting physical activity for children: review 7 - family and community interventions	It was not appropriate to use meta-analysis to synthesise the outcome data as interventions, methods and outcomes were heterogeneous (page 27).	
Needle and syringe programmes (PH18)	PH18 Needle and syringe programmes: review of effectiveness and cost effectiveness, revised full report October 2008	The results of the data extraction and quality assessment for each study of effectiveness are presented in structured tables and as a narrative summary (page 21).	nr

Social and emotional wellbeing in secondary education (PH20)	PH20 Social and emotional wellbeing in secondary education: effectiveness review	The heterogeneity of the interventions aim, design and outcome measures used preclude a meta-analysis of their results (page 61).	i,m,o
Reducing differences in the uptake of immunisations (PH21)	PH21 Reducing the differences in the uptake of immunisations: revised analysis of the evidence	Not stated	nr
	PH21 Reducing the differences in the uptake of immunisations: evidence review	Not stated	
Promoting mental wellbeing at work (PH22)	Promoting mental wellbeing at work: economic review	Given the heterogeneity of the interventions considered in the retained studies, this review is restricted to a narrative overview of those studies (page 10).	i
Alcohol-use disorders - preventing harmful drinking (PH24)	Screening and brief interventions for prevention and early identification of alcohol use disorders in adults and young people	Pre-specified outcomes are tabulated in evidence tables and presented within a narrative synthesis (page 41).	nr
	Interventions on control of alcohol price, promotion and availability for prevention of alcohol use disorders	Pre-specified outcomes were tabulated in evidence tables and are presented within a narrative synthesis (page 32).	
Prevention of cardiovascular disease (PH25)	PH25 prevention of cardiovascular disease: reviews and primary studies - 1 effectiveness	Synthesis was narrative and meta-analysis was not employed (page ix).	nr
	PH25 prevention of cardiovascular disease: reviews and primary studies - 2 effectiveness	Synthesis was narrative and meta-analysis was not employed (page ix).	
	PH25 prevention of cardiovascular disease: reviews and primary studies - 3 effectiveness	Synthesis was narrative and meta-analysis was not employed (page ix).	
Quitting smoking in pregnancy and following childbirth (PH26)	Quitting smoking in pregnancy and following childbirth: systematic review	The heterogeneity of the interventions aim, design and outcome measures used preclude a meta-analysis of their results (page 43).	i,m,o
	Quitting smoking in pregnancy and following childbirth: interventions to improve partner support and partner cessation during pregnancy	Due to heterogeneity of design among the studies, a narrative synthesis was conducted (page19).	
	Quitting smoking in pregnancy and following childbirth: rapid review of interventions to prevent relapse in pregnant ex-smokers	Overview of reviews. Summarised review level evidence	
Looked-after children and young people (PH28)	Review E1 - Transition support services	Because of the variation in variables, methods and measures used, it was not possible to conduct a meta-analytical review (page 24).	m,o,p
	Review E2 - Training and support for carers	Because of the variation in variables, methods and measures used, it was not possible to conduct meta-analysis. P20	
	Review E3 - Improving access to services	Because of the variation in variables, methods and measures used, it was not possible to conduct a meta-analytical review (page 20).	
	Review E4 - Correlates	Because of the variation in variables, methods and measures used, it was not possible to conduct a meta-analytical review (page 22).	
Strategies to prevent unintentional injuries among under-15s (PH29)	Review 1: International comparative analyses of injury prevention policies, legislation and other activities	Not stated	i
	Review 2: Risk factors for unintentional injuries among under 15s	In accordance with previous NICE Public Health correlates reviews, we undertook a qualitative approach to data synthesis rather than a formal pooling of outcomes using meta-analysis (page 22).	
	Review 3: Strategies and frameworks to prevent unintentional injury among under 15s – Road design	No formal quantitative pooling of effectiveness results was possible or desirable especially given the wide range of non-automated enforcement and other strategies in our review (page 20).	
	Review 4: Strategies and frameworks to prevent unintentional injury among under 15s	We used narrative synthesis methods rather than formal data pooling (page 24).	

	– Home Review 5: Strategies, frameworks and mass media to prevent unintentional injury among under 15s - Outdoor play and leisure	Narrative synthesis was used because quantitative data pooling was not possible (page 20).	
Preventing unintentional injuries among under-15s in the home (PH30)	Preventing unintentional injuries among under 15s in the home: Review of effectiveness and cost effectiveness	A formal meta-analysis was not conducted in view of the heterogeneity of interventions and measurement of outcomes (page 38).	i,o
Preventing unintentional road injuries among under-15s: road design (PH31)	Preventing unintentional road injuries among under 15s: road design: Review of effectiveness and cost effectiveness - Main Report	Due to the heterogeneity of the interventions studied, it was considered inappropriate to combine the studies statistically using a meta-analysis (page 30).	i
Skin cancer prevention: information, resources and environmental changes (PH32)	Skin cancer prevention: information, resources and environmental changes: Review 1: Effectiveness and cost effectiveness evidence review - phase 1 Skin cancer prevention: information, resources and environmental changes: Review 4: Effectiveness and cost effectiveness evidence review - phase 2	Where possible a narrative summary across similar studies was undertaken. Data were then grouped by setting and intervention category (changes to the built or natural environment, provision of sun protection resources, and multi-component interventions) and presented as a narrative synthesis (page 25).	i,m
Increasing the uptake of HIV testing among black Africans in England (PH33)	PH33 Increasing the uptake of HIV testing among black Africans in England: review of effectiveness and cost effectiveness	Nearly all the identified studies used Framework or Grounded Theory as the methodological approach to analysis. As a result, the review team decided that thematic meta-analysis would be the most appropriate method for synthesising the data (page 19).	m
Increasing the uptake of HIV testing among men who have sex with men (PH34)	PH34 Increasing the uptake of HIV testing among men who have sex with men: review of effectiveness, cost-effectiveness and barriers	The studies of effectiveness did not support meta-analysis and were synthesised narratively, as were the cost-effectiveness studies (page 22).	nr
Tuberculosis - hard-to-reach groups (PH37)	Review 2 - Evidence review on the effectiveness and cost-effectiveness of interventions aimed at identifying people with tuberculosis and/or raising awareness of tuberculosis among hard-to-reach groups Review 3 - Evidence review on the effectiveness and cost-effectiveness of interventions aimed at managing tuberculosis in hard-to-reach groups	The studies of effectiveness did not support meta-analysis and were synthesised narratively, as were the cost-effectiveness studies (page 25). In most cases, the studies of effectiveness did not support meta-analysis and were reported narratively, as were the cost-effectiveness studies (page 25)	nr
Smokeless tobacco cessation - South Asian communities (PH39)	PH39 Smokeless tobacco cessation - South Asian communities: evidence reviews 1 and 2	Meta-analysis and the use of forest plots was not appropriate for the studies included in this review since there was a high degree of heterogeneity across studies (page 47).	s

[+]Sources of heterogeneity (i=interventions, m=study method/design/setting, o=outcomes, p=study populations, s=heterogeneity studies and nr= not reported).

Appendix II: Between-study covariance for multi-arm studies reporting multiple outcomes

We show that under evidence consistency and the homogenous between-study covariance structure, $\sigma^2_{(bk)} = \sigma^2_m$ and $\rho^{mn}_{bk} = \rho^{mn}$, equation (7.3) can be extended to the multiple outcome settings by formulating the distribution of effects in a multi-arm study *i* with *p+1* arms reporting on $m = 1, 2, \cdots, M$ outcomes as follows:

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(bk_2)1} \\ \vdots \\ \delta_{i(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(bk_p)1} \\ \vdots \\ \delta_{i(bk_{p1})M} \end{pmatrix} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \begin{pmatrix} d_{i(bk_1)1} \\ \vdots \\ d_{i(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} d_{i(bk_2)1} \\ \vdots \\ d_{i(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{i(bk_p)1} \\ \vdots \\ d_{i(bk_{p1})M} \end{pmatrix} \end{pmatrix}, \mathbf{\Sigma}_{(Mp \times Mp)} = \begin{pmatrix} \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \cdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \vdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & & \ddots & \vdots \\ & & & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \end{pmatrix} \right)$$

Equation (A1)

where *p* is the number of treatment effect estimates. The corresponding marginal and conditional distributions for arm *j*, given the previous $1, 2, \cdots, (j-1)$ arms are:

$$\begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk_1)1} \\ \vdots \\ d_{(bk_1)M} \end{pmatrix}, \mathbf{\Sigma}_{(M \times M)} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_1^2 \end{pmatrix} \right) \qquad \text{for } j = 1$$

$$\begin{pmatrix} \delta_{i(bk_j)1} \\ \vdots \\ \delta_{i(bk_j)M} \end{pmatrix} \Bigg| \begin{pmatrix} \begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(bk_{21})1} \\ \vdots \\ \delta_{i(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(bk_{j-1})1} \\ \vdots \\ \delta_{i(bk_{j-1})M} \end{pmatrix} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk_j)1} + \frac{1}{j}\sum_{t=1}^{j-1}\left(\delta_{i(bk_t)1} - d_{(bk_t)1}\right) \\ \vdots \\ d_{(bk_j)M} + \frac{1}{j}\sum_{t=1}^{j-1}\left(\delta_{i(bk_t)M} - d_{(bk_t)M}\right) \end{pmatrix}, \Sigma' = \frac{(j+1)}{2j}\Sigma_{(M\times M)} \right) \quad \text{for } j = 2,3,\cdots,p \qquad \textbf{Equation (A2)}$$

Equations (A1) and (A2) are the multiple outcome versions of equations (7.3) and (7.4) of Chapter 7. These equations can be derived by noting that, a random effects between-study model for a multi-arm study *i* with *K* treatments labelled *A, B, C,…, K* reporting a total of *M* outcomes labelled *1, 2, …, M* can be specified as:

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(AB)1} \\ \vdots \\ \delta_{i(AB)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(AC)1} \\ \vdots \\ \delta_{i(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(AK)1} \\ \vdots \\ \delta_{i(AK)M} \end{pmatrix} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \begin{pmatrix} d_{i(AB)1} \\ \vdots \\ d_{i(AB)M} \end{pmatrix} \\ \begin{pmatrix} d_{i(AC)1} \\ \vdots \\ d_{i(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{i(AK)1} \\ \vdots \\ d_{i(AK)M} \end{pmatrix} \end{pmatrix}, \mathbf{\Sigma_{FULL}} \right) \qquad \textbf{Equation (A3)}$$

$$\Sigma_{\mathbf{FULL}} = \begin{pmatrix} \sigma^2_{(AB)1} & \cdots & \rho^{1M}_{(AB,AB)}\sigma_{(AB)1}\sigma_{(AB)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AB,AB)}\sigma_{(AB)1}\sigma_{(AB)M} & \cdots & \sigma^2_{(AB)M} \end{pmatrix} \begin{pmatrix} \rho^{11}_{(AB,AC)}\sigma_{(AB)1}\sigma_{(AC)1} & \cdots & \rho^{1M}_{(AB,AC)}\sigma_{(AB)1}\sigma_{(AC)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AB,AC)}\sigma_{(AB)1}\sigma_{(AC)M} & \cdots & \rho^{MM}_{(AB,AC)}\sigma_{(AB)M}\sigma_{(AC)M} \end{pmatrix} \cdots \begin{pmatrix} \rho^{11}_{(AB,AK)}\sigma_{(AB)1}\sigma_{(AK)1} & \cdots & \rho^{1M}_{(AB,AK)}\sigma_{(AB)1}\sigma_{(AK)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AB,AK)}\sigma_{(AB)1}\sigma_{(AK,AK)M} & \cdots & \rho^{MM}_{(AB,AK)}\sigma_{(AB)M}\sigma_{(AK)M} \end{pmatrix}$$

$$\begin{pmatrix} \sigma^2_{(AC)1} & \cdots & \rho^{1M}_{(AC,AC)}\sigma_{(AC)1}\sigma_{(AC)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AC,AC)}\sigma_{(AC)1}\sigma_{(AC)M} & \cdots & \sigma^2_{(AC)M} \end{pmatrix} \vdots \begin{pmatrix} \rho^{11}_{(AC,AK)}\sigma_{(AC)1}\sigma_{(AK)1} & \cdots & \rho^{1M}_{(AC,AK)}\sigma_{(AC)1}\sigma_{(AK)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AC,AK)}\sigma_{(AC)1}\sigma_{(AC)M} & \cdots & \rho^{MM}_{(AC,AK)}\sigma_{(AC)M}\sigma_{(AK)M} \end{pmatrix}$$

$$\ddots \qquad \vdots$$

$$\begin{pmatrix} \sigma^2_{(AK)1} & \cdots & \rho^{1M}_{(AK,AK)}\sigma_{(AK)1}\sigma_{(AK)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AK,AK)}\sigma_{(AK)1}\sigma_{(AK)M} & \cdots & \sigma^2_{(AK)M} \end{pmatrix}$$

Where $\delta_{i(Ak)m}$ and $d_{(Ak)m}$ are study-specific and mean effect of treatment k relative A (reference treatment) on outcome m in study i respectively and $\boldsymbol{\Sigma}_{\mathbf{FULL}}$ is the full $(K\text{-}1) \times (K\text{-}1)$ blocks of $M \times M$ within-treatment between-outcome covariance matrix. The parameters in $\boldsymbol{\Sigma}_{\mathbf{FULL}}$ have the following interpretation:

($\sigma^2_{(Ak)m}$) indicate the variance of the effect of treatment k ($k = B, C, \quad , K$) relative to A on outcome m across studies.

$\rho^{mn}_{(Ak,Ak)}$ indicate the correlation between $\delta_{i(Ak)m}$ and $\delta_{i(Ak)n}$ (i.e. the correlation between the effect of treatment k relative to A on outcome m and the effect of treatment k relative to A on outcome n ($m \neq n$)) specific to the Ak comparison.

$\rho^{mm}_{(Ah,Ak)}$ indicate the correlation between $\delta_{i(Ah)m}$ and $\delta_{i(Ak)m}$ (i.e. the correlation between the effect of treatment h relative to A on outcome m and the effect of treatment k relative to A ($h \neq k$) on outcome m because they share a common comparator A).

The diagonal block matrices in $\boldsymbol{\Sigma}_{\mathbf{FULL}}$ thus carry terms for the between-study variance ($\sigma^2_{(Ak)m}$) while the off-diagonal blocks carry terms for the between-study correlations. We make two assumptions to simplify and reduce the number of parameters in $\boldsymbol{\Sigma}_{\mathbf{FULL}}$. First, we assume homogenous

variances for intervention effects within outcomes [20]. This implies $\sigma^2_{(Ak)m} = \sigma^2_m$ and $\rho^{mm}_{(Ah,Ak)} = \frac{1}{2}$ as in the single outcome network meta-analysis case [20,34]. Second, we make the assumption of homogenous between-study correlations for the intervention effects from different outcomes. Under this assumption we can express $\rho^{mn}_{(Ah,Ah)}$ and $\rho^{mm}_{(Ak,Ak)}$ in terms of a common correlation parameter ρ^{mn} by noting that for any 3-treatment **(***A*, *h*, *k***)** configuration, the covariance between outcome *m* and *n* effects across studies can be expressed as a covariance between two sums under evidence consistency:

$$\begin{aligned}
&\left[\delta_{i(hk)m}, \delta_{i(hk)n}\right] = COV\left[\left(\delta_{i(Ak)m} - \delta_{i(Ah)m}\right), \left(\delta_{i(Ak)n} - \delta_{i(Ah)n}\right)\right] \\
&= COV\left[\delta_{i(Ak)m}, \delta_{i(Ak)n}\right] - COV\left[\delta_{i(Ak)m}, \delta_{i(Ah)n}\right] - COV\left[\delta_{i(Ah)m}, \delta_{i(Ak)n}\right] \\
&+ COV\left[\delta_{i(Ah)m}, \delta_{i(Ah)n}\right] \\
&= \left(\rho^{mn}_{(Ak,Ak)} + \rho^{mn}_{(Ah,Ah)} - 2\rho^{mn}_{(Ak,Ah)}\right)\sigma_m \sigma_n
\end{aligned}$$

Equation (A4)

The homogenous between-study correlation assumption implies $\rho^{mn}_{(Ah,Ah)} = \rho^{mn}_{(Ak,Ak)} = \rho^{mn}$ and $\rho^{mm}_{(Ah,Ak)} = \frac{1}{2}\rho^{mn}$ for the inequality $-1 \le \left(\rho^{mm}_{(Ak,Ak)} +^{mm}_{(Ah,Ah)} - 2\rho^{mm}_{(Ah,Ak)}\right) \le 1$ to hold. Substituting these expressions into equation (A1), we see that the between-study correlation terms equal ρ^{mn} in the diagonal block of matrices and $\frac{1}{2}\rho^{mn}$ in the off-diagonal block of matrices of in $\mathbf{\Sigma_{FULL}}$ leading to the following simplification of the between-study covariance matrix:

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(AB)1} \\ \vdots \\ \delta_{i(AB)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(AC)1} \\ \vdots \\ \delta_{i(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(AK)1} \\ \vdots \\ \delta_{i(AK)M} \end{pmatrix} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \begin{pmatrix} d_{(AB)1} \\ \vdots \\ d_{(AB)M} \end{pmatrix} \\ \begin{pmatrix} d_{(AC)1} \\ \vdots \\ d_{(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{(AK)1} \\ \vdots \\ d_{(AK)M} \end{pmatrix} \end{pmatrix}, \mathbf{\Sigma}_{(Mp\times Mp)} = \begin{pmatrix} \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \cdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \vdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & & \ddots & \vdots \\ & & & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \end{pmatrix} \right)$$

Equation (A5)

By relabeling the reference treatment *A* as *b*, ($\delta_{i(AB)1}, \quad , \delta_{i(AK)m}$) as $\left(\delta_{i(AB)1}, \cdots, \delta_{i(AK)m}\right)$ and ($d_{(AB)1}, \quad , d_{(AK)M}$) as $\left(\delta_{i(bk_1)1}, \cdots, \delta_{i(bk_j)M}\right)$, equation (A5) can be rewritten as equation (A1) which is the multivariate form of equation (7.3) of Chapter 7. Furthermore, we can use the law of total variance to decompose equation (A5) into univariate marginal and univariate series of conditional distributions as shown in equation (A2). To do so, it is helpful to rewrite equation (A5) using matrix algebra.

Let $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk_1})} = \left(\delta_{i(bk_1)}1 \quad \cdots \quad \delta_{i(bk_1)M}\right)^T$, $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk_2})} = \left(\delta_{i(bk_2)}1 \quad \cdots \quad \delta_{i(bk_2)M}\right)^T$, $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk_p})} = \left(\delta_{i(bk_p)}1 \quad \cdots \quad \delta_{i(bk_p)M}\right)^T$ and $\mathbf{d}_{(\mathbf{bk_1})}$, $\mathbf{d}_{(\mathbf{bk_1})}$ $\mathbf{d}_{(\mathbf{bk_1})}$ represent the corresponding vectors of mean treatment effects.

Similarly let $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ & \ddots & \vdots \\ & & \sigma_M^2 \end{pmatrix}$ so that equation (A5) can be written as

$$\begin{pmatrix} \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \\ \vdots \\ \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_p)} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \mathbf{d}_{(\mathbf{bk}_1)} \\ \vdots \\ \mathbf{d}_{(\mathbf{bk}_p)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \cdots & \frac{1}{2}\boldsymbol{\Sigma} \\ & \ddots & \vdots \\ & & \boldsymbol{\Sigma} \end{pmatrix} \right)$$

Equation (A6)

For *j = 1*, we have $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \sim \text{Normal}\left(\mathbf{d}_{(\mathbf{bk}_1)}, \boldsymbol{\Sigma}\right)$.

For *j=2*, we have $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_2)} \mid \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \sim \text{Normal}\left(\mathbf{d}_{(\mathbf{bk}_2)} + \left(\frac{1}{2}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} - \mathbf{d}_{(\mathbf{bk}_1)}\right), \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} - \left(\frac{1}{2}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\left(\frac{1}{2}\boldsymbol{\Sigma}\right)^T \right)$

which simplifies to $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_2)} \mid \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \sim \text{Normal}\left(\mathbf{d}_{(\mathbf{bk}_2)} + \frac{1}{2}\left(\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} - \mathbf{d}_{(\mathbf{bk}_1)}\right), \boldsymbol{\Sigma}_2 = \left(\boldsymbol{\Sigma} - \frac{1}{4}\boldsymbol{\Sigma}\right) = \frac{3}{4}\boldsymbol{\Sigma} \right)$ where $\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma}$ because the covariance matrix is symmetric.

For *j=3*, we have $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_3)} \mid \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \\ \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_2)} \end{pmatrix} \sim \text{Normal}\left(\mathbf{d}_{(\mathbf{bk}_3)} + \left(\frac{1}{4}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}_2^{-1}\sum_{t=1}^{j-1}\left(\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_t)} - \mathbf{d}_{(\mathbf{bk}_t)}\right), \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_2 - \left(\frac{1}{4}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}_2^{-1}\left(\frac{1}{4}\boldsymbol{\Sigma}\right)^T \right)$ which simplifies to

$$\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_3)} \mid \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_2)}, \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \sim \text{Normal}\left(\mathbf{d}_{(\mathbf{bk}_3)} + \frac{1}{3}\sum_{t=1}^{j-1}\left(\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_t)} - \mathbf{d}_{(\mathbf{bk}_t)}\right), \boldsymbol{\Sigma}_3 = \left(\boldsymbol{\Sigma}_2 - \frac{1}{12}\boldsymbol{\Sigma}\right) = \frac{4}{6}\boldsymbol{\Sigma} \right)$$

In general, for $j=p$, we have $\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_p)} \mid \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_1)} \\ \vdots \\ \boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_{p-1})} \end{pmatrix} \sim \text{Normal}\left(\mathbf{d}_{(\mathbf{bk}_\mathbf{p})} + \frac{1}{j}\sum_{t=1}^{j-1}\left(\boldsymbol{\delta}_{\mathbf{i}(\mathbf{bk}_t)} - \mathbf{d}_{(\mathbf{bk}_t)}\right), \boldsymbol{\Sigma}_p = \frac{(j+1)}{2j}\boldsymbol{\Sigma} \right)$ which is equation (A2).

Appendix III: WinBUGS Code for baseline risk adjustment models in Chapter 6

```
#Model
model{
for(i in 1:ns){
tmp1[i] <- id[i]
w[i,1] <-0
delta[i,1]<-0
bl[i,1]<-0                    #No covariate adjustment the non-intervention / control arm

#Baseline models
mu[i] ~ dnorm(0,.0001)                  #model 1: vague priors for trial baselines
#mu[i] ~ dnorm(mu.mean,taumu)           #model 2:normal distr. for trial baselines
#mu[i] ~dnorm(lambda[T[i]],taumu)       #model 3:Two normal dist.for trial baselines
#T[i] ~ dcat(P[])                       #Categorical variable for mixture

for(k in 1:na[i]) {
   #Model for binary outcome (example 1)
   #r[i,k] ~ dbin(p[i,k],n[i,k])                          #binomial likelihood
   #logit(p[i,k])<-mu[i] + delta[i,k]                     #model
   #rhat[i,k] <- p[i,k] * n[i,k]                          #expected value of the numerators
   #dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance
   #+ (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))

   #Model for continuous outcome (example 2)
   var[i,k] <- pow(se[i,k],2)                            #calculate variances
   prec[i,k] <- 1/var[i,k]                               #set precisions
   y[i,k] ~ dnorm(theta[i,k],prec[i,k])                  #Normal likelihood
   theta[i,k]<-mu[i] + delta[i,k]
   dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]   #Deviance contribution

  }

  sumdev[i]<-sum(dev[i, 1:na[i]])                        #summed residual deviance contribution for this trial

  for(k in 2:na[i]){
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])              #trial-specific LOR distributions
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]+ bl[i,k]   #mean of LOR distributions
      taud[i,k] <- tau *2*(k-1)/k                        #precision of LOR distributions

     w[i,k] <- (delta[i,k]-((d[t[i,k]]-d[t[i,1]])+bl[i,k]))   #adjustment,multi-arm RCTs
     sw[i,k] <-sum(w[i,1:k-1])/(k-1)                     #cumulative adjust,multi-arm trials
     bl[i,k] <-(beta[t[i,k]]-beta[t[i,1]])*(mu[i]-m.mu)  #baseline risk adjustment
   }
}

     ssumdev<-sum(sumdev[])#total residual deviance
     d[1]<-0
     beta[1]<-0

for(k in 2:nt){
    d[k] ~ dnorm(0,.0001)                  #vague prior for basic parameters
    beta[k] <- B                           #common covariate effect (Model A)
    #beta[k] ~ dnorm(B,tauB)               #exchangeable covariate effect (Model B)
    #beta[k] ~ dnorm(0,0.0001)             #independent covariate effect (Model C)
 }

    #mu.mean~ dnorm(0,0.0001)              #vague prior for mean baseline log-odds
    sd~dunif(0,30)                         #vague prior for between-study standard deviation
    tau<-1/pow(sd,2)                       #1/between-study variance (treatment effects)
    B~ dnorm(0,0.0001)                     #vague prior for common covariate effect
    #sdmu~dunif(0,30)                      #vague prior for random effects standard deviation
    #taumu<-1/pow(sdmu,2)                  #1/between-study variance (baseline log-odds)
    #sdB~dunif(0,30)                       #vague prior for random effects standard deviation
    #tauB<-1/pow(sdB,2)                    #1/between-study variance
}#END
```

Data file 1

list(nt=7,ns=13, m.mu=-1.32)

Data File 2

id[]	t[,1]	t[,2]	t[,3]	r[,1]	n[,1]	r[,2]	n[,2]	r[,3]	n[,3]	na[]
9007	1	2	NA	178	271	188	249	NA	1	2
26	1	2	NA	83	89	79	85	NA	1	2
9019	1	2	NA	72	74	140	144	NA	1	2
48	1	2	NA	54.5	55	55.5	56	NA	2	2
344	1	2	NA	4	57	6	60	NA	1	2
4	1	3	NA	68	82	79	83	NA	1	2
9002	1	3	4	147	149	171	173	160	163	3
9023	1	4	NA	70.26	79.58	74.07	80	NA	1	2
345	1	5	NA	683	738	712	762	NA	1	2
35	1	6	NA	88.42	248.37	128.16	248.37	NA	1	2
12	2	3	NA	14	47	19	49	NA	1	2
29	2	5	NA	44	49	43	45	NA	1	2
14	3	7	NA	22	101	20	104	NA	1	2
END										

#Initial values
list(d = c(NA,0,0,0,0,0,0), sd=1)

Appendix IV: WinBUGS Code for the multivariate NMAs in Chapter 7

```
Model {     #model 2b of Chapter 7

  # i = data point  (one for each arm of each study),
  # arm = study arm
  # s = study
  # m = outcome

  #Likelihood for arm level data
  #=======================
  for(i in 1:N1){
     tmp1[i] <- studyid[i]                                          # study  id not used in the model
     y[i,1:3] ~ dmnorm(mean.y[study[i],arm[i],1:3],omega[i,,])      # multivariate likelihood
     omega[i,1:3,1:3] <- inverse(cov.mat[i,,])                       # within-study precision matrix

     #define elements of within-study covariance matrix
     cov.mat[i,1,1] <-  pow(se[i,1],2)
     cov.mat[i,2,2] <-  pow(se[i,2],2)
     cov.mat[i,3,3] <-  pow(se[i,3],2)
     cov.mat[i,1,2] <-  se[i,1]*se[i,2]*cor[i,1]
     cov.mat[i,1,3] <-  se[i,1]*se[i,3]*cor[i,2]
     cov.mat[i,2,3] <-  se[i,2]*se[i,3]*cor[i,3]
     cov.mat[i,2,1] <-  cov.mat[i,1,2]
     cov.mat[i,3,1] <-  cov.mat[i,1,3]
     cov.mat[i,3,2]  <- cov.mat[i,2,3]

     for(m in 1:no){
         se[i,m] ~ dnorm(0, prec.se[m])I(0,)                            # input missing standard errors
         unif.a[i,m] <- mn.rhoW[m] - (sqrt(12)*se.rhoW[m]/2)        # parameter a of uniform distribution
         unif.b[i,m] <- mn.rhoW[m] +(sqrt(12)*se.rhoW[m]/2)         # parameter b of uniform distribution
         cor[i,m]   ~  dunif(unif.a[i,m], unif.b[i,m])              # within-study correlation model
     }
 }

  for(j in 1:ns){
     for(k in 1:NA[j]) {
        for(m in 1:no){
            mean.y[j,k,m] <- mu[j,m] + delta[j,k,m]                 # define study-specific treatment effects
        }
    }
 }

 #Random effects between-study model
 #=================================
 for(j in 1:ns) {
    for(m in 1:no) {
      delta[j,1,m] <-0            # delta in control arm to zero for all outcomes
      w[j,1,m]    <-0             # multi-arm adjustment in control group set to zero
    }

 for(k in 2:NA[j]){
   delta[j,k,1:no] ~ dmnorm(md[j,k,1:no],precBK[j,k,1:no,1:no])    #random effects model

     for(m in 1:no){
       for(mm in 1:no) {
            precBK[j,k,m,mm] <- prec[m,mm]*2*(k-1)/k              # between-study precision matrix
        }
      }
   }
 }

 #Consistency relations between basic parameters
 #=====================================
 for(i in 1:N2) {
   tmp2[i] <- studyid1[i]                        # temp variable to identify study id, not used
    for(k in 2:na[i]) {
        md[s[i],k,out[i]] <-  (d[out[i],t[i,k]] - d[out[i],t[i,1]])*equals(o[i],out[i])  + sw[s[i],k,out[i]]
        w[s[i],k,out[i]] <- (delta[s[i],k,out[i]] - (d[out[i],t[i,k]] - d[out[i],t[i,1]]))*equals(o[i],out[i])
```

```
            sw[s[i],k,out[i]] <- sum(w[s[i],1:k-1,out[i]])/(k-1)
        }
    }

    #Constraints
    # Effect in usual care arm is set to zero
# There are 9 interventions in total, but only 7 are trialled for each outcome, hence interventions 8 and 9 refer to the interventions were
        outcome information is not available.
    d[1,1] <- 0
    d[2,1] <- 0
    d[3,1] <- 0
    d[1,8] <- 0
    d[1,9] <- 0
    d[2,8] <- 0
    d[2,9] <- 0
    d[3,8] <- 0
    d[3,9] <- 0

    #Prior distributions and parameter to estimate
    prec[1:no,1:no] <- inverse(sigma[,])        #hash out if using inverse-wishart (model 2a)
    sd.se~ dunif(0, 2)

    for(m in 1:no) {
      prec.se[m] <- pow(sd.se,-2)
      sigma[m,m] <- pow(sd[m],2)      # hash out if using inverse-wishart (model 2a)
      sd[m] ~ dunif(0, 2)             # hash out if using inverse-wishart (model 2a)

      for(j in 1:ns){
        mu[j, m] ~ dnorm(0,0.001)
      }

      for(k in 2: nt.total[m]){
        or[m,k] <- exp(d[m,k])
          d[m,k]  ~ dnorm(0,0.001)
        }
    }

    #spherical parameterization (Wei and Higgins 2013)
    #hash out if using inverse-wishart (model 2a)
    pi <- 3.1415
    for(i in 1:2) {
        for(j in (i+1):no) {
            sigma[i,j] <- rho[i,j]*sd[i]*sd[j]
            sigma[j,i] <- sigma[i,j]
        g[j,i] <- 0
            a[i,j] ~ dunif(0, pi)
            rho[i,j] <- inprod(g[,i], g[,j])
      }
    }

    g[1,1] <- 1
    g[1,2] <- cos(a[1,2])
    g[2,2] <- sin(a[1,2])
    g[1,3] <- cos(a[1,3])
    g[2,3] <- sin(a[1,3])*cos(a[2,3])
    g[3,3] <- sin(a[1,3])*sin(a[2,3])

    #Inverse-Wishart prior (model 2a) hash
    #prec[1:no, 1:no] ~ dwish(R[1:no,1:no],no)
    #sigma[1:no,1:no] <- inverse(prec[,])

    #between-study standard deviations
    #sd[1] <- sqrt(sigma[1,1])
    #sd[2] <- sqrt(sigma[2,2])
    #sd[3] <- sqrt(sigma[3,3])

    #between-study correlations
    #rho[1,2] <- sigma[1,2]/(sd[1]*sd[2])
    #rho[1,3] <- sigma[1,3]/(sd[1]*sd[3])
    #rho[2,3] <- sigma[2,3]/(sd[2]*sd[3])
}
```

Model 2 of Chapter 7: Data file 1 of 3

```
list(
N1=45,                                  #no of datapoints
N2=66,                                  # no of studies x no of outcomes (22x3=66)
ns=22,                                  # no of studies
no=3,                                   #no of outcomes
nt.total =c(7,7,7),                     # no of interventions for outcomes 1, 2 and 3
mn.rhoW =c(0.184,-0.052,0.051),         #mean of within-study correlations from IPD
se.rhoW = c(0.118,0.064,0.059,1),       #se of within-study correlations from IPD
NA = c(2,2,2,2,2,  2,2,2,2,3,  2,2,2,2,2,  2,2,2,2,2,   2,2), # no. of arms in each study
#R = structure(.Data = c(1,0,0,  0,1,0, 0,0,1),.Dim = c(3,3)) # needed for model 2a
))
```

Model 2 of Chapter 7: Data file 2 of 3

studyid[]	study[]	arm[]	y[,1]	y[,2]	y[,3]	se[,1]	se[,2]	se[,3]
9007	1	1	0.649184	0.893818	NA	0.127948	0.279791	NA
9007	1	2	1.125568	1.270463	NA	0.147352	0.28292	NA
26	2	1	2.627081	0.996333	0.67634	0.422747	0.238854	0.224238
26	2	2	2.577688	1.245216	1.052092	0.423468	0.260352	0.247644
9019	3	1	3.583519	1.43848	1.714798	0.71686	0.297284	0.362093
9019	3	2	3.555348	1.466337	1.132514	0.507093	0.213504	0.213527
48	4	1	4.691348	1.363305	NA	1.420686	0.337883	NA
48	4	2	4.70953	2.100061	NA	1.42057	0.432522	NA
344	5	1	-2.584	-2.89037	NA	0.518525	0.593171	NA
344	5	2	-2.19722	-2.21557	NA	0.430332	0.42994	NA
203	6	1	NA	NA	-0.69315	NA	NA	0.181449
203	6	2	NA	NA	1.520952	NA	NA	0.222198
4	7	1	1.58045	0.395313	NA	0.293487	0.225192	NA
4	7	2	2.983154	0.899484	NA	0.512502	0.242107	NA
41	8	1	NA	NA	-1.3689	NA	NA	0.207978
41	8	2	NA	NA	-0.42652	NA	NA	0.187525
42	9	1	NA	-0.41651	-0.44425	NA	0.1663	0.166789
42	9	2	NA	0.377762	1.265666	NA	0.166221	0.197104
9002	10	1	4.297286	NA	NA	0.711901	NA	NA
9002	10	2	4.448516	NA	NA	0.71123	NA	NA
9002	10	3	3.976562	NA	NA	0.582738	NA	NA
279	11	1	NA	-0.61904	-1.38629	NA	0.331497	0.395285
279	11	2	NA	2.140066	2.140066	NA	0.528594	0.528594
9023	12	1	2.02004	1.440219	NA	0.34861	0.333812	NA
9023	12	2	2.524986	1.927793	NA	0.426773	0.393438	NA
49	13	1	NA	1.846879	NA	NA	0.152166	NA
49	13	2	NA	2.060979	NA	NA	0.165819	NA
345	14	1	2.519162	-0.04485	NA	0.140164	0.077344	NA
345	14	2	2.656055	0.124258	NA	0.146303	0.07612	NA
28	15	1	NA	NA	0.716309	NA	NA	0.212838
28	15	2	NA	NA	0.977271	NA	NA	0.22683
35	16	1	-0.59276	NA	NA	0.13252	NA	NA
35	16	2	0.064039	NA	NA	0.126971	NA	NA
9042	17	1	NA	NA	-2.03143	NA	NA	0.265889
9042	17	2	NA	NA	0.043172	NA	NA	0.169677
10001	18	1	NA	NA	1.921813	NA	NA	0.309077

10001	18	2	NA	NA	3.100092	NA	NA	0.323272
12	19	1	-0.85745	-0.12783	0.300105	0.318954	0.292326	0.29502
12	19	2	-0.45676	0.81831	0.916291	0.293198	0.309965	0.316228
29	20	1	2.174752	1.268511	NA	0.47194	0.377308	NA
29	20	2	3.068053	1.609438	NA	0.723364	0.387298	NA
24	21	1	NA	0.226982	NA	NA	0.092947	NA
24	21	2	NA	0.267138	NA	NA	0.091911	NA
14	22	1	-1.27841	-4.20469	NA	0.241066	0.822567	NA
14	22	2	-1.43508	-5.34233	NA	0.248807	1.417593	NA
END								

Model 2 of Chapter 7: Data file 3 of 3

studyid1[]	s[]	t[,1]	t[,2]	t[,3]	o[]	out[]	na[]
9007	1	1	2	NA	1	1	2
9007	1	1	2	NA	2	2	2
9007	1	1	2	NA	0	3	2
26	2	1	2	NA	1	1	2
26	2	1	2	NA	2	2	2
26	2	1	2	NA	3	3	2
9019	3	1	2	NA	1	1	2
9019	3	1	2	NA	2	2	2
9019	3	1	2	NA	3	3	2
48	4	1	2	NA	1	1	2
48	4	1	2	NA	2	2	2
48	4	1	2	NA	0	3	2
344	5	1	2	NA	1	1	2
344	5	1	2	NA	2	2	2
344	5	1	2	NA	0	3	2
203	6	1	2	NA	0	1	2
203	6	1	2	NA	0	2	2
203	6	1	2	NA	3	3	2
4	7	1	3	NA	1	1	2
4	7	1	3	NA	2	2	2
4	7	1	3	NA	0	3	2
41	8	1	3	NA	0	1	2
41	8	1	3	NA	0	2	2
41	8	1	3	NA	3	3	2
42	9	1	3	NA	0	1	2
42	9	1	3	NA	2	2	2
42	9	1	3	NA	3	3	2
9002	10	1	3	4	1	1	3
9002	10	1	3	4	0	2	3
9002	10	1	3	4	0	3	3
279	11	1	4	NA	0	1	2
279	11	1	4	NA	2	2	2
279	11	1	4	NA	3	3	2

9023	12	1	4	NA	1	1	2
9023	12	1	4	NA	2	2	2
9023	12	1	4	NA	0	3	2
49	13	1	4	NA	0	1	2
49	13	1	4	NA	2	2	2
49	13	1	4	NA	0	3	2
345	14	1	5	NA	1	1	2
345	14	1	5	NA	2	2	2
345	14	1	**8**	NA	0	3	2
28	15	1	**8**	NA	0	1	2
28	15	1	**8**	NA	0	2	2
28	15	1	5	NA	3	3	2
35	16	1	6	NA	1	1	2
35	16	1	6	NA	0	2	2
35	16	1	6	NA	0	3	2
9042	17	1	6	NA	0	1	2
9042	17	1	6	NA	0	2	2
9042	17	1	6	NA	3	3	2
10001	18	1	**9**	NA	0	1	2
10001	18	1	**9**	NA	0	2	2
10001	18	1	7	NA	3	3	2
12	19	2	3	NA	1	1	2
12	19	2	3	NA	2	2	2
12	19	2	3	NA	3	3	2
29	20	2	5	NA	1	1	2
29	20	2	5	NA	2	2	2
29	20	2	**8**	NA	0	3	2
24	21	4	6	NA	0	1	2
24	21	4	6	NA	2	2	2
24	21	4	6	NA	0	3	2
14	22	3	7	NA	1	1	2
14	22	3	7	NA	2	2	2
14	22	3	**9**	NA	0	3	2
END							

WinBUGS code for model 3 of Chapter 7

```
#Model 3 code

    Model {    #model 3

  #Likelihood for arm level data
  #=======================
  for(i in 1:N1){
     tmp1[i] <- studyid[i]                          # study  id not used in the model
     y[i,1:3] ~ dmnorm(mean.y[study[i],arm[i],1:3],omega[i,,])   # multivariate likelihood
     omega[i,1:3,1:3] <- inverse(cov.mat[i,,])                   # within-study precision matrix

     #define elements of within-study covariance matrix
     cov.mat[i,1,1] <-  pow(se[i,1],2)
```

```
      cov.mat[i,2,2] <-  pow(se[i,2],2)
      cov.mat[i,3,3] <-  pow(se[i,3],2)
      cov.mat[i,1,2] <-  se[i,1]*se[i,2]*cor[i,1]
      cov.mat[i,1,3] <-  se[i,1]*se[i,3]*cor[i,2]
      cov.mat[i,2,3] <-  se[i,2]*se[i,3]*cor[i,3]
      cov.mat[i,2,1] <-  cov.mat[i,1,2]
      cov.mat[i,3,1] <-  cov.mat[i,1,3]
      cov.mat[i,3,2] <-  cov.mat[i,2,3]

      for(m in 1:no){
          se[i,m] ~ dnorm(0, prec.se[m])I(0,)                    # input missing standard errors
          unif.a[i,m] <- mn.rhoW[m] - (sqrt(12)*se.rhoW[m]/2)    # parameter a of uniform distribution
          unif.b[i,m] <- mn.rhoW[m] +(sqrt(12)*se.rhoW[m]/2)     # parameter b of uniform distribution
          cor[i,m]   ~  dunif(unif.a[i,m], unif.b[i,m])          # within-study correlation model
      }
  }

   for(j in 1:ns){
      for(k in 1:NA[j]) {
         for(m in 1:no){
            mean.y[j,k,m] <- mu[j,m] + delta[j,k,m]          # define study-specific treatment effects
         }
      }
  }

  #Random effects between-study model
  #================================

              for(j in 1:ns)            {

      tmp2[j] <- studyid1[j]
      tmp3[j] <- s[j]

                for(m in 1:no)         {
         delta[j,1,m] <-0                                          #delta's in control arm to zero for all outcomes
              w[j,1,m] <-0                                         #multi-arm adjustment in control group set to zero
       }

       for(k in 2:na2[j])              {
  delta[j,k,1:no] ~ dmnorm(md[j,k,1:no],precBK[j,k,1:no,1:no])      #trial specific trt effects drawn from mvn distribution
  for(m in 1:no){
    md[j,k,m] <-  (d[m,t[j,k]] - d[m,t[j,1]])+ sw[j,k,m]         #consistency equations

              w[j,k,m] <- delta[j,k,m] - (d[m,t[j,k]] - d[m,t[j,1]])                    #multi-arm adjustemnt for treatment  k
               sw[j,k,m] <- sum(w[j,1:k-1,m])/(k-1)

     for(mm in 1:no)    {
                          precBK[j,k,m,mm] <- prec[m,mm]*2*(k-1)/k
               }
   }
 }
}

  #Constraints
  #There are 8 trts in total, but only 7 treatments are trialled for each outcome, hence 8 is code
  # when trt has not been considered for the outcome. also effect in usual care arm is set to zero
  d[1,1] <- 0
  d[2,1] <- 0
  d[3,1] <- 0

  #Prior distributions and parameter to estimate
  prec[1:no,1:no] <- inverse(sigma[,])        #hash out if using inverse-wishart (model 2a)
  sd.se~ dunif(0, 2)

   for(m in 1:no) {
     prec.se[m] <- pow(sd.se,-2)
     sigma[m,m] <- pow(sd[m],2)       #hash out if using inverse-Wishart (model 2a)
     sd[m] ~ dunif(0, 2)              #hash out if using inverse-Wishart (model 2a)
               for(j in 1:ns){
        mu[j, m] ~ dnorm(0,0.0001)
     }
   }

  #spherical parameterization (Wei and Higgins 2013)
  pi <- 3.1415
  for(i in 1:2) {
```

```
      for(j in (i+1):no) {
          sigma[i,j] <- rho[i,j]*sd[i]*sd[j]
        sigma[j,i] <- sigma[i,j]
        g[j,i] <- 0
              a[i,j] ~ dunif(0, pi)
              rho[i,j] <- inprod(g[,i], g[,j])
    }
  }

  g[1,1] <- 1
  g[1,2] <- cos(a[1,2])
  g[2,2] <- sin(a[1,2])
  g[1,3] <- cos(a[1,3])
  g[2,3] <- sin(a[1,3])*cos(a[2,3])
  g[3,3] <- sin(a[1,3])*sin(a[2,3])

# Borrowing information across outcomes
#==========================================
#intervention effects exponentiated and prior distributions
        for(k in 2: nt){
        for(m in 1:no) {
                        meanD[m,k-1] <- alpha[k-1] + gamma[m]          #outcome and intervention effects
                        d[m,k] ~ dnorm(meanD[m,k-1], prec.btw)         #trt effects
                        OR[m,k-1] <- exp(meanD[m,k-1])      # extrapolated effects of interest in model 3
         or[m,k] <- exp(d[m,k])              #shrunken estimates based on equation (7)
      }
        }

  for(m in 1:no) {gamma[m] ~ dnorm(0, 0.0001) }
  for(k in 1:(nt-1)) {alpha[k] ~ dnorm(0, 0.0001) }
  prec.btw <- pow(sd.btw,-2)
  sd.btw ~ dunif(0, 2)
}
#END
```

Model 3: Data file 1 of 3

```
list(
N=45,  #no of datapoints
ns=22,  # no of studies
no=3,    #no of outcomes
nt =9,
mn.rhoW =c(0.184,-0.052,0.051),      #mean of within-study correlations from IPD
se.rhoW = c(0.118,0.064,0.059,1),    #se of within-study correlations from IPD
NA = c(2,2,2,2,2,  2,2,2,2,3,  2,2,2,2,2,  2,2,2,2,2,   2,2),  # no. of arms in each study

)
```

Model 3: Data file 2 of 3 is the same as in model 2 of Chapter 7

Model 3: Data file 3 of 3

studyid1[]	s[]	t[,1]	t[,2]	t[,3]	na2[]
9007	1	1	2	NA	2
26	2	1	2	NA	2
9019	3	1	2	NA	2
48	4	1	2	NA	2
344	5	1	2	NA	2
203	6	1	2	NA	2
4	7	1	3	NA	2
41	8	1	3	NA	2
42	9	1	3	NA	2
9002	10	1	3	4	3
279	11	1	4	NA	2
9023	12	1	4	NA	2
49	13	1	4	NA	2

345	14	1	5	NA	2
28	15	1	6	NA	2
35	16	1	7	NA	2
9042	17	1	7	NA	2
10001	18	1	8	NA	2
12	19	2	3	NA	2
29	20	2	5	NA	2
24	21	4	7	NA	2
14	22	3	9	NA	2
END					

Appendix V: WinBUGS Code to fit decision analytic model

```
#-------------  DECISION MODEL FOR POISONING PREVENTION VERSION 1 ------------
#-------------  FELIX ACHANA (MARCH 2014) ------------------------------------
#-----------------------------------------------------------------------------

# NHS Perspective
# K= Intervention strategy
# S=Health states (1=Safe storage(SS),2=NO SS,3=SS/disability,4=NO SS/disability
# 5=death from fatal  fatal injury and 6=Death other causes

# N=Number of households
# C=Cycle
# T=Total number of years (time horizon)

# INTERVENTIONS
# 1Usual care
# 2Education
# 3Education + low/free equipment
# 4Education + low/free equipment + Home safety inspection
# 5Education + low/free equipment + Fitting
# 6Education + low/free equipment +Home safety inspection + Fitting
# 7Education + Home visit
# 8Equipment only


#---------------- NMA MODEL ------------------------------------------------
#-----------------------------------------------------------------------------

model{
  for(i in 1:NS) {
     tmp[i] <- id[i]
     w[i,1] <- 0
     delta[i,1] <- 0
     mu[i] ~ dnorm(0,.0001)

     #binomial likelihood model
     for(k in 1:na[i]){
        r[i,k] ~ dbin(pMTC[i,k],n[i,k])
        logit(pMTC[i,k]) <- mu[i] + delta[i,k]
     }

     #random effects model with multi-arm adjustment
     for(k in 2:na[i]){
        delta[i,k] ~  dnorm(md[i,k],taud[i,k])
        md[i,k]    <- d[t[i,k]]-d[t[i,1]]+sw[i,k]
        taud[i,k]  <- tau *2*(k-1)/k
        w[i,k]     <- (delta[i,k]-d[t[i,k]]+d[t[i,1]])
        sw[i,k]    <- sum(w[i,1:k-1])/(k-1)
    }
  }

  d[1]<-0                        #trt effect in placebo group set to zero
  sd~dunif(0,2)              #vague prior for re sd
  tau<-pow(sd,-2)

  for(k in 2:NT){
    d[k] ~ dnorm(0, 0.0001)  #vague priors for basic parameters
  }


  for(i in 1:NS){
    mu1[i] <- mu[i]*equals(t[i,1],1)  #Trt A baseline.
  }

  for(k in 1:NT){
    logit(pMTCfunc[k]) <- sum(mu1[])/NBS+d[k]
  }

  #Ranking
   for(k in 1:NT) {
```

```
   rk[k]<-NT+1 - rank(pMTCfunc[k],k)
   best[k]<-equals(rk[k],1)
 }

#Pairwise ORs
 for(c in 1:(NT-1)){
   for(k in (c+1):NT){
      lor[c,k] <- d[k] - d[c]
      log(or[c,k]) <- lor[c,k]
     # RR[c,k]<-or[c,k]/(1-mn.mu1+mn.mu1*or[c,k])
   }
 }




#---- DECISION MODEL ----------------------------------------------------------------------------------
#-------------------------------------------------------------------------------------------------------


#DATA ON PROBABILITIES
 pSafe1     ~  dbeta(pSafe1.a, pSafe1.b)        #Estimate 1:KCS community controls
 pSafe2     <- exp(mnSafe2)/(1.0+exp(mnSafe2))  #Estimate 2:Patel 2008 (see meta-analysis below)
 piSafe     <- pSafe1                           #set piSafe to estimate 1
#orIngest   ~  dlnorm(muLor, precLor)            #relative risk of accidental ingestion|SS
 orIngest   <- exp(lnOR)                         #relative risk of accidental ingestion|SS
 lnOR       ~  dnorm(mu.lnOR,prec.lnOR)          #log relative risk of accidental ingestion|SS
 pIngest[2] ~  dbeta(pIngest.a, pIngest.b)      #prob(accidental ingestion|NO SS)
 pIngest[1] <- pIngest[2]*orIngest
 pIngest[3] <- pIngest[2]*orIngest
 pIngest[4] <- pIngest[2]

 pAdmit     ~  dbeta(pAdmit.a, pAdmit.b)        #prob(admission following poisoning incident)
 pLong  ~  dbeta(pLong.a, pLong.b)              #prob(long stay, no complete recovery serious poisoning incident)
 pFatal     ~  dbeta(pFatal.a, pFatal.b)        #prob(admission ff poisoning incident)


#bSafe1: MA of baseline safety practices reported in Patel 2008
 for(i in 1:7){
    rSafe2[i] ~ dbin(pi.Safe2[i], nSafe2[i])
    logit(pi.Safe2[i]) <- delta.Safe2[i]
    delta.Safe2[i] ~ dnorm(mnSafe2, tauSafe2)
 }

#Prior distributions
 mnSafe2 ~ dnorm(0,0.001)
 sdSafe2 ~ dunif(0,2)
 tauSafe2 <- pow(sdSafe2,-2)


#----- PART 1: INTERVENTION MODEL (TIME t=1) ------------------------------------------------------------
#-------------------------------------------------------------------------------------------------------
 n1[1]<-N*piSafe                                #N households with SS at baseline
 n2[1]<-0
 n3[1]<-0
 n4[1]<-N*(1-piSafe)                            #N households with no SS at baseline

 for( k in 2:K){
   n1[k]<-N*piSafe                              #SS prior to intervention
   n2[k]<-N*(1-piSafe)*pAccept[k]*pMTCfunc[k]      #SS after intervention
   n3[k]<-N*(1-piSafe)*pAccept[k]*(1-pMTCfunc[k])  #NO SS after intervention
   n4[k]<-N*(1-piSafe)*(1-pAccept[k])              #NO SS, refuse intervention
 }

 for(k in 1:K){
   pi[1,k,1]<-n1[k]+n2[k]                       #state 1=SSM
   pi[1,k,2]<-n3[k]+n4[k]                       #state 2=NO SSM
   pi[1,k,3]<-0                                 #state 3=SSM:Chronic injury
   pi[1,k,4]<-0                                 #state 4=NO SSM:Chonic injury
   pi[1,k,5]<-0                                 #state 5=Death fatal injury
   pi[1,k,6]<-0                                 #state 6=Death other causes
   #sums to N
   CHECK[1,k]<-pi[1,k,1]+pi[1,k,2]+pi[1,k,3]+pi[1,k,4]+pi[1,k,5]+pi[1,k,6]
 }
```

```
#Account for households no longer having safe storage after 12 months
 for(k in 1:K){
   pSafe[k,1]<-  piSafe*(1-decay[k])
   pSafe[k,3]<-  pSafe[k,1]
   pSafe[k,2]<-  0
   pSafe[k,4]<-  0
   decay[k]  <-  0        #assumes households will continue to have safe storage
 }


#----- PART 2: PRE-SCHOOL MODEL(cYCLES 2 TO 5) ----------------------------------------------------
#-----------------------------------------------------------------------------------------------------
#Define pathways through the model
 for(c in 2:C){                    #cycles
   for(k in 1:K){                  #interventions
     for(s in 1:4){                #health states
       #No unintentional ingestion
       o1[c,k,s]<- (1-pIngest[s])*pDead[c]
        o2[c,k,s]<- (1-pIngest[s])*(1-pDead[c])*pSafe[k,s]
        o3[c,k,s]<- (1-pIngest[s])*(1-pDead[c])*(1-pSafe[k,s])

       #Ingestion but not admitted
       o4[c,k,s]<- pIngest[s]*(1-pAdmit)*pDead[c]
        o5[c,k,s]<- pIngest[s]*(1-pAdmit)*(1-pDead[c])*pSafe[k,s]
        o6[c,k,s]<- pIngest[s]*(1-pAdmit)*(1-pDead[c])*(1-pSafe[k,s])

       #Ingested, admitted short-inpatient stay (assumed for cases leading to complete recovery)
       o7[c,k,s]<-  pIngest[s]*pAdmit*(1-pLong)*pDead[c]
       o8[c,k,s]<-  pIngest[s]*pAdmit*(1-pLong)*(1-pDead[c])*pSafe[k,s]
       o9[c,k,s]<-  pIngest[s]*pAdmit*(1-pLong)*(1-pDead[c])*(1-pSafe[k,s])

       #Ingested, admitted long-inpatient stay (assumed for cases leading to fatal or chronic injury)
       o10[c,k,s]<- pIngest[s]*pAdmit*pLong*pFatal
       o11[c,k,s]<- pIngest[s]*pAdmit*pLong*(1-pFatal)*pDead[c]
       o12[c,k,s]<- pIngest[s]*pAdmit*pLong*(1-pFatal)*(1-pDead[c])*pSafe[k,s]
       o13[c,k,s]<- pIngest[s]*pAdmit*pLong*(1-pFatal)*(1-pDead[c])*(1-pSafe[k,s])

       #Check sums to 1
       TOT[c,k,s]<-o1[c,k,s]+o2[c,k,s]+o3[c,k,s]+o4[c,k,s]+o5[c,k,s]+o6[c,k,s]+o7[c,k,s]+
       o8[c,k,s]+o9[c,k,s]+o10[c,k,s]+o11[c,k,s]+o12[c,k,s]+o13[c,k,s]
     }
   }
 }

#Estimate transition probabilities between health states
 for(c in 2:C){
   for(k in 1:K){
     for(s in 1:2){
       lamb[c,k,s,1]<-o2[c,k,s]+o5[c,k,s]+o8[c,k,s]          #From state s to state 1
       lamb[c,k,s,2]<-o3[c,k,s]+o6[c,k,s]+o9[c,k,s]          #From state s to state 2
       lamb[c,k,s,3]<-o12[c,k,s]                                          #From state s to state 3
        lamb[c,k,s,4]<-o13[c,k,s]                                         #From state s to state 4
        lamb[c,k,s,5]<-o10[c,k,s]                                         #From state s to state 5
        lamb[c,k,s,6]<-o1[c,k,s]+o4[c,k,s]+o7[c,k,s]+o11[c,k,s] #From state s to state 6
     }

     for(s in 3:4){
        lamb[c,k,s,1]<-0                              #From state s to state 1
        lamb[c,k,s,2]<-0                              #From state s to state 2
       lamb[c,k,s,3]<-o2[c,k,s]+o5[c,k,s]+o8[c,k,s]+o12[c,k,s]  #From state s to state 3
        lamb[c,k,s,4]<-o3[c,k,s]+o6[c,k,s]+o9[c,k,s]+o13[c,k,s]  #From state s to state 4
        lamb[c,k,s,5]<-o10[c,k,s]                                          #From state s to state 5
        lamb[c,k,s,6]<-o1[c,k,s]+o4[c,k,s]+o7[c,k,s]+o11[c,k,s]  #From state s to state 6
     }

     for(s in 5:S){
        lamb[c,k,s,1]<- 0            #From state s to state 1
       lamb[c,k,s,2]<- 0            #From state s to state 2
        lamb[c,k,s,3]<- 0            #From state s to state 3
        lamb[c,k,s,4]<- 0            #From state s to state 4
        lamb[c,k,s,5]<- equals(s,5)    #From state s to state 5
        lamb[c,k,s,6]<- equals(s,6)    #From state s to 6
     }

     #Checks to ensure each row sums to 1
```

```
      for(s in 1:S){
        TOTAL[c,k,s]<-lamb[c,k,s,1]+lamb[c,k,s,2]+lamb[c,k,s,3]+lamb[c,k,s,4]+lamb[c,k,s,5]+lamb[c,k,s,6]
      }
    }
  }


#----------- DM PART 3: CYCLES 6 TO 100 ----------------------------------------------------------------------
#-----------------------------------------------------------------------------------------------------------

 for(c in C+1:T){
    for(k in 1:K){
      for(s in 1:4){
        lamb[c,k,s,1]<-(1-pDead[c])*equals(s,1)
         lamb[c,k,s,2]<-(1-pDead[c])*equals(s,2)
         lamb[c,k,s,3]<-(1-pDead[c])*equals(s,3)
         lamb[c,k,s,4]<-(1-pDead[c])*equals(s,4)
        lamb[c,k,s,5]<- 0
         lamb[c,k,s,6]<- pDead[c]
      }

      for(s in 5:S){
         lamb[c,k,s,1]<- 0
         lamb[c,k,s,2]<- 0
         lamb[c,k,s,3]<- 0
         lamb[c,k,s,4]<- 0
         lamb[c,k,s,5]<- equals(s,5)
         lamb[c,k,s,6]<- equals(s,6)
      }

      for(s in 1:S){
        TOTAL[c,k,s]<-lamb[c,k,s,1]+lamb[c,k,s,2]+lamb[c,k,s,3]+lamb[c,k,s,4]+lamb[c,k,s,5]+lamb[c,k,s,6]
      }
    }
  }

#Number of individuals in each state at time t>1
 for(c in 2:C){
    for(k in 1:K){
      for(s in 1:S){
         pi[c,k,s]<-inprod(pi[(c-1),k,],lamb[c,k, ,s])
      }

        CHECK[c,k]<-pi[c,k,1]+pi[c,k,2]+pi[c,k,3]+pi[c,k,4]+pi[c,k,5]+pi[c,k,6]  #Check sums to N
    }
  }

#Number of individuals in each state at time >C
 for(c in C+1:T){
    for(k in 1:K){
      for(s in 1:S){
        pi[c,k,s]<-inprod(pi[(c-1),k,],lamb[c,k, ,s])
      }

         CHECK[c,k]<-pi[c,k,1]+pi[c,k,2]+pi[c,k,3]+pi[c,k,4]+pi[c,k,5]+pi[c,k,6]  #Check sums to N
    }
  }


#---------------------- COSTS AND UTILITIES ------------------------------------------------------------
#-----------------------------------------------------------------------------------------------------------

#DM PART 1 COSTING
#Costs  and  AQLYS of each intervention at time point c=1
 for(k in 1:K){
    cInterv[k]<- cEdu[k]+cEqp[k]+cHSI[k]+cInstall[k]   #cost of each strategy
    c_n1[k]<-n1[k]*0                                                    #cost=0 for usual care
    c_n2[k]<-n2[k]*(cInterv[k]+cAccept)                                 #accept intervention (SSM after intervention)
    c_n3[k]<-n3[k]*(cInterv[k]+cAccept)                                 #accept intervention (NO SSM after intervention)
    c_n4[k]<-n4[k]                              #decline intervention (NO SSMt)
    ct[1,k]<-cFixed[k]+c_n1[k]+c_n2[k]+c_n3[k]+c_n4[k]  #total cost for each intervention
    ut[1,k]<-uPop[1]*N                          #total QALYs for each intervention
  }
```

```
#DM PART 2 COSTING
#Health sector costs estimates (ambulance costs weighted by probability of emergency transfer = 24.2% reported in HES 2012b)
 cMinor    <- 0.242*cAmb+cED[1]                    #cost of minor injury = costs of ED for cases not admitted
 cModerate <- (0.242*cAmb)+cED[2]+cAdmit[1]+cGP[1]      #cost of moderate injury = Ambulanace + ED(cases leading to adm) +
        Short-inpat. adm + 11 mins GP consult.
 cSevere   <- (0.242*cAmb)+cED[2]+cAdmit[2]+cGP[1]+cHV   #cost of severe injury = Ambulanace + ED(cases adm) + Short-inpat.
        adm + 11 mins GP consult.
 ctFatal   <- (0.242*cAmb)+cED[2]+cAdmit[2]+cFatal      #total cost of fatal injury = Ambulanace + ED(cases leading to adm) + long-
        inpat. adm
 cAmb ~ dgamma(cAmb.a, cAmb.b)                     #per episode cost paramedic ambulance unit
 cHV ~ dgamma(cHV.a, cHV.b)                        #per episode cost of health vistor contact
 cChronic ~ dgamma(cChro.a, cChro.b)               #annual costs of chronic ill-health or disability
 cGP.temp <- cGP[2]                          #cost of 17 mins GP consultation set to temp. variable as its not used in model

for( i in 1:2){
   cED[i] ~ dgamma(cED.a[i], cED.b[i])
   cAdmit[i] ~ dgamma(cAdmit.a[i], cAdmit.b[i])
 }

 for(k in 1:K){
   for(s in 1:4){
     cSafe[k,s] <-cSafe1              #annual cost of maintaining safe storage =0
   }
 }


for(c in 2:C){
   for(k in 1:K){
     for(s in 1:2){
       #No unintentional ingestion
       c_o1[c,k,s]<-o1[c,k,s]*0
        c_o2[c,k,s]<-o2[c,k,s]*cSafe[k,s]
        c_o3[c,k,s]<-o3[c,k,s]*0

       #Ingestion but not admitted
        c_o4[c,k,s]<-o4[c,k,s]*cMinor
        c_o5[c,k,s]<-o5[c,k,s]*(cMinor+cSafe[k,s])
        c_o6[c,k,s]<-o6[c,k,s]*cMinor

       #Ingested, admitted, short-stay (i.e. moderate injury)
       c_o7[c,k,s]<-o7[c,k,s]*cModerate
        c_o8[c,k,s]<-o8[c,k,s]*(cModerate+cSafe[k,s])
        c_o9[c,k,s]<-o9[c,k,s]*cModerate

        c_o10[c,k,s]<-o10[c,k,s]*ctFatal
       c_o11[c,k,s]<- o11[c,k,s]*cSevere
        c_o12[c,k,s]<- o12[c,k,s]*(cSevere+cSafe[k,s])
        c_o13[c,k,s]<- o13[c,k,s]*cSevere

       cost[c,k,s]<-c_o1[c,k,s]+c_o2[c,k,s]+c_o3[c,k,s]+c_o4[c,k,s]+c_o5[c,k,s]+c_o6[c,k,s]+c_o7[c,k,s]
                +c_o8[c,k,s]+c_o9[c,k,s]+c_o10[c,k,s]+c_o11[c,k,s]+c_o12[c,k,s]+ c_o13[c,k,s]
     }

     for(s in 3:4){
       c_o1[c,k,s]<-o1[c,k,s]*0
        c_o2[c,k,s]<-o2[c,k,s]*(cSafe[k,s]+cChronic)
        c_o3[c,k,s]<-o3[c,k,s]*(0+cChronic)
        c_o4[c,k,s]<-o4[c,k,s]*cMinor #dead
        c_o5[c,k,s]<-o5[c,k,s]*(cMinor+cSafe[k,s]+cChronic)
        c_o6[c,k,s]<-o6[c,k,s]*(cMinor+0+cChronic)

       c_o7[c,k,s]<-o7[c,k,s]*cModerate #dead
        c_o8[c,k,s]<-o8[c,k,s]*(cModerate+cSafe[k,s]+cChronic)
        c_o9[c,k,s]<-o9[c,k,s]*(cModerate+cChronic)

       c_o10[c,k,s]<-o10[c,k,s]*ctFatal
        c_o11[c,k,s]<- o11[c,k,s]*cSevere
        c_o12[c,k,s]<- o12[c,k,s]*(cSevere+cSafe[k,s]+cChronic)
        c_o13[c,k,s]<- o13[c,k,s]*(cSevere+cChronic)

       cost[c,k,s]<-c_o1[c,k,s]+c_o2[c,k,s]+c_o3[c,k,s]+c_o4[c,k,s]+c_o5[c,k,s]+c_o6[c,k,s]+c_o7[c,k,s]
                +c_o8[c,k,s]+c_o9[c,k,s]+c_o10[c,k,s]+c_o11[c,k,s]+c_o12[c,k,s]+ c_o13[c,k,s]
     }

       cost[c,k,5]<-0    #No costs associated with dead states
```

```
        cost[c,k,6]<-0    #No costs associated with dead states
    }
  }

#DM PART 3 COSTING
 for(c in C+1:T){
    for(k in 1:K){
        cost[c,k,1]<- 0
        cost[c,k,2]<- 0
        cost[c,k,3]<- cChronic
        cost[c,k,4]<- cChronic
        cost[c,k,5]<- 0
        cost[c,k,6]<- 0
    }
  }


#UTILITIES IN EACH STATE
#General UK population utility:1 to 100 yrs
 for(i in 1:T){
     uPop[i]~dnorm(mn.uPop[i], prec.uPop[i])
 }

 uMinor ~ dbeta(uInj.a[1], uInj.b[1])
 uModerate ~ dbeta(uInj.a[2], uInj.b[2])
 uSevere ~ dbeta(uInj.a[3], uInj.b[3])
 uChronic ~ dbeta(uChro.a, uChro.b)

 for(k in 1:K){
    u[1,k,1]<-uPop[1]
    u[1,k,2]<-uPop[1]
    u[1,k,3]<-uPop[1]
    u[1,k,4]<-uPop[1]
    u[1,k,5]<-0
    u[1,k,6]<-0

    for(c in 2:C){
      for(s in 1:2){
        #No unintentional ingestion
        u_o1[c,k,s]<-o1[c,k,s]*0
        u_o2[c,k,s]<-o2[c,k,s]*uPop[c]
        u_o3[c,k,s]<-o3[c,k,s]*uPop[c]

        #Ingestion but not admitted
        u_o4[c,k,s]<-o4[c,k,s]*0
        u_o5[c,k,s]<-o5[c,k,s]*(uPop[c]-uMinor)
        u_o6[c,k,s]<-o6[c,k,s]*(uPop[c]-uMinor)

        #Ingested, admitted, short-inpatient stay (complete recovery)
        u_o7[c,k,s]<-o7[c,k,s]*0
        u_o8[c,k,s]<-o8[c,k,s]*(uPop[c]-uModerate)
        u_o9[c,k,s]<-o9[c,k,s]*(uPop[c]-uModerate)

        #Ingested, admitted, long stay (no complete recovery or fatal injury assumed)
        u_o10[c,k,s] <- o10[c,k,s]*0
        u_o11[c,k,s] <- o11[c,k,s]*0
        u_o12[c,k,s] <- o12[c,k,s]*(uPop[c]-uSevere)
        u_o13[c,k,s] <-o13[c,k,s]*(uPop[c]-uSevere)

        u[c,k,s]<- u_o1[c,k,s]+u_o2[c,k,s]+u_o3[c,k,s]+u_o4[c,k,s]+u_o5[c,k,s]+u_o6[c,k,s]+u_o7[c,k,s]+
                u_o8[c,k,s] +u_o9[c,k,s]+u_o10[c,k,s]+u_o11[c,k,s] +u_o12[c,k,s]+u_o13[c,k,s]
      }

      for(s in 3:4){
        u_o1[c,k,s]<-o1[c,k,s]*0
        u_o2[c,k,s]<-o2[c,k,s]*(uPop[c]-uChronic)
        u_o3[c,k,s]<-o3[c,k,s]*(uPop[c]-uChronic)

        #Ingestion but not admitted
        u_o4[c,k,s]<-o4[c,k,s]*0
        u_o5[c,k,s]<-o5[c,k,s]*(uPop[c]-uMinor-uChronic)
        u_o6[c,k,s]<-o6[c,k,s]*(uPop[c]-uMinor-uChronic)

        #Ingested, admitted, short-inpatient stay (complete recovery)
        u_o7[c,k,s]<-o7[c,k,s]*0
```

```
        u_o8[c,k,s]<-o8[c,k,s]*(uPop[c]-uModerate-uChronic)
        u_o9[c,k,s]<-o9[c,k,s]*(uPop[c]-uModerate-uChronic)

        #Ingested, admitted, long stay (no complete recovery or fatal injury assumed)
        u_o10[c,k,s] <- o10[c,k,s]*0
        u_o11[c,k,s] <- o11[c,k,s]*0
        u_o12[c,k,s] <- o12[c,k,s]*(uPop[c]-uSevere-uChronic)
        u_o13[c,k,s] <-o13[c,k,s]*(uPop[c]-uSevere-uChronic)

        u[c,k,s]<- u_o1[c,k,s]+u_o2[c,k,s]+u_o3[c,k,s]+u_o4[c,k,s]+u_o5[c,k,s]+u_o6[c,k,s]+u_o7[c,k,s]+
                u_o8[c,k,s] +u_o9[c,k,s]+u_o10[c,k,s]+u_o11[c,k,s] +u_o12[c,k,s]+u_o13[c,k,s]
      }


      u[c,k,5]<- 0
      u[c,k,6]<- 0
    }
  }

#DM PART 3 UTILITIES
 for(k in 1:K){
   for(c in C+1:T){
        u[c,k,1]<- uPop[c]
        u[c,k,2]<- uPop[c]
        u[c,k,3]<- uPop[c]-uChronic
        u[c,k,4]<- uPop[c]-uChronic
        u[c,k,5]<- 0
        u[c,k,6]<- 0
   }
 }

#--------------------------- MODEL EVALUATION -------------------------------------
#-----------------------------------------------------------------------------------

#Costs in each cycle of model
 for(k in 1:K){
   for(c in 2:T){
      ct[c,k]<-inprod(pi[c,k, ],cost[c,k, ])/pow((1+d.rate),(c-1))
   }
 }

#Utlities in each cycle of model
 for(k in 1:K){
   for(c in 2:T){
      ut[c,k]<-inprod(pi[c,k, ],u[c,k, ])/pow((1+d.rate),(c-1))
 }

  TotC[k]<-sum(ct[,k])
  mean.C[k]<-TotC[k]/N

  TotU[k]<-sum(ut[,k])
  mean.U[k]<-TotU[k]/N
 }

 Cost.diff[2]<-mean.C[2]-mean.C[1]          #Intervention2 compared to usual care
 Cost.diff[3]<-mean.C[3]-mean.C[1]          #Intervention3 compared to usual care
 Cost.diff[4]<-mean.C[4]-mean.C[1]          #Intervention4 compared to usual care
 Cost.diff[5]<-mean.C[5]-mean.C[1]          #Intervention5 compared to usual care
 Cost.diff[6]<-mean.C[6]-mean.C[1]          #Intervention6 compared to usual care
 Cost.diff[7]<-mean.C[7]-mean.C[1]          #Intervention6 compared to usual care

 Util.diff[2]<-mean.U[2]-mean.U[1]          #Intervention2 compared to usual care
 Util.diff[3]<-mean.U[7]-mean.U[1]          #Intervention3 compared to usual care
 Util.diff[4]<-mean.U[4]-mean.U[1]          #Intervention4 compared to usual care
 Util.diff[5]<-mean.U[5]-mean.U[1]          #Intervention5 compared to usual care
 Util.diff[6]<-mean.U[6]-mean.U[1]          #Intervention6 compared to usual care
 Util.diff[7]<-mean.U[7]-mean.U[1]          #Intervention6 compared to usual care

#Cost-effectiveness
 for(b in 2:K){
   ICER[b]<-Cost.diff[b]/Util.diff[b]              #Iincremental cost-effectiveness ratio (ICER)
 }

 for(j in 1:J){
   Rc[j]<-(j-1)*2000
```

```
    for(k in 1:K){
      NB[k,j]<-Rc[j]*mean.U[k]-mean.C[k]    #Net monetary benefit
      pCE[k,j]<-equals(rank(NB[,j],k),NT)   #Probability CE for cost-effectiveness Acceptability Curves
    }
  }

}
```

APPENDIX VII: RESEARCH PAPERS

Research paper 1

The paper entitled 'An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial' has been published in the *Journal of Clinical Epidemiology* [PMID: 24388291].

Research paper 2

The paper entitled 'The effectiveness of different interventions to promote poison prevention behaviours in households with children: a network meta-analysis' has been submitted to *PLOS ONE* and is currently under review.

Research paper 3

The paper entitled 'Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis' has been published in *Statistics in Medicine* [PMID: 22865748].

Research paper 4

The paper entitled 'Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes' has been published in *BMC Medical Research Methodology*. http://www.biomedcentral.com/1471-2288/14/92.

ELSEVIER

Journal of Clinical Epidemiology ■ (2013) ■

Journal of Clinical Epidemiology

REVIEW ARTICLE

An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial

Felix Achana[a], Stephanie Hubbard[a], Alex Sutton[a], Denise Kendrick[b], Nicola Cooper[a,*]

[a]*Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK*

[b]*Division of Primary Care, Community Health Sciences, Faculty of Medicine & Health Sciences, University of Nottingham, Nottingham, NG7 2RD, UK*

Accepted 25 September 2013; Published online xxxx

Abstract

Objectives: To review the methods currently used to synthesize evidence in public health evaluations and demonstrate the availability of more sophisticated approaches.

Study Design and Setting: A systematic review of National Institute for Health and Care Excellence (NICE) public health appraisals published between 2006 and 2012 was performed to assess the methods used for the synthesis of effectiveness evidence. The ability of new developments in evidence synthesis methodology to address the challenges and opportunities present in a public health context is demonstrated.

Results: Nine (23%) of the 39 NICE appraisals included in the review performed pairwise meta-analyses as part of the effectiveness review with one of these also including a network meta-analysis. Of the remainder, 29 (74.4%) presented narrative summaries of the evidence only, and 1 (2.6%) appraisal did not present any review of effectiveness and/or cost-effectiveness evidence. Heterogeneity of outcomes, methods, and interventions were the main reasons given for not pooling the data. Exploration of quantitative synthesis methods shows that pairwise meta-analyses can be extended to incorporate individual participant data (when it is available), extend the number of interventions being compared using a network meta-analysis, and adjust for both subject- and summary-level covariates. All these can contribute to ensuring the analysis answers directly the policy-relevant questions.

Conclusion: More sophisticated methods in evidence synthesis should be considered to make evaluations in public health more useful for decision makers. © 2013 Elsevier Inc. All rights reserved.

Keywords: Public health evaluation; Network meta-analysis; Decision making; Meta-analysis; Systematic review

1. Introduction

Systematic reviews and economic evaluations conducted within a decision modeling framework are two important tools in health-care evaluation [1,2]. Systematic reviews with or without meta-analyses have been accepted as providing a transparent and consistent way of obtaining research evidence on effectiveness of interventions in a way that minimizes bias [3]. Decision analytical models offer an additional framework through which effectiveness evidence, ideally from a systematic review, may be integrated with other relevant evidence and information on resource utilization to derive comparative estimates of cost-effectiveness. By providing a framework for assessing effectiveness and cost-effectiveness, these methods enable policy-relevant questions such as which interventions represent the best use of scarce health-care resources to be answered [4].

A key component of a systematic review is how the evidence, on outcomes such as effectiveness and adverse events, is synthesized. Meta-analysis, when used in a systematic review to combine quantitative information from multiple well-conducted randomized controlled trials (RCTs), is considered at the top of the hierarchy of evidence for intervention effectiveness [5]. An alternative approach to evidence synthesis, when meta-analysis is considered

F.A. and S.H. are joint first authors.

Conflict of interest: None.

Funding: This study presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10231). The views expressed are those of the authors and are not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health.

* Corresponding author: Tel.: +44-(0)1162297267; fax: +44-(0) 1162297250.

E-mail address: njc21@le.ac.uk (N. Cooper).

0895-4356/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jclinepi.2013.09.018

What is new?

Key findings

- Quantitative synthesis is not carried out in the systematic reviews for most public health (PH) evaluations.
- When quantitative synthesis *is* done, it tends to use the simplest methods, for example, a fixed- or random-effects meta-analyses comparing two groups, which potentially limits the scope of the analysis.

What this adds to what was known?

- Demonstrates how more sophisticated synthesis methods can be used in PH appraisals to more realistically model the data and answer the relevant policy questions.

What is the implication and what should change now?

- Researchers working on PH evaluations should consider expanding their toolbox and using more sophisticated methods many of which have recently been developed, motivated, and applied in pharmaceutical evaluations.

inappropriate, is narrative synthesis (also referred to as qualitative synthesis [6]). In this approach, individual studies identified in the review are summarized using a variety of formats without combining results quantitatively [7].

Meta-analysis is widely applied in reviews of the effectiveness of clinical interventions, treatments, and medical device technologies where the interventions and health outcomes are usually well defined and evaluated in well-conducted RCTs [8]. In other fields of health-care evaluation, however, things may not always be as clear cut. A good example is public health (PH), where interventions are often more complex and less well defined than clinical interventions [9]. There may also be a lack of good-quality evidence, particularly from RCTs in PH, for a number of well-documented reasons [10,11] including limited generalizability of the findings of RCTs to the wider population due to highly selected study populations, a narrow definition of intervention strategies and outcomes, and a focus on the individual instead of the community that is of interest in PH. Even when feasible, many have argued that RCTs may not always be possible to conduct in PH for other reasons, for example, ethical concerns may be raised regarding not offering the control population a possibly beneficial intervention [10]. Also, many of the RCTs conducted in PH tend to be cluster randomized trials and hence have more complex designs that need adjusting for in the analysis. In addition, the best available PH evidence may often come from observational nonrandomized studies [9], despite the increased risk of bias associated with the lack of randomization. For these reasons, the use of quantitative evidence synthesis methods such as meta-analysis in PH raises a number of methodological challenges. These include (1) increased methodological heterogeneity and risk of bias as a result of including studies with different study designs (RCTs, cluster RCTs, controlled before-and-after studies, and other observational nonrandomized studies), (2) the interventions or "program" being evaluated is often described in little detail, (3) a wide range of outcomes measures are often used, which may be variously defined across studies, and (4) the use of intermediate and/or surrogate outcome measures.

There are growing calls for PH decision making to be based on the best available evidence whenever possible. For example, a 2004 Department of Health report [12] on improving health and reducing health inequalities in England called for economic evaluations of PH interventions to ensure judicious use of scarce resources. Following this report, the remit of the UK National Institute for Health and Care Excellence (NICE), which already evaluated pharmaceutical interventions, was expanded to include the development of guidance for PH based on sound appraisals of intervention effectiveness and cost-effectiveness [13]. Consequently, a number of PH appraisals have been produced by NICE since 2006 on a wide range of issues including smoking cessation, alcohol use, and, particularly of relevance to the example used in this article, unintentional injuries in children.

To help address specific methodological challenges and provide advice on the technical aspects of the appraisal development process, NICE published a manual of methods for PH evaluation in 2006 [14], which was subsequently updated in 2009 [15] (a further update was published in September 2012 [16] after this review was completed, but the guidance was not changed). The guidance recommended "Meta-analysis data may be used to produce a graph if the data (usually from RCTs) are sufficiently homogenous and if there are enough relevant and valid data from comparable (or the same) outcome measures. Where such data are not available, the synthesis may have to be restricted to a narrative overview of individual studies looking at the same question," "Before pooling or combining the results of different studies, the degree of heterogeneity in the data should be assessed to determine how the results have been affected by the circumstances in which studies were carried out," and "Publication bias [17,18] should be critically assessed and reported in the interpretation of the meta-analysis results." These recommendations match well to the challenges in systematic review/meta-analysis in PH highlighted by the Cochrane Collaboration [9] and the 2011 Institute of Medicine report on standards for systematic reviews [6].

In view of the aforementioned challenges facing PH evaluations and recommendations for synthesis of PH evidence contained in the NICE manuals of methods, a review of all NICE PH appraisals published since 2006 was conducted. The aim of this article is twofold: (1) to identify the current situation (ie, what is already done and/or not done) with regards to addressing problems in synthesis of PH evidence and (2) to illustrate the application of new synthesis methods (ie, beyond those recommended by NICE [14−16] and Cochrane [9]) including methods from other fields such as health technology assessment to PH evidence that we believe have the potential to address many of the challenges in PH evaluation as aforementioned and thus improve the quality of evidence syntheses in PH interventions.

2. Systematic review of NICE PH appraisals

2.1. Methods

Completed PH appraisals published between March 1, 2006 and September 25, 2012 were identified for inclusion in the review through the NICE Web site (http://www.nice.org.uk/Guidance/PHG/Published). Each PH appraisal consisted of a number of articles such as qualitative reviews, epidemiologic reviews, expert opinions, field reports, and other similar nonquantitative review reports, quantitative systematic reviews of intervention effectiveness and cost-effectiveness, and decision analytical modeling reports. These were retrieved from the "background information" sections and assessed for eligibility. The "how this guidance was produced" sections were also searched for relevant articles if none were identified under "background information." Articles meeting the inclusion criteria were systematic reviews of the quantitative effectiveness and cost-effectiveness evidence and/or decision analytical modeling reports. Qualitative evidence reviews, epidemiologic reviews, field reports, expert opinions, and other similar nonquantitative evidence review reports were excluded. In addition, the final appraisal/guidance documents developed for each PH appraisal area were also excluded as these did not contain relevant information on the conduct of the evidence synthesis and decision modeling, which is of interest in this review. All except two (PH1 and PH2) of the appraisals were published after the 2006 NICE manual of methods [14] so should have followed the guidance for quantitative effectiveness evidence synthesis techniques.

Information extracted from the retrieved articles was used to assess the methods used to synthesize the effectiveness evidence and subsequent incorporation of the evidence into the decision models (when developed) that informed the PH appraisal. The assessment criteria for the synthesis methods were

1. Type of systematic review—narrative summary vs. meta-analysis;
2. Included studies—RCT vs. observational (non-randomized) studies;
3. Methods used to synthesize the evidence (if undertaken), including specification of the statistical model (including fixed- and/or random-effects models), heterogeneity, publication bias, and the outcome measures used, as well as presentation of results; and
4. How evidence from the systematic review was used to inform any cost-effectiveness analysis.

2.2. Results of systematic review

Thirty-nine completed PH appraisals published since 2006 were identified from the NICE PH Web site. Within these 39 appraisals, 371 potentially relevant articles were retrieved, and after screening the titles and reading the introduction and/or abstract sections, 164 were excluded as they failed to meet the inclusion criteria. Fifty-two articles, identified as duplicates and supplementary appendices, were combined with the corresponding main report and counted as one article leaving a total of 155 articles for inclusion in this review. The median number of included articles per appraisal was 4 (range 0 to 10). [No relevant supporting document meeting our inclusion criteria existed for one appraisal (PH36—prevention and control of hospital infection).]

2.2.1. Type of review

Table 1 lists all 39 PH appraisals by summary of the evidence synthesis and cost-effectiveness analyses undertaken to inform each appraisal development. One appraisal (PH36) reported neither effectiveness and cost-effectiveness evidence reviews nor a decision model, two appraisals (PH33 and PH34) reported reviews of evidence but conducted no cost-effectiveness analysis, and the fourth appraisal (PH7) reported evidence reviews and decision models; however, no estimates of cost-effectiveness of interventions were presented.

Twenty-nine (74.4%) of the 39 appraisals contained systematic reviews in which only a narrative summary of the evidence was conducted, another seven (18%) conducted both narrative summary and meta-analysis, two appraisals (5%) conducted only meta-analysis, and one (2.6%) appraisal had no systematic review and hence no evidence synthesis. In the narrative summary approach, the review findings were summarized study by study in the text and through tables. Sometimes, forest plots were used to display results of primary studies, but no overall mean or pooled result was presented (see PH4 for an example). Eight of the 29 appraisals using only a narrative summary approach did not report the reasons for not pooling the data, 2 included only review-level evidence from the overview of reviews, and 19 cited heterogeneity as the reason why meta-analysis was not considered appropriate. The reported causes of heterogeneity are presented in Appendix at www.jclinepi.com.

ARTICLE IN PRESS

Table 1. NICE public health appraisals and summary of evidence synthesis methods and decision modeling used to inform their development

NICE public health appraisal title	Review of the effectiveness and cost-effectiveness evidence and decision analysis used to inform each appraisal					
	Systematic review of effectiveness (narrative summary)	Systematic review of effectiveness (at least one M-A)[a]	Cost-effectiveness reviews	Study quality	Decision model	Source of effectiveness estimate used in decision model[b]
Brief interventions and referral for smoking cessation (PH1)	✓	x^{nr}	✓	✓	✓	Published review
Four commonly used methods to increase physical activity (PH2)	✓	x^{nr}	✓	✓	✓	Individual study[1]
Prevention of sexually transmitted infections and under 18 conceptions (PH3)	✓	✓	✓	✓	✓	Published review
Interventions to reduce substance misuse among vulnerable young people (PH4)	✓	x^{o}	✓	✓	✓	Individual study[1]
Workplace interventions to promote smoking cessation (PH5)	✓	x^{m}	✓	✓	✓	Individual study[3]
Behaviour change (PH6)	✓	x^{nr}	✓	✓	✓	Individual study[4]
School-based interventions on alcohol (PH7)	✓	✓	✓	✓	✓	Individual study[1]
Physical activity and the environment (PH8)	✓	$x^{i,m,o}$	✓	✓	✓	Individual study[3]
Community engagement (PH9)	✓	$x^{i,m,o,p}$	✓	✓	x	Not applicable
Smoking cessation services (PH10)	✓	✓	✓	✓	✓	New meta-analysis
Maternal and child nutrition (PH11)	✓	x^{m}	✓	✓	✓	Individual study[5]
Social and emotional wellbeing in primary education (PH12)	✓	✓	✓	✓	✓	Individual study[5]
Promoting physical activity in the workplace (PH13)	✓	x^{nr}	✓	✓	✓	Individual study[5]
Preventing the uptake of smoking by children and young people (PH14)	✓	$x^{m,o}$	✓	✓	✓	Individual study[1]
Identifying and supporting people most at risk of dying prematurely (PH15)	✓	$x^{i,m,p}$	✓	✓	✓	Individual study[1]
Mental wellbeing and older people (PH16)	✓	$x^{i,m,o}$	✓	✓	✓	Individual study[1]
Promoting physical activity for children and young people (PH17)	✓	$x^{i,m,o}$	✓	✓	✓	Analyst estimate[4]
Needle and syringe programmes (PH18)	✓	x^{nr}	✓	✓	✓	Individual study[3]
Management of long-term sickness and incapacity for work (PH19)	✓	✓	✓	✓	✓	New meta-analysis
Social and emotional wellbeing in secondary education (PH20)	✓	$x^{i,m,o}$	✓	✓	✓	Individual study[2]
Reducing differences in the uptake of immunisations (PH21)	✓	x^{nr}	✓	✓	✓	Individual study[4]
Promoting mental wellbeing at work (PH22)	✓	x^{i}	✓	✓	✓	Individual study[3]
School-based interventions to prevent smoking (PH23)	✓	✓	✓	✓	✓	New meta-analysis
Alcohol-use disorders - preventing harmful drinking (PH24)	✓	x^{nr}	✓	✓	✓	Published review
Prevention of cardiovascular disease (PH25)	✓	x^{nr}	✓	✓	✓	Individual study[5]
Quitting smoking in pregnancy and following childbirth (PH26)	✓	$x^{i,m,o}$	✓	✓	✓	Published review
Weight management before, during and after pregnancy (PH27)	✓	✓	✓	✓	✓	Not clear[5] reported

Looked-after children and young people (PH28)	✓	$x^{m,o,p}$	✓	✓	✓	Individual study[3]
Strategies to prevent unintentional injuries among under-15s (PH29)	✓	x^{i}	✓	✓	✓	Individual study[3]
Preventing unintentional injuries among under-15s in the home (PH30)	✓	$x^{i,o}$	✓	✓	✓	Individual study[3]
Preventing unintentional road injuries among under-15s: road design (PH31)	✓	x^{i}	✓	✓	✓	Individual study[2]
Skin cancer prevention: information, resources and environmental changes (PH32)	✓	$x^{i,m}$	✓	✓	✓	Individual study[3]
Increasing the uptake of HIV testing among black Africans in England (PH33)	✓	x^{m}	✓	x	x	Not applicable
Increasing the uptake of HIV testing among men who have sex with men (PH34)	✓	x^{nr}	✓	x	x	Not applicable
Preventing type 2 diabetes—population and community interventions (PH35)	✓	✓	✓	✓	✓	New meta-analysis
Prevention and control of healthcare-associated infections (PH36)		x	x	x	x	Not applicable
Tuberculosis - hard-to-reach groups (PH37)	✓	x^{nr}	✓	✓	✓	Individual study[5]
Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38)	✓	✓	✓	✓	✓	New meta-analysis
Smokeless tobacco cessation - South Asian communities (PH39)	✓	x^{s}	✓	✓	✓	Published review

Abbreviations: NICE, National Institute for Health and Care Excellence; M-A, meta-analysis.

Ticks indicate a systematic review of evidence, meta-analysis, or decision model have been conducted, whereas x indicates analysis have not been conducted.

[a] Reported reason why meta-analysis was not done [i = heterogeneity of interventions, m = heterogeneity of methods, design, and settings, o = heterogeneity of outcome measures, p = heterogeneity of study populations, s = heterogeneity of studies (specific cause not reported), and nr = not reported or reported that studies do not support a meta-analysis].

[b] Selection of individual study estimate of intervention effect used in the decision model (1 = used a prespecified criteria reported in the decision model report, 2 = discussion with NICE or estimates selected based on quality grading of evidence using the guide manuals of methods, 3 = selected studies based on the relevance of the intervention to the decision problem, 4 = assumption/analyst estimated based on an assumption, and 5 = not clearly reported).

ARTICLE IN PRESS

2.2.2. Included studies—RCTs vs. nonrandomized studies

Two (PH23 and PH38) of the 38 appraisals (containing a systematic review) included evidence from RCTs only in the effectiveness review. The remaining 36 appraisals were informed by reviews of both randomized and observational (nonrandomized) evidence identified from individual study reports and/or published systematic review reports. All 38 appraisals (containing a systematic review) graded the quality of primary studies and assessed the applicability of the evidence adhering to the guidelines for PH appraisal methods [14,15].

2.2.3. Quantitative evidence synthesis

Only 9 of the 39 appraisals (23%) contained one or more systematic review with a meta-analysis (Table 2). In total, there were 10 systematic reviews and/or decision analytical modeling reports with at least one meta-analysis within the nine appraisal areas. (Note: PH10 has two systematic review reports in which a meta-analysis was conducted.) Four of the 10 meta-analyses included RCTs only, and six included both RCT and observational (non-RCT) studies. Six of the 10 meta-analyses were conducted on ‘‘final outcomes’’; that is, the main outcome measures on which the corresponding cost-effectiveness analyses were based (eg, PH10 Smoking abstinence). The remaining four meta-analyses were conducted on ‘‘intermediate outcomes’’ (eg, PH3 Uptake of Chlamydia screening in schools rather than prevention of chlamydia).

There was evidence that interventions may have been ‘‘lumped’’ [19,20] into two broad intervention groups to facilitate inclusion of more studies in 7 of the 10 reports with a meta-analysis. For example, in PH23, which investigated the effect of school-based interventions on alcohol consumption, seemingly different interventions (such as lessons delivered by teachers or other professionals as part of the curriculum; peer-led education by other pupils; external contributions from, for example, the police, life education center staff; and implementation of school policy—type interventions) were lumped together to form one ‘‘intervention group,’’ which was then compared with the no intervention control in a pairwise meta-analysis.

Seven of the 10 review reports conducted random-effects pairwise meta-analysis, one conducted fixed- and random-effects analyses, one conducted random-effects mixed treatment comparisons [20] (also referred to as network meta-analysis [21,22]—see later) alongside the pairwise analysis, and another one did not clearly present the statistical model used. Six of the 10 systematic reviews presented forest plots with heterogeneity statistics displayed on them, two (PH3 and PH1) presented forest plots without heterogeneity statistics, and one review (PH35) did not present a forest plot. Only one review (PH23) assessed publication bias using funnel plot and Egger’s test for asymmetry.

2.2.4. How the evidence from the systematic reviews was incorporated into the model

Thirty-five (89.7%) of the 39 appraisals were informed by cost-effectiveness evaluations contained in one or more decision analytical modeling reports (Table 1). Twenty-three (66%) of these used estimates of intervention effectiveness derived from individual studies identified in the systematic review to inform the decision analysis (reasons for using the studies selected given in Table 1), 5 (14%) used previously published systematic review results, another 5 (14%) used estimates from a meta-analysis of studies identified in the systematic review, 1 used expert opinion/analyst estimate, and another one did not clearly report the source(s) of the intervention effect.

3. Exposition of new synthesis methodology applied in a PH evaluation context

In this section, we outline new developments in evidence synthesis methodology, many of them motivated by the evaluation of medical interventions and others motivated specifically by challenges in PH. We also show how such methods can be applied in a PH context to help address challenges and opportunities that exist in this context and thus, in some situations, raise the quality bar (established in the first part of this article) for PH interventions.

We use, for illustration, a topic area in which the authors have actively been working for several years—accident prevention among preschool children at home. This area of accident prevention among children at home was recently appraised by NICE PH30 (Table 1) using only narrative summaries for the systematic review of intervention effectiveness and thus using estimates from individual trials to inform the cost-effectiveness analyses. We have found accident prevention to have many of the issues typical of PH appraisals including studies of different designs, heterogeneity in both study design [eg, specific nature of interventions, level of randomization (individual or cluster), etc.] and study results, and interest in differential treatment effects across degrees of population inequality such as accommodation type, proportion of black and minority ethnicity, and proportion of single-parent families.

The account discussed later follows an approximately chronological path and details the development and adaptation of methods to synthesize the evidence by making the best use of available data. In this study, we restrict our attention to strategies to reduce falls among children at home, in particular, to increase the possession of a fitted stair gate(s) in homes.

We start by discussing the analyses performed in a recently updated Cochrane review [23] of interventions to prevent unintentional injuries to children at home—pairwise meta-analysis, subgroup analyses to explore heterogeneity, and meta-regression incorporating individual participant data (IPD). We then present a network meta-

Table 2. Review of quantitative methods used to synthesize public health evidence for NICE public health appraisal

Appraisal title	Systematic review report title	Included RCTs only	Main outcome	Description of main outcome	Outcome measure: statistic	Type of synthesis	Model type	Lumping[a] of interventions	Presentation of results	Assessed publication bias	Software	Used result of M-A in decision model
Prevention of sexually transmitted infections and under 18 conceptions (PH3)	Review 2 - Review of evidence for the effectiveness of screening for genital chlamydial infection in sexually active young women and men	No	Intermediate	Uptake of proactive chlamydia screening using home-collected specimens	Screening response rate (%)	M-A	Random effects	No	FP/Txt	No	RevMan, Stata	No
School-based interventions on alcohol (PH7)	Alcohol and schools: effectiveness and cost-effectiveness review	No	Final	Alcohol use	Weighted mean difference	M-A	Random effects	Yes	FP/Txt	No	Not stated	No
Smoking cessation services (PH10)	Cut down to quit' with nicotine replacement therapies	Yes	Final	6 or more months' sustained abstinence	Relative risk and Cohen's d	M-A	Random effects	Yes	FP/T/Txt	No	RevMan	Yes
Smoking cessation services (PH10)	Final report	No	Final	6 or more months' sustained abstinence	Cohen's d	M-A	Fixed and random effects	Yes	FP/T/Txt	No	RevMan	No
Social and emotional wellbeing in primary education (PH12)	Teesside review	Yes	Intermediate	Social problem solving	Standardized mean difference	M-A	Random effects	Yes	FP/T	No	RevMan	No
Management of long-term sickness and incapacity for work (PH19)	PH19 Management of long-term sickness and incapacity for work: Economic analysis report	No	Yes	Number returning to work following sickness	Relative risk	M-A	Random	Yes	FP/T/Txt	No	RevMan	Yes
School-based interventions to prevent smoking (PH23)	School-based interventions to prevent smoking: quantitative effectiveness review	Yes	Final	Smoking uptake	Odds ratio	M-A	Random effects	Yes	FP/Txt	Yes	Stata	Yes
Weight management before, during and after pregnancy (PH27)	Weight management before, during and after pregnancy: evidence review	No	Intermediate	Number exceeding IoM[b] guidelines for healthy weight gain	Relative risk	M-A	Random effects	Yes	FP/T/Txt	No	RevMan	No

(*Continued*)

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 2. Continued

Appraisal title	Systematic review report title	Included RCTs only	Main outcome	Description of main outcome	Outcome measure: statistic	Type of synthesis	Model type	Lumping[a] of interventions	Presentation of results	Assessed publication bias	Software	Used result of M-A in decision model
Preventing type 2 diabetes - population and community interventions (PH35)	PH35 Preventing type 2 diabetes - population and community interventions: report on cost-effectiveness evidence and methods for economic modelling	No	Intermediate	Body mass index	Weighted mean difference	M-A	Not reported	Yes	T/Txt	No	Not reported	Yes
Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38)	Prevention of type 2 diabetes: systematic review & meta-analysis of lifestyle, pharmacological and surgical interventions	Yes	Final	Reduce progress to diabetes for people with IGT	Hazard ratio	M-A and NMA	Random effects	No	FP/TxT	No	RevMan (M-A) WinBUGS (NMA)	Yes

Abbreviations: NICE, National Institute for Health and Care Excellence; RCT, randomized controlled trial; M-A, pairwise meta-analysis; FP, forest plot; T, table; Txt, Text; IGT, impaired glucose tolerance; NMA, network meta-analysis.

[a] Lumping is a term used in the literature [19,20] to describe the tendency to aggregate or treat seemingly similar but disparate/different interventions as one intervention group, for example, to facilitate inclusion of many studies in a meta-analysis. A classic example is treating different doses of a drug as if they were the same treatment.

[b] American Institute of Medicine (IOM) guidelines on weight management in pregnancy.

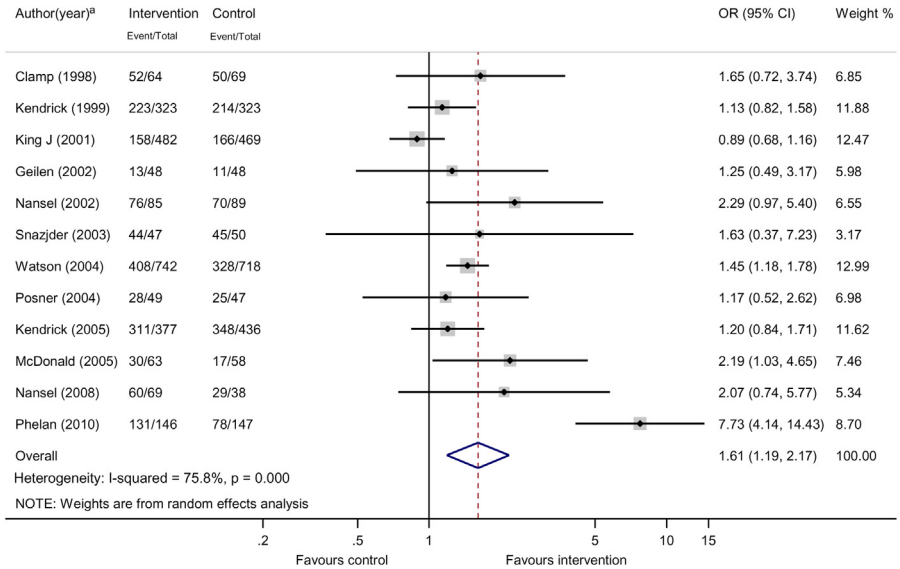


Fig. 1. Forest plot illustrating the findings of the random-effects meta-analysis of interventions aimed at increasing the uptake of safety equipment for the outcome ''possession of a fitted stair gate.'' [a] References for all the studies can be found in the updated Cochrane review [23]. OR, odds ratio; CI, confidence interval.

analysis that allows the interventions to be ranked and provides more informative evidence for a cost-effectiveness analysis.

3.1. Pairwise meta-analysis

A random-effects meta-analysis was used to synthesize the evidence for the possession of fitted stair gate(s) outcome, which comprised 12 studies [10 RCTs (2 clusters allocated) and 2 non-RCTs (1 cluster allocated)]. Because the original reporting of the cluster randomized studies had ignored the effect of clustering in their analysis, the meta-analysis was adjusted using external data to estimate the likely effects of such clustering on the certainty of the results [24]. Fig. 1 displays a forest plot of the results. Intervention arms were more likely to possess fitted stair gate(s) than the control arms [odds ratio (OR), 1.61; 95% confidence interval (CI): 1.19, 2.17]. Considerable heterogeneity was observed between study results (I^2 = 76%) [25].

3.2. Subgroup analyses

Potential sources of heterogeneity were explored using subgroup analyses based on a priori explanations, which were (1) whether the intervention included the provision of safety equipment, (2) follow-up period (up to and including 3 months and 4 or more months), (3) whether the intervention was delivered in a clinical setting or at home or community, (4) use of a randomized or nonrandomized design, and (5) study quality (allocation concealment, blinding of outcome assessment, and at least 80% follow-up in each treatment arm). Some of the heterogeneities were partly explained by different settings and the provision of stair gates, but significant heterogeneity remained in the different subgroups.

3.3. Meta-regression using IPD and summary data

In an attempt to explain further variability between study results—to address whether differential intervention effects could be discerned to be related to indicators of deprivation—and thus try and answer questions relating to inequalities in health care, a number of subject-level covariates were explored. To achieve this, the IPD were requested from the researchers responsible for all the relevant primary studies. By obtaining IPD, the power of meta-regression to explore subject-level covariates (eg, if the subject lived in owned or rented accommodation, etc.) is much increased over the use of summary data (eg, the percentage of subjects living in an owned house in a particular study) [26]; in fact, obtaining IPD is considered the gold standard way to carry out meta-analysis generally [27].

IPD were successfully obtained for approximately half of the studies across all types of injury prevention included in the review, with varying degree of success for the different injury prevention domains. But this partial success presented an analysis challenge. We wanted to not only use the IPD but also include the other studies in the analysis for which only summary data were available. This involved using a model developed for the original Cochrane review in this area [28], which essentially ''married'' summary and IPD meta-analysis models including covariates within a single analysis based on all available data [29]. This approach also accounted for the correct analysis of the cluster-allocated studies through appropriate reanalysis of the IPD (when available) and through utilization of adjustment

methods for the summary data as aforementioned. Importantly, using IPD allowed the use of data on outcomes that had not been reported in the articles; for example, some studies had reported composite measures of home safety and not individual safety practices, but the IPD included data on these individual safety practices.

For the possession of fitted stair gate(s) outcome, IPD were obtained for 10 of the 12 studies. Treatment interactions were investigated for child age, ethnic group, gender, family type (single or two parents), housing tenure (rented or owned), and parental unemployment. Most of the findings indicated little difference between the subgroups, except for the analysis of housing tenure, which combined the analyses of IPD for two cluster and five noncluster studies, and one study for which only summary data were available. The OR for intervention effect in non−owner-occupied households was 1.98 [95% credible interval (CrI, which is similar to a CI generated using Bayesian statistics): 1.48, 2.66], and in owner-occupied households, the OR was 1.22 (95% CrI: 0.96, 1.61), providing evidence to suggest that the intervention effect was larger in non−owner-occupied households (ratio of ORs, 1.62; 95% CrI: 1.18, 2.24).

It is interesting to note that such covariates could have been investigated without obtaining IPD through the use of meta-regression on summary-level covariates (ie, percentage of study participants in non−owner-occupied households), but such an analysis has much diminished power and is more prone to ecological/aggregation biases [30]. Running such an analysis on the same eight studies, but not using any IPD, produces an exponentiated regression coefficient of 1.01 (95% CI: 0.998, 1.022), indicating that there is no evidence of an increase in the odds of possession of fitted stair gate(s) for a one−percentage point increase in percentage of families living in non−owner-occupied household. This result is very different from the findings from the IPD analysis, which suggest that the odds of possessing fitted stair gate(s) are 62% higher among those in non−owner-occupied household than those in owner-occupied household.

3.4. Network meta-analysis

Our next refinement to the analysis, not included in the Cochrane review, came from concerns with the interpretability of the effect sizes from pairwise analyses of the type presented previously. We were aware that the interventions to increase the uptake of safety practices varied between studies (eg, interventions ranged from educational initiatives, through vouchers to reduce the price of equipment, through to the free provision and fitting of equipment), and therefore, by fitting the data into a meta-analysis framework of ‘‘intervention’’ vs. ‘‘usual care,’’ the interpretation of the resulting pooled effect was unclear—exactly what does the pooled effect relate to? This was especially important as the effectiveness results were to be used to inform the cost-effectiveness of injury prevention interventions evaluated via a decision model, which would require explicit interventions to be defined and costed. Thus, an analysis in which the different interventions were kept as unique was required. Once this was established, it became possible to include further relevant literature, known about but not used in the initial meta-analysis, in the analysis, namely, studies which compared different interventions to increase safety equipment uptake directly (but which had no ‘‘usual care’’ control group—hence their omission thus far). Further literature searches were conducted to identify all such studies. Network meta-analysis, which was being increasingly used in the evaluation of pharmaceuticals for funding bodies such as NICE [31], presented an analysis approach that would both keep interventions distinct and include trials with direct comparisons.

The meta-analysis of possession of fitted stair gate(s) outcome presented in the Cochrane review included all studies that compared a control group with an enhanced intervention group, but these controls and interventions varied considerably as outlined in Fig. 2A. In fact, seven distinct controls and interventions (including usual care) were identified across the included studies. To better understand the structure of the evidence base, when interventions are defined in this more refined way, a network diagram of the form presented in Fig. 2B [32] can be constructed. Network meta-analysis methods allow a simultaneous analysis of all the comparisons presented on the network. Table 3 [33] presents the ORs for the pairwise comparisons between the interventions produced both from the network meta-analysis and the direct comparisons from a trial or, when there was more than one trial, a pairwise meta-analysis of that particular comparison. In the network meta-analysis, the most intensive intervention (education + low-cost/free equipment + home safety inspection + fitting) was most effective for the possession of a fitted stair gate outcome compared with all other interventions. The probability that each intervention is best and the median rank (with uncertainty) of each intervention [31] calculated from the network meta-analysis are presented in Table 4. These data show that the most intensive intervention clearly had the highest probability (0.97) of being the most effective and a median rank of 1 (95% CI: 1, 2).

Although we believe such an analysis is more refined, interpretable, complete, and thus more helpful than the standard pairwise meta-analysis presented initially, it only considered summary study data, some of which were obtained from IPD, and did not include any potential treatment-modifying covariates. We had developed models to include covariates in network meta-analysis of summary data [34] (including a special model to deal with the inclusion of the control group event rate as a covariate in network meta-analysis [35], which is not illustrated here but potentially very useful in a PH context in which inequalities are of interest, particularly when IPD are not

ARTICLE IN PRESS

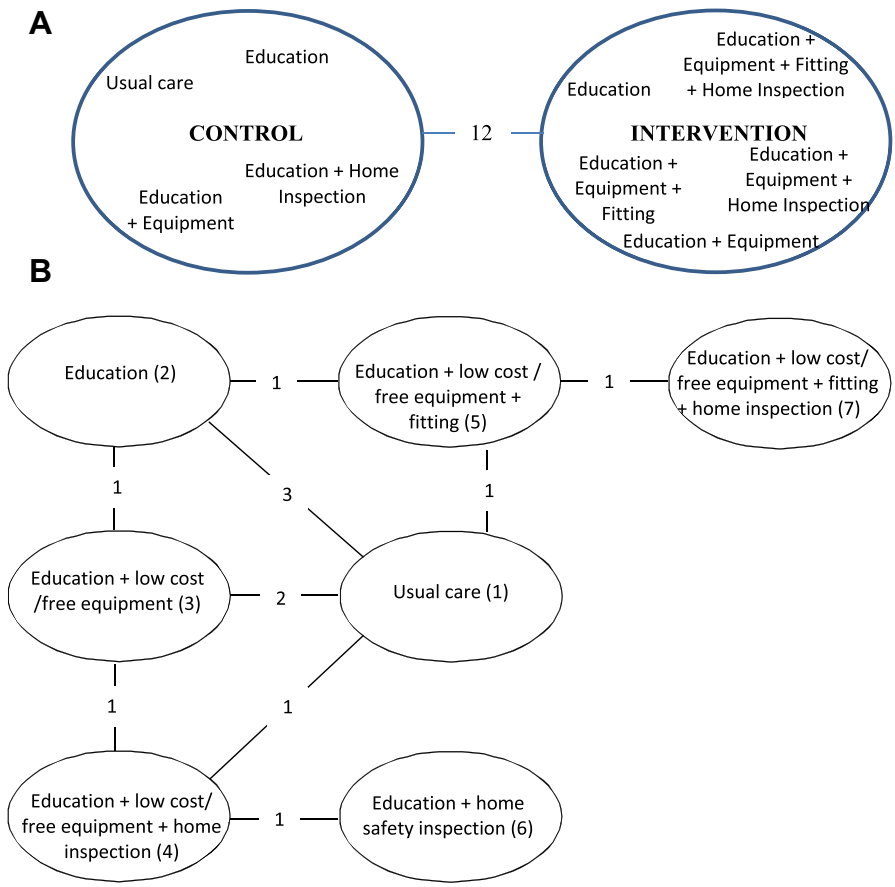


Fig. 2. Network diagrams indicating how intervention groups were defined and the number of studies in the (A) Cochrane review and (B) network meta-analysis.

available). We have also extended the network meta-analysis covariate model to allow the inclusion of IPD and thus subject-level covariates when possible [36].

4. Discussion

This review of completed NICE PH appraisals illustrates the current situation regarding the use of evidence synthesis methods to inform PH decision making in the United Kingdom. It identified that effectiveness evidence was mostly synthesized using narrative summaries and that quantitative synthesis was not carried out for most evaluations in PH systematic reviews. Of the 39 appraisals published since 2006, only 9 (23%) appraisals were informed by at least one systematic review with a meta-analysis. The other 30 appraisals may have refrained from meta-analysis because of a lack of randomized trials or heterogeneity in study design (ie, a mix of RCTs and non-RCTs). Moreover, systematic reviews opting for a quantitative summary tended to use the simplest methods such as fixed- or random-effects pairwise meta-analyses, which only enables comparison between two interventions at any one time and thus potentially limiting the scope of the analysis and the utility of the findings. These findings would seem to indicate that despite great advances in quantitative synthesis techniques, application in PH evaluation is still very much in its infancy and appears to lag behind other areas of health care such as the evaluation of clinical interventions. There are several reasons for this, not least due to the often heterogeneous nature of PH evidence including variations in many aspects of study design, including (1) the exact nature of the interventions, (2) outcome measures, (3) the wider scope of many PH research questions, and (4) the quantitative skills of the researchers involved.

Underlying our desire for PH reviews to become more quantitative, in the face of the challenges encountered, is a firm belief that a structured and transparent description and analysis of the decision question is desirable. Our review found that nearly 80% of NICE PH appraisals did not attempt a quantitative synthesis at all because of, what investigators believe but we want to challenge, insurmountable problems due to the heterogeneous nature of the evidence base. We believe that the more complex synthesis models, described in Section 3, can often more appropriately model the types of data commonly available in PH appraisals than carrying out less focused and detailed reviews of the literature.

NICE guidance states that "Meta-analysis data may be used to produce a graph if the data (*usually from RCTs*)

ARTICLE IN PRESS

Table 3. Results of a network meta-analysis (above stepped line) and pairwise meta-analysis (below stepped line) for possession of a fitted stair gate expressed as odds ratios (95% CrI) [a,b]

	Usual care (1)	Education (2)	Education + equipment (3)	Education + equipment + home inspection (4)	Education + equipment + fitting (5)	Education + home inspection (6)	Education + equipment + fitting + home inspection (7)
Usual care (1)		1.43 (0.90, 2.49)	1.63 (0.93, 3.03)	1.28 (0.69, 2.79)	1.52 (0.84, 3.38)	1.43 (0.56, 4.42)	7.80[c] (3.08, 21.3)
Education (2)	1.48 (0.97, 2.25)		1.14 (0.56, 2.23)	0.90 (0.41, 2.07)	1.07 (0.51, 2.41)	1.01 (0.33, 3.25)	5.46[c] (1.75, 16.12)
Education + equipment (3)	1.92[c] (1.05, 3.51)	1.17 (0.52, 2.63)		0.78 (0.38, 1.77)	0.94 (0.42, 2.41)	0.88 (0.32, 2.80)	4.77[c] (1.56, 15.18)
Education + equipment + home inspection (4)	1.13 (0.82, 1.58)		1.25 (0.49, 3.17)		1.20 (0.45, 3.25)	1.12 (0.52, 2.49)	6.13[c] (1.75, 18.71)
Education + equipment + fitting (5)	1.45[c] (1.18, 1.79)	1.63 (0.37, 7.23)				0.94 (0.27, 3.28)	5.07[c] (1.47, 15.93)
Education + home inspection (6)				1.12 (0.86, 1.47)			5.48[c] (1.23, 20.73)
Education + equipment + fitting + home inspection (7)	7.73[c] (4.14, 14.43)						

Abbreviation: CrI, credible interval.

Blank cells indicate that no direct evidence on specific pairwise comparisons was available.

[a] Values above the stepped line are results from the NMA; those below the line are direct estimates from a trial or, when more than one was available, a meta-analysis.

[b] Column and row headings signify intervention or comparison (intervention number).

[c] Significant at the 5% level.

Table 4. Assessment of which intervention is best for possession of a fitted stair gate

Intervention	Possession of a stair gate	
	Probability treatment is best	Median treatment rank (95% CrI)
Usual care (1)	0.00	7 (5, 7)
Education (2)	0.002	4 (2, 7)
Education + equipment (3)	0.004	3 (2, 7)
Education + equipment + home inspection (4)	0.001	5 (2, 7)
Education + equipment + fitting (5)	0.008	4 (2, 7)
Education + home inspection (6)	0.013	4 (2, 7)
Education + equipment + fitting + home inspection (7)	0.97	1 (1, 2)

is sufficiently homogenous'' (Section 5.4.4.2 n NICE guidance 2012 [16]). For PH reviews, the evidence from RCTs is often limited, and the best available evidence may be from non-RCTs, which reviewers may be reluctant to pool because of the risk of bias (Cochrane chapter 13 [9], Valentine and Thompson[37], and Moher et al. [38]). However, provided reviewers quality assess non-RCTs (as they would RCTs) to identify well-conducted studies, to limit confounding by selection bias, then meta-analysis can be considered.

Although concluding the evidence base to be ''too heterogeneous for meta-analysis'' may be better than carrying out a naive simple meta-analysis, not being able to present a quantitative analysis severely restricts the utility of the review, particularly for decision making. Exploring heterogeneity and attempting to account for it should be part of the analysis, and greater awareness of modern methods, and greater expertise in using them, will yield fruit for future PH reports. There are several other reasons why conducting a meta-analysis may not be advisable, however, for example, a small number of studies may mean that statistical heterogeneity is underestimated; some studies are too biased to draw a conclusion from them; there is evidence of publication bias; and insufficient reporting of outcomes.

We acknowledge that although softwares to undertake pairwise meta-analysis are widely available (eg, RevMan, Comprehensive Meta-Analysis), analyses such as the most complex ones previously described require advanced statistical expertise in evidence synthesis to implement (and some groundwork regarding the Bayesian theory underpinning such an approach may be required by nonstatistical PH specialists). Our software package of choice is WinBUGS. This is a freely available Bayesian simulation package [39] and is extremely powerful for fitting models not immediately available in other packages. (It even allows economic decision models to be included in the same program as the synthesis model, allowing a truly comprehensive assessment [40].) With the recent publication of the NICE technical support documents on evidence synthesis methodology [32,41,42] including all WinBUGS code to implement the models, together with more widely available specialist training courses and the new introductory WinBUGS book [43], the time is ripe for getting to grips with the more complex evidence synthesis methodologies currently being embraced by health technology appraisals [1,44]. A detailed discussion of specific technical challenges in Bayesian random-effects synthesis models is available elsewhere [45].

This article is limited to only considering NICE PH appraisals in the review and does not claim to have all the answers to all evidence synthesis challenges that exist in PH evaluation. For example, none of the above analyses considers directly the influence of the study quality/validity of the individual studies going into an analysis, although others are doing work in other contexts that could be adapted, for example, including different, both observational and randomized, evidence [46].

Regarding the specific injury prevention context, for the analyses presented previously, even when categorizing the interventions into seven distinct groups, there is still residual heterogeneity in intervention definition, for example, education may be a leaflet designed for the prevention of an injury at home, it may also include a face-to-face interview, a computer-based questionnaire producing tailored advice based on the user answers, and so forth. We are developing further modeling extensions including how to extrapolate across a series of evidence networks to allow information sharing on the effectiveness of interventions in promoting other safety practices for the prevention of falls. We hope such analyses will be more efficient and robust than individual analyses of each outcome. Note that all the data considered only relate to an increase in the *uptake of safety practice* and not to reduction in accidents per se. Therefore, a further initiative is to develop models, which extend those presented to include the direct evidence between safety practices and injury data. This problem is similar to the use of surrogate end points in clinical evaluation, and we plan to adapt methods developed there.

PH evaluations are notoriously messy and complex, with many factors to consider. But if a decision has to be made, explicit, transparent, and appropriate analysis of the data should be preferred to current alternatives. Just as evaluations of clinical interventions are becoming more sophisticated, we think there is a pressing need to do the same for

PH contexts and we hope this article can contribute to the initiation of such an initiative.

Appendix

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2013.09.018.

References

[1] Cooper NJ, Peters J, Lai MCW, Juni P, Wandel S, Palmer S, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value Health 2011;14(2):371−80.

[2] Novielli N, Cooper NJ, Abrams KR, Sutton AJ. How is evidence on test performance synthesized for economic decision models of diagnostic tests? A systematic appraisal of health technology assessments in the UK since 1997. Value Health 2010;13(8):952−7.

[3] Higgins JPT, Green SE. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration 2011.

[4] Drummond MF, Sculpher M, Torrance GW, O'Brien B, Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford, UK: Oxford University Press; 2005.

[5] Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A. An encouraging assessment of methods to inform priorities for updating systematic reviews. J Clin Epidemiol 2009;62:241−51.

[6] Eden J, Levit L, Berg A, Morton S. Finding what works in health care: standards for systematic reviews. 2012. Available at http://www.nap.edu/catalog.php?record_id=13059. Accessed January 1, 2013.

[7] Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing methodological guidance on the conduct of narrative synthesis in systematic reviews. Evaluation 2009;15(1):49−73.

[8] Weatherly H, Drummond M, Claxton K, Cookson R, Ferguson B, Godfrey C, et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. Health Policy 2009;93(2−3):85−92.

[9] Armstrong R, Waters E, Doyle Je. Chapter 21: reviews in health promotion and public health. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: John Wiley & Sons; 2008.

[10] Rosen L, Manor O, Engelhard D, Zucker D. In defense of the randomized controlled trial for health promotion research. Am J Public Health 2006;96:1181−6.

[11] Petticrew M, Egan M. Relevance, rigour and systematic reviews. In: Popay J, ed. Moving beyond effectiveness in evidence synthesis. Methodological issues in the synthesis of diverse sources of evidence. 2006. Available at http://www.nice.org.uk/niceMedia/docs/Moving_beyond_effectiveness_in_evidence_synthesis2.pdf. Accessed June 6, 2012.

[12] Wanless D. Securing good health for the whole population: final report—February 2004 (Chapter 6). 2004. Available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4074426. Accessed June 6, 2012.

[13] Chalkidou K, Culyer A, Naidoo B, Littlejohns P. Cost-effective public health guidance: asking questions from the decision-maker's viewpoint. Health Econ 2008;17(3):441−8.

[14] NICE. Methods for the development of NICE public health guidance. 2006. Available at http://www.nice.org.uk/media/FB9/59/PHMethodsManual2006.pdf. Accessed January 1, 2013.

[15] NICE. Methods for the development of NICE public health guidance. 2nd ed. 2009. Available at http://www.nice.org.uk/media/CE1/F7/CPHE_Methods_manual_LR.pdf. Accessed January 1, 2013.

[16] NICE. Methods for the development of NICE public health guidance. 3rd ed. 2012. Available at http://publications.nice.org.uk/methods-for-the-development-of-nice-public-health-guidance-third-edition-pmg4. Accessed January 1, 2013.

[17] Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical-trials. Control Clin Trials 1987;8:343−53.

[18] Sutton A, Abrams K, Jones DR, Sheldon TA, Song F. Methods for meta-analysis in medical research. Chichester, UK: Wiley; 2000.

[19] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897−900.

[20] Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol 2010;63:875−82.

[21] Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. Value Health 2010;13(8):976−83.

[22] Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21:2313−24.

[23] Kendrick D, Young B, Mason-Jones AJ, Ilyas N, Achana FA, Cooper NJ, et al. Home safety education and provision of safety equipment for injury prevention. Cochrane Database Syst Rev 2012;(9):CD005014.

[24] Kendrick D, Sutton A, Smith S, Mulvaney C, Woods A, Watson M, et al. Does the effect of home safety education, with or without the provision of safety equipment, differ by social group? Systematic review and meta-analysis. Report to the Department of Health 2006.

[25] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557−60.

[26] Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86−94.

[27] Stewart LA. Practical methodology of meta-analyses (overviews) using updated individual patient data. Stat Med 1995;14:2057−79.

[28] Kendrick D, Coupland C, Mulvaney C, Simpson J, Smith SJ, Sutton A, et al. Home safety education and provision of safety equipment for injury prevention. Cochrane Database Syst Rev 2007;(1):CD005014.

[29] Sutton AJ, Kendrick D, Coupland CAC. Meta-analysis of individual- and aggregate-level data. Stat Med 2008;27:651−69.

[30] Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010;340:c221.

[31] Dias S, Sutton AJ, Ades AE, Welton NJ. A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making 2013;33:607−17.

[32] Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011. Available at http://www.nicedsu.org.uk. Accessed January 1, 2013.

[33] Hubbard S, Sutton A, Kendrick D, Young B, Achana F, Cooper N. Network meta-analysis to evaluate the effectiveness of interventions to prevent falls injuries in children under 5. Leicester, UK: University of Leicester; 2013.

[34] Cooper NJ, Kendrick D, Achana F, Dhiman P, He Z, Wynn P, et al. Network meta-analysis to evaluate the effectiveness of interventions to increase the uptake of smoke alarms. Epidemiol Rev 2012;34:32−45.

[35] Achana FA, Cooper NJ, Dias S, Lu GB, Rice SJC, Kendrick D, et al. Extending methods for investigating the relationship between

ARTICLE IN PRESS

treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Stat Med 2013;32:752−71.

[36] Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Stat Med 2012;31:3516−36.

[37] Valentine JC, Thompson SG. Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. Res Synth Methods 2013; 4(1):26−35.

[38] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet 1999;354: 1896−900.

[39] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000;10(4):325−37.

[40] Cooper NJ, Abrams KR, Sutton AJ, Turner D, Lambert PC. A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer. J R Stat Soc Ser A Stat Soc 2003;166(3):389−405.

[41] Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU technical support document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011. Available at http://www.nicedsu.org.uk. Accessed November 8, 2011.

[42] Dias S, Welton NJ, Sutton AJ, Ades AE. NICE technical support document 4: inconsistency in networks of evidence based on randomised trials. 2011. Available at http://www.nicedsu.org.uk. Accessed January 1, 2013.

[43] Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book: a practical introduction to Bayesian analysis. Boca Raton, FL; CRC Press; 2012.

[44] Cooper NJ, Spiegelhalter D, Bujkiewicz S, Dequen P, Sutton A. A systematic review of the deliberate and accidental use of Bayesian methods in health technology assessment. Int J Technol Assess Health Care 2013; 29:336−42.

[45] Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009; 172:137−59.

[46] Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc 2009;172:21−47.

The effectiveness of different interventions to promote poison prevention behaviours in households with children: a network meta-analysis

Felix Achana[1], Alex J Sutton[2], Denise Kendrick[3], Persephone Wynn[3], Ben Young[3], David R Jones[2], Stephanie Hubbard[2], Nicola J Cooper[2]

[1] Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK
[2]Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, UK
[3]Division of Primary Care, University of Nottingham, Nottingham, NG7 2RD, UK

ADDRESS FOR CORRESPONDENCE

Prof Denise Kendrick

Division of Primary Care

Room 1303 Tower Building

University Park

Nottingham

NG7 2RD

UK.

E-mail: denise.kendrick@nottingham.ac.uk

Tel: +44 (0) 115 846 6914

Fax: +44 (0) 115 8466904

Abstract

Background

There is evidence from 2 previous meta-analyses that interventions to promote poison prevention behaviours are effective in increasing a range of poison prevention practices in households with children. The published meta-analyses compared any intervention against a “usual care or no intervention” which potentially limits the usefulness of the analysis to decision makers. We aim to use network meta-analysis to simultaneously evaluate the effectiveness of different interventions to increase prevalence of safe storage of i) Medicines only, ii) Other household products only, iii) Poisons (both medicines and non-medicines), iv) Poisonous plants; and v) Possession of poison control centre (PCC) telephone number in households with children.

Methods

Data on the effectiveness of poison prevention interventions was extracted from primary studies identified in 2 newly-undertaken systematic reviews. Effect estimates were pooled across studies using a random effects network meta-analysis model.

Results

28 of the 47 primary studies identified were included in the analysis. Compared to *usual care intervention*, the intervention with *education and low cost/free equipment* elements was most effective in promoting safe storage of medicines (odds ratio 2.51, 95% credible interval 1.01 to 6.00) while interventions with *education, low cost/free equipment, home safety inspection and fitting* components were most effective in promoting safe storage of other household products (2.52, 1.12 to 7.13), safe storage of poisons (11.10, 1.60 to 141.50) and possession of PCC number (38.82, 2.19 to 687.10). No one intervention package was more effective than the others in promoting safe storage of poisonous plants.

Conclusion

The most effective interventions varied by poison prevention practice, but *education alone* was not the most effective intervention for any poison prevention practice. Commissioners and providers of poison prevention interventions should tailor the interventions they commission or provide to the poison prevention practices they wish to promote.

Keywords

Poison prevention, network meta-analysis, childhood

Highlights

- Network meta-analysis is useful for comparing multiple injury-prevention interventions
- More intensive poison prevention interventions were more effective than *education alone*
- *Education and low cost/free equipment* was most effective in promoting safe storage of medicines
- *Education, low cost/free equipment, home safety inspection and fitting* was most effective in promoting safe storage of household products and poisons
- *Education, low cost/free equipment and home inspection* were most effective in promoting possession of a poison control centre number.
- None of the intervention packages was more effective than the others in promoting safe storage of poisonous plants.

Introduction

Globally poisonings result in approximately 45,000 deaths[1] and approximately 2.4 million disability adjusted life years (DALYS) lost (GBD 2010) each year in children and young people aged 0-19 years. They are a particular problem in young children with 13% of deaths[1] and 11% of the DALYs lost (GBD 2010) occurring in children aged 0-4 years. Each year, poisonings result in approximately 25,000 emergency department attendances in the UK in 0-4 year olds [2], 63,000 ED attendances in the USA for drug poisoning alone in the 0-5 year olds[3] and more than 1.2 million calls to poison control centres each year following poisoning in the under 5s in the USA[4]. Poisonings among 0-15 year olds have been estimated to cost the NHS more than £2 million each year[5]. In the US despite an estimated saving of $7 to $15 for every $1 spent on poison control centres[6], non-fatal poisonings resulted in $48 million medical costs for hospitalisations and ED attendances in 2005 for the under 5s[7].

There is evidence from 2 previous meta-analyses that interventions to promote poison prevention behaviours are effective in increasing a range of poison prevention behaviours [8,9]. The first meta-analysis in 2001 found a modest effect (statistical significance not reported) of interventions in a clinical setting on safe storage of cleaning products[9]. More recent meta-analyses found home safety education, with or without the provision of safety equipment was effective in increasing safe storage of medicines, safe storage of non-medicinal products and increasing availability of poison control centre numbers[8]. These meta-analyses compared any intervention against a “usual care or no intervention” comparison group. The interventions comprised various combinations of education, home safety inspection, provision of free or low cost safety latches for cabinets, drawers or cupboards and fitting of safety latches. Some of these interventions were aimed only at preventing poisonings, but most aimed to prevent a range of injuries and also included the provision of education and other items of safety equipment. Health care commissioners and housing providers, amongst others, have to make decisions about the “best” intervention for preventing

poisoning and this requires comparisons between these multiple interventions. Therefore meta-analyses which have been conducted to date that combine effects across all interventions and compare against usual care or no intervention can only make a limited contribution to these decisions.

Network meta-analysis (NMA) methods[10,11] (also known as mixed treatment comparison[12,13]) extend standard (pair-wise) meta-analysis to allow simultaneous comparison of all evaluated interventions within a single coherent analysis. Therefore all interventions can be compared with one another, including comparisons not evaluated within any of the primary studies. These analyses are increasingly being used in health technology assessment to help decide on the optimal intervention for a particular condition[14]. The objective of this research was to evaluate the effectiveness of different interventions to promote poison prevention behaviours by households with children. We believe this is the first application of NMA in this area of injury prevention.

Methods

Study identification

For the NMAs, data on the effectiveness of poison prevention interventions was extracted from primary studies identified in 2 newly-undertaken systematic reviews: a systematic review of reviews [15] and a systematic review of primary studies published since the most comprehensive systematic review [16].

The review included systematic reviews and meta-analyses of experimental study designs (randomised controlled trials (RCTs), non-RCTs and controlled before-and after (CBA) studies) and controlled observational studies (case control and cohort studies), and primary studies of experimental or controlled observational designs published since the most recently published comprehensive systematic review. Studies including children aged 0-19 years and their families that

provided interventions to promote poison prevention behaviours were included. Legislative interventions to reduce poisonings were excluded. Interventions to promote possession of ipecac were reported in many papers included in the systematic review, but were excluded from the NMA as use of ipecac is no longer recommended[17,18].

We searched MEDLINE, Embase, CINAHL, ASSIA, PsycINFO and Web of Science from date of inception to January 2012. We searched a range of other electronic sources in January 2013 and hand searched the journal Injury Prevention (March 1995 – January 2012) and abstracts from 1st-10th World Conferences on Injury Prevention and Control (1989 – 2010). Reference lists of included reviews and primary studies were searched for relevant citations. Full-text articles were retrieved regardless of language and translated where necessary. The search terms were adapted for study design and the same sources were searched from 2001 – January 2012 to identify primary studies. The search strategy used to search Medline and adapted as necessary for other databases is shown in Supplementary material S1 and the other sources searched are given in supplementary material S2.

Titles and abstracts of articles were scanned independently by 2 researchers to identify articles to retrieve in full. Where an article appeared to be eligible based on the title, but an abstract was unavailable, it was retrieved in full. Full articles were independently reviewed by 2 researchers using a standard form listing inclusion criteria. Disagreement between researchers was dealt with by referral to a 3rd member of the research team and consensus-forming discussions.

Data was extracted onto a standard form which recorded data on study design, participants, interventions, outcomes and numerators and denominators in each treatment arm. Data were extracted by 2 researchers independently and compared. Any discrepancies were investigated by referral back to the original article by a senior member of the research team. Authors were asked to

supply individual participant data (IPD) or unpublished aggregated data where the published data did not report numerators and or denominators or intra-class correlation coefficients (ICCs) for clustered data. Where studies did not adjust for clustered allocation of intervention, we estimated the effective sample size based on the design effect using published ICCs[19] or ICCs estimated from IPD where the author provided it.

The quality of included primary studies was assessed in terms of the following criteria: allocation concealment, blinding of outcome assessment and completeness of follow up for randomised studies, and blinding of outcome assessment, completeness of follow up and balance of confounders between treatment arms for non-randomized studies. Non-randomised studies were considered to be balanced in terms of confounders if the prevalence of confounders did not differ by more than 10 percent between the treatment arms, and unclear with respect to balance of confounders if the intervention and control groups were matched on various characteristics but no data was provided to judge the adequacy of this matching. The quality of controlled observational studies (case control and cohort studies) was assessed using the Newcastle-Ottawa scale[20].

Statistical methods

The five poisoning prevention outcomes considered in the NMAs were i) safe storage of medicines (Yes/No), ii) safe storage of other household products (Yes/No), iii) safe storage of poisons (Yes/No), iv) safe storage of poisonous plants (Yes/No), and iv) possession of poison control centre (PCC) telephone number (Yes/No). The safe storage of poisons outcome refers to storage of any potentially toxic substance and includes studies where the reported outcome included both medicines and other household products (i.e. where the outcomes were not reported separately). Safe storage was defined as storing potentially toxic substances (medicinal or non-medicinal) at

adult eye level and/or in locked cupboards/drawers/cabinets where they are inaccessible to children [16].

For each of the outcomes, NMA[11] was implemented to enable the comparison of all interventions with one another using all the available data in a connected network of studies; thus allowing comparisons of interventions not directly compared in studies but linked through a connected network of studies (indirect evidence). For example, a comparison of the following 4 interventions; usual care, education, equipment giveaway and home inspection, could be achieved using studies containing the following pair-wise comparisons, usual care *vs*. education, education *vs*. equipment giveaway, equipment giveaway *vs*. home inspection. However, if only studies of usual care *vs*. education and equipment giveaway *vs*. home inspection existed then the network would be *disconnected*; in such cases the analysis would be limited to performing only direct pairwise comparisons. For randomised trials, NMA preserves the within-study randomised treatment comparison of each trial while combining all available comparisons between interventions. Such analyses assume that there is consistency across evidence. For example, if an equipment giveaway arm had been included in the studies of usual care *vs.* education the estimate of education *vs.* equipment giveaway would be consistent across studies (i.e. the underlying estimates are assumed to be identical or exchangeable depending on whether fixed or random effects are assumed) with education *vs.* equipment giveaway.

For these analyses, a standard NMA random effects model with a binary outcome [11,12] was fitted to the data. We obtained pooled estimates of intervention effects for all combinations of pair-wise comparisons from the NMAs and for completeness we also present estimates from the head-to-head evidence for each pair-wise comparison where available. Effectiveness estimates are presented as odds ratios and summarised using forest plots developed by Tan et al[21]. Interventions were ranked based on absolute intervention effects (derived using a underlying rate based on the usual

care arms) and the probability that each intervention is best for a particular outcome[11,12] was calculated.

To assess the goodness of fit of the model to the data, the posterior mean residual deviance (defined as the difference between the deviance for the fitted model and the deviance for the saturated model, where the deviance measures the fit of the model to the data points using the likelihood function[22] was calculated. Under the null hypothesis that the model provides an adequate fit to the data, it is expected that the posterior mean residual deviance would have a mean equal to the number of unconstrained data points[23,24].

The between-study standard deviation parameter, τ was used to quantify the heterogeneity of the network (i.e. the variability in treatment effects within pair-wise comparisons above that expected by chance)[25]. The degree of heterogeneity was assessed as reasonable, high or extremely high based on guidelines for interpreting τ on the log-odds ratio scale suggested by Spiegelhalter *et al.* [26]. These state that values of τ from 0.1 to 0.5 may be considered as indicating a reasonable degree of heterogeneity, 0.5 to 1 as high and values above 1 as very extreme heterogeneity. In NMA it is important to assess consistency between the 'direct' and 'indirect' evidence of the dataset; this was evaluated using a method based on 'node splitting'[27] which calculates the probability that the mean treatment effect estimates based on the direct evidence (i.e. studies that directly compared the two treatments under consideration) exceeds the mean treatment effect estimates based on the indirect evidence (i.e. the remaining studies in the network under the consistency assumption). A 2-sided p-value was then derived (using the formula 2 x minimum(*prob*, *1-prob*)[27]. Note that only pairs of interventions that are part of a closed loop in the network of interest have both direct and indirect evidence available[13] and therefore can be assessed for consistency.

Sensitivity analysis

As the quality of included studies varied, analyses for all outcomes except safe storage of poisonous plants were repeated restricted to data obtained from RCTs only. It was not possible to conduct this repeat analysis for safe storage of poisonous plants as only 3 studies provided data for this outcome. All of the analyses were conducted using a Markov chain Monte Carlo method and fitted in the WinBUGS software[28]. Further technical details of the analysis together with the WinBUGS code are available from the corresponding author. The analysis and subsequent reporting adhere to the PRISMA statement (supplementary material S3) guidelines[29] and the implied criteria for reporting the results of NMA outlined in Bafeta et al[30].

Results

The process of selection of studies is shown in Fig. 1. One hundred and eighty two papers were assessed for inclusion. This included 125 papers from the search for systematic reviews and 57 from the search for primary studies published since the review we considered to be most comprehensive [31] which was published in 2001. In total 47 primary studies were identified for inclusion in the review, of which 27 were selected for inclusion in at least one of the NMAs (S4 Table). One study[32] divided patients into a randomised and a quasi-randomised study groups and analysed the two groups separately. This study was therefore counted as two separate studies, thus increasing the total number of studies included in the NMAs to 28. A detailed description of the characteristics of included studies have been published elsewhere [15,16]. A table of studies excluded from the NMA is given in the accompanying supplementary material (S5 Table). Summary characteristics of the 28 studies included in the NMAs together with their study quality which was observed to be variable across studies are reported in the accompanying supplementary material (S4 Table). Twenty (71%)

of the 28 studies were RCTs and 8 (29%) were non-RCTS. Overall, the following 9 intervention packages were evaluated across the 5 networks and no single study compared all interventions directly:

1) Usual care - including usual safety education or no education (UC)
2) Education - more than usual safety education (E)
3) Education + free/low cost equipment (E+FE)
4) Education + free/low cost equipment + home safety inspection (E+FE+HSI)
5) Education + free/ low cost equipment + fitting (E+FE+F)
6) Education + home safety inspection (E+HSI)
7) Education + free/low cost equipment + fitting +home safety inspection (E+FE+HSI+F)
8) Education + home visit as part of Healthy Steps for Young Children program (E+HV) and
9) Free/low cost equipment only (FE).

Fig. 1. PRISMA flow chart for the systematic overview of reviews and systematic review of primary studies.

The free/low cost equipment component of interventions varied between studies [15,16] and included items such as a smoke alarm, batteries, cabinet and window locks, fire guards and stair gates, among others. A detailed list of the equipment reported by each study is presented in are reported in the accompanying supplementary material (S4 Table). Fitting refers to installation of safety equipment by for example a researcher or professional as part of the intervention package [33,34].

Storage of medicines

Thirteen of the 28 studies compared the effectiveness of 7 interventions to promote safe storage of medicines (Panel A, Fig. 2). Eleven (85%) studies were RCTs and 2 (15%) were non-RCTs (S4 Table). One study[35] reported a 100% event rate in both treatment and control arms. This study was excluded from the analysis as it contributed no information on relative effectiveness of the interventions that is of interest in the analysis.

Fig. 2. Network Diagrams of interventions to increase safety practices to prevent poisonings in pre-school children in the home. PCC = poison control centre telephone number. Nodes/oval circles represent an intervention (E = education, F = Fitting, FE = low cost/free equipment, HSI = Home safety inspection, HV = Home visit). For example E+FE+HSI = Education + low cost/free equipment + home visit intervention. The lines connecting any two nodes represent the pairwise comparison. The numbers on each line represent the total number of studies and the number of non-RCTs (in brackets) contributing to each pairwise comparison.

Pooled estimates of 21 possible pairwise comparisons between the 7 interventions, together with the available direct within-trial estimates, are reported in Fig. 3. The results show that home safety interventions increase safe storage of medicines with education and low cost/free equipment the most likely to be effective (probability best = 0.39), with an estimated odds ratio compared to usual care of 2.51 (95% CrI: 1.01 to 6.00). When the effect of study design on the NMA results was assessed, by repeating the above analysis using only data from the 11 RCTs, the results were similar, although for this analysis the network was limited to only 6 interventions (i.e. excluding the intervention education, low cost/ free equipment and home safety inspection).

Fig. 3. Network meta-analysis results for safe storage of medicines. H-H trials refer to the number of head-to-head trials available for the specified pairwise comparison.

Storage of other household products

Fifteen studies evaluated 7 interventions for promoting safe storage of other household products (Panel B, Fig. 2) of which 11 (73%) studies were RCTs and 4 (27%) were non-RCTs (S4 Table). One study (Dershewitz 1977) reported zero events (i.e. none of the households surveyed safely stored other household products) in the equipment only (9) intervention arm. To facilitate inclusion of this study in the analysis, a continuity correction was applied by adding 0.5 and 1 to the denominator and numerator.

The NMA estimated the 21 possible pairwise comparisons between the 7 interventions trialled across the included studies (Fig. 4). The most intensive intervention (education, low cost/free equipment, home safety inspection and fitting) was most likely to be effective (probability best = 0.37), with an estimated odds ratio compared to usual care of 2.59 (95% CrI: 0.59 to 15.16).

Fig. 4. Network meta-analysis results for safe storage of other household products. H-H trials refer to the number of head-to-head trials available for the specified pairwise comparison.

The effect of study design on the NMA results was assessed by repeating the above analysis using only data from the 11 RCTs limiting the network to 6 interventions (i.e. excluding education only). The results changed slightly but the most intensive intervention was still most likely to be the most effective (probability best = 0.56) closely followed by the intervention education, low cost/ free equipment and home safety inspection (probability best = 0.44).

Safe storage of poisons

Nine studies provided data on the effectiveness of 5 interventions to increase safe storage of poisons in households with children (Panel C, Fig. 2). Six (67%) studies were RCTs and 3 (33%) were non-RCTs (S4 Table).

The NMA estimated the 10 possible pairwise comparisons between the 5 interventions trialled across the included studies (Fig. 5). There was evidence to suggest that the most intensive intervention (i.e. education, low cost /free equipment, home safety inspection and installation) was most effective in promoting the number of households with storage of poisons compared to usual care intervention (Probability best = 0.78; OR=11.10, 95% CrI= 1.60 to 141.50).

Fig. 5. Network meta-analysis results for safe storage of poisons. H-H trials refer to the number of head-to-head trials available for the specified pairwise comparison.

Repeating the analysis using only data from the 6 RCTs identified both education and low/free equipment (Probability best 0.38), and education, low cost/free equipment, home safety inspection and installation (Probability best 0.36) to be the most effective at promoting the number of households with storage of poisons compared to usual care intervention.

Safe storage of poisonous plants

Three RCTs, one of which is the 3-arm study [36] provided data on 5 interventions for storage of poisonous plants (Panel D, Fig. 2; S4 Table). The NMA estimated the 10 possible pairwise comparisons between the 5 interventions trialled across the included studies (Fig. 6). There was no evidence that any of the intervention was more likely to be effective than the others at promoting safe storage of poisonous plants.

Fig. 6. Network meta-analysis results for safe storage of poisonous plants. H-H trials refer to the number of head-to-head trials available for the specified pairwise comparison.

Possession of a PCC number

Ten studies evaluated 7 interventions to promote uptake of PCC number (Panel D, Fig. 2). 7 (70%) studies were RCTs and 3 (30%) were non-RCTs (S4 Table). There was evidence that the intervention education, low cost/free equipment and home safety inspection was more effective than usual care in increasing uptake of PCC number (Probability best = 0.76; OR=39.25, 95% CrI 2.19 to 687.10) (Fig. 7). When the effect of study design on the NMA results was assessed by repeating the above analysis using only data from the 7 RCTs based on 5 interventions (i.e. the following 2 interventions were excluded from the network: education and home safety inspection, and education and home visit) the results were very similar.

Fig. 7. Network meta-analysis results for possession of poison control centre number. H-H trials refer to the number of head-to-head trials available for the specified pairwise comparison.

Evaluation of models

Overall, the NMA models fitted the data well with the posterior mean residual deviance being close to the number of data points in each network (Table 1). The between study standard deviations for each of the NMA models are reported in Table 1 and indicate moderate between-study heterogeneity for storage of medicines and poisons, high heterogeneity for other household products and extremely high for possession of PCC number. The uncertainty in the estimation of the heterogeneity parameter reflects the relatively low number of studies providing direct evidence for each pairwise comparison. Where both direct and indirect evidence was available, consistency was checked for closed loops (excluding loops formed by multi-arm studies) in the network, using the

node-split method. There was no evidence of inconsistency between the direct and indirect evidence in all networks; that is all the p-values were not statistically significant at the 5% significance level (S6 Table).

Table 1: Evaluation of model fit

Outcome	No. of studies	Residual deviance	Posterior median of the between-study standard deviation, τ and 95% CrI in brackets
Safe storage of medicines	12	23.5 (cf 24 data points)	0.331 (0.013 to 1.239)
Safe storage of other household products	15	30.9 (cf 30 data points)	0.561 (0.128, 1.270)
Safe storage of poisons	10	21.0 (cf 21 data points)	0.361 (0.029, 1.436)
Safe storage of poisonous plants	3	6.6 (cf 7)	1.00 (0.003 to 3.818
Possession of a PCC number	10	19.5 (cf 20 data points)	1.165 (0.574, 1.926)

Discussion

Principal findings

In this NMA, we have been able to compare the different interventions evaluated with one another for promoting poison prevention behaviours by households with children. This analysis has allowed comparisons of strategies not addressed within any of the individual primary studies. The findings showed that more intensive interventions are more effective than education alone for each of the poison prevention practices we evaluated. Education and low cost/free equipment was most effective in promoting safe storage of medicines; education, low cost/free equipment, home safety inspection and fitting was most effective in promoting safe storage of other household products and poisons; and education, low cost/free equipment and home inspection was most effective in

promoting possession of a PCC number. There was no evidence that any of the interventions was more effective than the others at promoting safe storage of poisonous plants.

Strengths and limitations

NMA is a useful synthesis tool for comparing multiple injury prevention interventions which are often complex and multi-faceted, and where the number of studies evaluating the same comparisons is small. NMA enables interventions to be ranked in terms of their effectiveness in promoting safety practices providing results which are more likely to be useful to policymakers, service commissioners and providers when making choices between multiple alternatives than multiple pairwise meta-analyses.

We did not find evidence of inconsistency between direct evidence and indirect evidence in our analyses, although the power to detect inconsistency will have been limited by sparse data, particularly for analyses involving very few studies. The inclusion of non-randomised study designs allowed us to include a greater number of studies in our analysis, but also resulted in the inclusion of studies with greater potential for bias. Eight (29%) of the 28 studies included in our analysis were non-RCTs. Of these, only 2 were assessed as not balanced or unclear in terms of the distribution of confounders between study-arms was unclear is reported (S4 Table). Sensitivity analyses restricting analyses to RCTs produced similar results suggesting our findings were robust to exclusion of non-randomised studies. The quality of studies included in our analyses (assessed in terms of allocation concealment (RCTs only), blinded outcome assessment, balance of confounders (non-RCTs only) and completeness of follow-up) was variable. It was not possible to explore the impact of the individual measures of quality on our results since such an analysis would be extremely limited due to the large number of parameters being estimated in the NMA relative to the number of studies and may even lead to disconnected networks.

Although NMA allows interventions to be classified into more categories than standard pairwise meta-analysis, there is, inevitably, still some “lumping” of interventions within these categories. For example, education may differ in intensity across studies; that is, from a leaflet or brochure distributed by post, to intensive face-to-face classes teaching home safety. Subcategorising the interventions further, to avoid “lumping”, is reliant on detailed information being reported in the primary study publications. However, in the case of poison prevention education, insufficient detail was often reported to enable further sub categorisation.

Comparisons with existing work

Our findings are consistent with findings from two previous pairwise meta-analyses. DiGuiseppi found interventions promoting “child-proofing” the home delivered in clinical settings had a modest effect (odds ratio 1.8, statistical significance not reported) on safe storage of cleaning products substances[37]. The seconds meta-analysis by the authors of this paper[8], found that education, with or without the provision of safety equipment was effective in increasing safe storage of medicines (OR 1.53, 95% CI 1.27-1.84), safe storage of household products (OR 1.55, 95% CI 1.22- 1.96) and, increasing availability of poison control centre numbers (OR 3.30, 95% CI 1.70-6.39). Our findings extend those from the previous meta-analyses by demonstrating which elements of multifaceted interventions are most effective. Furthermore, one of the previous meta-analyses failed to find significant effects of education, with or without the provision of safety equipment on keeping (unspecified) poisons (OR 0.57, 95%CI 0.31-1.07) or plants out of reach (OR 1.18, 0.40-3.48), but we now demonstrate that some poison prevention interventions are effective in promoting these safety practices.

The effect sizes in our NMA for safe storage of medicines, other household products and availability of the poison control centre number are all larger than the effect sizes found in the pairwise meta-analyses previously reported[8,9]. It is likely that, by reducing clinical heterogeneity of interventions, our NMAs may explain some of the statistical heterogeneity in effect sizes found in previous pairwise meta-analyses. Our findings also suggest meta-analyses combining all interventions, (which include less intensive, and as we have shown, less effective interventions) may underestimate the effect of more intensive interventions.

Implications for practice and research

Our findings suggest that the "best" intervention for increasing a range of poison prevention practices are more intensive interventions. These include, at a minimum, education and providing equipment, but for some poison prevention practices the most effective intervention requires education, equipment provision and fitting and home safety inspection. The most effective intervention varied by poison prevention practice, so commissioners and providers of poison prevention interventions should tailor the interventions they commission or provide to the poison prevention practices they wish to promote. Knowing which interventions are most effective is important, but is only part of the information required to commission or provide poison prevention and cost-effectiveness is an essential part of any decision making process. The effect sizes from this NMA will be used in subsequent decision analyses to determine the most cost effective interventions for increasing poison prevention practices and these analyses will be presented elsewhere. Such an analysis is vital to determine which interventions provide best value for money, as more intensive interventions, which we have shown to be the most effective, will also be the most expensive.

Despite 28 studies being included in at least one NMA, the maximum number included in any NMA was 15 and many comparisons contained only a small number of studies. Further studies are therefore required to increase precision of effect estimates, to increase power to explore effects by study quality and inconsistency between direct and indirect evidence of effectiveness. In addition, a more detailed description of the intervention in future studies, in particular of the content of the educational elements of interventions would be helpful in allowing a finer subcategorisation and exploration of individual educational components. Methods to incorporate individual level data into NMA analyses are now available [38], and these would be useful for exploring whether the effect of interventions vary by characteristics of study population (e.g. deprivation) and the potential impact of interventions on inequalities in prevention practices.

Conclusions

Network meta-analysis has demonstrated that the most effective interventions varied by poison prevention practice, with more intensive interventions being more effective than education alone for each poison prevention practice. Education and the provision of home safety equipment are important components for all poison prevention practices. Home safety inspections are more important for promoting safe storage of non-medicinal poisons and plants and for possession of PCC numbers. Commissioners and providers of poison prevention interventions should tailor the interventions they commission or provide to the poison prevention practices they wish to promote.

Acknowledgements

Author Affiliations

Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, England (Nicola J. Cooper, Felix Achana, Alex J. Sutton, David R Jones); Division of Primary Care Research, Floor

13, Tower Building, University Park, University of Nottingham, NG7 2RD, England (Denise Kendrick, Persephone Wynn, Ben Young).

Funding source

This study presents independent research by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10231). The views expressed are those of the author(s) and are not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health.

Competing interest

DK is an author on 1 of the primary studies included in the NMA. DK is an author of 2 systematic reviews and DK & AJS are authors of one meta-analysis included in the systematic review.

References

1. World Health Organisation, UNICEF (2008) World report on child injury prevention.
2. The Royal Society for the prevention of accidents (2002) Home & Leisure Accident Surveillance System.
3. Xiang Y, Weiyan Z, Xiang H, Smith, G.A. (2012) ED visits for drug-related poisoning in the United States, 2007. American Journal of Emergency Medicine 30: 293-301.
4. Bronstein AC, Spyker DA, Cantilena LR, Green JL, Rumack BH, et al. (2011) 2010 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 28th Annual Report. Clinical Toxicology 49: 910-941.
5. CAPT: Child Accident Prevention Trust The case for action on childhood poisoning.
6. Spiller HA, Griffith JRK (2009) The value and evolving role of the US poison control center system. Public Health Reports 124: 359-363.
7. Centers for disease control and prevention Cost of injury report.
8. Kendrick D, Young B, Mason-Jones AJ, Ilyas N, Achana FA, et al. (2012) Home safety education and provision of safety equipment for injury prevention. Cochrane Database of Systematic Reviews The Cochrane Library 9.
9. DiGuiseppi C, Goss CW, Higgins JPT (2001) Interventions for promoting smoke alarm ownership and function. Cochrane Database of Systematic Reviews: 1-77.
10. Lumley T (2002) Network meta-analysis for indirect treatment comparisons. Statistics in Medicine 21: 2313-2324.
11. Dias S, Sutton AJ, Ades AE, Welton NJ (2013) Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making 33: 607-617.
12. Caldwell DM, Ades AE, Higgins JPT (2005) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 331: 897-900.
13. Lu G, Ades AE (2004) Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine 23: 3105-3124.
14. Cooper NJ, Sutton AJ, Lai M, Peters J, Abrams KR, et al. (2011) How valuable are multiple treatment comparison methods in economic decision modelling? Value in Health 14: 371-380.
15. Wynn P, Zou K, Young B, Majsak-Newman G, Hawkins A, et al. (2014) (In press). Prevention of childhood poisoning in the home: overview of systematic reviews and a systematic review of primary studies. International Journal of Injury Control and Safety Promotion.
16. Kendrick D, Young B, Mason-Jones AJ, Ilyas N, Achana FA, et al. (2012) Home safety education and provision of safety equipment for injury prevention. Cochrane Database of Systematic Reviews.
17. American Academy of Pediatrics Committee on Injury V, and Poison Prevention, (2003) Poison treatment in the home. American Academy of Pediatrics Committee on Injury, Violence, and Poison Prevention. Pediatrics 112: 1182-1185.
18. Benson BE, Hoppu K, Troutman WG, Bedry R, Erdman A, et al. (2013) Position paper update: gastric lavage for gastrointestinal decontamination. Clinical Toxicology 51: 140-146.
19. Kendrick D, Coupland C, Mulvaney C, Simpson J, Smith S, et al. (2007) Home safety education and provision of safety equipment for injury prevention. Cochrane Database of Systematic Reviews 1: 1-124.
20. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, et al. (Accessed November 2013) The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://wwwohrica/programs/clinical_epidemiology/oxfordhtm: Ottawa Hospital Research Institute.

21. Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, et al. (2014) Novel presentational approaches were developed for reporting network meta-analysis. Journal of Clinical Epidemiology.
22. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64: 583-639.
23. Congdon P (2003) Applied Bayesian modelling. Chichester: Wiley.
24. McCullagh P, Nelder JA (1989) Generalized linear models. London: Chapman and Hall.
25. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Chichester: Wiley.
26. Spiegelhalter DJ, Abrams KR, Myles JP (2004) Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: John Wiley & Sons.
27. Dias S, Welton NJ, Caldwell DM, Ades AE (2010) Checking consistency in mixed treatment comparison meta-analysis. Statistics in Medicine 29: 932-944.
28. Spiegelhalter D, Thomas A, Best N, Lunn D (2003) WinBUGS user manual: Version 1.4. Cambridge: MRC Biostatistics Unit.
29. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339: b2535.
30. Bafeta A, Trinquart L, Seror R, Ravaud P (2014) Reporting of results from network meta-analyses: methodological systematic review. British Medical Journal 348.
31. Towner E, Dowswell T, Mackereth C, Jarvis S (2001) What works in preventing unintentional injuries in children and young adolescents? An updated systematic review . . London: NHS Health Development Agency.
32. Minkovitz CS, Hughart N, Strobino D, Scharfstein D, Grason H, et al. (2003) A practice-based intervention to enhance quality of care in the first 3 years of life the - Healthy steps for young children program. Journal of the American Medical Association 290: 3081-3091.
33. King WJ, Klassen TP, LeBlanc J, Bernard-Bonnin AC, Robitaille Y, et al. (2001) The effectiveness of a home visit to prevent childhood injury. Pediatrics 108: 382-388.
34. Phelan KJ, Khoury J, Xu Y, Liddy S, Hornung R, et al. (2011) A randomized controlled trial of home injury hazard reduction: the HOME injury study. Arch Pediatr Adolesc Med 165: 339-345.
35. Kelly B, Sein C, McCarthy PL (1987) Safety Education in a Pediatric Primary Care Setting. Pediatrics 79: 818-824.
36. Babul S, Olsen L, Janssen P, McIntee P, Raina P (2007) A randomized trial to assess the effectiveness of an infant home safety programme. International Journal of Injury Control and Safety Promotion 14: 109-117.
37. DiGuiseppi C, Higgins JPT (2000) Systematic review of controlled trials of interventions to promote smoke alarms. Archives of Disease in Childhood 82: 341-348.
38. Saramago P, Sutton AJ, Cooper NJ, Manca A (2012) Mixed treatment comparisons using aggregate and individual participant level data. Statistics in Medicine 31: 3516-3536.

Legends for supplementary material

Research Article

Statistics in Medicine

Received 9 February 2012, Accepted 3 July 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5539

Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis

Felix A. Achana,[a*†] Nicola J. Cooper,[a] Sofia Dias,[b] Guobing Lu,[b] Stephen J. C. Rice,[c] Denise Kendrick[d] and Alex J. Sutton[a]

Baseline risk is a proxy for unmeasured but important patient-level characteristics, which may be modifiers of treatment effect, and is a potential source of heterogeneity in meta-analysis. Models adjusting for baseline risk have been developed for pairwise meta-analysis using the observed event rate in the placebo arm and taking into account the measurement error in the covariate to ensure that an unbiased estimate of the relationship is obtained. Our objective is to extend these methods to network meta-analysis where it is of interest to adjust for baseline imbalances in the non-intervention group event rate to reduce both heterogeneity and possibly inconsistency. This objective is complicated in network meta-analysis by this covariate being sometimes missing, because of the fact that not all studies in a network may have a non-active intervention arm. A random-effects meta-regression model allowing for inclusion of multi-arm trials and trials without a 'non-intervention' arm is developed. Analyses are conducted within a Bayesian framework using the WinBUGS software. The method is illustrated using two examples: (i) interventions to promote functional smoke alarm ownership by households with children and (ii) analgesics to reduce post-operative morphine consumption following a major surgery. The results showed no evidence of baseline effect in the smoke alarm example, but the analgesics example shows that the adjustment can greatly reduce heterogeneity and improve overall model fit. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: network meta-analysis; mixed-treatment comparison; baseline risk; underlying risk; MCMC; meta-regression

1. Introduction

In meta-analyses of clinical trials, differences in patient-level or trial/study-level characteristics often give rise to variation in treatment effect estimates between studies, also called heterogeneity [1]. Between-study variance in the treatment effects is usually taken into account through including a parameter for the residual heterogeneity in a random-effects meta-analysis [1, 2]. A random-effects model quantifies the degree of heterogeneity but does not explain it. To explain the source of the heterogeneity, patient-level and study-level characteristics are sometimes included in the analysis as covariates [1, 2]. A trial-level covariate of interest as a possible source of heterogeneity is the 'baseline risk' or the underlying risk of the disease. The baseline risk reflects the burden of disease in a study population and defines the average risk of a patient to experience the outcome of interest if they have not been treated [3]. It is potentially an important proxy for a number of unmeasured (and even measured) patient-level characteristics such as age, sex, medical history and disease severity that collectively influence a patient's

[a]*Department of Health Sciences, University of Leicester, Leicester LE1 7RH, U.K.*
[b]*School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, U.K.*
[c]*Centre for Reviews and Dissemination, University of York, York YO10 5DD, U.K.*
[d]*Division of Primary Care Research, Floor 13, Tower Building, University Park, Nottingham, NG7 2RD, U.K.*
**Correspondence to: Felix A. Achana, Department of Health Sciences, University of Leicester, Leicester LE1 7RH, U.K..*
[†]*E-mail: faa7@le.ac.uk*

Copyright © 2012 John Wiley & Sons, Ltd.

response to treatment [4]. In addition to heterogeneity, baseline imbalances between trials may also give rise to inconsistency [i.e. variability in the treatment effect between pairwise contrasts [5] in a network meta-analysis (NMA)]. Therefore, adjusting for it may have the benefit of reducing both heterogeneity and inconsistency in NMA and improve the overall model fit.

Various measures have been used for the baseline risk in meta-analyses. Examples include the observed event rate in the placebo or non-active intervention arm, the observed placebo arm log odds and the average of the observed event rates in the placebo and treatment arms [6–8]. However, including observed measures of baseline risk in a meta-regression can be problematic because of the measurement error in both response (i.e. treatment effect) and explanatory variables and functional relationship between the two [4]. The problem has received considerable attention in the literature with several authors proposing alternative model-based solutions. Examples include the methods of McIntosh [9], Walter [8], Thomson *et al.* [4], Sharp and Thomson [10], Arends *et al.* [11] and van Houwelingen *et al.* [12]. However, these methods are directly applicable mainly in pairwise meta-analysis. Our objective is to extend these methods to NMA where it may be of interest to adjust for baseline imbalance in the underlying risk across studies. The main reason for doing this would be to reduce between-study heterogeneity and possible inconsistency in the direct and indirect trial evidence on pairwise comparisons. This objective is complicated by missing data, due to the fact that not all studies in a network may have a placebo or non-active treatment control and thus an observed covariate value.

A review of the pairwise meta-analysis methods for investigating the relationship between treatment effect and baseline risk is presented in Section 2. In Section 3, we present an approach that primarily extends the methods of Thompson *et al.* [4, 10] and Arends *et al.* [11] from pairwise to NMA where it is of interest to adjust for the baseline risk. The method we present complements previous general multivariate meta-regression models suggested for NMA [5, 13–16] by allowing for the following: (i) alternative distributional assumptions to be made about the nature of the 'true' unobserved baseline risk measure and (ii) the inclusion of trials without a non-active treatment control and hence no baseline risk measure while allowing for the treatment × covariate interactions to be exchangeable or even different (i.e. as many regression coefficients as there are treatment effects). Section 4 presents application of the method to two recently published NMAs [17, 18]. The first example has a binary outcome and examines effectiveness of home safety education interventions to promote ownership of functional smoke alarm in households with children [17]. The second example has a continuous outcome measure and examines the effectiveness of analgesic treatments in reducing post-operative morphine consumption in adult patients following major surgery [18]. In Section 6, we discuss the results and the findings from the example datasets followed by the strengths and the limitations of the approach outlined in this paper.

2. Review of baseline risk models for pairwise meta-analysis

Both Sharp and Thomson [10] and Arends *et al.* [11] present good introductions to baseline risk adjustment and a detailed review of available methods for pairwise meta-analysis. We have summarised important features of six of the methods that we consider most relevant to our modelling approach for NMA in Table I. A common feature in these methods is to model the relationship of interest in three parts, although this was only stated explicitly by Arends *et al.* [11]. This involves specifying in any order the following: (i) an appropriate likelihood for the data; (ii) a regression model relating the 'true' treatment effect as an explanatory variable and the 'true' baseline risk as the covariate; and (iii) a model for the distribution of the baseline risk across studies.

Differences between approaches have mostly arisen from slightly different strategies adopted for each part of the model. For example, Thompson *et al.* [4], Arends *et al.* [11] and Sharp and Thompson [10] assumed a binomial likelihood for a binary outcome, whereas McIntosh [9], Walter [8] and van Houwelingen *et al.* [12] used a normal distribution to model a binary outcome measure (e.g. log odds or log-odds ratio). Approximating a log-odds ratio with a normal distribution can be mathematically and computationally convenient, but the normality assumption may be inappropriate if there are trials in the meta-analysis with zero or small numbers of events [10]. Secondly, except for the method of Walter [8], all the other methods assumed random study-specific effects. Walter's [8] model is a fixed effect in that no allowance is made for any residual heterogeneity other than that explained by the baseline risk, although we believe expecting residual heterogeneity is more realistic in most applications where it is of interest to adjust for the baseline risk.

Copyright © 2012 John Wiley & Sons, Ltd.

Table I. Summary of methods for modelling the relationship between treatment effect and baseline risk in pairwise meta-analysis with a binary outcome.

Method	Outcome data	Likelihood model	Distribution of baseline risk	Method of estimation	Further notes
Method 1 (RE): Walter [8]	Arm level	**Two normal distributions**: observed treatment and control group log odds with normal errors	None	ML or WSL with bias correction	Gives only fixed effects results as no allowance for excess heterogeneity. Narrow standard errors for regression slope β
Method 2 (RE): McIntosh [9]	Trial level	**Bivariate normal approximation**: log-OR and control group log odds assumed bivariate normal with known variance and covariance matrix estimated from data	Normal (RE on baseline risk)	ML and Bayesian	Bivariate normal assumption may be inappropriate if there are trials with small number of events. May result in more extreme estimates of slope β and lower estimates of between-study heterogeneity σ^2 Normality of baseline risk across trials may be hard to justify
Method 3 (RE): van Houwelingen *et al.* [12]	Arm level	**Bivariate normal approximation**: for binary outcome data, treatment and control group log odds *or* observed treatment effect and control group log odds assumed bivariate normal with known covariance matrix estimated from data	Normal (RE on baseline risk)	EM algorithm and SAS Proc Mix	
Method 4a (RE): Thompson *et al.* [4, 10], Arends *et al.* [11]	Arm level	**Exact binomial model**: observed number of events in each treatment arm assumed binomial	Fixed, flat prior	Bayesian	Eliminates need for zero-cell corrections Useful in situations where trials with small sample sizes are included in the meta-analysis
Method 4b (RE): Arends *et al.* [11]	Arm level	**Exact binomial model**: observed number of events in each treatment arm assumed binomial	Normal (RE)	Bayesian	
Method 4c (RE): Arends *et al.* [11]	Arm level	**Exact binomial model**: observed number of events in each treatment arm assumed binomial	Mixture of two normal distributions	Bayesian	Eliminates need for zero-cell corrections Useful in situations where trials with small sample sizes are included in the meta-analysis Flexible distributional assumptions for the baseline risk measure

ML, maximum likelihood; WLS, weighted least square; Log-OR, log-odds ratio; MA, meta-analysis; FE, fixed effects; RE, random effects

Copyright © 2012 John Wiley & Sons, Ltd.

Finally, the approaches outlined in Table I make different assumptions about the distribution of the 'true' unobserved baseline risk across studies [9, 19]. Although some models assumed a vague or minimally informative normal prior distribution (e.g. [4, 10] and also in [11]), a common parametric formulation is to assume that the baseline risk is normally distributed across trials as in McIntosh [9], van Houwelingen *et al.* [12] and also Arends *et al.* [11]. Additionally, Arends *et al.* [11] also proposed a more flexible model for the distribution of the baseline risk comprising a mixture of two normal distributions with different means but common between-study variance. Whether or not to assume a model for the baseline risk is a much debated issue with as yet no clear consensus among methodologists [10–12, 20]. More recently, Ghidey *et al.* [19, 21] proposed semi-parametric models for the distribution of the baseline risk as well as models that do not make any distributional assumptions. In the next section, we aim to develop methods for the baseline risk adjustment in NMA that incorporate the different assumptions about the baseline risk distribution across studies to assess the effect of these assumptions on parameter estimates.

3. Network meta-analysis with baseline risk covariate

3.1. Network meta-analysis model with no covariate adjustment

Suppose in a meta-analysis of $i = 1, 2, \ldots, N$ trials, we have $k = A, B, C, \cdots, NT$ treatments being compared with one another where NT is the total number of treatments under consideration. Take treatment A as the overall baseline or reference treatment of the entire network. For a binary outcome, we assume r_{ik} events occur out of n_{ik} patients in treatment arm k of trial i according to a binomial distribution with underlying event probability p_{ik}. Standard random-effects NMA for a binary outcome with no covariate can be specified using logistic regression [22, 23],

$$\begin{aligned} &r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik}) \text{ with } \theta_{ik} = \text{logit}(p_{ik}); \\ &\theta_{ik} = \begin{cases} \mu_{ib} & k = b; \quad b \in \{A, B, C, \cdots\} \\ \mu_{ib} + \delta_{ibk} & k > b; \quad b \in \{A, B, C, \cdots\} \end{cases} \\ &\delta_{ibk} = d_{bk} + \varepsilon_{ibk} \text{ with } \varepsilon_{ibk} \sim N\left(0, \sigma_{bk}^2\right) \\ &Note: d_{AA} = 0, k > b \text{ implies treatment } k \text{ comes alphabetically after } b \end{aligned} \tag{1}$$

where θ_{ik} is a continuous measure of the treatment effect in arm k of trial i (log odds in the case of a binary outcome), μ_{ib} is the effect of baseline treatment b (log odds) in trial i and ε_{ibk} denotes a random effect indicating that the trial-specific effects (log-odds ratios) of treatment k relative to b, δ_{ibk}, are normally distributed with mean d_{bk} and between-study variance σ_{bk}^2. The fundamental assumption underlying random-effects NMA is that the treatments effects are exchangeable across the entire network of trials regardless of whether or not treatments b and k are included in trial i [22, 24]. Validity of this assumption means that the pooled treatment effects, d_{bk}, can further be expressed as functions of basic parameters taken with reference to treatment A (i.e. $d_{bk} = d_{Ak} - d_{Ab}$) [25]. Effect estimates from trials with more than two treatment groups will be correlated through sharing a common comparator treatment. The correlation is taken into account by assuming homogenous variances $\left(\sigma_{bk}^2 = \sigma^2\right)$ so that the covariance is equal to $\sigma^2/2$ [23]. Alternatively, heterogeneous variance models have also been proposed [26]. Modelling is conducted within the framework of Bayesian analysis using MCMC simulation through the WinBUGS software [27, 28] with minimally informative prior distributions specified for d_{Ak}, μ_{ib} and σ.

3.2. Extending the network meta-analysis to include a covariate for the baseline risk

Using the 'true' but unobserved non-active control or placebo group log odds μ_{iA} (i.e. for b = treatment A) in trial i as a measure of the baseline risk, the trial-specific treatment effects in Equation (1) can be made to depend on the baseline risk through the following regression:

$$\begin{aligned} &\delta_{ibk} = d_{bk} + \beta_{bk}\left(\mu_{iA} - \bar{\mu}\right) + \varepsilon_{ibk}; \quad \varepsilon_{ibk} \sim N\left(0, \sigma_{bk}^2\right) \\ &Note: d_{AA}, \beta_{AA} = 0 \end{aligned} \tag{2}$$

where δ_{ibk} and σ_{bk}^2 are defined as in Equation (1), d_{bk} is the mean effect of treatment k relative to baseline treatment b adjusted for the baseline risk and β_{bk} is the change in the log-odds ratio of an event

Copyright © 2012 John Wiley & Sons, Ltd.

per unit change in the baseline risk for treatment k relative to b at the mean baseline risk across trials. We centred the baseline risk covariate on $\bar{\mu}$, the observed mean log odds in the non-active control group (treatment A), to improve convergence of the model [29]. For trials with an active treatment control (baseline treatment $b \neq A$), we make use of the substitution $d_{bk} = d_{Ak} - d_{Ab}$ under consistency of evidence arising from the exchangeability assumption [25] to express Equation (2) as

$$\begin{aligned} \delta_{ibk} &= (d_{Ak} - d_{Ab}) + (\beta_{Ak} - \beta_{Ab}) \times (\mu_{iA} - \bar{\mu}) + \varepsilon_{ibk} \\ \varepsilon_{ibk} &\sim N\left(0, \sigma_{bk}^2\right) \end{aligned} \tag{3}$$

All variables in Equation (3) have the same interpretation as in the previous equations. Although treatment A is not actually included in trial i of Equation (3), the fundamental assumption on exchangeability means that treatment arms can be assumed to be missing at random without loss to efficacy [25,30]. This allows us to imagine that there would still be a baseline risk in trials without treatment A and, thus, borrow strength from other trials. Therefore, no new parameters are needed for including for example a B versus C trial, and all other aspects of the model will remain the same. For multi-arm trials, the model takes the form of a multivariate regression to accommodate the within-study correlations between effect estimates arising from such trials. The multivariate form of Equation (2) with bold characters denoting vectors and matrices is given by

$$\boldsymbol{\delta_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{\varepsilon_i}; \quad \boldsymbol{\varepsilon_i} \sim MVN\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \tag{4}$$

where $\boldsymbol{\delta_i}$ with elements $\delta_{i,1}\delta_{i,2}\cdots\delta_{i,NT_i-1}$ for trial i is now a vector of relative effects (e.g. log-odds ratios), NT_{i-1} is the total number of treatment effects in trial i, $\boldsymbol{\varepsilon_i}$ is a vector of random effects associated with trial i and $\boldsymbol{\Sigma}$ is a variance–covariance matrix (as defined for the network in Equation (5)). The design matrix $\boldsymbol{X_i}$ contains the covariate information with entries indicating the treatment effects being estimated in trial i, and $\boldsymbol{\beta}$ is a vector of regression coefficients including the intercept and slope terms [31]. Following Salanti *et al.* [13] as an example, consider a network of four trials with three treatments A, B and C in which trial 1 is AB (i.e. A versus B), trial 2 is AC, trial 3 is ABC and trial 4 is BC (i.e. no non-active control). With treatment A taken as the overall baseline treatment, assuming homogenous variances $\left(\text{i.e.}\sigma_{bk}^2 = \sigma^2\right)$, we can write Equation (4) in full for this network as

$$\begin{pmatrix} \delta_{1,AB} \\ \delta_{2,AC} \\ \delta_{3,AB} \\ \delta_{3,AC} \\ \delta_{4,BC} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \mu_{1A} - \bar{\mu} & 0 \\ 0 & 1 & 0 & \mu_{2A} - \bar{\mu} \\ 1 & 0 & \mu_{3A} - \bar{\mu} & 0 \\ 0 & 1 & 0 & \mu_{3A} - \bar{\mu} \\ -1 & 1 & -(\mu_{4A} - \bar{\mu}) & \mu_{4A} - \bar{\mu} \end{pmatrix} \begin{pmatrix} d_{AB} \\ d_{AC} \\ \beta_{AB} \\ \beta_{AC} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,AB} \\ \varepsilon_{2,AC} \\ \varepsilon_{3,AB} \\ \varepsilon_{3,AC} \\ \varepsilon_{4,BC} \end{pmatrix}; \tag{5}$$

$$\begin{pmatrix} \varepsilon_{1,AB} \\ \varepsilon_{2,AC} \\ \varepsilon_{3,AB} \\ \varepsilon_{3,AC} \\ \varepsilon_{4,BC} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & \sigma^2/2 & 0 \\ 0 & 0 & \sigma^2/2 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix} \right)$$

where $\boldsymbol{\beta} = (d_{AB}\, d_{AC}\, \beta_{AB}\, \beta_{AC})^{\mathrm{T}}$ is the 4×1 matrix of regression coefficients representing the pooled effects of treatments B and C relative to treatment A and the effect of baseline risk on treatment effect estimates. All that remains is to specify models for the distribution of the 'true' baseline risk across trials and distribution of the regression coefficients. These are presented in the next section.

3.3. Models for the baseline risk and treatment × covariate interactions

As stated in the review of previous models (Section 2), there is no consensus in the literature about what form of distribution the baseline risk should take. We follow the example of Arends *et al.* [11] to specify models from three different assumptions about the distribution of the baseline risk as follows:

1. Model 1 assumes that baseline risk is independent or unconstrained so that each trial has its own baseline risk measure. This is equivalent to specifying a vague normal prior distribution for the baseline risk across trials: $\mu_{i,A} \sim N(0, 10^3)$
2. Model 2 assumes that the baseline risk across trials is drawn from a normal distribution with common mean and between-study variance: $\mu_{i,A} \sim N\left(\bar{\mu}, \sigma_{\mu}^2\right)$. Prior distributions are specified for $\bar{\mu}$ and σ_{μ}: $\bar{\mu} \sim N(0, 10^3)$ and $\sigma_{\mu} \sim \text{Uniform}(0, 100)$

Copyright © 2012 John Wiley & Sons, Ltd.

3. Model 3 assumes that the baseline risk is drawn from a mixture of two normal distributions with a common between-study variance: $\mu_{i,A} \sim p_1 N\left(\bar{\mu}_1, \sigma_\mu^2\right) + (1 - p_1) N\left(\bar{\mu}_2, \sigma_\mu^2\right)$ with prior distributions $\bar{\mu}_1, \bar{\mu}_2 \sim N(0, 10^3)$, $\sigma_\mu \sim \text{Uniform}(0, 100)$;

$$p_1 \sim \text{ddirch}(\alpha_c = 1); c = 1, 2$$

Similar to the models for the distribution of the baseline risk, we also specify models for the interaction terms from the following three assumptions described by Cooper *et al.* [5]:

A. Common effect treatment × covariate interactions: $\beta_{AK} = B; B \sim N(0, 10^3)$
B. Exchangeable treatment × covariate interactions: $\beta_{AK} \sim N\left(B, \sigma_B^2\right)$; $B \sim N(0, 10^3)$ and $\sigma_B \sim \text{Uniform}(0, 100)$ and
C. Independent and unrelated treatment × covariate interactions: $\beta_{AK} \sim N(0, 10^3)$

Therefore, a total of nine models are fitted on the basis of the combination of assumptions about distribution of the baseline risk and the treatment × covariate interactions (slopes):

Model A1: Unconstrained baseline risk and common slope
Model A2: Normal distribution for baseline risk and common slope
Model A3: Mixture distribution for baseline risk and common slope
Model B1: Unconstrained baseline risk and exchangeable slopes
Model B2: Normal distribution for risk and exchangeable slopes
Model B3: Mixture distribution for baseline risk and exchangeable slopes
Model C1: Unconstrained baseline risk and independent slopes
Model C2: Normal distribution for baseline risk and independent slopes
Model C3: Mixture distribution for baseline risk and independent slopes

3.4. Goodness of fit and model selection

In the applications that follow, adequacy of model fit to the data was assessed through the residual deviance where a model is judged to adequately fit the data if the residual deviance closely matches the actual number of unconstrained data points available [27]. The overall goodness of fit and model selection criteria were based on the deviance information criteria, a measure of model fit that penalises model complexity [27], where the model with the lowest deviance information criteria is generally preferred and differences of 3 or 5 are considered significant [32].

4. Application examples

4.1. Example 1: functional smoke alarm data

The data come from a published NMA [17] and consist of 20 randomised and non-randomised studies that evaluated the effectiveness of home safety education to increase ownership of functioning smoke alarm (FSA) systems in households with children. The outcome of interest is whether or not a household had an FSA. Thus, each study supplied arm-level data on the number of households with an FSA and the total number of households surveyed. We used the FSA data here to illustrate application of the method to binary outcome data. Table AI of Appendix A displays the full data, with Figure 1A displaying a network diagram for the seven interventions and 40 data points from the 20 studies. The baseline or non-intervention arm is the usual-care intervention. Seven of the 20 studies did not have a usual-care intervention and therefore had no baseline risk covariate. [Note: In the data-coding step, these seven studies are included by using NA to represent missing information on the number of events in the usual-care arm (see Appendix C data for WinBUGS)]. Baseline functioning smoke alarm ownership in the remaining 13 studies ranged from about 3% to about 96%. We also detected evidence of significant inconsistency [17] using the method of node splitting [25]. Hence, it is of interest to know whether baseline differences in FSA ownership across studies can explain the heterogeneity and inconsistency. For this example, where the outcome is binary, we assumed a binomial likelihood for the arm-level data and fitted NMA without covariate adjustment (model 0) on the basis of the model defined by Equation (1). We then investigated the relationship between intervention effect and baseline FSA ownership using the methods described in Section 3.2. The covariate was centred on the observed mean baseline log odds of

Copyright © 2012 John Wiley & Sons, Ltd.

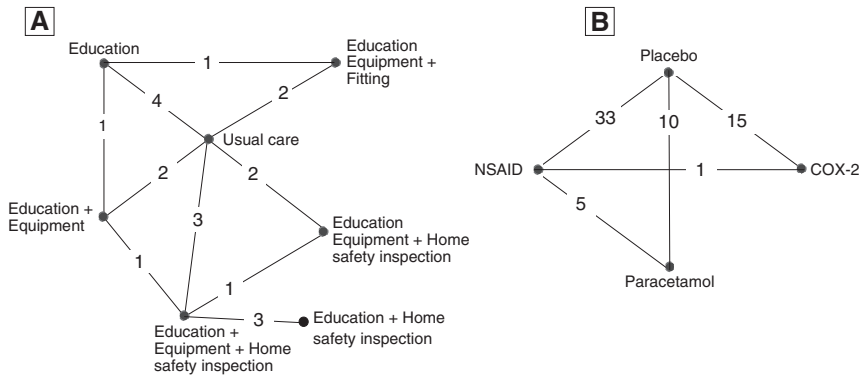


Figure 1. Intervention network for (A) possession of functional smoke alarm and (B) post-operative pain relief.

0.81 (calculated outside WinBUGS) for FSA ownership in the 13 studies with a usual-care arm. In total, 10 models were fitted (the nine models described in Section 3.3 in addition to the unadjusted model) using MCMC simulation in the WinBUGS software [27]. A modified version of the NMA code from Dias *et al.* [15] was used and is given in Appendix B. The following prior distributions were used and intended to be minimally informative:

$$\sigma, \sigma_{\mu} \text{ and } \sigma_{B} \sim \text{Uniform}(0, 100)$$

$$\beta_{Ak}, B, d_{Ak}, \mu_{ib}, \bar{\mu} \sim N(0, 10^3)$$

Models were run for 100 000 iterations, discarding the first 30 000 iterations as burn-in samples to ensure convergence of the MCMC sampler. There was evidence of poor convergence for the models that assumed separate/independent treatment × covariate interactions (models C1, C2 and C3). This may be due to the following: (i) seven out of 20 studies not having a usual-care intervention arm, and (ii) hence, relatively few data points compared with the number of parameters needing to be estimated [5]. Therefore, parameter estimates from models C1 to C3 are not presented in the results in Section 5.1.

4.2. Example 2: pain relief data

The second dataset consists of 56 RCTs with 116 data points from a published Health Technology Assessment report [18]. The report examined the effectiveness of three non-opioid analgesics [paracetamol, non-steroidal anti-inflammatory drugs (NSAIDs) or cyclooxygenase 2 (COX-2) inhibitors] and placebo in reducing morphine consumption following major surgery in adults. The outcome of interest is the amount of morphine in milligrammes consumed over a 24-h period (continuous outcome). Each study provided arm-level information on the number of patients together with the mean 24-h morphine consumption and its standard deviation (SD). The treatment network is given in Figure 1B. The dataset is presented in Table AII of Appendix A, in order of increasing mean morphine consumption in the placebo group (baseline risk). Two trials have no placebo, and four of the trials are three-arm studies. There is considerable variability in the 24-h morphine consumption in the placebo arm of trials ranging from a low of 8.6 mg (SD 5.2 mg) to a high of 142 mg (SD 80 mg). The average across the placebo group is 45.26 mg. Therefore, a sensitivity analysis was conducted in the original report [18] to investigate the effect of this baseline imbalance in morphine use on the treatment effect estimates. To include the two studies that did not have placebo, the original analysis in the published report was first carried out without these studies to derive an estimate of baseline morphine consumption for the two trials. The derived estimates were then included in the sensitivity analysis that adjusted for the baseline morphine use. In our analysis, however, we make use of the exchangeability assumption mentioned earlier [Equation (3)] to include the trials without a placebo arm and thus baseline risk measure. [Note: In the data-coding step, these two studies are included by using NA to represent missing mean morphine consumption and 1 for its standard error (see data for WinBUGS in Appendix D)].

Copyright © 2012 John Wiley & Sons, Ltd.

Because 24-h morphine consumption is a continuous outcome, we replace the binomial likelihood and logistic regression model in Equation (1) with a normal distribution for the observed arm-specific outcome (i.e. mean 24-h morphine), $\hat{y}_{ik}$ in treatment arm k of trial i:

$$\begin{aligned} &\hat{y}_{ik} \sim N\left(\theta_{ik}, S_{ik}^2\right), \qquad i = 1, 2, \cdots, 56 \text{ and } k = 1, 2, \cdots, 4 \\ &\theta_{ik} = \begin{cases} \mu_{ib} & k = b; \quad b \in \{1, 2, 3, 4\} \\ \mu_{ib} + \delta_{ibk} & k > b; \quad b \in \{1, 2, 3, 4\} \end{cases} \\ &\delta_{ibk} = d_{bk} + \beta_{bk} \times (\mu_{i1} - \bar{\mu}) + \varepsilon_{ibk} \text{ with } \varepsilon_{ibk} \sim N(0, \sigma^2) \end{aligned} \tag{6}$$

where θ_{ik} is the 'true' unobserved mean morphine consumption in treatment arm k of trial i with variance S_{ik}^2 assumed known but estimated from the data [12]. We centred the baseline morphine consumption μ_{i1} on 45.26 mg, the average consumption across the placebo arms. All other aspects of the modelling assumptions and model fit remain the same as in example 1 including the specification of minimally informative prior distributions as follows:

$$\sigma, \sigma_\mu \text{ and } \sigma_B \sim \text{Uniform}(0, 100)$$

$$\beta_{bk}, B, d_{bk}, \mu_{ib}, \bar{\mu} \sim N(0, \ 10^3)$$

The MCMC simulations were run using WinBUGS for 50 000 iterations, discarding the first 20 000 iterations as burn-in samples to ensure convergence. The results are presented in Section 5.2.

5. Results

5.1. *Example 1: functioning smoke alarm model*

In this example, the log-odds ratio was regressed on the 'true' control group log odds (usual-care intervention) taken as a measure of baseline risk. Table II displays estimates of the residual heterogeneity σ, treatment × covariate interactions (regression slopes) and model fit statistics excluding the three models (C1, C2 and C3), which showed evidence of non-convergence. Firstly, different assumptions about the distribution of the baseline risk did not seem to greatly affect estimates of the treatment × covariate interaction terms in this case. The slope of the regression lines is slightly steeper when minimally informative prior distribution was assumed for the baseline risk (models A1 and B1) than in models that assumed a normal baseline distribution (model A2) or a mixture of two normal distributions (model B2). Secondly, the posterior credible intervals for the slope terms included 0 in all models, indicating that none of these is statistically significant. Therefore, baseline imbalance in smoke alarm distribution across studies was not significantly related to effectiveness of home safety education to promote FSA ownership in households with children (provided this analysis is powered appropriately for effects under investigation). Consequently, the heterogeneity and also the inconsistency were not significantly reduced in all models that adjusted for the baseline risk compared with the unadjusted model (Table II).

When we use the posterior mean residual deviance as a measure of model fit to the data (Table II), both adjusted and unadjusted models predicted values close to the 40 unconstrained data points in the FSA data, indicating that these models fit the data equally well. Because baseline risk appears to be unrelated to intervention effect, there was very little difference to choose between these models, and we report only the results from the common slope or treatment × covariate interaction models for convenience. Posterior median estimates of the slope are −0.08 (95% Credible Interval (CrI); −0.41 to 0.28) from model A1, −0.03 (95% CrI; −0.41 to 0.35) from model A2 and −0.03 (95% CrI; −0.39 to 0.34) from model A3, which all indicate non-significant decrease in intervention effectiveness with increasing baseline FSA ownership.

5.2. *Example 2: pain relief model*

For pain relief data, the treatment effect, expressed as the mean difference in 24-h morphine use, was regressed on the 'true' but unobserved 24-h mean morphine consumption in the placebo group (taken as baseline risk measure). There were no problems with convergence of the MCMC simulations, and all nine models described in Section 3.3 were fitted in addition to the unadjusted model. We present parameter estimates of interest and model fit statistics in Table III. Firstly, estimates of the regression slopes

Copyright © 2012 John Wiley & Sons, Ltd.

Table II. Network meta-analysis with baseline risk adjustment applied to functional smoke alarm data.

Regression slopes	Model 0: unadjusted model Median (95% CrI)	Model A1: unconstrained baseline; common slope Median (95% CrI)	Model A2: baseline normally distributed; common slope Median (95% CrI)	Model A3: baseline mixture of l; two norma common slope Median (95% CrI)	Model B1: unconstrained baseline; exchangeable slopes Median (95% CrI)	Model B2: baseline normally distributed; exchangeable slopes Median (95% CrI)	Model B3: mixture model; exchangeable slopes Median (95% CrI)
Common β	—	−0.08(−0.41, 0.28)	−0.03 (−0.41, 0.35)	−0.03 (−0.39, 0.34)	—	—	—
Education (β_2)					−0.13 (−0.42, 0.22)	−0.09 (−0.39, 0.25)	−0.09 (−0.40, 0.27)
Education + equipment (β_3)					0.19 (−0.59, 1.38)	0.26 (−0.57, 1.54)	0.25 (−0.57, 1.50)
Education + equipment + HSI (β_4)					−1.08 (−2.75, 0.20)	−1.20 (−2.81, 0.16)	−1.16 (−2.63, 0.28)
Education + equipment + fitting (β_5)					0.26 (−0.32, 1.05)	0.35 (−0.25, 1.20)	0.35 (−0.28, 1.22)
Education + HSI (β_6)					−0.07 (−3.34, 3.02)	−0.07 (−3.45, 3.08)	−0.08 (−3.55, 2.54)
Education + equipment + fitting + HSI (β_7)					0.09 (−1.69, 2.48)	0.19 (−2.01, 2.66)	0.165 (−1.98, 2.38)
Mean random effects β	—	—	—	—	−0.09 (−1.55, 1.28)	−0.07 (−1.58, 1.42)	−0.07 (−1.63, 1.30)
Residual heterogeneity σ	0.77 (0.34, 1.47)	0.83 (0.39, 1.56)	0.84 (0.40, 1.59)	0.84 (0.40, 1.59)	0.59 (0.16, 1.35)	0.57 (0.167, 1.30)	0.59 (0.14, 1.38)
SD for random effects β	—	—	—	—	0.88 (0.07, 3.31)	1.02 (0.12, 3.45)	0.97 (0.09, 3.41)
Model fit statistics							
Residual deviance ($\overline{D}$)	41.72	41.49	40.86	40.99	40.33	39.98	40.04
Effective number of parameters	35.28	35.85	35.70	36.65	35.53	34.97	35.091
Deviance information criteria	77.00	77.34	76.56		75.86	74.95	75.131

Estimates of residual heterogeneity σ and regression coefficients; β_k of intervention k relative to usual care (control intervention) measuring the relationship between intervention effect (log-odds ratio) and baseline risk (control group log odds).

β_2, interaction term for education relative to usual care; CrI, Credible interval; HSI, home safety inspection; SD, standard deviation in treatment effect estimate.

Copyright © 2012 John Wiley & Sons, Ltd.

Table III. Network meta-analysis with baseline risk adjustment applied to pain relief data.

Regression slopes	Model 0: unadjusted model Mean (95% CrI)	Model A1: unconstrained baseline; common slope Mean (95% CrI)	Model A2: baseline normally distributed; common slope Mean (95% CrI)	Model A3: baseline mixture of two normal; common slope Mean(95% CrI)	Model B1: unconstrained baseline; exchangeable slopes Mean(95% CrI)
Common β		−0.34 (−0.41, −0.27)	−0.31 (−0.38, −0.23)	−0.31 (−0.38, −0.23)	—
Paracetamol (β_2)					−0.22 (−0.39, 0.00)
NSAIDs (β_3)					−0.36 (−0.44, −0.28)
COX-2 (β_4)					−0.34 (−0.450, −0.18)
Random effects mean β					−0.30 (−0.93, 0.34)
Residual heterogeneity σ	5.44 (4.50, 5.98)	3.19 (2.15, 4.47)	3.19 (2.14, 4.51)	3.22 (2.15, 4.56)	3.20 (2.15, 4.49)
SD for random effects β	—	—	—	—	0.35 (0.01, 2.18)
Model fit statistics					
Residual deviance ($\overline{D}$)	124	119.5	121.90	121.1	117.60
Effective number of parameters	90.63	84.11	81.97	82.31	85.27
Deviance information criteria	214.63	202.61	203.87	203.41	202.87

Estimates of the residual heterogeneity σ and the regression coefficients; β_k of intervention k relative to usual care (control intervention) measuring the relationship between treatment effect (mean difference in 24-h morphine use in milligrammes) and placebo group morphine use in milligrammes.
COX-2, cyclooxygenase 2; CrI, credible interval; NSAIDs, non-steroidal anti-inflammatory drugs; SD, standard deviation.

from the nine adjusted models were all negative, suggesting evidence of increasing treatment effect with increasing baseline morphine consumption. Estimate of the common regression slope is −0.34 (95% CI; −0.41 to −0.27) for unconstrained baseline model (model A1) and −0.31 (95% CI; −0.38 to −0.23) for models with normal distribution (model A2) and a mixture of two normal distributions (model A3) for baseline risk. Similar estimates of the relationship between treatment effect and baseline risk were also obtained from the independent and exchangeable slope models, but only the effects of NSAIDs and COX-2 are statistically significant. Again, the three modelling assumptions about the distribution of the baseline risk seem to have very little impact on treatment × covariate interactions. Figure 2 plots treatment effects versus baseline 24-h morphine use from the model with independent/separate slopes (model C1) for paracetamol, NSAIDs and COX-2. The plot shows the following: (i) evidence of increasing effectiveness with increasing baseline morphine use for all three classes of analgesics; (ii) NSAIDS and COX-2 are increasingly more effective than paracetamol at higher baseline morphine use; and (iii) little difference between NSAIDs and COX-2. The vertical distance between the line of no effect and each treatment regression line gives an estimate of the treatment effect relative to placebo at a given baseline morphine consumption. Similarly, the relative effectiveness of any two analgesics at a given baseline morphine consumption can be obtained from the plot as the vertical distance between the two regression lines. Secondly, adjusting for the baseline risk reduced the residual heterogeneity and improved the overall model fit. From Table III, the posterior mean estimate of the residual heterogeneity σ is 5.44 mg (95% CI; 4.5 to 5.98) in the unadjusted model and 3.48 mg (95% CI; 3.24 to 4.57) in model C1, the adjusted model with the least reduction in heterogeneity. Compared with the unadjusted models, model C1 has at least a 40% reduction in between-study heterogeneity.

Copyright © 2012 John Wiley & Sons, Ltd.

Table III. *Continued.*

Regression slopes	Model B2: baseline normally distributed; exchangeable slopes Mean(95% CI)	Model B3: mixture model; exchangeable slopes Mean(95% CI)	Model C1: unconstrained baseline; separate slopes Mean(95% CI)	Model C2: baseline normally distributed; separate slopes Mean(95% CI)	Model C3: mixture model; separate slopes Mean(95% CI)
Common β	—	—	—	—	—
Paracetamol (β_2)	−0.18 (−0.35, 0.04)	−0.19 (−0.36, 0.01)	−0.16 (−0.36, 0.04)	−0.15 (−0.34, 0.06)	−0.13 (−0.34, 0.10)
NSAIDs (β_3)	−0.34 (−0.42, −0.25)	−0.33 (−0.42, −0.25)	−0.36 (−0.45, −0.28)	−0.34 (−0.43, −0.26)	−0.34 (−0.43, −0.26)
COX-2 (β_4)	−0.27 (−0.43, −0.09)	−0.27 (−0.42, −0.09)	−0.35 (−0.53, −0.19)	−0.25 (−0.44, −0.06)	−0.26 (−0.44, −0.05)
Random effects mean β	−0.26 (−1.00, 0.47)	−0.27 (−0.98, 0.48)			
Residual heterogeneity σ	3.13 (2.06, 4.50)	3.13 (2.04, 4.50)	3.28 (2.20, 4.57)	3.16 (2.06, 4.51)	3.20 (2.06, 4.53)
SD for random effects β	0.39 (0.01, 2.37)	0.36 (0.01, 2.32)	—	—	—
Model fit statistics					
Residual deviance ($\overline{D}$)	121.10	120.30	116.40	120.60	119.70
Effective number of parameters	82.58	82.57	85.56	82.58	82.96
Deviance information criteria	2	202.87	201.964	203.18	202.66

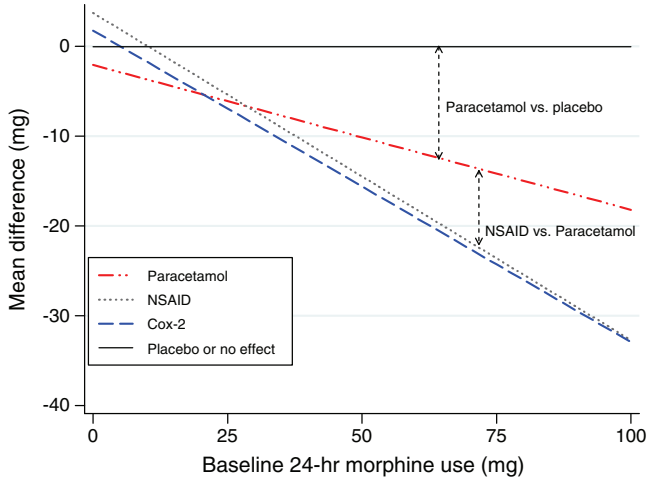


Figure 2. Network meta-analysis adjusting for baseline morphine use assuming independent slopes for different treatment effects and unconstrained baseline risk. COX-2, cyclooxygenase-2; NSAIDs, non-steroidal anti-inflammatory drugs.

6. Discussion

We have shown how methods for baseline risk covariate adjustment can be extended from pairwise to NMA when it is of interest to account for differences in underlying risk across trial populations. This

Copyright © 2012 John Wiley & Sons, Ltd.

type of analysis can help identify potential treatment effect modifiers, which may give rise to heterogeneity in effect estimates and/or inconsistency in the direct and indirect evidence on pairwise contrasts in a network of trials. The pain relief example shows how adjusting for baseline risk can greatly reduce heterogeneity and improve overall model fit. Similar results and conclusions have been reported before, for example, by Lu *et al.* [33] in an NMA at multiple follow-up times where the baseline effects were adjusted at different follow-up points. However, there was no evidence of baseline effect in the FSA example, and the inconsistencies identified by Cooper *et al.* [17] were not resolved by baseline risk adjustment. In meta-analyses of studies evaluating complex and or public health interventions such as the FSA data, interventions may not always be clearly defined, and studies are often of variable quality and conducted in populations with different characteristics. These factors can introduce heterogeneity in both meta-analyses of clinical trials and studies of complex and or public health interventions, but the problem is more pronounced in public health. The FSA network included both RCTs and non-randomised observational studies, both of which are of variable quality. Although care was taken to categorise the interventions appropriately, 'lumping' of interventions within categories could not be completely ruled out [17]. Lumping of interventions creates relative contrasts that are unevenly distributed across contrast and have been cited as a possible source of heterogeneity and inconsistency in NMA [34].

The main advantage of the approach described in this paper is that the models can be easily implemented by making simple modifications to freely available WinBUGS code for NMA [35] (see code in Appendix B). Specifying the models in WinBUGS, and analysing them using MCMC simulation, is beneficial as it allows the adjustment to be carried out without excluding trials with missing placebo or no treatment control group (hence no baseline risk covariate). The imputation step is implemented automatically in WinBUGS through the model jointly specified by the likelihood and prior distribution placed on the 'baseline risk' (described in Section 3.3). Because parameters are considered as random variables within the Bayesian framework requiring a distribution [36], the 'missing covariate' is treated as any other unknown parameter to be estimated under exchangeability [37, p. 117]. Alternatively, the analysis can also be carried outside a Bayesian framework using multivariate meta-analysis methods (for example, [16]) fitted in standard statistical software or self-written programs. However, validity of the results obtained from either classical or Bayesian analyses will depend on appropriateness of the assumption that the non-active intervention arm of studies without a baseline risk is missing at random.

Fitting models with separate and or exchangeable regression slopes described by Cooper *et al.* [5], in addition to the common slope model, can be useful for assessing the appropriateness of these assumptions. For example, the common slope assumption can be tested by first calculating the difference between estimates of any two slopes in the separate slope model followed by a *probability* that this difference is greater than 0 using the step function in WinBUGS [27]. A two-sided P-value can then be derived using the formula $2 \times \text{minimum}(\textit{probability}, 1 - \textit{probability})$ [25]. However, as shown by the FSA example, fitting models with separate/independent slopes may not always be feasible, possibly because of limited availability of data. In those circumstances, the exchangeable slope or even common slope models can be considered as a compromise [5]. Under the exchangeable regression slope assumption, power is improved by borrowing strength across regression slopes, which shrinks treatment effect estimates towards each other. This can have policy implications especially in a decision-making context where manufacturers of alternative interventions may see the effectiveness of their products 'shrink' towards that of the competitor. Also, the exchangeable slope assumption can reduce heterogeneity in the effect estimates (σ), but the regression slopes themselves can be quite variable as illustrated by the pain relief example where the σ_B's are larger than σ. This shows that the regression slopes are much more variable than the treatment effects, and therefore, a common regression coefficient may not be the best model for this example.

Finally, going back to the review of pairwise meta-analysis models presented in Section 2, a much debated issue in modelling the relationship between treatment effect and baseline risk has been whether or not to assume a parametric distribution for the baseline risk and what form if any such a distribution should take. Ghidey *et al.* [19] examined the issue in a recently published methods review paper using real and simulated data for pairwise meta-analysis. The simulated results found no difference between models that assumed normality for the baseline risk and those that did not, with both models producing robust/unbiased estimates of the regression slope when the baseline risk is normally distributed across studies [19]. However, the estimate of the regression slope was found to be less biased under the functional modelling approach when normality of the baselines was violated, but the relative difference in bias was small. The results from the approach outlined in this paper for NMA appear consistent with

Copyright © 2012 John Wiley & Sons, Ltd.

the findings of Ghidey *et al.* [19] and also Arends *et al.* [11]. Estimates of regression slopes from both FSA and pain relief examples were slightly less negative and tended to shrink towards 0 in models that assumed normally distributed baselines (models A2, A3, B2 and B3 in Table II and models A2–A3, B2–B3 and C2–C3 in Table III) compared with the unconstrained or minimally informative prior distributions for the baseline risk (models A1–C1). The effect of different distributional assumptions about the baseline risk was, however, very minimal as both unconstrained and normally distributed baseline risk models produced practically identical estimates of regression slopes.

Appendix A

Table AI. Example 1 data: effectiveness of home safety education to promote possession of a functional smoke alarm.

Study	Comparison (control versus intervention)	No. of smoke alarms/total no. households surveyed (control arm versus intervention arm)	Proportion with smoke alarm in usual-care group (baseline risk) (%)
Mock 2003	Usual care versus education	10/297 vs. 18/308	3
DiGuiseppi 2002	Usual care versus education + equipment + fitting	5/30 vs. 8/44	17
Miller 1982	Usual care versus education + equipment	46 (9.34)/105 (21.31) vs. 61 (12.38)/108 (21.92)	44
Sangvai 2007	Usual care versus education + equipment + HSI	5/10 vs. 16/17	50
Hendrickson 2002	Usual care versus education + equipment + HSI	26/40 vs. 37/38	65
Schwrz 1993	Usual care versus education + equipment + fitting + HSI	816/1060 vs. 866/902	77
Bulzacchelli 2009	Usual care versus education	55/71 vs. 109/139	77
Phelan 2010	Usual care versus education + equipment + fitting + HSI	112/138 vs. 130/140	81
Watson 2005	Usual care versus education + equipment + fitting	619/737 vs. 692/764	84
Clamp 1998	Usual care versus education + equipment	71/82 vs. 81/83	87
Gielen 2005	Usual care versus education	325/375 vs. 345/384	87
Kendrick 1999	Usual care versus education + equipment + HSI	321 (245.62)/363 (277.76) vs. 325 (248.68)/361 (276.23)	88
Gielen 2001	Usual care versus education	54 (52.02)/56 (53.95) vs. 77 (74.18)/80 (77.07)	96
Matthews 1988	Education + equipment + HSI versus education + HSI	6/12 vs. 6/12	NA
Johnston 2000	Education + equipment + HSI versus education + HSI	211 (20.05)/211 (21.15) vs. 136 (31)/133	NA
Gielen 2002	Education + equipment versus education + equipment + HSI	47 (44.20)/56 (52.66) vs. 47 (44.21)/58 (54.54)	NA
Barone 1988	Education versus education + equipment	34 (20.08)/38 (22.45) vs. 39 (23.04)/41 (24.22)	NA
Sznaider 2003	Education versus education + equipment + fitting	6/50 vs. 27/47	NA
King 2001	Education + equipment + HSI versus education + HSI	394/469 vs. 406/482	NA
Harvey 2004	Education + equipment + HSI versus education + equipment + fitting + HSI	997/1545 vs. 1421/1583	NA

Figures in brackets adjusted for clustering; adjusted figures used in analysis; usual-care arm not available.
HSI, home safety inspection; equipment, low cost/free equipment.

Copyright © 2012 John Wiley & Sons, Ltd.

Table AII. Example 2 data: reduction in 24-h post-operative morphine consumption.

Study author	Placebo (*N*/mean/SD)	Paracetamol (*N*/mean/SD)	NSAID (*N*/mean/SD)	COX-2 (*N*/mean/SD)
Inan 2007	20/8.55/5.18		20/5.4/4.3	
Cakan 2008	20/12.45/7.02	20/11.25/8.42		
Peduto 1998	10/14.90/15.1		10/17.4/12.7	
Fogarty Mack 2001	29/17.40/8.8			30/12.6/6.5
Cheng 2004	15/19.50/8.3		15/16.9/6.5	
Xuerong 2008	47/20.10/12.8	42/12.1/9.9		
Kvalsvik 2003	30/21.10/11	30/16.8/8.4		
Varrassi 1994	47/21.70/19.88		48/15/13.2	
Cassinelli 2008	12/22.10/18		13/8/7.5	
Fong 2008	20/27/7.2			40/12.5/5.76
Karaman 2006	20/29.7/3.8		40/23/3.41	
Hsu 2003	48/30.8/19.4		45/20.9/14.9	
Munro 1998	19/30.9/22.6		18/17.4/15.5	
Riest 2008	80/31.3/21.8			240/25.93/20.73
Trampitsch 2003	22/31.6/3.91		44/28.33/4.24	
Fletcher 1997	15/32.9/25.2	15/28/20.3	15/25.7/17	
Durmus 2003	20/34.9/10.35			20/25.6/5.92
El-Halafawy 2004	30/35.5/12.6			30/25.5/8.3
De Decker 2001	15/36.5/20.3		45/23.53/15.82	
Chau-in 2008	15/36.6/8.9			34/26.8/10.42
Fayaz 2004	20/37/15		17/27/12	
Plummer 1996	49/38/20		55/32/18	
Thompson 2000	18/38.2/20.8		18/33.2/16.9	
Vandermeulen 1997	256/39.2/27.6		258/34.6/25.9	
Delbos 1995	30/43.1/15.9	30/34.5/12.7		
Hernandez-Palazon 2001	21/43.3/15.3	21/26/12.2		
Hubbard 2003	63/43.5/18.7			126/33.97/17.77
Ready 1994	45/44/26		96/31.96/23.23	
Siddiqui 2008	100/44.2/8.2			100/35.1/7
Colquhoun 1989	15/44.8/24		15/44.6/20.7	
Jirarattanaphochai 2008	60/45.2/21			60/28/14.1
Martinez 2007	21/47/27			41/25.54/12.32
Owen 1986	31/48.2/25.1		29/39.1/17.1	
Rao 2000	19/51/23.86		20/33/16.03	
Tang 2002	18/51/27			37/33.51/20.21
Alexander 2002	32/51.6/22.2		67/41.34/27.11	
Rowe 1992	14/51.6/28.1		13/36.3/23.82	
Cobby 1999	21/54.9/28.3	24/35/20.4	20/32.7/27.4	
Blackburn 1995	29/55/22		30/43/17	
Hegazy 2003	15/55.1/12		15/36.6/9	
Moodie 2008	41/56.5/30.73		82/46.05/37.48	
Sinatra 2005	52/57.4/52.3	99/39.56/32.57		
Malan 2003	65/57.5/31.83			116/40.35/35.62
Balestrieri 1997	66/58.1/24.9		133/44.01/24.26	
Sevarino 1992	12/58.75/58.89		23/30/45.6	
Hodsman 1987	31/59/27.84		31/38/22.27	
Schug 1998	26/59.5/42.3	25/50.3/40.1		
Etches 1995	78/64.2/38.6		79/39.6/26.7	
Siddik 2001	20/66.7/20	20/61.1/23	20/36/18	
Ng 2003	17/72/27.22			19/54/23.86
Gillies 1987	18/78/38.18		39/53.97/30.53	
Pertunnen 1992	15/80.4/43.37		15/32.4/25.17	
Celik 2003	20/93/6		20/63/6	
Lee 2008	18/141.5/74.9			38/73.56/59.42
Alhashemi 2006	NA	22/65/30	23/58/25	
Munishankar 2008	NA	24/54.5/28.5	25/44.1/24.4	

COX-2, cyclooxygenase 2; mean, mean 24-h morphine consumption (mg); *N*, sample size; NSAID, non-steroidal anti-inflammatory drug; SD, standard deviation.

Copyright © 2012 John Wiley & Sons, Ltd.

Appendix B. Modified WinBUGS code for baseline risk adjustment (taken from Dias *et al.*[15])

```
model{
for(i in 1:ns){
tmp1[i] <- id[i]
w[i,1] <-0
delta[i,1]<-0
bl[i,1]<-0                          #No covariate adjustment the non-intervention / control arm

#Baseline models
#mu[i] ~ dnorm(0,.0001)             #model 1: vague priors for trial baselines
mu[i] ~ dnorm(mu.mean,taumu)        #model 2:normal distribution for trial baselines
#mu[i] ~dnorm(lambda[T[i]],taumu)   #model 3:Two normal distribution for trial baselines
#T[i] ~ dcat(P[])                   #Categorical variable for mixture

for (k in 1:na[i]) {
#Model for binary outcome (example 1 (smoke alarm data)
r[i,k] ~ dbin(p[i,k],n[i,k])                                        #binomial likelihood
logit(p[i,k])<-mu[i] + delta[i,k]                                   #model
 rhat[i,k] <- p[i,k] * n[i,k]                                       #expected value of the numerators
dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))              #Deviance
+ (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))

#Model for continuous outcome (example 2)
#var[i,k] <- pow(se[i,k],2)                                                #calculate variances
#prec[i,k] <- 1/var[i,k]                                                   # set precisions
#y[i,k] ~ dnorm(theta[i,k],prec[i,k])                                      #Normal likelihood
#theta[i,k]<-mu[i] + delta[i,k] #+ bl[i,k]
#dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]             #Deviance contribution
}

sumdev[i]<-sum(dev[i, 1:na[i]])                        #summed residual deviance contribution for this trial
for (k in 2:na[i]) {
delta[i,k] ~ dnorm(md[i,k],taud[i,k])                        #trial-specific LOR distributions
md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]+ bl[i,k]         #mean of LOR distributions
taud[i,k] <- tau *2*(k-1)/k                                  #precision of LOR distributions

w[i,k] <- (delta[i,k]-((d[t[i,k]]-d[t[i,1]])+bl[i,k]))       #adjustment,multi-arm RCTs
sw[i,k] <-sum(w[i,1:k-1])/(k-1)                              #cumulative adjustment,multi-arm trials


#baseline risk adjustment
bl[i,k] <-(beta[t[i,k]]-beta[t[i,1]])*(mu[i]-m.mu)
}
}

ssumdev<-sum(sumdev[])#total residual deviance
d[1]<-0
beta[1]<-0
for (k in 2:nt){
d[k] ~ dnorm(0,.0001)                       #vague prior for basic parameters
beta[k] <- B                                #common covariate effect (Model A)
#beta[k] ~ dnorm(B,tauB)                    #exchangeable covariate effect (Model B)
#beta[k] ~ dnorm(0,0.0001)                  #independent covariate effect (Model C)
 }

tau<-1/pow(sd,2)                            #1/between-study variance (treatment effects)
sd~dunif(0,100)                             #vague prior for between-study standard deviation

mu.mean~ dnorm(0,0.0001)                    #vague prior for mean baseline log-odds
taumu<-1/pow(sdmu,2)                        #1/between-study variance (baseline log-odds)
sdmu~dunif(0,100)                           #vague prior for random effects standard deviation

B~ dnorm(0,0.0001)                          #vague prior for common covariate effect


#tauB<-1/pow(sdB,2)                        #1/between-study variance
#sdB~dunif(0,100)
}
```

Copyright © 2012 John Wiley & Sons, Ltd.

Appendix C. Functional smoke alarm data coding for WinBUGS (data with decimals are from cluster randomised trials and have been adjusted for clustering)

```
Intervention code
1=Usual Care
2=Education
3=Education +low cost/free equipment
4=Education + low cost/free equip + home safety inspection
5=Education + free equipment + fitting
6=Education + home safety inspection
7=Education + low cost/free equip + fitting + home safety inspection)

list(nt=7,ns=20,m.mu=0.81)
```

id[]	t[,1]	t[,2]	t[,3]	r[,1]	n[,1]	r[,2]	n[,2]	r[,3]	n[,3]	na[]
1	1	2	NA	10	297	18	308	NA	1	2
2	1	5	NA	5	30	8	44	NA	1	2
3	1	3	NA	9.33631	21.3111	12.3808	21.92	NA	1	2
4	1	4	NA	5	10	16	17	NA	1	2
5	1	4	NA	26	40	37	38	NA	1	2
6	1	7	NA	816	1060	866	902	NA	1	2
7	1	2	NA	55	71	109	139	NA	1	2
8	1	7	NA	112	138	130	140	NA	1	2
9	1	5	NA	619	737	692	764	NA	1	2
10	1	3	NA	71	82	81	83	NA	1	2
11	1	2	NA	325	375	345	384	NA	1	2
12	1	4	NA	245.619	277.757	248.68	276.226	NA	1	2
13	1	2	NA	52.02	53.95	74.18	77.07	NA	1	2
14	**1**	**4**	**6**	**NA**	**1**	**6**	**12**	**6**	**12**	3
15	**1**	**4**	**6**	**NA**	**1**	**20.0499**	**21.1515**	**31**	**31**	3
16	**1**	**3**	**4**	**NA**	**1**	**44.1995**	**52.6633**	**44.1995**	**54.5441**	3
17	**1**	**2**	**3**	**NA**	**1**	**20.0827**	**22.4454**	**23.036**	**24.2174**	3
18	**1**	**2**	**5**	**NA**	**1**	**6**	**50**	**27**	**47**	3
19	**1**	**4**	**6**	**NA**	**1**	**394**	**469**	**406**	**482**	3
20	**1**	**4**	**7**	**NA**	**1**	**997**	**1545**	**1421**	**1583**	3
END										

Initial Values example 2 (Functional smoke alarm data)

```
list(
B=0, d = c(NA,0,0,0,0,0,0), mu = c(0,0,0,0,0,  0,0,0,0,0,   0,0,0,0,0,  0,0,0,0,0), sd=1,
delta = structure(.Data = c(
NA,0,NA,NA,0, NA,NA,0,NA,NA, 0,NA,NA,0,NA, NA,0,NA,NA,0, NA,NA,0,NA,NA, 0,NA,NA,0,NA,
NA,0,NA,NA,0, NA,NA,0,NA,NA, 0,0,NA,0,0, NA,0,0,NA,0, 0,NA,0,0,NA, 0,0,NA,0,0),
.Dim = c(20,3)),

r = structure(.Data = c(
NA,NA,NA,NA,NA, NA,NA,NA,NA,NA, NA,NA,NA,NA,NA, NA,NA,NA,NA,NA, NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA, NA,NA,NA,NA,NA, NA,NA,NA,NA,0, NA,NA,0,NA,NA, 0,NA,NA,0,NA,
NA,0,NA,NA,0, NA,NA,0,NA,NA),.Dim = c(20,3)),

#Additional initials for normal baseline (model B)
sdmu=1, mu.mean=0,

#Additional initials for two normal distributions model (model C)
#eta = 0, lambda = c(0,NA), P=c(0.5,0.5),

#Additional initials for exchangeable and independent regression slopes model
#beta = c(NA,0,0,0,0,0,0)
)
```

Copyright © 2012 John Wiley & Sons, Ltd.

Appendix D. Example 2 (post-operative analgesics data for WinBUGS)

Intervention codes (1= placebo, 2 = paracetamol, 3 = NSAID and 4 = COX-2) list(ns = 56, nt = 4, m.mu = 45.3)

id[]	t[,1]	y[,1]	se[,1]	t[,2]	y[,2]	se[,2]	t[,3]	y[,3]	se[,3]	na[]
1	1	58.1	3.0649781	3	44.01	2.1036088	NA	NA	NA	2
2	1	44.8	6.1967734	3	44.6	5.344717	NA	NA	NA	2
3	1	32.9	6.506612	2	28	5.2414375	3	25.7	4.3893811	3
4	1	51.6	3.9244426	3	41.34	3.3120136	NA	NA	NA	2
5	1	14.9	4.7750393	3	17.4	4.0160926	NA	NA	NA	2
6	1	72	6.6018197	4	54	5.4738594	NA	NA	NA	2
7	1	55	4.0852974	3	43	3.1037612	NA	NA	NA	2
8	1	93	1.3416408	3	63	1.3416408	NA	NA	NA	2
9	1	54.9	6.1755663	2	35	4.1641326	3	32.7	6.1268263	3
10	1	43.1	2.9029296	2	34.5	2.3186922	NA	NA	NA	2
11	1	64.2	4.3705894	3	39.6	3.0039847	NA	NA	NA	2
12	1	43.3	3.3387337	2	26	2.6622583	NA	NA	NA	2
13	1	59	5.0002116	3	38	3.9998101	NA	NA	NA	2
14	1	30.8	2.8001488	3	20.9	2.2211609	NA	NA	NA	2
15	1	21.1	2.008316	2	16.8	1.5336232	NA	NA	NA	2
16	1	55.1	3.0983867	3	36.6	2.32379	4	35.2	2.1430508	3
17	1	17.4	1.634119	4	12.6	1.1867322	NA	NA	NA	2
18	1	36.5	5.2414375	3	23.53	2.3583064	NA	NA	NA	2
19	1	34.9	2.3143304	4	25.6	1.3237522	NA	NA	NA	2
20	1	43.5	2.3559785	4	33.97	1.5830774	NA	NA	NA	2
21	1	57.5	3.9480256	4	40.35	3.307234	NA	NA	NA	2
22	1	30.9	5.1847956	3	17.4	3.653385	NA	NA	NA	2
23	1	48.2	4.5080931	3	39.1	3.1753903	NA	NA	NA	2
24	1	20.1	1.8670719	2	12.1	1.5276032	NA	NA	NA	2
25	1	80.4	11.198086	3	32.4	6.4988661	NA	NA	NA	2
26	1	38	2.8571429	3	32	2.4271195	NA	NA	NA	2
27	1	44	3.8758512	3	31.96	2.3709019	NA	NA	NA	2
28	1	21.7	2.899796	3	15	1.9052559	NA	NA	NA	2

Copyright © 2012 John Wiley & Sons, Ltd.

Appendix D. *Continued.*

id[]	t[,1]	y[,1]	se[,1]	t[,2]	y[,2]	se[,2]	t[,3]	y[,3]	se[,3]	na[]
29	1	51.6	7.5100409	3	36.3	6.6064793	NA	NA	NA	2
30	1	58.75	17.000079	3	30	9.5082573	NA	NA	NA	2
31	1	66.7	4.472136	2	61.1	5.1429563	3	36	4.0249224	3
32	1	78	8.9991123	3	53.97	4.8887125	NA	NA	NA	2
33	1	59.5	8.2957125	2	50.3	8.02	NA	NA	NA	2
34	1	51	6.363961	4	33.51	3.3225035	NA	NA	NA	2
35	1	38.2	4.902607	3	33.2	3.9833682	NA	NA	NA	2
36	1	39.2	1.725	3	34.6	1.6124636	NA	NA	NA	2
37	1	51	5.4738594	3	33	3.584417	NA	NA	NA	2
38	1	29.7	0.8497058	3	23	0.5391683	NA	NA	NA	2
39	1	22.1	5.1961524	3	8	2.0801257	NA	NA	NA	2
40	1	36.6	2.2979701	4	26.8	1.7870153	NA	NA	NA	2
41	1	35.5	2.3004347	4	25.5	1.5153657	NA	NA	NA	2
42	1	12.45	1.5697197	2	11.25	1.8827692	NA	NA	NA	2
43	1	27	1.6099689	4	12.5	0.910736	NA	NA	NA	2
44	1	8.55	1.1582832	3	5.4	0.9615092	NA	NA	NA	2
45	1	45.2	2.7110883	4	28	1.8203022	NA	NA	NA	2
46	1	47	5.891883	4	25.54	1.9240607	NA	NA	NA	2
47	**1**	**NA**	**1**	**2**	**54.5**	**5.8175381**	**3**	**44.1**	**4.88**	**3**
48	1	56.5	4.7992197	3	46.05	4.1389736	NA	NA	NA	2
49	1	31.6	0.8336148	3	28.33	0.6392041	NA	NA	NA	2
50	1	37	3.354102	3	27	2.9104275	NA	NA	NA	2
51	**1**	**NA**	**1**	**2**	**65**	**6.3960215**	**3**	**58**	**5.2128604**	**3**
52	1	31.3	2.4373141	4	25.93	1.3381157	NA	NA	NA	2
53	1	44.2	0.82	4	35.1	0.7	NA	NA	NA	2
54	1	57.4	7.2527051	2	39.56	3.2734082	NA	NA	NA	2
55	1	19.5	2.1430508	3	16.9	1.6782928	NA	NA	NA	2
56	1	141.5	17.654099	4	73.56	9.6391968	NA	NA	NA	2
END										

Copyright © 2012 John Wiley & Sons, Ltd.

#Initial values for example 2 (Post operative analgesics data)

```
list(
#Common slope and independent baseline model (Model A1)
B =0, d = c(NA,0,0,0), sd=1,
mu = 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,  0,0,0,0,0,  0),

delta = structure(.Data = c(
NA,0,NA,NA,0,   NA,NA,0,0,NA,  0,NA,NA,0,NA,   NA,0,NA,NA,0,  NA,NA,0,NA,NA,  0,0,NA,0,NA,  NA,0,NA,NA,0,
NA,NA,0,NA,NA,  0,NA,NA,0,NA,  NA,0,0,NA,0,  NA,NA,0,NA,NA,  0,NA,NA,0,NA,  NA,0,NA,NA,0,  NA,NA,0,NA,NA,
0,NA,NA,0,NA,  NA,0,NA,NA,0,  NA,NA,0,NA,NA,  0,NA,NA,0,NA,  NA,0,0,NA,0,  NA,NA,0,NA,NA,  0,NA,NA,0,NA,
NA,0,NA,NA,0,  NA,NA,0,NA,NA,  0,NA,NA,0,NA,  NA,0,NA,NA,0,  NA,NA,0,NA,NA,  0,NA,NA,0,NA,  NA,0,NA,NA,0,
0,NA,0,NA,NA,  0,NA,NA,0,NA,  NA,0,0,NA,0,  NA,NA,0,NA,NA,  0,NA,NA,0,NA,   NA,0,NA),.Dim = c(56,3)),

y = structure(.Data = c(
NA, NA, NA,NA, NA,  NA,NA,NA,NA,NA,  NA, NA,NA,NA,NA,   NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA, NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,   NA,NA,NA,NA,NA,   NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,   NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,   NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  NA,NA,NA, 0,NA,  NA,NA,NA,NA,NA,  NA,NA,NA,NA,NA,  0,NA,NA,NA,NA,  NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,  NA,NA, NA),.Dim = c(56,3)),

# Additional initials for normal baseline (model B)
sdmu=1, mu.mean=0,

# Additional initials for two normal distributions model (model C)
#eta = 0, lambda = c(0,NA), P=c(0.5,0.5),

#Additional initials for # exchangeable and independent regression slopes model
#beta = c(NA,0,0,0)
)
```

Acknowledgements

The authors would like to thank Tony Ades for useful discussions relating to the statistical aspects of the research presented. They would also like to thank Catriona McDaid, Emma Maund and Nerys Woolacott for their contribution to the effectiveness of non-opioid analgesics dataset.

This study presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10231). The views expressed are those of the author(s) and are not necessarily those of the National Health Service (NHS), the NIHR or the Department of Health, UK.

References

1. Sutton A, Abrams K, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research*. Wiley: Chichester, 2000.
2. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Wiley: Hoboken, 2009.
3. Higgins JPT, Green SE (eds). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from "http://www.cochrane-handbook.org" [Accessed 10 December 2011].
4. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**:2741–2758.
5. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. 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* 2009; **28**:1861–1881.
6. Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine* 1992; **11**:2077–2082.
7. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *BMJ* 1996; **313**:735–738.
8. Walter SD. Variation in baseline risk as an explanation of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**:2883–2900.
9. McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* 1996; **15**:1713–1728.
10. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in Medicine* 2000; **19**:3251–3274.
11. Arends LR, Hoes AW, Lubsen J, Grobbee DE, Stijnen T. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Statistics in Medicine* 2000; **19**:3497–3518.
12. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
13. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008; **17**:279–301.
14. Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of Clinical Epidemiology* 2009; **62**:857–864.

Copyright © 2012 John Wiley & Sons, Ltd.

15. Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment, 2011. Available from: http://www.nicedsu.org.uk [Accessed 8 November 2011].
16. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine* 2010; **29**:3046–3067.
17. Cooper NJ, Kendrick D, Achana F, Dhiman P, He Z, Wynn P, Le Cozannet E, Saramago P, Sutton AJ. Network meta-analysis to evaluate the effectiveness of interventions to increase the uptake of smoke alarms. *Epidemiologic Reviews* 2011; **34**:32–45.
18. McDaid C, Maund E, Rice S, Wright K, Jenkins B, Woolacott N. Paracetamol and selective and non-selective nonsteroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery: a systematic review. *Health Technology Assessment* 2010; **14**:1–153.
19. Ghidey W, Stijnen T, van Houwelingen HC. Modelling the effect of baseline risk in meta-analysis: a review from the perspective of errors-in-variables regression. *Statistical Methods in Medical Research* 2011. DOI: 10.1177/0962280211412244.
20. van Houwelingen H, Senn S. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1999; **18**:110–115.
21. Ghidey W, Lesaffre E, Stijnen T. Semi-parametric modelling of the distribution of the baseline risk in meta-analysis. *Statistics in Medicine* 2007; **26**:5434–5444.
22. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials, 2011. Available from: http://www.nicedsu.org.uk [Accessed 8 November 2011].
23. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004; **23**:3105–3124.
24. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; **331**:897–900.
25. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* 2010; **29**:932–944.
26. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; **101**:447–459.
27. Spiegelhalter D, Thomas AB, Best N, Lunn D. WinBUGS User Manual Version 1.4 January 2003. Upgraded to Version 1.4.3, 2007.
28. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.
29. Draper NR, Smith H. *Applied Regression Analysis*, 3rd edn. Academic Press: London, 1998.
30. Lu GB, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; **10**:792–805.
31. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; **17**:2537–2550.
32. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (B)* 2002; **64**:1–34.
33. Lu G, Ades AE, Sutton AJ, Cooper NJ, BriggS AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics in Medicine* 2007; **26**:3681–3699.
34. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *Journal of Clinical Epidemiology* 2010; **63**:875–882.
35. Dias S, Welton NJ, Sutton AJ, Ades AE, 2011. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011. Available from: http://www.nicedsu.org.uk. [Accessed 8 November 2011].
36. Ntzoufras I. *Bayesian Modelling Using WinBUGS*. John Wiley & Sons: New Jersey, 2009.
37. Mason AJ. Bayesian methods for modelling non-random missing data mechanisms in longitudinal studies. *PhD Thesis*, Imperial College, UK, 2009.

Copyright © 2012 John Wiley & Sons, Ltd.

Achana et al. BMC Medical Research Methodology 2014, 14:92
http://www.biomedcentral.com/1471-2288/14/92

BMC
Medical Research Methodology

RESEARCH ARTICLE Open Access

Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes

Felix A Achana[1*], Nicola J Cooper[1], Sylwia Bujkiewicz[1], Stephanie J Hubbard[1], Denise Kendrick[2], David R Jones[1] and Alex J Sutton[1]

Abstract

Background: Network meta-analysis (NMA) enables simultaneous comparison of multiple treatments while preserving randomisation. When summarising evidence to inform an economic evaluation, it is important that the analysis accurately reflects the dependency structure within the data, as correlations between outcomes may have implication for estimating the net benefit associated with treatment. A multivariate NMA offers a framework for evaluating multiple treatments across multiple outcome measures while accounting for the correlation structure between outcomes.

Methods: The standard NMA model is extended to multiple outcome settings in two stages. In the first stage, information is borrowed across outcomes as well across studies through modelling the within-study and between-study correlation structure. In the second stage, we make use of the additional assumption that intervention effects are exchangeable between outcomes to predict effect estimates for all outcomes, including effect estimates on outcomes where evidence is either sparse or the treatment had not been considered by any one of the studies included in the analysis. We apply the methods to binary outcome data from a systematic review evaluating the effectiveness of nine home safety interventions on uptake of three poisoning prevention practices (safe storage of medicines, safe storage of other household products, and possession of poison centre control telephone number) in households with children. Analyses are conducted in WinBUGS using Markov Chain Monte Carlo (MCMC) simulations.

Results: Univariate and the first stage multivariate models produced broadly similar point estimates of intervention effects but the uncertainty around the multivariate estimates varied depending on the prior distribution specified for the between-study covariance structure. The second stage multivariate analyses produced more precise effect estimates while enabling intervention effects to be predicted for all outcomes, including intervention effects on outcomes not directly considered by the studies included in the analysis.

Conclusions: Accounting for the dependency between outcomes in a multivariate meta-analysis may or may not improve the precision of effect estimates from a network meta-analysis compared to analysing each outcome separately.

Keywords: Network meta-analysis, Mixed treatment comparisons, Multiple outcomes, Multivariate, WinBUGS

Background

Meta-analysis or the quantitative synthesis of evidence, usually from systematic reviews, has become a popular tool in healthcare evaluations [1,2]. Largely driven by a desire for more realistic synthesis of complex healthcare evidence, increasingly sophisticated methodology has been developed. One area of meta-analysis that has seen significant methodological development is the application of multivariate statistical methods for the comparison of treatments on two or more endpoints (usually known as multivariate meta-analysis) [3-8]. These methods are appealing because many studies and systematic reviews focus on broad health effects and therefore typically report several outcome measures [4,6,9]. In such instances, the multivariate approach offers some advantages over separate univariate analyses including the ability to account for the inter-relationship between

* Correspondence: faa7@le.ac.uk
[1]Biostatistics Group, Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK
Full list of author information is available at the end of the article

BioMed Central

© 2014 Achana et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

outcomes and borrow strength across studies as well as across outcomes [10] through modelling the correlation structure [7,11]. This can potentially reduce outcome reporting bias [12] and the uncertainty with which intervention effects are estimated. Additionally, in a decision making context where the synthesis is meant to inform a health economic evaluation, accounting for the correlations between effect estimates on different outcomes is important as the dependence between outcomes may have implication for estimating quality of life or economic consequences associated with treatment [13]. An example is the situation where a particularly effective treatment for a disease condition is associated with a large side effect profile. Ignoring information about the inter-relationships between beneficial and 'side effect' endpoints in such instances may have implications for quantifying the benefits associated with treatment.

When summarising effectiveness evidence, correlations between the effectiveness estimates typically arise at either within-study and or between-study levels. At the within-study level, correlations arise mainly due to differences in patient-level characteristics. They are rarely reported in the published literature and usually have to be estimated from external sources such as individual patient level data if available or elicited from expert opinion [8,11,14,15]. At the between-study level, correlations arise from i) differences in the distribution of patient-level characteristics across studies, in which case they will be related to the within-study correlations and/or ii) differences in the distribution of other study-level characteristics such as study design, population and baseline disease severity [16]. The within-study correlations thus give an indication of the association between multiple endpoints within a study while the between-study correlations indicate how the underlying true study-specific effects on different outcomes vary jointly across studies.

A second area of rapid methodological development is network meta-analysis (NMA) [17], also known as mixed treatment comparison meta-analysis [18-20] or multiple treatment meta-analysis [21-23]. NMA methods extend standard pairwise meta-analysis to enable simultaneous comparison of multiple treatments while maintaining randomisation of individual studies [18]. The method enables 'direct' evidence (i.e. evidence from studies directly comparing two interventions of interest) and 'indirect' evidence (i.e. evidence from studies that do not compare the two interventions directly) to be pooled under the assumption of evidence consistency [24]. Estimates of intervention effects can then be obtained, including effects between treatments not directly compared within any one individual study [19]. NMA methods thus provide a coherent framework for appraising all available evidence relevant to a specific decision problem. The results from such analyses are increasingly being used to inform economic evaluations in healthcare decision making where coherent decisions (about judicious use of scarce resource) need to be made based on sound appraisal of all available evidence.

Approaches to extend NMA methodology to multiple outcome settings have been proposed in the literature [13,25-27], initially focusing on mutually exclusive competing risk outcomes [13] or a single outcome measured at multiple time points [26,28]. More recently, Efthimiou *et al.* [14] proposed a method for modelling multiple correlated outcomes in networks of evidence with binary outcome measures. The proposed method accounts for both the within-study and between-study correlation structure and includes a strategy for eliciting expert opinion to inform the within-study correlations. This paper contributes to the growing literature on the simultaneous evaluation of correlated outcomes. We do this in two stages. In the first stage (labelled as model 2 in the remainder of the paper), information is borrowed across studies as well as across outcomes through modelling the correlations between effectiveness estimates on different outcomes. In the second stage (labelled as model 3 in the remainder of the paper), additional information is borrowed across outcomes based on ideas for combining evidence across human and animal studies originally proposed by DuMouchel and Harris [29] and also revisited by Jones *et al.* [30]. The proposed second stage analysis methods allows: i) disconnected treatments to be incorporated as nodes in a network of evidence and ii) prediction of intervention effects for outcomes where evidence from primary studies is either sparse or not directly available from any one study included in the analysis. The motivating application area is injury prevention in children where a broad array of outcomes and intervention packages have been evaluated with the aim of increasing safety practices around the home (to ultimately reduce household injuries).

The remainder of this paper is structured as follows: the example dataset is first described followed by a Methods section describing the statistical models developed and implementation of the models. These are followed by sections presenting the results of applying the methods to the motivating dataset and a discussion.

Dataset

The example data comes from a recently updated Cochrane systematic review of home safety education and provision of safety equipment for injury prevention in children [31]. The models developed in this paper are applied to a subset of the review evidence relating to the prevention of poisoning injuries. Table 1 presents the data from 22 studies for the following outcomes:

a) Safe storage of medicines
b) Safe storage of other household products (e.g. cleaning products) and
c) Possession of a poison control centre (PCC) telephone number.

Table 1 Summary of the available evidence

			Outcome information (no. of events/no. of households in control versus (vs.) treatment arm)		
Comparison	First author and year of publication	IPD	Safe storage of medicines	Safe storage of other household products	Possession of a PCC number
Usual care (1) *vs.* Education (2)	Gielen 2007	Yes	178/271 vs. 188/249[ǂ]	44/62 vs. 57/73[ǂ]	
	Nansel 2002	Yes	83/89 vs. 79/85	65/89 vs. 66/85	59/89 vs. 63/85
	Nansel 2008	Yes	72/74 vs. 140/144[†]	59/73 vs. 117/144[†]	50/59 vs. 90/119[†]
	Kelly B 1987	No	54/54 vs. 55/55	43/54 vs. 49/55	
	McDonald 2005	No	4/57 vs. 6/60	3/57 vs. 6/61	
	Kelly N 2003	No			45.56/136.68 vs. 112.95/137.63[*]
Usual care (1) *vs.* Education + free/low cost safety equipment (3)	Clamp 1998	Yes	68/82 vs. 79/83	49/82 vs. 59/83	
	Woolf 1987	No			29/143 vs. 47/119
	Woolf 1992	No		60/151 vs. 89/150	59/151 vs. 117/150
Usual care (1) *vs.* Education + equipment (3) *vs.* Education + equipment + home safety inspection (4)	Babul 2007	Yes	147/149 vs. 171/173 vs. 160/163		
Usual care (1) *vs.* Education + equipment + home safety inspection (4)	Hendrickson 2002	Yes		14/40 vs. 34/38	8/40 vs. 34/38
	Swart 2008	No	70.26/79.58 vs. 74.07/80[*]	46.86/57.96 vs. 50.87/58.27[*]	
	Kendrick 1999	Yes		317/367 vs. 322/363	
Usual care (1) *vs.* Education + equipment + fitting (5)	Watson 2005	Yes	683/738 vs. 712/762	327/669 vs. 368/693	
Usual care (1) *vs.* Education + home safety inspection (6)	Petridou 1997	No			67.26/100.12 vs. 71.08/97.83[*]
Usual care (1) *vs.* Education + equipment + home safety inspection + fitting (7)	Schwarz D 1993	No	88.42/248.37 vs. 128.16/248.37[*]		
	Phelan 2011	No			16/138 vs. 71/139
Usual care (1) vs. Home visit (8)	Johnson 2006	No			82/91 vs. 222/232[†]
Education (2) *vs.* Education + equipment (3)	Posner 2004	Yes	14/47 vs. 19/49	22/47 vs. 34/49	27/47 vs. 35/49
Education (2) vs. Education + equipment + fitting (5)	Sznajder 2003	Yes	44/49 vs. 43/45	32/41 vs. 40/48	
Education + equipment + home safety inspection (4) vs. Education + equipment + home safety inspection + fitting (7)	King J 2001	No		261/469 vs. 273/482	
Education + equipment (3) vs. Equipment (9)	Dershewitz 1979	No	22/101 vs. 20/104	1/101 vs. 0/104	

Treatment abbreviation and codes:
Usual care = UC (1).
Education = E (2).
Education + free/low cost equipment = E + FE (3).
Education + equipment + home safety inspection = E + FE + HSI (4).
Education + equipment + fitting = E + FE + F (5).
Education + home safety inspection = E + HSI (6).
Education + equipment + home safety inspection + fitting = E + FE + HSI + F (7).
Education + home visit = E + HV (8).
Free/low cost equipment = FE (9).
*Effective sample size reported for cluster randomised studies after adjusting clustering, hence not whole numbers (details given in Kendrick *et al.* 2012 [31]).
[ǂ]The IPD for Gielen 2007 shows information on safe storage of medicines and safe storage of other household products was collected from different sets of households in this study (i.e. all the households that provided information for storage of medicines had missing data for safe storage of other household products and vice versa). Hence the Gielen 2007 IPD was not used to estimate the within-study correlations. [†]The intervention arms of Nansel 2008 and Johnson 2006 [32] comprises two groups that received different versions of a home safety intervention. The two versions were considered to be similar, hence combined into one intervention group for the analysis reported here.

Thirteen of the 22 studies considered at least two of the three outcomes. Of these, 8 considered storage of medicines and storage of other household products, 2 considered storage of other household products and possession of a PCC telephone number, and 3 considered all three outcome measures. Individual patient data (IPD) were available for 8 of the 13 studies, of which 7 were in a format suitable for the analysis reported here as explained by the footnotes in Table 1.

We classified the interventions trialled in the 22 studies into 9 relatively homogenous treatment packages:

1) Usual care (UC)
2) Education (E)
3) Education + provision of free/low cost equipment (E + FE)
4) Education + provision of free/low cost equipment + home safety inspection (E + FE + HSI)
5) Education + provision of free/low cost equipment + fitting of equipment (E + FE + F)
6) Education + home safety inspection (E + HSI)
7) Education + provision of free/low cost equipment + home safety inspection + fitting of equipment (E + FE + HSI + F)
8) Education + home visit (E + HV)
9) Provision of free/low cost equipment (FE).

Figure 1 shows the comparisons between the interventions that were made by individual studies and the number of comparisons in each network. All studies compared 2 intervention strategies, except Babul *et al.* (2007) [33] which compared 3 strategies. Data on each outcome was not available for all interventions; i.e. for the storage of medicines and other household products outcomes, interventions E + HSI and E + HV were not investigated in any of the included studies, and for possession of a PCC number interventions, E + FE + F and FE were not available.

Methods

In this section, we first present the NMA statistical model for one binary outcome measure and then extend it to

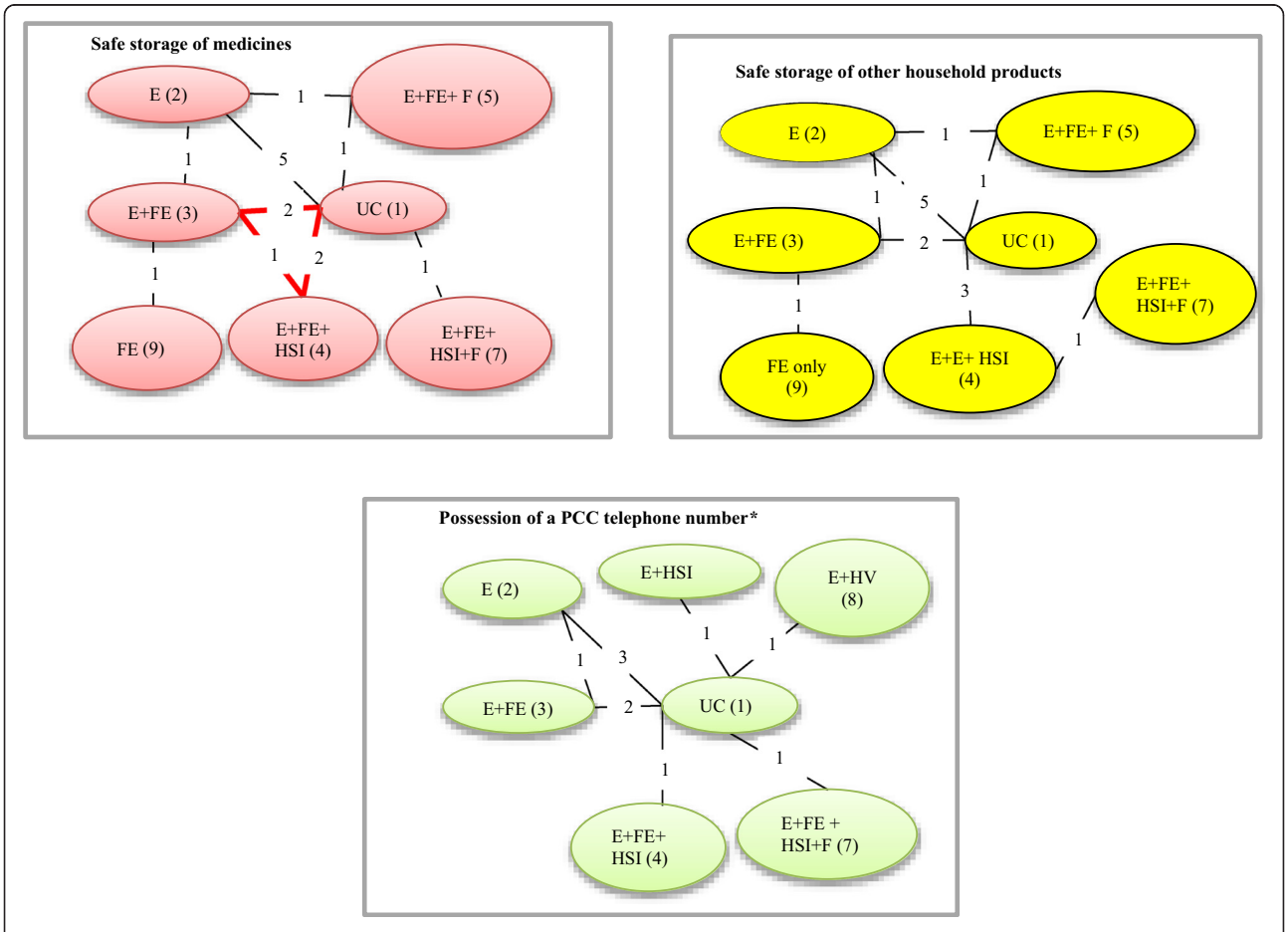


Figure 1 Intervention networks for the poisoning prevention outcomes (thick red lines indicate multi-arm comparisons). *PCC= poison control centre telephone number.

compare multiple interventions across multiple outcomes. Throughout the paper, we refer to these single and multiple outcome models as univariate and multivariate NMAs respectively. Where studies report multiple outcomes, these will not be independent as each household provides information on the different outcome measures within intervention arms. The multivariate model takes this correlation structure into account by allowing the intervention effects measured by one outcome to be correlated with the intervention effects measured by other outcomes.

Model 1: Univariate NMA

Given arm-level binary data of the form presented in Table 1, a random effects NMA may be specified using the method of Lu and Ades [20]. It is assumed that the occurrence of r_{ik} events out of a total of n_{ik} households in the kth-arm ($k = A,B,C,\ldots,$) of the ith-study follow a binomial distribution with underlying event probability p_{ik}:

$$\begin{aligned} & r_{ik} \sim \text{Binomial}(p_{ik},\ \ n_{ik});\ \ \text{logit}(p_{ik}) = \theta_{ik} \\ & \theta_{ik} = \begin{cases} \mu_{ib} & \text{if} \quad k = b \\ \mu_{ib} + \delta_{i(bk)} & \text{if} \quad k > b \end{cases} \\ & b = A, B, C, \end{aligned} \tag{1}$$

$$\begin{aligned} & \delta_{i(bk)} \sim \text{Normal}\left(d_{(bk)} = d_{(Ak)} - d_{(Ab)}\ ,\ \ \sigma^2_{(bk)}\right) \\ & \text{Note}:\ d_{AA} = 0 \end{aligned} \tag{2}$$

where μ_{ib} is a study-specific baseline effect (i.e. the log-odds for the control group in study i with baseline treatment b), $\delta_{i(bk)}$ is a study-specific log-odds ratio, $d_{(bk)}$ is the pooled effect of treatment k relative to treatment b (a quantity usually of interest in a meta-analysis) and $\sigma^2_{(bk)}$ is the between-study variance or heterogeneity parameter. Random effects NMA assumes that intervention effects are exchangeable across the network regardless of whether or not treatments b and k are included in study i [18]. This assumption implies that the pooled effects $d_{(bk)}$, can be expressed as functions of basic parameters with reference to a common comparator or baseline treatment (i.e. $d_{(bk)} = d_{(Ak)} - d_{(Ab)}$) [24]. Throughout this paper, we take usual care (UC) intervention to be the reference or 'baseline' treatment (i.e. UC is taken as treatment A of equation (2) above). Multi-arm studies (i.e. studies with more than 2 treatment groups) present a special problem in network meta-analysis because they produce evidence on multiple treatment effects that are correlated through sharing a common reference or 'baseline' treatment. Under a homogenous variance assumption ($\sigma^2_{(bk)} = \sigma^2$), the covariance between any two effects that share a common reference treatment is $\frac{\sigma^2}{2}$ [20]. The homogeneous variance assumption allows for the distribution of effects (in a study with an arbitrary number of arms) to be expressed as a univariate marginal distribution and a series of univariate conditional distributions. Specifically, for the ith-study with $p+1$ arms and p treatment effect estimates relative to the reference treatment, if

$$\begin{pmatrix} \delta_{i(bk_1)} \\ \delta_{i(bk_2)} \\ \vdots \\ \delta_{i(bk_p)} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk_1)} \\ d_{(bk_2)} \\ \vdots \\ d_{(bk_p)} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{2} & \cdots & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \sigma^2 & \cdots & \frac{\sigma^2}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma^2}{2} & \frac{\sigma^2}{2} & \cdots & \sigma^2 \end{pmatrix} \right) \tag{3}$$

then the marginal and conditional univariate distributions for arm j, given the previous $1, \cdots, (j-1)$ arms are:

$$\begin{aligned} & \delta_{i(bk_1)} \sim \text{Normal}\left(d_{(bk_1)}\ ,\ \ \sigma^2\right)\ \text{for}\ j = 1. \\ & \delta_{i(bk_j)} \Big| \begin{pmatrix} \delta_{i(bk_1)} \\ \vdots \\ \delta_{i(bk_{j-1})} \end{pmatrix} \sim \text{Normal}\left(d_{(bk_j)} + \frac{1}{j} \sum_{t=1}^{j-1} \left(\delta_{i(bk_t)} - d_{(bk_t)}\right), \frac{(j+1)}{2j} \sigma^2 \right) \\ & \qquad \text{for}\ j = 2, \ldots, p \end{aligned} \tag{4}$$

The analysis is conducted within a Bayesian framework requiring prior distributions to be specified for all model parameters. Accordingly, we specified minimally informative prior distributions corresponding to a Normal ($0,10^3$) prior distribution for the pooled mean effects relative to usual care, $d_{(Ak)}$ and the study-specific baseline effects, μ_{ib} and a Uniform (0,2) prior distribution for the between-study standard deviation on the log odds ratio scale σ [34].

Model 2: Multivariate NMA

We extend the univariate NMA model defined above to the multiple outcomes settings in order to account for correlations between intervention effects on different outcomes. The method presented here follows from Ades *et al.* (2010) NMA with competing risks model [13] where only the within-study correlations are taken into account. We extend their method to account for the between-study correlation as well.

Note that in Ades *et al.* (2010), a multinomial likelihood was appropriate as the three binary outcomes (relapse during treatment for Schizophrenia, discontinuation because of intolerable side effects, and discontinuation for other reasons) are mutually exclusive and event probabilities sum to 1 across outcomes. A multinomial likelihood will not be appropriate for our example dataset because each household can have one, two or all three outcomes simultaneously so that the event probabilities do not sum to 1

across outcomes. Instead, we use a Normal distribution on the log-odds scale to take account of the within-study correlations between outcomes. We assume that in each study i and for each k-th arm, the estimates $\hat{\theta}_{ikm}$ of the observed log-odds of an event on the mth outcome ($m = 1, 2, \cdots, M$) jointly follow a multivariate normal distribution:

$$\begin{pmatrix} \hat{\theta}_{ik1} \\ \vdots \\ \hat{\theta}_{ikM} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \theta_{ik1} \\ \vdots \\ \theta_{ikM} \end{pmatrix}, \mathbf{S}_{ik} = \begin{pmatrix} s_{ik1}^2 & \cdots & r_{ik}^{1M} s_{ik1} s_{ikM} \\ & \ddots & \vdots \\ & & s_{ikM}^2 \end{pmatrix} \right) \tag{5}$$

$$\begin{pmatrix} \theta_{ik1} \\ \vdots \\ \theta_{ikM} \end{pmatrix} = \begin{cases} \begin{pmatrix} \mu_{ib1} \\ \vdots \\ \mu_{ibM} \end{pmatrix} & \text{if } k = b \\ \begin{pmatrix} \mu_{ib1} + \delta_{i(bk)1} \\ \vdots \\ \mu_{ibM} + \delta_{i(bk)M} \end{pmatrix} & \text{if } k > b \end{cases}$$

$$\text{for } b = A, B, C, \ldots$$

where (μ_{ib1}, μ_{ib2}, $\cdots$, μ_{ibM}) and ($\delta_{i(bk)1}$, $\delta_{i(bk)2}$, $\cdots$, $\delta_{i(bk)M}$) represent vectors of *'true'* baseline and study-specific effects in study i with baseline treatment b respectively. The quantities $\left(\hat{\theta}_{ik1}, \hat{\theta}_{ik2}, \cdots, \theta_{ikM}\right)$ and (θ_{ik1}, θ_{ik2}, $\cdots$, θ_{ikM}) represent vectors of observed and *'true'* log-odds of response in arm k of study i and $\mathbf{S}_{ik}$ is the associated within-study covariance matrix usually assumed known but estimated in practice from the data [35]. We calculated $\left(\hat{\theta}_{ik1}, \hat{\theta}_{ik2}, \cdots, \hat{\theta}_{ikM}\right)$ and the diagonal elements of $\mathbf{S}_{ik}$ using standard formulae for log-odds and variance of the log-odds [2]. We applied a continuity correction by adding 0.5 to the numerators and 1 to the denominators of studies with 0% or 100% event rate in one of the treatment arms [36,37]. The off-diagonals of $\mathbf{S}_{ik}$ were calculated from estimates of within-study correlations r_{ik}^{mn} between outcomes m and n ($m \neq n$) in arm k of study i obtained from studies with IPD (see Additional file 1: Table S1). The method used to estimate the correlations from the IPD is described in the implementation section below.

When summarising evidence across multiple endpoints, it is common to encounter instances where some studies do not report information for all outcomes of interest leading to incomplete vectors with missing study-specific effects for the outcomes not reported [5,10]. Such studies can be included in our model under the assumption that the effects for outcomes not reported are missing at random. When implemented using the WinBUGS software, the missing study effects and standard errors are coded as NA in the data, a strategy previously outlined in Bujkiewicz *et al.* [10] and Dakin *et al.* [28]. This enables WinBUGS to automatically 'impute' values for the missing information under missing at random assumption with predicted distributions.

We refer to equation (5) as the within-study model and the model describing the distribution of the *'true'* effects across studies (presented below) as the between-study model following standard terminology in multivariate meta-analysis [5,6,10,11,38,39]. For the network of two-arm trials, the between-study model for the ith study is thus given by:

$$\begin{pmatrix} \delta_{i(bk)1} \\ \vdots \\ \delta_{i(bk)M} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk)1} = d_{(Ak)1} - d_{(Ab)1} \\ \vdots \\ d_{(bk)M} = d_{(Ak)M} - d_{(Ab)M} \end{pmatrix}, \ \mathbf{\Sigma}_{(bk)} \right)$$
$$\mathbf{\Sigma}_{(bk)} = \begin{pmatrix} \sigma_{(bk)1}^2 & \cdots & \rho_{bk}^{1M} \sigma_{(bk)1} \sigma_{(bk)M} \\ & \ddots & \vdots \\ & & \sigma_{(bk)M}^2 \end{pmatrix} \tag{6}$$

where the *'true'* effects $\delta_{i(bk)m}$ ($m = 1, 2, \cdots, M$) jointly follow a Normal distribution with mean effects $d_{(bk)m}$. The parameters in equation (6) have the same interpretation as in equation (2) except that they are now specific to each outcome. The covariance matrix $\Sigma_{(bk)}$ contains terms for the between-study variances, $\sigma_{(bk)m}^2$ for each outcome m and the between-study correlations ρ_{bk}^{mn} between effects measured by outcome m and n ($m \neq n$) specific to each k versus b comparison. Fitting the full model would thus require a large number of possibly multi-arm studies in order to make $\mathbf{\Sigma}_{(bk)}$ identifiable [5,13]. The number of parameters in $\Sigma_{(bk)}$, can however be reduced if reasonable assumptions can be made about the covariance structure. In particular, most practical applications of NMA methods involve the assumption of a common between-study variance across treatment arms, often referred to as a homogenous variance assumption [18,40,41]. Therefore, to simplify $\mathbf{\Sigma}_{(bk)}$ we make the additional assumption in this context of a common between-study correlation ($\rho_{bk}^{mn} = \rho^{mn}$) leading to the following simplified between-study covariance structure for two-arm studies:

$$\begin{pmatrix} \delta_{i(bk)1} \\ \vdots \\ \delta_{i(bk)M} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk)1} = d_{(Ak)1} - d_{(Ab)1} \\ \vdots \\ d_{(bk)M} = d_{(Ak)M} - d_{(Ab)M} \end{pmatrix}, \Sigma_{(M \times M)} \right)$$
$$\mathbf{\Sigma}_{(M \times M)} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M} \ \sigma_1 \sigma_M \\ & \ddots & \vdots \\ & & \sigma_M^2 \end{pmatrix} \tag{7}$$

where, as in the univariate case, σ_m represent the common between-study standard deviation or heterogeneity parameter specific to outcome m.

To include multi-arm studies in our model, we extend equations (3) and (4) to the multiple outcome setting. We show in Appendix A, that under evidence consistency and the homogenous between-study covariance structure ($\sigma_{(bk)m}^2 = \sigma_m^2$ and $\rho_{bk}^{mn} = \rho^{mn}$), equation (3) can be extended

to the multiple outcome settings by formulating the distribution of effects in a multi-arm study i with $p+1$ arms reporting on $m = 1, 2, \cdots, M$ outcomes as follows:

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(bk_2)1} \\ \vdots \\ \delta_{i(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(bk_p)1} \\ \vdots \\ \delta_{i(bk_p)M} \end{pmatrix} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \begin{pmatrix} d_{(bk_1)1} \\ \vdots \\ d_{(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} d_{(bk_2)1} \\ \vdots \\ d_{(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{(bk_p)1} \\ \vdots \\ d_{(bk_p)M} \end{pmatrix} \end{pmatrix}, \ \mathbf{\Sigma}_{(Mp \times Mp)} \right) \tag{8}$$

$$\mathbf{\Sigma}_{(Mp \times Mp)} = \begin{pmatrix} \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \cdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \vdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & & \ddots & \vdots \\ & & & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \end{pmatrix}$$

where p is the number of treatment effect estimates. The corresponding marginal and conditional distributions for arm j, given the previous $1, \cdots, (j-1)$ arms are:

$$\begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} d_{(bk_1)1} \\ \vdots \\ d_{(bk_1)M} \end{pmatrix}, \ \mathbf{\Sigma}_{(M \times M)} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\ \sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{M1}\ \sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \right) \text{for } j = 1$$

$$\begin{pmatrix} \delta_{i(bk_j)1} \\ \vdots \\ \delta_{i(bk_j)M} \end{pmatrix} \Bigg| \begin{pmatrix} \begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(bk_2)1} \\ \vdots \\ \delta_{i(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(bk_{j-1})1} \\ \vdots \\ \delta_{i(bk_{j-1})M} \end{pmatrix} \end{pmatrix} \sim \text{Normal} \left(\begin{pmatrix} d_{(bk_j)1} + \frac{1}{j}\sum_{t=1}^{j-1}\left(\delta_{i(bk_t)1} - d_{(bk_t)1}\right) \\ \vdots \\ d_{(bk_j)M} + \frac{1}{j}\sum_{t=1}^{j-1}\left(\delta_{i(bk_t)M} - d_{(bk_t)M}\right) \end{pmatrix}, \mathbf{\Sigma}' = \frac{(j+1)}{2j}\mathbf{\Sigma}_{(M \times M)} \right)$$
$$\text{for } j = 2, \ldots, p \tag{9}$$

To complete model 2, μ_{ibm} and $d_{(1k)m}$ are given minimally informative prior distributions:

$$\mu_{ibm},\ d_{(1k)m} \sim \text{Normal}\left(0,\ \ 10^3\right)$$

Prior distributions also need to be specified for $\mathbf{\Sigma}_{(M \times M)}$ which, in general, is non-trivial because of the positive definite constraint. Initially we specified an Inverse-Wishart distribution [42]:

$$\mathbf{\Sigma}_{(M \times M)} \sim \text{Inverse} - \text{Wishart}(\mathbf{K}, M)$$

where $\mathbf{K}$ is $M \times M$ scale matrix and M is the total number of outcomes. Specifying minimally informative Inverse-Wishart prior distribution is, however, problematic, especially when the amount of data is small relative to the dimensions of $\mathbf{\Sigma}_{(M \times M)}$ as is the case for our example data. Therefore, to allow for flexibility in formulating a prior distribution for $\mathbf{\Sigma}_{(M \times M)}$, we also followed a strategy outlined by Lu and Ades (2009) [43] and more recently by Wei and Higgins (2013) [39] to express $\mathbf{\Sigma}_{(M \times M)}$ in terms of a diagonal matrix of standard deviations $\mathrm{V}^{1/2}$ and squared positive semi-definite matrix of correlations $\mathbf{R}$ based on a separation strategy (Barnard *et al.* [44]):

$$\mathbf{\Sigma}_{(M \times M)} = \ \mathbf{V}^{1/2}\ \mathbf{R}\ \ \mathbf{V}^{1/2}$$

where the off-diagonal elements of **R** contain correlation terms and diagonal elements equal 1. Lu and Ades [43] and also Wei and Higgins [39] showed that **R** can be written as $\mathbf{R} = \mathbf{L}^{\mathbf{T}}\mathbf{L}$ using Cholesky decomposition where **L** is an upper triangular matrix. The spherical parameterization technique [39,43] can be used to express **R** in terms of sine and cosine functions of the elements in **L**. Using this later technique, we specified Uniform (0,π) prior distributions for the spherical coordinate φ_{mn} in our model to ensure that elements of the correlation matrix **R** lie in the interval (−1,1). Finally, the elements of $\mathbf{V}^{1/2}$ correspond to the between-study standard deviation terms in $\mathbf{\Sigma}_{(M \times M)}$ and are given independent Uniform (0,2) prior distributions as in the univariate case (model 1).

Model 3: Borrowing strength across interventions and outcomes

From Table 1, it can be seen that none of the studies had considered the interventions E + HSI and E + HV for storage of medicines and storage of other household products. Similarly, interventions E + FE + F and FE were not trialled by any of the included studies on possession of a PCC number. To estimate the full set of 24 basic intervention effects relative to usual care from 9 interventions on 3 outcomes, we applied methods originally proposed by DuMouchel and Harris [29] and revisited by DuMouchel and Groer [45] and Jones *et al.* [30]. We assume that the pooled effects of treatment k relative to usual care intervention $d_{(Ak)m}$, can be expressed as a sum of treatment-specific effect α_k and outcome-specific effect γ_m. This assumption replaces the minimally informative prior distribution N(0, 10^3) specified for $d_{(Ak)m}$ in model 2 with:

$$d_{(Ak)m} \sim \text{Normal}\left(\alpha_k + \gamma_m, \quad \tau^2\right) \qquad (10)$$
$$k = 2,\ 3,\ \cdots K;\ m = 1, 2,\ M$$

where K is the total number of treatments being evaluated across M outcomes, and for $k = 1$ (i.e. reference treatment A), $d_{(Ak)m}$ equal to zero. Note that on the logarithmic scale, this would imply that the ratio of any intervention effects is constant across outcomes as the γ_m cancel, i.e.

$$d_{(bk)m} = \left(d_{(Ak)m} - d_{(Ab)m}\right) \sim \text{Normal}\left(\alpha_k - \alpha_b, 2\tau^2\right) \qquad (11)$$

Equation (10) thus embodies an assumption of equal or constant relative potency of treatments across outcomes which imply exchangeability of the relative effects between the non-reference/baseline treatments indicated by equation (11). For our example dataset, this implies that missing intervention effects for comparisons with the usual care intervention can be predicted directly from equation (10) as a linear combination of γ_m and α_k assuming that each treatment effect relative to usual care is reported on at least one outcome. The missing intervention effects between non- reference/baseline treatments if required can similarly be predicted directly from the model as linear combinations of the intervention effects relative to usual care.

The parameter τ controls the accuracy of the constant relative potency assumption. Values of τ close to zero would thus indicate a high degree of confidence (and support from the data) in the parallelism of effect profiles across outcomes and the constant relative potency assumption. Conversely, larger values of τ would indicate otherwise.

Multi-arm studies are included in model 3 based on equations (8) and (9) in the same way as in model 2. To complete model 3, the parameters α_k, γ_m and τ are given minimally informative prior distributions. For the mean effects, this is a normal distribution with zero mean and large variance:

$$\alpha_k,\ \ \gamma_m \sim \text{Normal}\left(0,\ 10^3\right)$$

We give τ a Uniform (0, 2) prior distribution, considered to be minimally informative on the log-odds ratio scale. Sensitivity analyses were conducted to assess the impact of specifying alternative prior distributions for τ that are also considered minimally informative [46]:

i. Normal prior distribution centred on 0 with large variance and constrained to be positive, $\tau \sim$ N (0, 10^2), $\tau \geq 0$
ii. Gamma prior distribution placed on the precision: $\tau^{-2} \sim$ Gamma(0.001, 0.001).

The results of the sensitivity analyses are presented in Additional file 1: Figure S1.

There is a limitation to the number of data (i.e. intervention effects relative to the usual care) on outcomes allowed to be missing for the model hyper-parameters to be identifiable. For K interventions and M outcomes, there will be $(K-1) \times M$ equation (10) that are used to estimate a total of $(K-1) + M$ hyper-parameters (i.e. $(K-1)$ of α_k and M of γ_m hyper-parameters). Therefore no more than $((K-1) \times M) - ((K-1) + M)$ missing values in total are allowed. For example, for $K = 3$ treatments and $M = 2$ outcomes, data has to be available on both outcomes for both treatment comparisons with the baseline when the prior distributions are non-informative. When large number of data on outcomes is missing, placing informative prior distributions on the hyper-parameters can improve convergence.

Implementation of the models

We fitted a total of four models, models 1 and 3 as described above and two versions of model 2. In model 2a,

we specified an inverse-Wishart prior distribution for the between-study covariance matrix $\mathbf{\Sigma}_{(M\times M)}$ whilst in model 2b, we specified a prior distribution for $\mathbf{\Sigma}_{(M\times M)}$ based on the separation strategy. All four models allowed for multi-arm trials to be included in the analysis. To fit the multivariate NMA models, the quantities $(\hat{\theta}_{ik1},\ \hat{\theta}_{ik2},\ \hat{\theta}_{ik3})$ and the diagonals of $\mathbf{S}_{ik}$ were estimated using standard $2\times2-table$ formulae [2]. Next, we obtained estimates of the within-study correlations from the IPD studies using the following three methods: i) Pearson correlation coefficient between the observed outcome events ii) Bootstrapping as described in Daniel and Hughes (1998), and iii) Generalised Estimating Equations (See details in Additional file 1). All three methods produced identical estimates of the correlations between pairs of outcome specific log-odds of event from each IPD study (Additional file 1: Table S1). Therefore, we formulated informative prior distributions for the correlation terms in $\mathbf{S}_{ik}$ of equation (5) using the estimates of the correlations between the observed outcome events (Pearson correlation) as follows:

$$r_{ik}^{mn} \sim \text{Uniform}(a^{mn}, b^{mn}) \text{ with}$$

$$a^{mn} = \overline{r^{mn}} - \left(\frac{\sqrt{12\times var(r^{mn})}}{2}\right) \text{ and}$$

$$b^{mn} = \overline{r^{mn}} + \left(\frac{\sqrt{12\times var(r^{mn})}}{2}\right)$$

where r_{ik}^{mn} is the within-study correlation between the outcomes m and n effects measured on the log-odds scale in arm k of study i, and $\overline{r^{mn}}$ and $var\ (r^{mn})$are the mean and variance of the within-study correlation between outcomes m and n effects measured by IPD respectively.

We also assessed consistency of the evidence within each network using a method based on node splitting [24]. We found no evidence of conflict between the direct and indirect sources (Additional file 1: Table S2) in all three outcome networks.

We fitted all models described above in WinBUGS [47] using Markov Chain Monte Carlo (MCMC) simulations. The univariate models were fitted separately for each outcome using WinBUGS code available from Dias *et al.* [48]. The WinBUGS code for the multivariate models is provided in the appendices 1 and 2 of the Additional file 1. Convergence was assessed by examination of the trace and autocorrelation plots and the Rubin-Gelman statistic after running 400 000 simulations and discarding the first 200 000 samples as 'burn in samples'.

Results

Univariate and multivariate analyses

Parameters of interest were the posterior median estimate (and 95% credible intervals) of the pooled intervention effects relative to the usual care intervention and the posterior median estimate (and 95% credible intervals) of the between-study standard deviation and correlation terms. Summary forest plots displaying effectiveness estimates relative to usual care on the odds ratio (OR) scale are presented in Figure 2. It can be seen that, all 4 models produced broadly similar estimates when the treatment effect is not extreme compared to the other effect estimates for the same outcome. Compared to the univariate analysis, the multivariate models produced noticeably less extreme estimates of intervention effects. This can be seen in the effect of FE + HSI (3) on possession of PCC number being shifted towards the line of no effect from an OR of 39.35 (95% CrI 2.37 to 732.30) in model 1 to 23.55 (95% CrI 1.39, to 456.80) in model 2a, 20.37 (95% CrI 0.72, to 706.00) in model 2b and 4.20 (95% CrI 1.59 to 13.16) in model 3. Similarly, the OR for FE (9) on safe storage of other household products shifted from 0.37 (95% CrI 0.00 to 15.10) in model 1 to 1.81 (95% CrI 0.63, to 5.37) in model 3.

Posterior median and 95% credible intervals of the between-study standard deviations and correlations are presented in Table 2. The posterior medians of the between-study correlations from the multivariate models were small and estimated with considerable uncertainty (i.e. all had large variances). Estimates of the between-study standard deviations were broadly similar for the univariate NMA (model 1) and the multivariate NMA using the separation strategy (model 2b), and relatively high for multivariate NMA using the inverse-Wishart prior distribution (model 2a).

Borrowing strength across outcomes

It can be seen from Figure 2 that the effect of E + HSI and E + HV relative to usual care intervention on safe storage of medicines and safe storage of other household products, and E + FE + F and FE on possession of a PCC telephone number were only estimated in model 3 as none of the studies had trialled these interventions on the respective outcomes. In this model, estimates of relative effects between non- reference/baseline treatments were assumed to be exchangeable across outcomes, which enabled estimates to be obtained for all outcomes by predicting effects where the interventions have not been considered for the particular outcome of interest. For the intervention/outcome pair where data from trials were available, this extrapolation step had the additional effect of producing more precise estimates of the treatment effect in comparison to the models that do not assume exchangeability effects across outcomes.

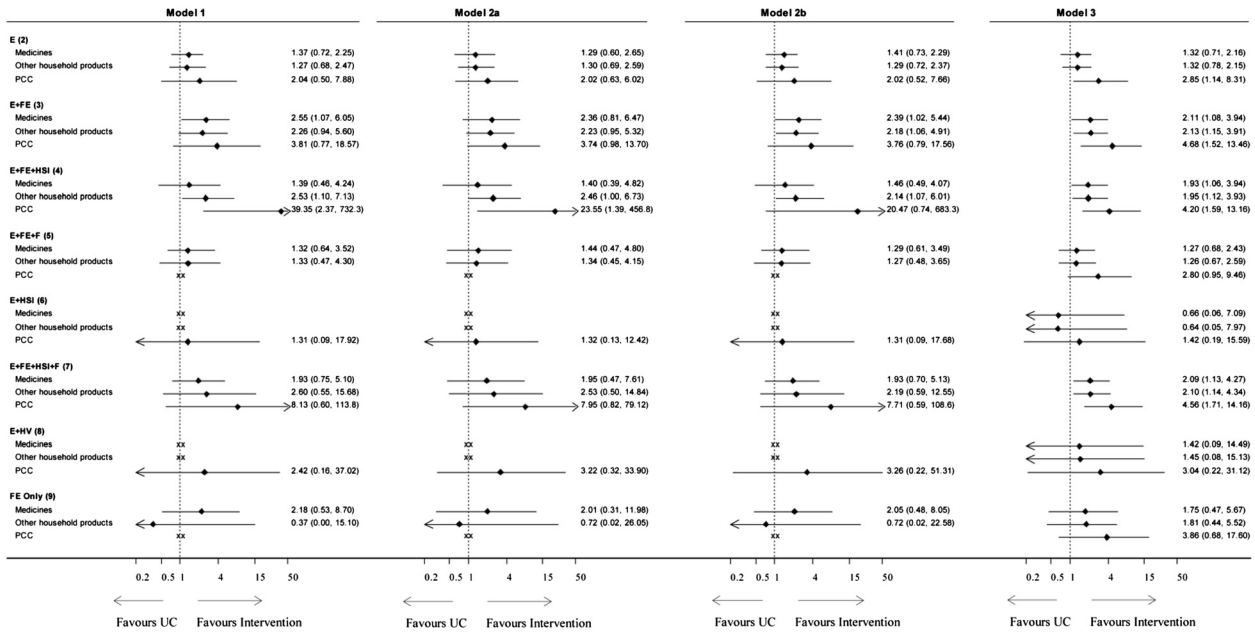


Figure 2 Summary forest plot of intervention effects relative to usual. Outcomes are safe storage of medicines, safe storage of other household products and possession of a PCC telephone number. Model 1: Univariate NMA. Model 2a: Multivariate NMA (Wishart prior distribution). Model 2b: Multivariate NMA (separation strategy). Model 3: Multivariate NMA allowing for the relative effects between non-usual care interventions to be exchangeable across outcomes. Effect estimate for which direct study data was not available are indicated by xx on the forest plot. Intervention components: E = Education, FE = low cost/free equipment, HSI = Home safety inspection, HV = Home visit and F= Fitting of equipment.

Table 2 Posterior median and 95% credible intervals of the between-study standard deviation and correlation parameters

Parameter	Description/prior distribution	Model 1: univariate	Model 2a: Multivariate using inverse-Wishart prior distribution for $\Sigma_{(M \times M)}$	Model 2b: Multivariate using a separation strategy to specify priors for elements of $\Sigma_{(M \times M)}$	Model 3: Multivariate with extrapolation of effects across outcomes
σ_1	Between-study standard deviation: safe storage of medicines	0.26 (0.03, 1.02)	0.58 (0.33, 1.18)	0.27 (0.01, 1.08)	0.23 (0.01, 0.80)
σ_2	Between-study standard deviation: safe storage of other household products	0.56 (0.13, 1.27)	0.62 (0.35, 1.15)	0.47 (0.04, 1.18)	0.31 (0.01, 0.81)
σ_3	Between-study standard deviation: PCC	1.16 (0.57, 1.93)	0.94 (0.53, 1.99)	1.18 (0.57, 1.93)	1.08 (0.58, 1.85)
τ	Primary analysis: τ ~Uniform *(0, 2)*	-	-	-	0.10 (0.01, 0.53)
τ	Sensitivity analysis: $\tau \sim$ Normal (0, 10^2), $\tau \geq 0$	-	-	-	0.11 (0.00, 0.56)
τ	Sensitivity analysis: $\tau^2 \sim$ Inverse – Gamma (0.001, 0.001)	-	-	-	0.08 (0.02, 0.36)
ρ^{12}	Between-study correlation [medicines, other household products]	-	0.03 (−0.73, 0.76)	0.05 (−1.00, 1.00)	0.45 (−0.99, 1.00)
ρ^{13}	Between-study correlation [medicines, PCC]	-	0.06 (−0.80, 0.81)	0.20 (−1.00, 1.00)	0.50 (−0.98, 1.00)
ρ^{23}	Between-study correlation [Other household products, PCC]	-	0.08 (−0.81, 0.83)	0.13 (−0.97, 0.98)	0.60 (−0.87, 0.99)

The posterior median and 95% credible intervals of intervention effects relative to usual care were unaffected by placing alternative minimally informative prior distributions on τ (Additional file 1: Appendix 2). The posterior median and credible intervals for τ (Table 2) were similarly not sensitive to the choice of prior distribution placed on τ in the primary and sensitivity analyses. The posterior median estimates of τ were all close to zero, which suggest that assumptions about the parallelism of effect profiles across outcomes is supported by the data. The estimates of τ would thus suggest with 95% probability, that based on the information in our example dataset, the extrapolation model could be accurate to within a factor of about $e^{(2 \times 0.10)} = 1.24$ (95% CrI 1.01 to 2.87).

Discussion

We have presented methods for simultaneous comparison of multiple treatments across multiple outcome measures while preserving the internal randomisation of individual studies. Our method may be viewed as an extension of Ades *et al.'s* (2010) NMA with competing risks paper [13] wherein only the within-study correlation is taken into account. We have extended their method to account for the dependency between outcome effects across studies as well as within-studies.

In this particular application of the multivariate approach, accounting for the correlation between outcomes alone (models 2a and 2b) did not reduce the uncertainty around estimates of intervention effects compared to analysing each outcome separately (model 1). Assuming that intervention effects are exchangeable across outcome did however lead to a modest reduction in uncertainty around effectiveness estimates (model 3).

The between-study correlations were estimated with considerable uncertainty (Table 2) and appear to have little impact on overall effect estimates. This may be because the between-study correlation arises due to, among other things, differences in study-level characteristics that also give rise to between-study heterogeneity in a meta-analysis. Based on a criterion outlined in Spiegelhalter *et al.* [49] the posterior median estimates of the between-study standard deviations, σ_1 and σ_2 on the log odds ratio scale (Table 2) could be interpreted as indicating evidence of low to moderate heterogeneity for storage of medicines and storage of other household products outcomes. Only the estimates for possession of poison control centre number exhibited a considerable degree of heterogeneity. Consequently, the posterior medians of the between-study correlations were small. There was therefore very little gain (in terms of increasing the precision of estimates) from formulating the between-study covariance structure described for the analysis presented here. Accounting for the between-study correlation is likely to be beneficial in situations where the between-study variance (heterogeneity) is large relative to within-study variances.

We opted to incorporate the within-study correlation through the arm-specific effects (log-odds) rather than the study-specific treatment difference (log-odds ratio) as is often done in multivariate meta-analysis [3,4,15,38]. This approach greatly simplifies the likelihood for multi-arm studies because treatment arms can be considered independent as a consequence of randomisation. Hence, there is no requirement to account for the additional correlations between effect estimates which share a common comparator treatment in the model likelihood [50]. The arm-based approach is also likely to be useful when (as is typical with many practical application of multivariate meta-analysis) the within-study correlations are not available [10,12,15,51] and have to be obtained from an external source such as expert opinion [14]. In such situations, formulating questions about correlations between outcome-specific event probabilities (which can be used directly in an arm-based approach) is more likely to be intuitive and easily understood by non-statistician healthcare experts than questions about correlations between intervention effects. It is acknowledged however, that the correlations between the intervention effects if required can easily be obtained from the correlations between the outcomes [14,51].

At the between-study level, we assumed a common correlation structure across treatments in addition to the common variance assumption underlying most practical application of NMA methods. The common correlation assumption implies that if several separate multivariate meta-analyses were conducted with the same outcomes, each with a different set of k versus b comparison, the assumption is that the between-study correlations would be the same across the different sets of bk comparisons. We suggested this structure to simplify the covariance structure and reduce the number of parameters in the model. Appropriateness of such modelling assumptions would need to be considered carefully and assessed when it is feasible to do so.

Initially we specified an inverse-Wishart distribution for the between-study covariance matrix $\mathbf{\Sigma}_{(M \times M)}$. However, we believe this prior distribution to be influential due to the small number of studies in our example dataset relative to the number of outcomes. Under these conditions, the inverse-Wishart prior distribution produced upwardly-biased estimates of σ_1 and σ_2 and downward bias in the estimate for σ_3 when compared to the corresponding estimates obtained from the univariate model (Table 2). These findings are consistent with observations in the univariate case where the use of a

Gamma prior distribution (which is the univariate analogue of the Inverse-Wishart prior distribution) can lead to an overestimation of the heterogeneity parameter when the true value is close to 0 [46,52]. As an alternative to an inverse-Wishart prior distribution therefore, we followed the spherical decomposition technique suggested by Lu and Ades [43]. This parameterization offered greater flexibility in formulating independent prior distributions for the standard deviation and correlation terms in $\mathbf{\Sigma}_{(M \times M)}$.

An obvious limitation to implementation of the multivariate models presented in this paper is the limited availability of data including i) the problem of missing within-study correlations and ii) the requirement for a relatively large number of studies to estimate all model parameters. The problem of missing within-study correlations has traditionally hampered the widespread application of multivariate meta-analysis [7,10,15]. In our example, IPD was available from a proportion of the included studies and we have used correlations estimated from the IPD to formulate informative prior distributions for the within-study model. Alternative approaches to dealing with missing within-study correlations when IPD is not available include: i) using the observed correlation from the summary study-specific effects [12], ii) eliciting information about the correlations from external sources such as clinical experts [14] and iii) specifying 'vague' prior distributions for analysis conducted within a Bayesian framework [6].

The second data issue concerns the number of studies needed to estimate the full unstructured between-study covariance matrix presented in equation (6). We anticipate a large number of multi-arm studies reporting across the three outcomes will be needed to identify $\mathbf{\Sigma}_{(bk)}$ and estimate all model parameters. This can be problematic considering the fact that most applications of network meta-analysis typically include mostly two-arm studies with very small numbers of multi-arm studies. Even with the simplification of the between-study covariance matrix given in equation (5), a relatively large number of studies in comparison to the total number of outcomes being considered may still be needed. We are unable to answer the question of how many studies should be considered large enough for a NMA with multiple outcomes. As a guide, Wei and Higgins [39] recently estimated 15, 27 and 42 studies as a minimum for multivariate pairwise meta-analysis with two, three and four-outcomes respectively. Hence, we believe an even larger number of studies will be required for the NMA with multiple outcomes.

Another limitation of the multivariate models presented here is that they rely on the normal approximation to binomial distribution to incorporate the within-study correlations in the model. The normal approximation frequently fails and may not provide adequate fit to the data in the presence of studies with zero or a small number of events, necessitating use of continuity corrections. We were unable to use the exact binomial distribution as our primary interest was to develop models for summary binary data where outcomes are not mutually exclusive, and where it is not reasonable to assume that within-study correlations are zero so that the likelihood factorises easily as in Arends *et al.* [3]. Further methodological investigations into modelling multivariate summary data that is not normally distributed will therefore be useful. An example is provided in Chu *et al.* [53] where parameterization of the within-study model enabled the special case of diagnostic sensitivity and specificity to be jointly modelled with disease prevalence using a trivariate binomial likelihood. In the interim, an alternative formulation which bypasses the need for approximating normal distributions is to directly model the IPD where this is available. This will require extending Saramago *et al.'s* [54] NMA model with aggregate and individual participant level data from single outcome to multiple outcome settings.

We assessed the consistency of each outcome network separately using the method of node splitting [24]. We found no evidence of conflict between the direct and indirect sources on pairwise contrasts that have both sources of evidence in model 1. We did not assess the consistency of the multivariate estimates partly because we are unaware of current methods for carrying out this type of assessment. We are investigating extensions of the node-split method to multiple outcome networks and investigate the effect of jointly synthesising evidence across multiple endpoints on evidence consistency in a simulation study.

Our initial motivation for a multiple outcome NMA was to estimate intervention effects for all the outcomes, including effects of interventions on outcomes not considered by any of the studies included in the analysis. This requires the correlation structure between effects on multiple outcomes to be appropriately modelled and also ensuring the mechanism of "borrow strength" across outcomes through the assumption of exchangeability of the random effect across outcomes. This implies a priori assumption that outcomes are related but different and that there is no way of knowing the order of magnitude of effects on outcomes. If this assumption does not hold, it may potentially lead to worse or more biased effectiveness estimates. In our example, the outcomes are similar and measured on the same scale. It would be clearly inappropriate to assume that intervention effects are exchangeable across outcomes that are different in some important respects such as

being measured on different scales (e.g. where one outcome reports a weighted mean difference and another outcome reports a log-odds ratio) as such estimates will differ in terms of the precision with which they are estimated.

Conclusion

Our aim in this paper was to present methods for simultaneous comparison of multiple treatments across multiple outcome measures while preserving the internal randomisation of individual studies. Application of the method to the poison prevention data yielded similar point estimates of treatment effect to those obtained from a univariate NMA but the uncertainty around the multivariate estimates increased or decreased depending on the prior distribution specified for the between-study covariance structure. The proposed method followed the usual hierarchical approach to multivariate meta-analysis where correlations between outcomes are modelled at the within-study and or between-study levels.

Appendix A

Between-study covariance for multi-arm studies reporting multiple outcomes

For a multi-arm study i with K treatments labelled A, B, C, …, K reporting a total of M outcomes labelled *1, 2, …, M*. A random effects between-study model can be represented as:

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(AB)1} \\ \vdots \\ \delta_{i(AB)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(AC)1} \\ \vdots \\ \delta_{i(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(AK)1} \\ \vdots \\ \delta_{i(AK)M} \end{pmatrix} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \begin{pmatrix} d_{i(AB)1} \\ \vdots \\ d_{i(AB)M} \end{pmatrix} \\ \begin{pmatrix} d_{i(AC)1} \\ \vdots \\ d_{i(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{i(AK)1} \\ \vdots \\ d_{i(AK)M} \end{pmatrix} \end{pmatrix}, \boldsymbol{\Sigma}_{\mathbf{FULL}} \right) \tag{A1}$$

Where $\delta_{i(Ak)m}$ and $d_{(Ak)m}$ are study-specific and mean effect of treatment k relative A (reference treatment) on outcome m in study i respectively and $\boldsymbol{\Sigma}_{\mathbf{FULL}}$ is the full (K-1) × (K-1) blocks of $M \times M$ within-treatment between-outcome covariance matrix. The parameters in $\boldsymbol{\Sigma}_{\mathbf{FULL}}$ have the following interpretation:

$(\sigma^2_{(Ak)m})$ indicate the variance of the effect of treatment k ($k = B, C, \cdots, K$) relative to A on outcome m across studies.

$\rho^{mn}_{(Ak,Ak)}$ indicate the correlation between $\delta_{i(Ak)m}$ and $\delta_{i(Ak)n}$ (i.e. the correlation between the effect of treatment k relative to A on outcome m and the effect of treatment k relative to A on outcome n ($m \neq n$)) specific to the Ak comparison.

$\rho^{mm}_{(Ah,Ak)}$ indicate the correlation between $\delta_{i(Ah)m}$ and $\delta_{i(Ak)m}$ (i.e. the correlation between the effect of treatment h relative to A on outcome m and the effect of treatment k relative to A ($h \neq k$) on outcome m because they share a common comparator A).

The diagonal block matrices in $\boldsymbol{\Sigma}_{\mathbf{FULL}}$ thus carry terms for the between-study variance ($\sigma^2_{(Ak)m}$) while the off-diagonal blocks carry terms for the between-study covariance. We make two assumptions to simplify and reduce the number of parameters in $\boldsymbol{\Sigma}_{\mathbf{FULL}}$. First, we assume homogenous variances for intervention effects within outcomes [20]. This implies $\sigma^2_{(Ak)m} = \sigma^2_m$ and $\rho^{mm}_{(Ah,Ak)} = \frac{1}{2}$ as in the single outcome network meta-analysis case [20,34]. Second, we make the assumption of homogenous between-study correlations for the intervention effects from different outcomes. Under this assumption we can express $\rho^{mn}_{(Ah,\,Ah)}$ and $\rho^{mn}_{(Ak,\,Ak)}$ in terms of a common correlation parameter ρ^{mn} by noting that for any 3-treatment (A, h, k) configuration, the covariance between outcome m and n effects across studies can be expressed as a covariance between two sums under evidence consistency:

$$\begin{aligned} \left[\delta_{i(hk)m}, \delta_{i(hk)n}\right] &= COV\left[\left(\delta_{i(Ak)m} - \delta_{i(Ah)m}\right), \left(\delta_{i(Ak)n} - \delta_{i(Ah)n}\right)\right] \\ &= COV\left[\delta_{i(Ak)m}, \delta_{i(Ak)n}\right] - COV\left[\delta_{i(Ak)m}, \delta_{i(Ah)n}\right] \\ &\quad - COV\left[\delta_{i(Ah)m}, \delta_{i(Ak)n}\right] + COV\left[\delta_{i(Ah)m}, \delta_{i(Ah)n}\right] \\ &= \left(\rho^{mn}_{(Ak,Ak)} + \rho^{mn}_{(Ah,Ah)} - 2\rho^{mn}_{(Ak,Ah)}\right)\sigma_m \sigma_n \end{aligned} \tag{A2}$$

$$\boldsymbol{\Sigma}_{\mathbf{FULL}} = \begin{pmatrix} \sigma^2_{(AB)1} & \cdots & \rho^{1M}_{(AB,AB)}\sigma_{(AB)1}\sigma_{(AB)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AB,AB)}\sigma_{(AB)1}\sigma_{(AB)M} & \cdots & \sigma^2_{(AB)M} \end{pmatrix} \begin{pmatrix} \rho^{11}_{(AB,AC)}\sigma_{(AB)1}\sigma_{(AC)1} & \cdots & \rho^{1M}_{(AB,AC)}\sigma_{(AB)1}\sigma_{(AC)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AB,AC)}\sigma_{(AB)1}\sigma_{(AC)M} & \cdots & \rho^{MM}_{(AB,AC)}\sigma_{(AB)M}\sigma_{(AC)M} \end{pmatrix} \cdots \begin{pmatrix} \rho^{11}_{(AB,AK)}\sigma_{(AB)1}\sigma_{(AK)1} & \cdots & \rho^{1M}_{(AB,AK)}\sigma_{(AB)1}\sigma_{(AK)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AB,AK)}\sigma_{(AB)1}\sigma_{(AK,AK)M} & \cdots & \rho^{MM}_{(AB,AK)}\sigma_{(AB)M}\sigma_{(AK)M} \end{pmatrix}$$

$$\begin{matrix} \begin{pmatrix} \sigma^2_{(AC)1} & \cdots & \rho^{1M}_{(AC,AC)}\sigma_{(AC)1}\sigma_{(AC)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AC,AC)}\sigma_{(AC)1}\sigma_{(AC)M} & \cdots & \sigma^2_{(AC)M} \end{pmatrix} & \vdots & \begin{pmatrix} \rho^{11}_{(AC,AK)}\sigma_{(AC)1}\sigma_{(AK)1} & \cdots & \rho^{1M}_{(AC,AK)}\sigma_{(AC)1}\sigma_{(AK)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AC,AK)}\sigma_{(AC)1}\sigma_{(AC)M} & \cdots & \rho^{MM}_{(AC,AK)}\sigma_{(AC)M}\sigma_{(AK)M} \end{pmatrix} \\ \ddots & & \vdots \\ & & \begin{pmatrix} \sigma^2_{(AK)1} & \cdots & \rho^{1M}_{(AK,AK)}\sigma_{(AK)1}\sigma_{(AK)M} \\ \vdots & \ddots & \vdots \\ \rho^{1M}_{(AK,AK)}\sigma_{(AK)1}\sigma_{(AK)M} & \cdots & \sigma^2_{(AK)M} \end{pmatrix} \end{matrix}$$

The homogenous between-study correlation assumption implies $\rho^{mn}_{(Ah,Ah)} = \rho^{mn}_{(Ak,Ak)} = \rho^{mn}$ and $\rho^{mn}_{(Ak,Ah)} = \frac{1}{2}\rho^{mn}$ for the inequality $-1 \leq \left(\rho^{mn}_{(Ak,Ak)} + \rho^{mn}_{(Ah,Ah)} - 2\rho^{mn}_{(Ak,Ah)}\right) \leq 1$ to hold. Substituting these expressions into equation (A1), we see that the between-study correlation terms equal ρ^{mn} in the diagonal block of matrices and $\frac{1}{2}\rho^{mn}$ in the off-diagonal block of matrices of in $\mathbf{\Sigma_{FULL}}$ leading to the following simplificationfollowing simplification of the between-study covariance matrix:

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(AB)1} \\ \vdots \\ \delta_{i(AB)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(AC)1} \\ \vdots \\ \delta_{i(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(AK)1} \\ \vdots \\ \delta_{i(AK)M} \end{pmatrix} \end{pmatrix} \sim Normal \left(\begin{pmatrix} \begin{pmatrix} d_{(AB)1} \\ \vdots \\ d_{(AB)M} \end{pmatrix} \\ \begin{pmatrix} d_{(AC)1} \\ \vdots \\ d_{(AC)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{(AK)1} \\ \vdots \\ d_{(AK)M} \end{pmatrix} \end{pmatrix}, \mathbf{\Sigma}_{(Mp \times Mp)} \right)$$

$$\text{with} \quad \mathbf{\Sigma}_{(Mp \times Mp)} = \begin{pmatrix} \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \cdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} & \vdots & \frac{1}{2}\begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \\ & & \begin{pmatrix} \sigma_1^2 & \cdots & \rho^{1M}\sigma_1\sigma_M \\ \vdots & \ddots & \vdots \\ \rho^{1M}\sigma_1\sigma_M & \cdots & \sigma_M^2 \end{pmatrix} \end{pmatrix} \tag{A3}$$

Finally by relabeling the reference treatment A as b, $(\delta_{i(AB)1}, \cdots, \delta_{i(AK)m})$ as $\left(\delta_{i(bk_1)1}, \cdots, \delta_{i(bk_j)M}\right)$ and $(d_{(AB)1}, \cdots, d_{(AK)M})$ as $\left(d_{(bk_1)M}, \cdots, d_{(bk_j)M}\right)$, equation (A3) can be rewritten as equation (A4).

$$\begin{pmatrix} \begin{pmatrix} \delta_{i(bk_1)1} \\ \vdots \\ \delta_{i(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} \delta_{i(bk_2)1} \\ \vdots \\ \delta_{i(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \delta_{i(bk_p)1} \\ \vdots \\ \delta_{i(bk_p)M} \end{pmatrix} \end{pmatrix} \sim \text{Normal} \left(\begin{pmatrix} \begin{pmatrix} d_{(bk_1)1} \\ \vdots \\ d_{(bk_1)M} \end{pmatrix} \\ \begin{pmatrix} d_{(bk_2)1} \\ \vdots \\ d_{(bk_2)M} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} d_{(bk_p)1} \\ \vdots \\ d_{(bk_p)M} \end{pmatrix} \end{pmatrix}, \mathbf{\Sigma}_{(Mp \times Mp)} \right) \tag{A4}$$

Additional file

Additional file 1: (WinBUGS Code for model 2): Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes.

Abbreviations

CrI: Credible interval; IPD: Individual patient data; MCMC: Markov Chain Monte Carlo; NMA: Network meta-analysis; PCC: Poison centre control number.

Competing interests

The authors declare that they have no competing interests.

Author's contribution

FAA drafted the initial manuscript, conducted the analysis and coordinated contributions from all authors in drafting the final manuscript. NJC designed the study and revised the manuscript to improve general readability. SB contributed to developing the statistical analysis and interpretation of the results. SJH contributed to preparing the data for analysis, drafting the network diagrams and revision of the initial and final drafts. DK provided the data for analysis, defined the outcomes and interventions and revision of the final drafts.DRJ participated in designing the study and revision of the final manuscript. AJS conceived the original idea and participated developing the statistical model. All authors read and approved the final manuscript.

Acknowledgements

FAA is funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research programme (Reference Number RP-PG- 0407–10231) to pursue a PhD in Public Health Evaluation of Interventions.
SB is supported by the MRC Methodology Research Programme (grant number MR/L009854/1).

Author details

[1]Biostatistics Group, Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK. [2]Division of Primary Care Community Health Sciences, Faculty of Medicine & Health Sciences, University of Nottingham, Nottingham NG7 2RD, UK.

Received: 8 January 2014 Accepted: 2 July 2014
Published: 21 July 2014

References

1. Cooper NJ, Sutton AJ, Ades AE, Paisley S, Jones DR, W. G. O. U. o. E. i. E. D: **Models, Use of evidence in economic decision models: Practical issues and methodological challenges.** *Health Economics* 2007, **16**(12):1277–1286.
2. Sutton A, Abrams K, Jones DR, Sheldon TA, Song F: *Methods for Meta-analysis in Medical Research.* Chichester: Wiley; 2000.
3. Arends LR, Voko Z, Stijnen T: **Combining multiple outcome measures in a meta-analysis: an application.** *Statistics in Medicine* 2003, **22**(8):1335–1353.
4. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA: **Meta-analysis of multiple outcomes by regression with random effects.** *Statistics in Medicine* 1998, **17**(22):2537–2550.
5. Jackson D, Riley R, White IR: **Multivariate meta-analysis: Potential and promise.** *Statistics in Medicine* 2011, **30**(20):2481–2498.
6. Nam IS, Mengersen K, Garthwaite P: **Multivariate meta-analysis.** *Statistics in Medicine* 2003, **22**(14):2309–2333.
7. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR: **Bivariate random-effects meta-analysis and the estimation of between-study correlation.** *BMC Medical Research Methodology* 2007, **7**.
8. Riley RD, Thompson JR, Abrams KR: **An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown.** *Biostatistics* 2008, **9**(1):172–186.
9. Kendrick D, Coupland C, Mulvaney C, Simpson J, Smith SJ, Sutton A, Watson M, Woods A: **Home safety education and provision of safety equipment for injury prevention.** *Cochrane Database Systematic reviews* 2007, **1**:CD005014.
10. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, Abrams KR: **Multivariate meta-analysis of mixed outcomes: a Bayesian approach.** *Statistics in Medicine* 2013, **32**(22):3926–3943.
11. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR: **An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes.** *Statistics in Medicine* 2007, **26**(1):78–97.
12. Kirkham JJ, Riley RD, Williamson PR: **A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews.** *Statistics in Medicine* 2012, **31**(20):2179–2195.
13. Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T: **Network meta-analysis with competing risk outcomes.** *Value Health* 2010, **13**(8):976–983.
14. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bagos P, Salanti G: **An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios.** *Statistics in Medicine* 2014, **33**(13):2275–2287.
15. Riley RD: **Multivariate meta-analysis: the effect of ignoring within-study correlation.** *Journal of the Royal Statistical Society Stat Soc* 2009, **172**:789–811.
16. Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, Popay J: **Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews.** *Evaluation* 2009, **15**(1):49–73.
17. Lumley T: **Network meta-analysis for indirect treatment comparisons.** *Statistics in Medicine* 2002, **21**(16):2313–2324.
18. Caldwell DM, Ades AE, Higgins JP: **Simultaneous comparison of multiple treatments: combining direct and indirect evidence.** *British Medical Journal* 2005, **331**(7521):897–900.
19. Caldwell DM, Welton NJ, Ades AE: **Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency.** *Journal of Clinical Epidemiology* 2010, **63**(8):875–882.
20. Lu G, Ades AE: **Combination of direct and indirect evidence in mixed treatment comparisons.** *Statistics in Medicine* 2004, **23**(20):3105–3124.
21. Salanti G, Ades AE, Ioannidis JP: **Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial.** *Journal of Clinical Epidemiology* 2010, **64**(2):163–171.
22. Salanti G, Higgins JP, Ades AE, Ioannidis JP: **Evaluation of networks of randomized trials.** *Statistical Methods in Medical Research* 2008, **17**(3):279–301.
23. Salanti G, Marinho V, Higgins JP: **A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered.** *Journal of Clinical Epidemiology* 2009, **62**(8):857–864.
24. Dias S, Welton NJ, Caldwell DM, Ades AE: **Checking consistency in mixed treatment comparison meta-analysis.** *Statistics in Medicine* 2010, **29**(7–8):932–944.
25. Hong H, Carlin BP, Chu H, Shamliyan TA, Wang S, Kane RL: *Methods Research Report: A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons*; 2013. Avaialable online from http://www.ncbi.nlm.nih.gov/books/NBK116689/pdf/TOC.pdf.
26. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM: **Meta-analysis of mixed treatment comparisons at multiple follow-up times.** *Statistics in Medicine* 2007, **26**(20):3681–3699.
27. Welton NJ, Cooper NJ, Ades AE, Lu G, Sutton AJ: **Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B.** *Statistics in Medicine* 2008, **27**(27):5620–5639.
28. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S: **Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension.** *Statistics in Medicine* 2011, **30**(20):2511–2535.
29. DuMouchel WH, Harris JE: **Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species.** *Journal of the American Statistical Association* 1983, **78**(382):293–308.
30. Jones DR, Peters JL, Rushton L, Sutton AJ, Abrams KR: **Interspecies extrapolation in environmental exposure standard setting: A Bayesian synthesis approach.** *Regulatory Toxicology Pharmacology* 2009, **53**(3):217–225.
31. Kendrick D, Young B, Mason-Jones AJ, Ilyas N, Achana FA, Cooper NJ, Hubbard SJ, Sutton AJ, Smith S, Wynn P, Mulvaney C, Watson MC, Coupland C: **Home safety education and provision of safety equipment for injury prevention.** *Cochrane Database Systematic reviews* 2012. 9.
32. Johnston BD, Huebner CE, Anderson ML, Tyll LT, Thompson RS: **Healthy steps in an integrated delivery system - Child and parent outcomes at 30 months.** *Archives of Pediatrics and Adolescent Medicine* 2006, **160**(8):793–800.

33. Babul S, Olsen L, Janssen P, McIntee P, Raina P: **A randomized trial to assess the effectiveness of an infant home safety programme.** *International Journal of Control and Safety Promotion* 2007, **14**(2):109–117.
34. Dias S, Sutton AJ, Ades AE, Welton NJ: **A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials.** *Medical Decision Making* 2012. **33**(5):607–17
35. van Houwelingen HC, Arends LR, Stijnen T: **Advanced methods in meta-analysis: multivariate approach and meta-regression.** *Statistics in Medicine* 2002, **21**(4):589–624.
36. Dershewitz RA, Williamson JW: **Prevention of Childhood Household Injuries - Controlled Clinical-Trial.** *American Journal of Public Health* 1977, **67**(12):1148–1153.
37. Kelly B, Sein C, McCarthy PL: **Safety education in a pediatric primary care setting.** *Pediatrics* 1987, **79**(5):818–824.
38. Mavridis D, Salanti G: **A practical introduction to multivariate meta-analysis.** *Statistical Methods in Medical Research* 2013, **22**(2):133–158.
39. Wei Y, Higgins JPT: **Bayesian multivariate meta-analysis with multiple outcomes.** *Statistics in Medicine* 2013, **32**(17):2911–2934.
40. Carter P, Achana F, Troughton J, Gray LJ, Khunti K, Davies MJ: **A Mediterranean diet improves HbA1c but not fasting blood glucose compared to alternative dietary strategies: a network meta-analysis.** *Journal of Human Nutrition and Dietetics* 2014, **27**(3):280–297.
41. Cooper NJ, Sutton AJ, Lu G, Khunti K: **Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation.** *Archives Internal Medicine* 2006, **166**(12):1269–1275.
42. Spiegelhalter D, Thomas A, Best N, Lunn D: *WinBUGS User Manual*. Available from http://www.politicalbubbles.org/bayes_beach/manual14.pdf, 2003. Accessed 10th December 2013.
43. Lu GB, Ades A: **Modeling between-trial variance structure in mixed treatment comparisons.** *Biostatistics* 2009, **10**(4):792–805.
44. Barnard J, McCulloch R, Meng XL: **Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage.** *Statistica Sinica* 2000, **10**(4):1281–1311.
45. DuMouchel W, Groer PG: **A Bayesian Methodology for Scaling Radiation Studies from Animals to Man.** *Health Physics* 1989, **57**:411–418.
46. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR: **How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS.** *Statistics in Medicine* 2005, **24**(15):2401–2428.
47. Lunn DJ, Thomas A, Best N, Spiegelhalter D: **WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility.** *Statistical Computing* 2000, **10**(4):325–337.
48. Dias S, Welton NJ, Sutton AJ, Ades AE: *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-analysis of Randomised Controlled Trials*; 2011. Available from http://www.nicedsu.org.uk.
49. Spiegelhalter DJ, Abrams KR, Myles JP: *Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Statistics in Practice.* Chichester: John Wiley & Sons; 2004.
50. Franchini AJ, Dias S, Ades AE, Jansen JP, Welton NJ: **Accounting for correlation in network meta-analysis with multi-arm trials.** *Research Synthesis Methodology* 2012, **3**(2):142–160.
51. Wei Y, Higgins JPT: **Estimating within-study covariances in multivariate meta-analysis with multiple outcomes.** *Statistics in Medicine* 2013, **32**(7):1191–1205.
52. Gelman A: **Prior distributions for variance parameters in hierarchical models(Comment on an Article by Browne and Draper).** *Bayesian Analysis* 2006, **1**(3):515–533.
53. Chu HT, Nie L, Cole SR, Poole C: **Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection.** *Statistics in Medicine* 2009, **28**(18):2384–2399.
54. Saramago P, Sutton AJ, Cooper NJ, Manca A: **Mixed treatment comparisons using aggregate and individual participant level data.** *Statistics in Medicine* 2012, **31**(28):3516–3536.

doi:10.1186/1471-2288-14-92
Cite this article as: Achana *et al.*: Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Medical Research Methodology* 2014 **14**:92.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central