A FRAMEWORK OF METHODOLOGIES FOR DESIGNING NEW TRIALS BASED ON THE POWER OF UPDATED EVIDENCE SYNTHESIS MODELS WHICH INCLUDE THE NEW TRIAL

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

By

Agampodi Shanthi Jeewaka Ruwan Mendis
B.Sc. (Sp) Hons, M.Sc.

Department of Health Sciences
University of Leicester

2014
A framework of methodologies for designing new trials based on the power of updated evidence synthesis models, which include the new trial

A S J R Mendis B.Sc.(Hons), M.Sc.

The consideration of power of a clinical trial to detect the treatment effect is integral in the planning and designing stage of any trial. Recent research however argues that the power of the subsequent evidence synthesis model including the new trial should receive attention in the design stage of any new trial besides the power of the new trial in isolation. This thesis begins with a survey aiming to assess the extent of the use of previous evidence in the design stage of a new trial. The survey concludes that there is a lack of such use of previous evidence, particularly in determining the sample size of the new trial. The findings of the survey set the background and the context of the methodology developed in this thesis.

The thesis aims to develop a simulation based methodological framework for designing new trials based on the power of the subsequent evidence synthesis model including the new trial, using several evidence synthesis models. The purpose of this methodology is to offer guidance to researchers in computing the sample size of a new trial on the basis that the new trial will eventually be a part of an evidence synthesis model. The variety of evidence synthesis methods includes standard meta-analysis methods, indirect comparison methods, mixed treatment comparison methods and meta-regression methods.

This approach treats the pooled effect size of the initial evidence synthesis model to be the mean of the predictive distribution of the new trial. The predictive distribution of the new trial provides the distribution of an effect size of a new trial that is deemed sufficiently similar to the existing trials to be eligible for the synthesis. Under each evidence synthesis method, we develop various models to design new trials by varying the variance of the predictive distribution. The power of the new trial is shown to increase with increased variance in the predictive distribution. In contrast, the power of the updated evidence synthesis is shown to decrease with the increased variance of the predictive distribution. The new trials designed using fixed effects principles yield the lowest power. However, the updated evidence synthesis model including the new trial designed using the fixed effects principles shows the highest power. This is a common phenomenon noticed in all evidence synthesis methods explored. The framework also includes a component that develops a methodology to design new trials using Bayesian meta-analysis principles. The reasons for the differences found in power results of the Bayesian and frequentist approaches are investigated.

The variance of the predictive distribution of a new trial clearly influences both the power of the new trial and the updated evidence synthesis. Trialists are advised to adapt the fixed effects method in designing new trials, whenever the assumption of a common true effect is possible. The Bayesian meta-analysis method developed here does not produce sufficient power in the updated meta-analysis and the methodology requires further refinements.
Acknowledgements

I would like to thank my supervisors Alex J Sutton and David R Jones for having given me the means to start this PhD. Needless to say, this thesis would have never been possible without the guidance of them. I am greatly beholden to them for putting so much faith in me from the very first day. I have greatly benefitted from David’s expertise and experience and from Alex’s enthusiasm for research.

Further, I acknowledge the department of Health Science of University of Leicester for sponsoring me during my doctoral programme and funding me for a Bayesian course in Venice, Italy.

I am deeply indebted to my mother, brother and sister and friends for their unconditional support during my spinal injury which delayed the completion of my PhD. Finally, I would have got nowhere without the help of Thilini, my unconventional wife and incredible friend. Ultimately, this thesis is dedicated to the memory of my father who sadly left us 19 years ago.
# Table of contents

1 Introduction ..............................................................................................................1
  1.1 Background........................................................................................................1
  1.2 Aims of the Thesis ............................................................................................5
  1.3 Thesis outline .....................................................................................................7

2 Literature Review on evidence synthesis methods .................................................9
  2.1 Evidence Based Medicine .................................................................................9
  2.2 Systematic Reviews .........................................................................................9
  2.3 Meta-Analysis ...................................................................................................11
  2.4 Different outcome measures ..........................................................................12
    2.4.1 Binary outcomes .........................................................................................12
      2.4.1.1 Odds ratio ..............................................................................................14
    2.4.2 Heterogeneity .............................................................................................15
      2.4.2.1 Clinical heterogeneity ..........................................................................18
      2.4.2.2 Statistical heterogeneity ......................................................................18
    2.4.3 Fixed effects method of meta-analysis ....................................................19
    2.4.4 Random effects method of meta-analysis ...............................................20
  2.5 Indirect comparison meta-analysis ..................................................................23
    2.5.1 Statistical methods for indirect comparisons ...........................................25
      2.5.1.1 Statistical methods for indirect comparisons using aggregate data ......25
      2.5.1.2 Modelling approaches using generalised linear mixed model ..........26
      2.5.1.3 Bayesian and likelihood based approach .............................................26
  2.6 Mixed treatment comparison / Network meta-analysis / Mixed treatment meta-analysis ........................................................................................................27
  2.7 Meta-regression ...............................................................................................30
    2.7.1 Residual heterogeneity .............................................................................32
    2.7.2 Fixed effects meta-regression ..................................................................33
    2.7.3 Random effects meta-regression ..............................................................33
    2.7.4 Limitations of meta-regression .................................................................35

2.8 Summary .............................................................................................................36

3 Literature review on designing new trials ..............................................................37
  3.1 Randomised controlled trials .........................................................................37
3.2 Sample size calculation ................................................................. 37
3.3 Statistical power ......................................................................... 40
3.4 Shortcomings of current practices in designing RCTs .................. 43
3.5 Designing new trials based on existing evidence ...................... 45
3.6 Summary .................................................................................. 51
4 Assessing the current practice of using existing evidence in designing RCTs ........................................................................... 52
4.1 Introduction .............................................................................. 52
4.2 Motivation ................................................................................ 52
4.3 Objectives ................................................................................ 53
4.4 Methods .................................................................................. 53
4.4.1 Data collection ...................................................................... 55
4.4.1.1 Basic level ....................................................................... 56
4.4.2 Statistical analysis ................................................................ 57
4.4.2.1 Yes/No level ................................................................... 57
4.4.2.2 Count level ..................................................................... 57
4.4.2.3 Summary level ............................................................... 57
4.5 Results ..................................................................................... 58
4.6 Discussion ................................................................................ 63
4.7 Conclusions ............................................................................ 65
5 Designing future trials based on existing evidence using frequentist meta-analysis methods .......................................................... 67
5.1 Introduction .............................................................................. 67
5.2 Motivation ................................................................................ 68
5.3 Objectives ................................................................................ 68
5.4 Introduction to meta-analysis .................................................... 69
5.5 Existing evidence ...................................................................... 70
5.5.1 Outcome of interest ............................................................... 71
5.5.2 Definition of an event ............................................................. 71
5.5.3 Estimating the events rate in the control group ..................... 71
5.6 Methods .................................................................................. 72
5.6.1 Null hypothesis tested and power in the updated meta-analysis ........................................................................ 72
5.6.2 Procedure ............................................................................ 72
5.7 Designing a new trial based on existing evidence using a fixed effects meta-analysis model ......................................................... 77
5.7.1 Fixed effect meta-analysis model ....................................................... 77
5.7.2 Initial fixed effects meta-analysis using existing evidence ............. 77
5.7.3 Specification of the predictive distribution of a new trial .......... 79
  5.7.3.1 The mean of the predictive distribution of the new trial ....... 79
  5.7.3.2 The variance of the predictive distribution of the new trial .... 79
5.7.4 Prediction model 1 – Designing a new trial based on fixed effects meta-analysis method ................................................................. 79
  5.7.4.1 Rationale .................................................................................. 79
  5.7.4.2 Methods ................................................................................. 80
  5.7.4.3 Results – Power of the updated fixed effects meta-analysis ... 81
5.8 Designing a new trial based on existing evidence using the random effects meta-analysis model ................................................................. 82
  5.8.1 Random effects meta-analysis model ........................................... 82
  5.8.2 Initial random effects meta-analysis of existing evidence ........ 83
  5.8.3 Random effects prediction models .............................................. 84
  5.8.4 Specification of the predictive distribution of a new trial ........... 85
    5.8.4.1 The mean of the predictive distribution of the new trial ....... 85
    5.8.4.2 The variance of the predictive distribution of the new trial ...... 85
    5.8.4.3 Hypothetical models ............................................................... 86
  5.8.5 Prediction model 2 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution is specified only using within study variation) .................................................. 87
    5.8.5.1 Rationale ............................................................................. 87
  5.8.6 Prediction model 3 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution is specified using the between study variations).................................................. 87
    5.8.6.1 Rationale ............................................................................. 87
  5.8.7 Prediction model 4 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution includes both within and between study variations- Normal distribution approach) 88
    5.8.7.1 Rationale ............................................................................. 88
  5.8.8 Prediction model 5 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution of a new
effect size includes both within and between study variance - t-distribution approach) .......................................................... 89

5.8.8.1 Rationale ................................................................. 89
5.8.9 Methods ........................................................................ 90

5.8.10 Results of the power of the updated meta-analysis of prediction model 2-5 ................................................................. 91

5.8.11 Sample size of the new trial in between 300 and 500 ............ 93

5.9 Power of the new trial in isolation ........................................ 98
5.9.1 Objectives ...................................................................... 98
5.9.2 Methods ......................................................................... 99
5.9.2.1 Null hypothesis and power ........................................... 99
5.9.2.2 Procedure ................................................................... 99
5.9.3 Results .......................................................................... 100

5.10 Comparison – The power of the new trial alone and that of the updated meta-analysis following the inclusion of the new trial ................. 102

5.11 Critical evaluation of the results ........................................... 104
5.11.1 Power of the new trial in isolation of the fixed effects model and the random effects model accounted for only within study variation .... 105

5.11.2 Increased power in the fixed effects updated meta-analysis following the inclusion of the new trial to the existing meta-analysis .... 109

5.11.3 Decreased power in the random effects updated meta-analysis following the inclusion of the new trial into the existing meta-analysis .... 110

5.11.4 Zero power of the updated random effects meta-analysis for smaller sample sizes of the new trial ........................................ 111

5.11.5 Increased power in the updated meta-analysis following the inclusion of the new trial based on the prediction model accounted both within and between study variation ................................. 113

5.11.6 Larger sample sizes of the new trial .................................. 113

5.12 Discussion ....................................................................... 115
5.12.1 Key findings of this chapter ........................................... 115
5.12.2 Related research .......................................................... 117
5.12.3 Limitations of the methodology ...................................... 118
5.12.4 Potential Future work ................................................................. 119

5.13 Conclusions .................................................................................. 120

6 Designing future trials based on existing evidence using indirect comparison meta-analysis methods ........................................................................ 122

6.1 Introduction ....................................................................................... 122

6.2 Motivation .......................................................................................... 123

6.3 Objectives ........................................................................................... 123

6.4 Introduction to indirect comparison meta-analysis .......................... 124

6.4.1 Naive method of indirect comparison ........................................... 124

6.4.2 Adjusted method of indirect comparison (Bucher’s method) ....... 125

6.4.3 Underlying assumptions ................................................................. 126

6.4.3.1 Homogeneity assumption ......................................................... 126

6.4.3.2 Similarity assumption ............................................................... 126

6.5 Existing evidence ............................................................................... 127

6.5.1 Trials comparing LEV add-on therapy and PLC ....................... 128

6.5.1.1 Initial list of trials comparing LEV and PLC ......................... 128

6.5.1.2 Updated list of trials comparing LEV and PLC ................. 128

6.5.2 Trials comparing TPM add-on therapy and PLC .................... 131

6.5.3 Definition of an event ................................................................. 132

6.5.4 Estimating the events rate in the control group ......................... 132

6.6 Methods ............................................................................................. 133

6.6.1 Null hypothesis tested and power in the context of indirect comparisons ................................................................. 133

6.6.2 Procedure ..................................................................................... 134

6.7 Designing a new head to head trial using indirect comparison principles ......................................................................................... 136

6.7.1 Designing a new trial based on existing evidence using indirect comparison meta-analysis methods (fixed effects method) ......... 137

6.7.1.1 Initial fixed effects meta-analysis .......................................... 137

6.7.1.1.1 Initial fixed effects meta-analysis of trials comparing LEV add-on therapy with PLC .................................................. 137

6.7.1.1.2 Initial fixed effects meta-analysis of trials comparing PLC with TPM add-on therapy .................................................. 138
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.7.1.2</td>
<td>Specifying the predictive distribution of a new head-to-head trial</td>
</tr>
<tr>
<td>6.7.1.2.1</td>
<td>The mean of the predictive distribution of the new head-to-head trial</td>
</tr>
<tr>
<td>6.7.1.2.2</td>
<td>The variance of the predictive distribution of the new head-to-head trial</td>
</tr>
<tr>
<td>6.7.1.3</td>
<td>Prediction model 6: Designing a new head-to-head trial using fixed effects indirect comparisons meta-analysis method</td>
</tr>
<tr>
<td>6.7.1.3.1</td>
<td>Rationale</td>
</tr>
<tr>
<td>6.7.1.3.2</td>
<td>Method</td>
</tr>
<tr>
<td>6.7.1.3.3</td>
<td>Results</td>
</tr>
<tr>
<td>6.7.2</td>
<td>Designing a new head-to-head trial based on indirect evidence using random effects meta-analysis model</td>
</tr>
<tr>
<td>6.7.2.1</td>
<td>Initial random effects meta-analysis</td>
</tr>
<tr>
<td>6.7.2.1.1</td>
<td>Initial random effects meta-analysis of trials comparing PLC with LEV add-on therapy</td>
</tr>
<tr>
<td>6.7.2.1.2</td>
<td>Initial random effects meta-analysis of trials comparing PLC with TPM add-on therapy</td>
</tr>
<tr>
<td>6.7.2.2</td>
<td>Specifying the predictive distribution of a new head-to-head trial</td>
</tr>
<tr>
<td>6.7.2.2.1</td>
<td>Mean of the predictive distribution of a new head-to-head trial</td>
</tr>
<tr>
<td>6.7.2.2.2</td>
<td>Variance of the predictive distribution of a new head-to-head trial</td>
</tr>
<tr>
<td>6.7.2.3</td>
<td>Prediction model 7: Designing a new head-to-head trial using random effects indirect comparisons meta-analysis methods (predictive distribution accounted for only the within study variation)</td>
</tr>
<tr>
<td>6.7.2.3.1</td>
<td>Rationale</td>
</tr>
<tr>
<td>6.7.2.4</td>
<td>Prediction model 8: Designing a new head-to-head trial using random effects indirect comparisons meta-analysis methods (The variance of the predictive distribution accounted for only the between study variation)</td>
</tr>
<tr>
<td>6.7.2.4.1</td>
<td>Rationale</td>
</tr>
<tr>
<td>6.7.2.5</td>
<td>Prediction model 9: Designing a new head-to-head trial based on random effects indirect comparisons meta-analysis methods (The predictive distribution accounted for both within and between study variance: mean heterogeneity approach)</td>
</tr>
</tbody>
</table>
7.7.1.2.1 Rationale

7.7.1.2.2 Methods

7.7.1.2.3 Results

7.7.2 Designing new head-to-head trial using random effects MTC meta-analysis methods

7.7.2.1 Specifying the predictive distribution of a new head-to-head trial

7.7.2.2 Prediction model 12: Designing a new head-to-head trial using random effects MTC methods: The predictive distribution includes only the within study variation

7.7.2.3 Prediction model 13: Designing a new head-to-head trial using random effects MTC methods: The predictive distribution includes only the between study variation

7.7.2.4 Prediction model 14: Designing a new head-to-head trial using random effects MTC methods: The predictive distribution accounts for both the within and between study variations: The mean heterogeneity approach

7.7.2.5 Prediction model 15 - Designing a new head-to-head trial using random effects MTC methods: The predictive distribution accounts for both the within and between study variations: The residual heterogeneity approach

7.7.3 Methods

7.7.4 Power results of prediction models 11-15

7.8 Critical evaluation of the results

7.9 Discussion

7.10 Conclusions

8 Designing future trials based on existing evidence using meta-regression methods

8.1 Introduction

8.2 Motivation

8.3 Objectives

8.4 Theory used

8.4.1 Meta-regression model
8.4.1.1 Fixed effects meta-regression model ................................................. 200
8.4.1.1.1 Differences to the fixed effects meta-analysis model ........ 201
8.4.1.1.2 Practical limitations of the fixed effects meta-regression model
........................................................................................................... 201
8.4.1.2 Random effects meta-regression model ............................................ 201
8.4.1.3 Weighted regression and residual heterogeneity ......................... 203
8.4.1.3.1 The weighted Least Square Estimation .................................. 203
8.4.2 The Hat matrix and leverage ............................................................. 204
8.5 Existing evidence .................................................................................. 206
8.5.1 Gathering existing evidence .............................................................. 206
8.5.2 Updating existing evidence ................................................................. 207
8.5.2.1 Searching for new trials ............................................................... 207
8.5.2.2 Incorporation of new trials into the existing evidence .............. 208
8.5.3 The consistency of the existing evidence included in the analysis 211
8.5.3.1 Different patient population and treatment periods ................. 211
8.5.4 Transforming three arm trials into two-arm trials ....................... 214
8.5.4.1 The correlation in the treatment effects of three-arm trials .... 215
8.5.5 Event of interest ................................................................................. 216
8.5.5.1 Definition of an event ................................................................. 216
8.5.5.2 Estimating the event rate in the control group ......................... 217
8.6 Development of the Methodology ....................................................... 217
8.6.1 Null hypothesis tested and power in the context of the meta-regression
.............................................................................................................. 218
8.7 Phase 1: The consistency of summary measures computed using
meta-analysis and meta-regression models ............................................. 219
8.7.1 Meta-analysis .................................................................................... 220
8.7.1.1 Model 1 - Random Effects meta-analysis that estimates the
heterogeneity using the Mantel-Haenszel method .......................... 220
8.7.1.2 Model 2 - Random effects meta-analysis that estimates the
heterogeneity using the inverse variance method ......................... 221
8.7.2 Meta-regression models without covariates ................................. 222
8.7.2.1 Model 3: Meta-regression based on method of moments
estimation and with Knapp and Hartung modification ................. 223
8.7.2.2 Model 4: Meta-regression based on method of moments estimation and without Knapp and Hartung modification......................... 223
8.7.2.3 Model 5: Meta-regression based on Residual Maximum Likelihood and with Knapp and Hartung modification....................... 223
8.7.2.4 Model 6: Meta-regression based on Residual Maximum Likelihood and without Knapp and Hartung modification.................... 224
8.7.3 Critical evaluation........................................................................................................... 224
8.8 Phase 2: The synthesis of existing evidence ...................................................... 226
  8.8.1 Objectives ............................................................................................................... 226
  8.8.2 Trials published before the end of year 2000 ................................................. 227
    8.8.2.1 Existing evidence ...................................................................................... 227
    8.8.2.2 Results ........................................................................................................ 228
  8.8.3 Trials published before the end of year 2006 ........................................... 230
    8.8.3.1 Existing evidence ...................................................................................... 230
    8.8.3.2 Results ........................................................................................................ 231
  8.8.4 Trials published before the end of year 2009 ........................................... 233
    8.8.4.1 Existing evidence ...................................................................................... 233
    8.8.4.2 Results ........................................................................................................ 234
  8.8.5 Critical evaluation........................................................................................................... 235
8.9 Phase 3: Methodology for designing a single new trial ...................................... 236
  8.9.1 Designing a new two-arm trial with specific covariates............................. 236
    8.9.1.1 Objectives ....................................................................................................... 236
    8.9.1.2 Existing evidence ...................................................................................... 237
    8.9.1.3 Methods ........................................................................................................ 237
    8.9.1.4 Prediction models.......................................................................................... 241
      8.9.1.4.1 Prediction model 16: Designing a new trial between PLC and LEV 1000 ......................................................................................... 241
      8.9.1.4.2 Prediction model 17: Designing a new trial between PLC and LEV 2000 ......................................................................................... 241
      8.9.1.4.3 Prediction model 18: Designing a new trial between PLC and LEV 3000 ......................................................................................... 241
    8.9.1.5 Results .............................................................................................................. 242
    8.9.1.6 Extrapolation .................................................................................................... 243
8.9.1.7 Critical evaluation................................................................. 245
8.9.1.7.1 Exploring the association between the predicted effect size and the leverage ........................................... 246
8.9.1.7.2 Exploring the association between the leverage and the p-value ................................................................. 249
8.9.1.7.3 Exploring the association between the predicted effect size and the p-value ........................................... 250
8.9.1.7.4 Exploring the association between the predicted effect size and the dose ......................................................... 252
8.9.1.7.5 Summary........................................................................ 253

8.9.2 Phase 4: Designing new three-arm trials using meta-regression methods ................................................................. 256
8.9.2.1 Objectives ....................................................................... 256
8.9.2.2 Existing evidence ................................................................. 256
8.9.2.3 Methods ........................................................................ 256
8.9.2.4 Prediction models................................................................. 259
8.9.2.4.1 Prediction model 19: Designing a new trial comparing PLC and LEV 1000 with 2000...................................................... 260
8.9.2.4.2 Prediction model 20: Designing a new trial comparing PLC and LEV 1000 with 3000...................................................... 260
8.9.2.4.3 Prediction model 21: Designing a new trial comparing PLC and LEV 2000 with 3000...................................................... 260
8.9.2.5 Results........................................................................ 260
8.9.2.6 Critical evaluation................................................................. 263
8.9.2.6.1 Key findings.................................................................. 263
8.9.2.6.2 Factors influencing the power of the updated meta-regression after including a three-arm trial ....................... 264
8.9.2.6.3 Recommended trial designs............................................. 266
8.9.2.6.4 Future work..................................................................... 266

8.10 Phase 5: Methodology to design multiple new trials ............. 267
8.10.1 Designing two trials of two arms using meta-regression methods ................................................................. 268
8.10.1.1 Objective ....................................................................... 268
8.10.1.2 Existing evidence ................................................................. 268
8.10.1.3 Methods ........................................................................ 268
8.10.1.4 Prediction models................................................................. 269
  8.10.1.4.1 Prediction model 22: Designing two new trials comparing PLC and LEV 1000 ...................................................... 269
  8.10.1.4.2 Prediction model 23: Designing two new trials comparing PLC and LEV 2000 ...................................................... 269
  8.10.1.4.3 Prediction model 24: Designing two new trials comparing PLC and LEV 3000 ...................................................... 269

8.10.1.5 Results .................................................................................. 270

8.10.2 Phase 6: Designing five trials using meta-regression methods 273
  8.10.2.1 Objectives ........................................................................... 273
  8.10.2.2 Existing evidence ................................................................. 273
  8.10.2.3 Methods ............................................................................. 273
  8.10.2.4 Prediction models ............................................................... 274
    8.10.2.4.1 Prediction model 25: Designing five new trials comparing PLC and LEV 1000 ...................................................... 274
    8.10.2.4.2 Prediction model 26: Designing five new trials comparing PLC and LEV 2000 ...................................................... 274
    8.10.2.4.3 Prediction model 27: Designing five new trials comparing PLC and LEV 3000 ...................................................... 275
  8.10.2.5 Results ................................................................................. 275

8.11 A dose-specific comparison considering power of varying number of trials ....................................................... 278
  8.11.1 LEV 1000 .............................................................................. 278
  8.11.2 LEV 2000 .............................................................................. 280
  8.11.3 LEV 3000 .............................................................................. 281

8.12 Critical evaluation .................................................................... 283

8.13 Discussion .................................................................................. 285
  8.13.1 Key findings of this chapter ..................................................... 285
    8.13.1.1 A single two arm trial ......................................................... 285
    8.13.1.2 Three arm trials ................................................................. 285
    8.13.1.3 Multiple two-arm trials ...................................................... 286
  8.13.2 Related similar work ............................................................ 286
  8.13.3 Strengths and limitations of this methodology ..................... 287
  8.13.4 Potential future work ........................................................... 289
9.9.5.2 The sensitivity analysis on different prior distributions of the between study standard deviation .................................................. 313

9.9.5.2.1 Prediction model 29 - Designing a new trial based on random effects Bayesian meta-analysis methods with an inverse gamma prior on $\tau$ .............................. 314
9.9.5.2.2 Prediction model 30 - Designing a new trial based on random effects Bayesian meta-analysis methods with an uniform prior on $\tau$ 316

9.9.5.3 Critical evaluation of the sensitivity analysis ...................... 317

9.10 Designing a new trial based on existing evidence using Bayesian random effects meta-regression model .................................................. 319

9.10.1 Existing evidence ............................................................... 319
9.10.2 Null hypothesis .................................................................. 319
9.10.3 Event rate in the control group ............................................ 319
9.10.4 Prediction model 31: Designing a new trial based on random effects Bayesian meta-regression methods .................................................. 320

9.10.4.1 Rationale ........................................................................ 320
  9.10.4.1.1 Centring the covariates .............................................. 320
  9.10.4.1.2 Bayesian meta-regression model ................................. 321
  9.10.4.1.3 Prediction distribution of the new trial .......................... 321
  9.10.4.1.4 Specification of prior distributions .............................. 322
9.10.4.2 Methods ......................................................................... 323
9.10.4.3 Results ........................................................................... 323

9.11 Critical evaluation ................................................................. 327

9.12 Discussion ............................................................................ 329
  9.12.1 Key findings of this chapter ............................................... 329
  9.12.2 Related similar work .......................................................... 329
  9.12.3 Critical evaluation of the results ......................................... 330
  9.12.4 Future work ...................................................................... 331

9.13 Conclusions ......................................................................... 333

10 The comparison and reconciliation of discrepancies in power estimates yielded using frequentist and Bayesian random effects meta-analysis ........................................ 335

10.1 Introduction .......................................................................... 335
10.2 Motivation ............................................................................ 336
10.3 Objectives ............................................................................ 336
10.4 Methods ........................................................................................................................................... 336

10.4.1 Designing Bayesian models restricting the uncertainty in the estimation of $\tau$ ........................................................................................................................................... 337

10.4.1.1 Model 1 – Prior distribution of $\tau$ is centred on the mean of $\tau$ of the frequentist random effects analysis ............................................................................................................................. 339

10.4.1.2 Model 2 – Prior distribution of $\tau$ is centred on the mean of $\tau$ of the Bayesian random effects analysis ............................................................................................................................. 339

10.4.1.3 Model 3 – Prior distribution of $\tau$ is centred on the median of $\tau$ of the Bayesian random effects analysis ............................................................................................................................. 340

10.4.2 Results ............................................................................................................................................. 341

10.4.3 Critical evaluation .............................................................................................................................. 342

10.4.4 The comparison of the frequentist confidence limits and Bayesian credible limits .................................................................................................................................................... 343

10.4.4.1 The lower confidence / 2.5 % credible limit of the pooled Odds ratio .............................................................................................................................................................................. 344

10.4.4.1.1 Overall comparisons ...................................................................................................................... 346

10.4.4.2 The upper confidence / 97.5 % credible limit .................................................................................... 347

10.4.4.2.1 Overall comparisons ...................................................................................................................... 349

10.4.4.3 The mean of the pooled odds ratio .................................................................................................. 350

10.4.4.3.1 Overall comparisons ...................................................................................................................... 352

10.4.4.4 Critical evaluations ......................................................................................................................... 353

10.4.5 Rationalising the bi-modality of the Bayesian and frequentist distributions of the pooled OR of the updated MA ........................................................................................................................................ 354

10.4.5.1 Exploring the distributions of the pooled odds ratio and the pooled log odds ratio of the updated MA ........................................................................................................................................... 355

10.4.5.2 Exploring the distributions of the treatment group event rate in the new trial .............................................................................................................................................................................. 359

10.4.5.3 Exploring the distributions of the heterogeneity of the updated MA .............................................................................................................................................................................. 361

10.4.5.4 Exploring the distributions of the odds ratio and the log odds ratio of the new trial .............................................................................................................................................................................. 363

10.4.5.5 Exploring the scatter plots between the pooled OR of the updated MA and OR of the new trial ........................................................................................................................................... 366

10.4.5.6 Exploring the scatter plots between the heterogeneity of the updated MA and the OR of the new trial ........................................................................................................................................... 368
<table>
<thead>
<tr>
<th>10.4.5.7 Exploring the scatter plots between the pooled OR and the heterogeneity of the updated MA</th>
<th>370</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5 Critical evaluation</td>
<td>374</td>
</tr>
<tr>
<td>10.6 Discussion</td>
<td>377</td>
</tr>
<tr>
<td>10.7 Conclusions</td>
<td>382</td>
</tr>
<tr>
<td>11 General discussion and conclusions</td>
<td>383</td>
</tr>
<tr>
<td>11.1 Key findings</td>
<td>383</td>
</tr>
<tr>
<td>11.2 Related research</td>
<td>386</td>
</tr>
<tr>
<td>11.3 Critical appraisal of the proposed methodology</td>
<td>387</td>
</tr>
<tr>
<td>11.3.1 Limitations of the proposed methodology</td>
<td>389</td>
</tr>
<tr>
<td>11.4 Recommendations and Future work</td>
<td>390</td>
</tr>
<tr>
<td>11.4.1 Future applications</td>
<td>390</td>
</tr>
<tr>
<td>11.4.2 Future methodological aspect</td>
<td>391</td>
</tr>
<tr>
<td>11.5 Conclusions</td>
<td>393</td>
</tr>
<tr>
<td>12 Appendixes</td>
<td>395</td>
</tr>
<tr>
<td>13 Bibliography</td>
<td>397</td>
</tr>
</tbody>
</table>
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBM</td>
<td>Evidenced Based Medicine</td>
</tr>
<tr>
<td>FE</td>
<td>Fixed Effects</td>
</tr>
<tr>
<td>HPD</td>
<td>Highest Posterior Density</td>
</tr>
<tr>
<td>IPD</td>
<td>Individual Patient Data</td>
</tr>
<tr>
<td>LEV</td>
<td>Levetiracetam</td>
</tr>
<tr>
<td>MA</td>
<td>Meta-Analysis</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte-Carlo</td>
</tr>
<tr>
<td>MR</td>
<td>Meta-Regression</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>PLC</td>
<td>Placebo</td>
</tr>
<tr>
<td>RE</td>
<td>Random Effects</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised Controlled Trial</td>
</tr>
<tr>
<td>SR</td>
<td>Systematic Review</td>
</tr>
<tr>
<td>TPM</td>
<td>Topiramate</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

1.1 Background

Medical research can only be done as a partnership between the medical professionals and the public. Although medical research saves human lives all over the world, it poses a risk to humans as well. Part of this risk associated with medical research is unavoidable. For example, a risk is always posed on trial participants of the control group when a less superior new treatment is compared with the standard treatment. However, the concern is of the risk that could be avoidable but incurred due to bad research practice (Young and Horton 2005). The bad research is not only inappropriately conducted research but also unnecessary research, conducted research that remains unpublished and published research with no justification of its relevance and existence (Fergusson, Glass et al. 2005).

Randomised Controlled Trials (RCTs) that are irrelevant and unable to justify their contribution to the given field are not only unnecessary and unethical but also a waste of resources, which could have been invested in more worthy work (Fergusson, Glass et al. 2005). They also abuse the volunteers and patients who place trust in trial investigators. Such research contributes little or nothing to the scientific record (Chalmers 2005). Researchers have a moral responsibility towards the research participants to make sure their study is the most appropriate design.
RCTs are the major type of research that informs health care decisions. To serve this purpose, RCTs should be designed and reported in the light of totality of existing evidence (Clarke 2004). The public and other healthcare decision makers seem to entrust RCTs as a basis of informing health care decisions since RCTs are increasingly encouraged to be designed and reported in the light of similar existing evidence (Clarke, Hopewell and Chalmers 2010).

Researchers who do not consider existing evidence before embarking on design a new trial indirectly avoids to learn from the successes and failures of previous trials (Clarke 2004). They also take a risk of answering a research question for which the answer is already known (Clarke 2004). Duplicate research is not only a failure of the integrity of the researchers but also of the institutes, which funded the research, the ethical bodies that approve the research and importantly the journals which carrying on publishing them, although they contribute nothing to the existing evidence base (Chalmers 2005). Authors claim that theirs is the first should explain how they reach this conclusions (Clark and Horton 2011).

The structure of any scientific research should be conceptualised in terms of four questions: ‘Why did you start? What did you do? What answers did you get? What does it mean anyway?’ (Hill 1965). To reflect this idea the common structure of a scientific paper comprises an Introduction, Materials and Methods, Results and a Discussion of the findings. The introduction section should indicate and justify the requirement for designing the trial whereas the discussion section should interpret their findings and highlight how their findings contribute to existing evidenced base (Clarke, Hopewell and Chalmers 2007).
In fact, almost every aspect of the clinical trial design could be informed by previous evidence (Cooper, N. J., Jones and Sutton 2005) including the process of identifying, refining and justifying the hypothesis; estimating and justifying the sample size; defining and justifying the primary outcomes; and recognizing and avoiding pitfalls of previous work (Egger, Davey and Altman 2000). In designing new clinical trials, the use of previous evidence is of prime importance particularly to avoid making flawed sample size calculations resulting in underpowered trials (Brasher and Brant 2007). However, relatively little confirmation is found that new trials designs are currently informed by existing evidence (Cooper, N. J., Jones and Sutton 2005).

Systematic reviews and especially meta-analyses are widely acknowledged as a basis for new research initiatives. They strive to prevent duplicate trial designs by ensuring new trials are built on the lessons from previous findings and set the results of new trials in the context (Clarke, Hopewell and Chalmers 2007). The meta-analysis (MA) in particular can provide a quantitative and explicit framework to inform the design of new trials in areas where some evidence exists. Although unarguably MA provides the highest level of evidence regarding the effectiveness of interventions, there is a lack of evidence to claim that when designing individual clinical trials, future meta-analyses are formally considered.

The conventional approach of justifying the development of a further new trial is often based on the consensus about the effectiveness of the intervention from a large MA of clinical trials. This conventional approach has been criticised for not considering the impact of the new trials on existing evidence (Ferreira, Herbert
et al. 2012). Often, existing relevant evidence is overlooked in the planning and designing stages of new trials, despite their importance and ability to inform various aspects of the trial design. The current practice of designing new clinical trials has been criticised for not adequately considering relevant previous research evidence (Robinson and Goodman 2011).

With the rapid increase in the number of reviews included in the Cochrane library, it has never been easier to find a context to which a new research would readily fit in (Clarke and Chalmers 1998). But evidence is available to claim that systematic reviews are under-utilised in practice to inform new trial design (Cooper, N. J., Jones and Sutton 2005). Their study reveals only 8% of trials used relevant Cochrane reviews to inform the design of a new trial.

Even though, to a certain degree, (According to Cooper and Jones et al 2005, 33% of new studies have been influenced by a review) future research is informed by previous research, how exactly previous research is used to underpin the designing of new clinical trials is often not transparent (Cooper, N. J., Jones and Sutton 2005). The previous research literature, particularly review results, could be qualitatively used to inspect results to identify clear gaps in evidence base, as well as quantitatively to calculate sample size of new trials (Sutton, A.J. , Cooper and Abrams 2003).

The lack of citation of previous evidence is not caused due to space limitations in journals (Robinson and Goodman 2011). With increasing popularity of electronic publishing, the barrier of space limitation has been removed (Clarke 2004). Hence, the lack or limited consideration of existing evidence in designing
new trials can mainly be attributable to the absence of an explicit and coherent framework of methodologies, rather than a philosophical objection to the concept (Sutton, A. J., Cooper et al. 2007).

Accordingly, the overall objective of this thesis is to develop a framework of methodologies based on evidence synthesis methods, to inform the design of future trials. The next section explicitly outlines the key aims of the thesis.

### 1.2 Aims of the Thesis

Section 1.1 above demonstrates and justifies the need for this particular piece of research. This section sets out the key aims of the thesis. The ultimate goal of the thesis is to contribute to the development of valid methodologies of designing future trials using evidence synthesis methods of existing evidence via both frequentist and Bayesian simulation methods (Sutton, A. J., Cooper et al. 2007). The project comprises following elements of aims to achieve the overall goal.

1. To investigate and assess the extent to which existing evidence is currently employed in practice to inform the design of a new RCT. Moreover, it is aimed to establish the forms and the frequency of use of existing evidence to inform new trial design. It is further expected to identify which aspects of trial design are often informed by existing evidence.

2. To extend the methodology developed by Sutton et al (Sutton, A. J., Cooper et al. 2007) to design future trials based on existing evidence
using MA methods, focusing on the design aspect of sample size
calculation of a new trial. To explore the possibility of using various other
evidence synthesis methods such as Indirect Comparison (IC) methods
and Mixed Treatment Comparison methods (MTC), to inform the new
trial design, in situations where numerous treatment regimens exist for a
given clinical condition (Caldwell, D.M., Ades and Higgins 2005).

3. To explore the potential of using Meta-Regression (MR) models to
develop a framework to identify potential treatment modifiers and to
incorporate treatment modifiers as covariates in the regression model to
inform the new trial design.

4. To examine the use of Bayesian evidence synthesis methods particularly
the meta-analysis and meta-regression methods to inform the planning
and designing of future trials. Moreover, to compare and contrast the
findings of the frequentist and the Bayesian approaches.

5. To develop an array of STATA commands to facilitate the computation of
statistical power of updated evidence synthesis models, following the
inclusion of new trial/(s).
1.3 Thesis outline

Subsequent to Chapter 1, which briefly discusses the background to the problem and setting up the aims of the thesis, Chapter 2 reviews the literature pertaining to evidence synthesis methods.

The Chapter 3 further reviews the literature concerning the designing aspects of clinical trials with special focus on sample size calculations. The statistical power is given a special attention since the focus in this thesis is to design trials to achieve a certain power of the updated MA including the new trial.

The Chapter 4 explicitly assesses the current practices of using evidence synthesis methods in relation to designing new clinical trials via a survey. Two sets of samples of articles reporting randomised controlled trials are drawn and scrutinised with an intention to identify the extent to which previous evidence are used to inform the planning and design of each trial.

The Chapter 5 uses frequentist MA methods, both fixed effects (FE) and random effects (RE) models, to develop methodologies to inform the next trial design. A simulation-based approach is adopted to compute the power of the updated MA including the new trial. To enable a comparison of power results, the power of the individual trials is considered as well. Moreover, this chapter examines the sensitivity of power to the changes of the variance component in the predictive distribution.

The framework of methodology is further expanded in Chapter 6 by examining the possibility of using IC meta-analysis methods to inform the design of future
trial design. The MTC meta-analysis methods are used to develop methodologies to design future trials in Chapter 7.

The Chapter 8 implements and develops the methodology to design future trials using MR methods. This enables to design a new trial to include a single covariate with specific value, for example dose.

The use of Bayesian evidence synthesis methods to underpin the framework of methodologies developed to inform the design of future trials is explored in Chapter 9. Further, it attempts to explore the discrepancies of the findings between the frequentist and Bayesian approaches. The chapter 10 compares and reconciles the power results of frequentist and the Bayesian MA models.

Chapter 11 concludes the thesis by highlighting the key findings of the thesis, relating the findings to other work in the field. Further, the potential benefits and limitations of the proposed methodologies are discussed. In addition to that, future research to address the areas, which has not been investigated in this thesis, is outlined. Some additional information can be found in the appendixes.

Following the outline of objectives and aims of the thesis and its structure, an extensive review of the literature on evidence synthesis methods is carried out in Chapter 2.
CHAPTER 2

Literature Review on evidence synthesis methods

2.1 Evidence Based Medicine

Evidence Based Medicine (EBM), the practice of making decisions about the care of individual patients based on the current best evidence, has now become commonplace (Sackett, Rosenberg et al. 1996). EBM entails integrating the individual clinical expertise with available best external clinical evidence from systematic research. The best available clinical evidence is acquired mainly through reviewing and summarising patient centred individual clinical research. When there is more than one study on a given topic, where no one supersedes others, a review is often taken place, which carefully considers evidence and put forward an answer to the therapeutic questions in concern (Slavin 1995). The process of drawing the best available evidence and produce quantitative summaries is called evidence synthesis. Evidence synthesis enables to identify areas of agreements in science and areas where discrepancies exist and further research is required (Mosteller and Colditz 1996). It facilitates bringing together the findings from separately conducted studies.

2.2 Systematic Reviews

EBM often requires consideration of the totality of the available evidence when evaluating interventions and new technologies. The explosion in the scientific
publications demands a formalised approach and generally accepted set of scientific tools to review, summarise and combine data from a large set of published studies. SR aims to tackle these issues by offering a formal mechanism to synthesis evidence from multiple studies.

SR is a scientific tool used to appraise, summarise and results and implementation of quantities of research (Green, S. 2005). “Systematic Review is an overview of primary studies those used explicit and reproducible methods” (Greenhalgh 1997). It offers a systematic approach in identification and evaluation of research. Moreover, it offers objective interpretations and reproducible conclusions. Using SRs provides the busy policy makers easy access to the evidence. Performing a SR requires not only a thorough search of the relevant literature but also a protocol describing the inclusion and exclusion criteria, which is explicit and independent from the final results of primary studies.

The Cochrane collaboration conduct SRs to provide well informed decisions about the effectiveness of healthcare interventions, using their specific methodology, and published in the Cochrane library (Chalmers 1993). Depending on whether the evidence is adequately similar enough to be combined, and the subsequent results could be interpreted sensibly, a MA, which is the statistical synthesis may be conducted.
2.3 **Meta-Analysis**

The field of statistics has provided a set of tools to combine data from different studies, although at the outset the circumstances of the application are not well defined. Gene Glass in 1976 introduced a new ground breaking methodology of synthesising findings from multiple trials (Glass 1976), where he referred to MA as the analysis of analysis. MA is the statistical analysis of a collection of analytic results with an aim to pool the findings (DerSimonian, Rebecca and Laird 1986). It is a statistical procedure that integrates the results of several combinable studies (Huque 1988). MA has been defined as ‘the rubric used to describe the quantitative methods for combining evidence across studies’ (Muncer, Taylor and Craigie 2002).

MA provides the highest level of evidence regarding the effectiveness of interventions. Although people use SR and MA interchangeably, MA is an optional component of a SR. MA has become an increasing practice in medical research over the past few decades, in establishing the overall effect of a specific treatment regime from several alternative clinical trials/studies examining the efficacy of similar treatment protocols. Many medical interventions have been implemented based on the results of the MA of a SR of randomised controlled trials (RCT).

The aim of undertaking a MA may be much broader from that of conducting individual clinical trials (Thompson, S.G. 1994). Individual clinical trials attempt to investigate the effectiveness of a particular treatment regimen on a selected group of patients for specified time duration on a pre specified outcome measure. In contrast MA aims to combine all available evidence of a particular
clinical condition by including trials with a variety of treatment regimens, types of patients and outcomes (Thompson, S.G. 1994).

It is inevitable that subjectivity to a certain degree is involved in a scientific review, even though a praiseworthy attempt to secure the objectivity is made. Subjective judgements may involve in deciding the relevant studies to be added to MA and in determining whether and how heterogeneity should be investigated (Thompson, S.G. 1994).

2.4 Different outcome measures

Outcomes need to be measured from individual trials before they can be quantitatively combined. Three types of outcome measures are found in practice depending on the data they derived from, i.e. binary, continuous and ordinal data. Those based on binary data such as if the patient is alive or dead, diseased or non-diseased are discussed in section 2.4.1 as methodological framework developed in forthcoming chapters are based on the odds ratio (OR) scale. Outcome measures based on continuous data such as blood pressure and outcomes based on ordinal data, for example, disease severity scale are not discussed in detail here.

2.4.1 Binary outcomes

Two types of binary data are found in the medical literature namely descriptive and comparative. When the interest lies on non-comparative studies, such as non-controlled trials, the outcome measure is descriptive as opposed to comparative, e.g. odds of an event.
The vast majority of health care trials are comparative in nature. Hence, they report comparative measures in which the outcome of two groups is being compared. RCTs are inherently comparative. So are many observational studies, such as case-control studies. Due to this comparative nature of the outcomes measured in most trials/studies, binary data has become the most common type of data in health research. In situations where comparative binary outcome measures are considered, a 2x2 table for each study is often available in the original report or possible to construct. This discloses all the information required to compute the commonly used outcome measures. A typical 2x2 table producing the results of a RCT is presented in figure 2-1 below.

Figure 2-1: A typical 2 x 2 table of a RCT of binary outcome measures.

<table>
<thead>
<tr>
<th></th>
<th>Failure/Dead</th>
<th>Success/Alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Control</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

When the data from prospective trials are available in the form of a 2 x 2 table, binary measures such as risk ratios, ORs and risk differences can easily be computed. Although the OR is less intuitive than the risk ratio, it inherits important statistical properties those make the OR the best choice in MA (Emerson 1994).

Hence, OR is deemed to be the most suitable outcome measure to base the methodologies developed in this thesis (Sutton, A. J., Cooper et al. 2007). However, the methodologies developed in this thesis could be extended to be
able to work with other outcome indices as well. A detailed explanation of the OR is given in the forthcoming section.

2.4.1.1 Odds ratio

The odds ratio (OR) is the ratio of the odds between the treatment and the control group. Odds of an event is the ratio between the number of patients having and not having the event (Borenstein, M., Hedges et al. 2009). Transformation of OR into log OR is recommended in the synthesis, because the distribution of the log OR is closely approximated by a normal distribution (Sutton, A., Abrams et al. 2000). OR can be defined as below in equation 2.1, where a, b, c and d represent the appropriate cells in figure 2.1.

\[
OR = \frac{ad}{bc} \tag{2.1}
\]

In a RCT settings where a new treatment and a control is compared, and the outcome of interest considered undesirable (e.g. death), an OR estimate less than one indicates that the new treatment is superior than the control and OR of greater than one implies that the new treatment is not as good as the control.

The large sample variance of the log OR (Fleiss 1993) is given by

\[
var_{\ln}(OR) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \tag{2.2}
\]

In situations where there are no events or all events in either of the treatment arms, equation 2.2 is undefined. A continuity correction is required such as adding 0.5 to each cell of figure 2.1 before calculating the OR.
The point estimates of effect sizes from different studies are varied mainly due to sampling variation, which is always found in every estimate. Often the variability of the effect size estimates exceed that could have occurred from the sampling variation alone, but differences exist between studies. This extra variability is called heterogeneity and it needs special attention in the synthesis. The section 2.4.2 below discusses exclusively about heterogeneity with special attention to identifying and quantifying heterogeneity as well as to deal with it when conducting a MA.

2.4.2 **Heterogeneity**

Heterogeneity is referred to the surplus variation in the true treatment effect over the variation that is expected from sampling errors (Thompson, S. G. and Sharp 1999). Heterogeneity is the variation between the true effect sizes of primary studies. Heterogeneity causes the observed effect size to be overestimated. (DerSimonian, Rebecca and Laird 1986). The observed effect sizes are partly spurious as they incorporate the true variation along with the random error (Borenstein, M., Hedges et al. 2009). The quantification of heterogeneity requires partitioning the observed variation into these two components.

Heterogeneity could be present when all or most of the effect sizes are in the same direction but their magnitudes are different, or when the effect sizes are in different directions (Sutton, A., Abrams et al. 2000). Several tests based on $\chi^2$ and $F$ statistics have been proposed but the Cochran’s test based on $Q$ statistic (Cochran 1954) is widely used to assess the presence of heterogeneity, when
the number of studies is small. Under the null hypothesis that the true treatment
effects of all primary studies are the same, $Q$ statistic is approximately take a $\chi^2$
distribution with $k-1$ degrees of freedom (where $k$ is the number of trials in the
MA). However, this test of homogeneity has also been proven to be of low
statistical power (Hardy and Thompson 1998) when the number of combined
studies is small. Therefore, even a reasonable level of genuine heterogeneity
may not be identified as statistically significant. To overcome this, it is now a
customary practice in MA to use a significance level of 0.10 following Fleiss’s
recommendation (Fleiss 1986).

Arguably, the test of homogeneity is not relevant as given the diversity of the
MA, it is inevitable that heterogeneity is present in trials irrespective of the
results of the test of homogeneity (Thompson, S. G. and Higgins 2002). Since
the low power of the test based on $Q$ statistic, a visual inspection of graphical
representations of data is complementary to the statistical tests. These indicate
which studies are the sources of greatest heterogeneity and possible outliers,
etc. The plot of normalised $z$ scores (Greenland 1987), forest plot and L’abbe
plots (L’Abbe, Detsky and O’Rourke 1987) are useful as a exploratory tool in
MA.

An investigation into the sources of heterogeneity would improve the clinical
and scientific relevance of the results of a MA (Thompson, S.G. 1994). The
differences in the design and the conduct of studies cause the heterogeneity. In
RCT settings, the differences in the underlying risk (Brand and Kragt 1992), the
size of the dose of the treatment, the compliance rate, the length of follow up
(Thompson, S. G. 1993) , etc cause to change the effect estimates and hence
increases the heterogeneity. Hughes et al (Hughes, Freedman and Pocock 1992) and Green et al (Green, S. J., Fleming and Emerson 1987) have highlighted the impact of early stopping rules on the heterogeneity. Heterogeneity is caused by patient level covariates too.

Although Cochrane’s Q statistics is widely accepted as a measure of dispersion between studies, Higgins et al (Higgins, J. P. T., Thompson et al. 2003) developed a new index called $I^2$ index which measures the proportion of the heterogeneity out of the total observed variation. Based on a simulation which compared Q statistic with $I^2$ index, Huedo-Medina et al (Huedo-Medina, Sánchez-Meca et al. 2006) report that the utility of $I^2$ index should complement Q statistic and in the synthesis of small number of studies both metrics suffer from the problem of low power.

Different methods have been proposed to deal with heterogeneity. The use of random effects MA model, discussed in detail in the section 2.4 is commonplace in the presence of the heterogeneity. Incorporation of patient level and/or study level covariates to explain the heterogeneity via a MR analysis is discussed in the section 2.7. If the studies are not sensible to combine due to heterogeneity, a subgroup analysis is proven to be useful technique which groups similar studies.

The sources of heterogeneity inducing clinical differences need to be investigated in order to interpret overall results in a clinically meaningful manner. Based on the aspect of studies that causes studies to be different, heterogeneity has two major classifications namely clinical and statistical
heterogeneity. These are discussed in detail in sections 2.4.2.1 and 2.4.2.2 below.

2.4.2.1 **Clinical heterogeneity**

Clinical conditions of almost every clinical trial are considerably different from one another. Important design aspects like patient selection criteria, baseline severity of the disease, treatment techniques and regimens, management of intermediate outcomes, and the duration of follow up are usually substantially different between trials. This subsequently leads for the overall effect estimate to be heterogeneous and thus referred as ‘clinical heterogeneity’ (Thompson, S.G. 1994).

2.4.2.2 **Statistical heterogeneity**

Clinical trials are prone to have methodological differences in the process of randomisation and handling withdrawals, etc. Due to those methodological differences of unknown or unrecorded characteristics, the results of individual trials become incompatible and this is called ‘statistical heterogeneity’ (Thompson, S.G. 1994). Statistical heterogeneity may be caused because of using undue scale to measure treatment effects. For example use of absolute measures rather than relative measures can induce statistical heterogeneity (Thompson, S.G. 1994). Uncontrolled or non-randomised studies those usually with greater potential to bias, are usually prone to greater heterogeneity.

Publication bias is another potential cause of statistical heterogeneity.

Depending on the assumption that the heterogeneity is present or not, two models of MA are proposed, namely the fixed effects and the random effect
model. Detailed discussions of the fixed and random effects models are given in the section 2.4.3 and 2.4.4 respectively.

### 2.4.3 Fixed effects method of meta-analysis

The fixed effect method of MA assumes that all studies in the synthesis are estimating the same underlying effect size, hence the term ‘fixed’. The factors affecting the effect size are assumed to be the same in each study. The reason for the variation in the observed effect size is only the random error inherent in each study.

In the presence of heterogeneity, the validity of the assumption of estimating one single underlying treatment effect is questionable. The true underlying effect size (μ) is estimated by combining primary studies. The pooled effect size estimate (\( \hat{\mu} \)) is assumed to be normally distributed (Hardy and Thompson 1998) when deriving the confidence interval for the underlying true effect size (μ).

The general fixed effect model is introduced by Birge (Birge 1932) and Cochrane (Cochran 1937), which is based on assigning weights to each study proportional to their precision (inverse variance) to compute an overall effect size. This method is commonly known as inverse variance-weighted method. The variance of the pooled underlying effect is minimised when the weights assigned are inversely proportional to the variance of each study.

Specific methods to combine ORs using FE model have been developed. Mantel and Haenszel (Mantel and Haenszel 1959) presented a method of combining OR, which is commonly known as Mantel-Haenszel method for
combing OR. Sato (Sato 1990) developed a method directly work on OR scale as opposed to log OR. Peto proposed a modification to Mantel-Haneszel method, which can be used even when the cells in 2 x 2 table of individual studies are zero (Peto, Pike et al. 1977). Conditional and unconditional maximum likelihood estimates of the pooled OR have been developed. Emersion (Emerson 1994) conducted an exclusive review of traditional and the computer intensive newer methods.

2.4.4 Random effects method of meta-analysis

The synthesis of primary studies should be conducted using the random effects MA model if the test of homogeneity described in section 2.4.2 is statistical significant. DerSimonian and Laird (DerSimonian, Rebecca and Laird 1986) developed the random effects MA model, which assumes that the true underlying treatment effects ($\theta_i$) vary in each study. All studies are assumed to be a random sample from a population of studies. The underlying effect sizes vary about the mean ($\mu$) of this distribution. It is this mean that is deemed to be the combined effect estimate in the random effect MA model. (Borenstein, M., Hedges and Rothstein 2007)

The random effects MA model deals with two levels of variation issues, namely the within study and between study variations. The number of subjects within each trial is associated with within study variation ($\sigma_i^2$) and the total number of studies on the other hand is associated with between study variation ($\tau^2$). Both these elements are integral in estimating the precise combined effects. Each trial is used to obtain an estimate of the true effect size specific to a particular
population. All observed effect size estimates ($Y_i$) are then pooled to estimate the mean of the true effects ($\mu$).

A common between study variation ($\tau^2$) is assumed in the random effects MA scenarios. The estimate of the between study variation ($\hat{\tau}^2$) which is derived from the data is found to be a very good representation of the true heterogeneity. True heterogeneity parameter is unknown and conventionally the estimate ($\hat{\tau}^2$) is used in place of the parameter. In situations where the number of studies is less, the uncertainty in the estimation is greater (Thorlund, Wetterslev and Gluud 2008).

Four different methods; the Weighted and Un-Weighted Least Squares (WLS, UWLS) and Maximum and Restricted Maximum Likelihood (ML and REML) have been proposed to estimate the RE model. A weighted non iterative approach based on methods of moment is developed by DerSimonian and Laird (DerSimonian, Rebecca and Laird 1986) to estimate the pooled effect of the RE model. This method requires each study to be assigned a weight inversely proportional to the sum of both within and between study variances. Using this method reduces the relative weightings assigned to larger studies, and increases the relative weightings given to the smaller studies.

The ML and REML estimation methods assume the normality of the distribution of the underlying effects parameters. DerSimonian and Laird compared these four estimation methods and claimed that ML estimates of $\tau^2$ is slightly lower than that of REML and WLS methods (Sutton, A., Abrams et al. 2000). This is
because MLE procedures do not adjust the pooled effect estimate for the fact \( \hat{\tau}^2 \) is estimated from data itself.

The confidence interval of the overall effect estimate of the RE model is markedly wider than that of the FE counterpart, because it creates extra width as the between study variation is also accounted for. Conducting a fixed effects MA in the presence of the heterogeneity causes the confidence interval of the overall treatment effect to be artificially narrower (Thompson, S. G. and Pocock 1991).

Methods have been proposed to incorporate the uncertainty caused by estimating the between study variation in to the random effects MA model. Hardy and Thompson (Hardy and Thompson 1996) used a profile likelihood method, which assumes the normality of the data and calculated the confidence regions. Biggerstaff and Tweedie (Biggerstaff and Tweedie 1997) developed a new approach that considers the variation in the estimation of \( (\hat{\tau}^2) \) in assigning weights to individual studies. The Bayesian approach automatically accounts for the variation in the estimation of \( \tau^2 \).

Concerns have been raised of the RE model about the inaccurate estimation of \( \hat{\tau}^2 \) when the number of studies is low (Hardy and Thompson 1996). It is further criticised based on making unjustified distributional assumptions (Peto 1987). The RE model is prone to publication bias as it assigns relatively larger weights on smaller studies (Greenland 1994).

The goodness of fit of the MA model to the data is very rarely carried out in practice. This is partially because of the lack in methodological development in
that area. Moreover, MA is embraced as an integral part of the data processing procedure than simply a model fitting exercise.

RCTs are the most valid design to evaluate the efficacy between competing treatments regimes. Direct head-to-head trials are not always available between the competing interventions. IC methods are increasingly used in MA to compare the competing interventions based on a common comparator. The next section reviews the literature pertaining to IC meta-analysis.

### 2.5 Indirect comparison meta-analysis

In EBM a large proportion of health interventions lack of direct evidence that links the health intervention to the health outcome of interest (Eddy, D.M. Hasselblad, V. Shachter, R. 1992). This is primarily because the drug licensing process requires the investigators to demonstrate the relative efficacy of the drug in concern only with placebo. This requirement enables drug developers to evade comparing their drug with other competing drugs. However it is clinically important to identify the most superior intervention out of all available intervention options used to treat in a particular health condition, especially in the decision making process as to what treatment should be administered in a particular clinical condition.

RCTs comparing relative effectiveness of every available competing intervention are not often found. Results of those direct head-to-head trials with active competitors, may lead to undermine the particular drug. Therefore, head-to-head trails are not always encouraged by drug developers. Therefore, it is
now increasingly common to carry out IC to evaluate the relative efficacy of competing interventions.

The results of the IC are interpreted as if they are results from direct comparisons within RCTs. In situations where both direct and indirect evidence are available it is recommended to consider both options separately and direct evidence should take precedence over the indirect evidence (Yazdanpanah, Sissoko et al. 2004). However, if the direct evidence is not adequate to undertake a statistical analysis, it might require the need for borrowing strength from IC. Higgins and Whitehead have introduced the concept of borrowing strength from indirect evidence (Higgins, J. P. T. and Whitehead 1996).

Song et al conclude (Song, F., Altman et al. 2003) that the results of the adjusted IC more often agree with the results of head-to-head RCTs than not. They further suggested that the validity of the findings of the adjusted IC lies on internal validity and the similarity of the included trials. In contrast to that a recent research by the same authors (Song, F., Xiong et al. 2011) recommended assessing the consistency of the direct and indirect evidence before combining. They observed significant inconsistency between the direct and indirect evidences, when the analyses include a few trials, subjectively assessed outcomes and statistically significant results in either direct or indirect comparison.

Limited consideration to IC is received in the decision making process due to their relative uncertainty (Ades, Anthony E., Madan and Welton 2011). Decisions about the approval of biologics are made based on the placebo-
controlled trials. Therefore, the utility of the evidence of indirect comparison is to rule out treatments that would otherwise be in use.

To reduce the bias caused by different patient characteristics across trials, Signorovitch et al (Signorovitch, Wu et al. 2010) developed a methodology, which makes use of all available data by adjusting the average patient characteristics in trials with IPDs to match those reported with aggregate data.

2.5.1 **Statistical methods for indirect comparisons**

Glenny and the colleagues (Glenny, Altman et al. 2005) conducted a survey to identify and evaluate the statistical methods available to conduct an indirect comparison. These statistical methods are derived from methods to investigate heterogeneity in MA. Three statistical methods have been suggested in indirect comparison scenarios.

2.5.1.1 **Statistical methods for indirect comparisons using aggregate data**

The very first methodological discussion of the context of indirect comparison is found in Eddy et al (Eddy, D.M. Hasselblad, V. Shachter, R. 1992), although earlier examples of indirect methods applications are found in the literature. Bucher and colleagues (Bucher, Guyatt et al. 1997) gives an explicit account of the generic adjusted IC method in which they illustrated the methodology specifically using a OR outcome to MA of trials. This generic method is subsequently used in relative risk, risk ratio (Hasselblad and Kong 2001) and hazard ratio outcome measures (Packer, Antonopoulos et al. 2001). The utility
of the adjusted IC method is discussed by Fisher (Fisher, Gent and Buller 2001) in the assessment of the superiority of a drug over placebo, when placebo controlled trials have not been conducted.

### 2.5.1.2 Modelling approaches using generalised linear mixed model

If Individual Patient Data (IPD) is available, an IC can also be analysed in a regression framework using a generalised linear model (GLM) too. In binary outcomes the 2 x 2 frequency table reflects IPD but for continuous and survival outcomes IPD needs to be requested from original authors. A logistic regression approach has been discussed by Hasselblad (Hasselblad 1998), when conducting IC using IPD. These approaches can be used in a simple IC scenario and could readily be extended to deal with scenarios that are more complex.

### 2.5.1.3 Bayesian and likelihood based approach

Bayesian approaches offer more flexibility in analysing IC. Higgins and Whitehead (Higgins, J. P. T. and Whitehead 1996) implemented a Bayesian approach to evaluate the relative efficacy of two drugs based on a common comparator. The confidence profile method (Eddy, D.M. Hasselblad, V. Shachter, R. 1992) is often considered a fully Bayesian approach, which can be used to synthesis any sort of evidence pertaining to a particular question. Any study making different treatment comparisons can be incorporated into this model. Importantly, this model can explicitly model any bias. Ades (Ades, A.E.
2003) also presented a similar approach of synthesising different kind of evidence in his multi-parameter evidence synthesis model.

### 2.6 Mixed treatment comparison / Network meta-analysis / Mixed treatment meta-analysis

In clinical situations where more treatment options are available, the question often raises is that which one is superior and how to rank them based on their relative benefits in order to choose the best. This triggers the development of ‘umbrella systematic review’ (Bialy 2006) which then leads to the invention of the MTC. The MTC is a generalisation of standard pair wise MA. It is also termed as ‘Mixed Treatment meta-analysis’ and/or ‘Network meta-analysis’ (Salanti, G., Higgins et al. 2007).

The MTC offers a platform to combine both direct and indirect evidence of relative efficacy of competing treatments. When the direct evidence is inconclusive and IC is a possibility, then combining both indirect and direct evidence in a MTC framework will strength the inferences of the relative efficacy of treatments (Lu and Ades 2004). These inferences are more uncertain and associated with wider confidence intervals (Song, F., Glenny and Altman 2000).

In clinical conditions to which multiple treatment options are available, direct head-to-head comparisons are often available only for a subset of treatments. The MTC provides a mechanism to perform simultaneous IC between the
treatments for which direct comparisons do not exist. This makes it possible to obtain estimates of all relative efficacies of all possible pair wise comparisons. Each treatment is then ranked based on their efficacy. The MTC is arguably an extension to the IC when more than two treatments are involved and it offers the flexibility of dealing with trials with multiple arms.

Comparing different treatment options constitute a network. Methodology for combining data from network of trials is developed via network MA. The concept of network geometry and the asymmetry have been introduced by Salanti et al (Salanti, G. , Higgins et al. 2007). The concept of network geometry that is the pattern of direct comparison within a network, has been explored elsewhere (Salanti, G., Kavoura and Ioannidis 2008). A network meta-analytic technique, using linear mixed models is presented by Lumley (Lumley 2002), to estimate both the heterogeneity in the effect of any given treatment and the inconsistency in the evidence of any given pair of treatments.

In an attempt to extend their initial model Lu and Ades (Lu and Ades 2006) developed a method to incorporate more than one reference treatment into the MTC framework. Moreover, it offers a method to estimate the inconsistency between the direct and indirect evidence. Salanti introduced the concept of network asymmetry, (Salanti, G. , Higgins et al. 2007) which refers to the extent to which specific treatments or specific treatment comparisons appear heavily in the network of MTC. Two methods of evaluating asymmetry have been proposed. The first method focuses to ascertain whether a specific comparison appears in the network more than it is expected by chance. The second method
is aimed to ascertain whether a specific treatment appears more than others in the network.

Caldwell (Caldwell, D.M., Ades and Higgins 2005) used the results of the MTC to determine the probability that a specific treatment being the greatest for a particular condition using MCMC methods. The commentary following the ad hoc network MA meeting work group (Li, Puhan et al. 2011) suggests the requirement of further methodological developments to demonstrate the validity of the evidence and to ease the interpretation of the findings. They further suggested the need for rigorously designed and carefully conducted SRs to enable validating the findings of network MA.

In an attempt to evaluate the comparative validity of the IC with MTC approach, O'Regan and colleagues (O'Regan, Ghement et al. 2009) conclude that although both adjusted IC and MTC yield similar results and benefits in simple situations where a common comparator is involved, the MTC is favoured over the IC in more complex scenarios.

The MTC methods are not without their inherent limitations. Due to the complexity, the utility of these models is limited. The nature of Bayesian approach involves a subjective judgement in specifying prior distributions and the validity of the judgments are questionable. Despite the elegance of the model presented by Lu and Ades, critics have been raised as to the model demanding information that is not available.
2.7 **Meta-regression**

Exploring the sources of heterogeneity is one of the main difficulties faced by a meta-analyst (Colditz, Burdick and Mosteller 1995). MA often makes very little or perhaps no attempt to explain the reason for obtaining disparate treatment effects and to explore the factors which cause disparate treatment effects. Moreover, it does not make any adjustments to allow for the fact that it combines different populations (Simmonds and Higgins 2007). Reporting standards and guidelines for both RCTs (Moher, D., Cook et al. 2000) and observational settings (Stroup, Berlin et al. 2000) require, wherever possible, to assess the degree of heterogeneity, to explore the possible sources of heterogeneity. In failing this, heterogeneity should be incorporated into the analysis to reflect in the estimates and policy decisions.

Understanding possible causes of heterogeneity and incorporating it into the analysis could improve the clinical relevance and scientific value of the results of the MA (Berlin 1995). Increasing interest has been noticed in last two decades about using MA to investigate and estimate the potential treatment covariate interactions. In situations where only aggregated or summary data of study level is available, Meta Regression (MR) is proven to be the effective statistical method to explore the treatment-covariate interactions (Simmonds and Higgins 2007).

Berkey proposed a method to incorporate a covariate to explain the heterogeneity in the synthesis of 2 x 2 table (Berkey, Hoaglin et al. 1995). Thompson (Thompson, S. G. and Sharp 1999) compared different forms of weighted regression and random effects logistic regression methods, and
concluded that methods allow for additive component of heterogeneity should be used.

The MR can be an extension to the random effects MA, which determines the extent to which the between study variation is explained by the study level characteristics (Lau, Ioannidis and Schmid 1998). The MR is essentially an attempt to combine meta-analytic principals with regression ideology (Sutton, A. J. and Higgins 2008). The MR attempts to relate one or more study/patient characteristics with the observed effect sizes (Thompson, S. G. and Higgins 2002). Example of such characteristics (study level covariates) may be the average duration of follow up, a geographical measure of each study, or some measures specific to study, etc.

The MR also be seen as an attempt to determine and quantify the association between the observed treatment effects (Dependent variable) and study level covariates (Independent variables) (Lau, Ioannidis and Schmid 1998). It fits in to the regression framework when attempting to predict effect sizes for given study level covariate. Identification of the association of patient level covariates to the treatment effect could assist in determining the type of patients who would benefit or harm from a particular intervention. This would eventually aid to individualise treatments for specific groups of patients.

The techniques to explore the association between the study specific design characteristics and the treatment effects are straightforward. However, investigations of patient level characteristics to explain the heterogeneity is not trivial. This is simply because often only the aggregated or average summaries
of character of interests are included in trial reports, such as a percentage of patients with a characteristic of concern, etc (Lambert, Sutton et al. 2002).

In simple regression, a unit of observation is the individual patient, whereas in MR a unit of observation could be either a study or a sub group. Sometimes, the unit of assessment could be an arm (either treatment or control) or an arm crossed with an outcome (all patients in the treatment group with mortality). MR is most useful when a relatively larger number of studies are combined. When the number of studies is relatively smaller, having several covariates to explain the heterogeneity in the regression model produces spurious results. As a rule of thumb, at least 10 studies are required for each covariate included in the regression model (Borenstein, M., Hedges et al. 2009).

Even if the attention is restricted to RCTs, the nature of the covariates is inherently observational because it is not possible to randomised patient to one covariate value or another. Hence interpretation and inference of MR inherits all the difficulties attached to observational studies (Dias, Sutton et al. 2011).

### 2.7.1 Residual heterogeneity

Although the covariates or treatment modifiers explore the extent to which the between study variation can be explained, it is highly unlikely that covariates individually or in combination explain all associated heterogeneity. Unexplained heterogeneity is always left with in this process and it is called the residual heterogeneity (Thompson, S. G. and Higgins 2002). Various estimation methods including empirical Bayes estimates and restricted maximum likelihood have been developed to quantify the amount of residual heterogeneity.
2.7.2 **Fixed effects meta-regression**

Greenland originally introduced the fixed effects method of MR (Greenland 1987). Hedges subsequently implemented a fixed effect MR model, which could include discrete or continuous predictor variables. This model assumes the approximate normality of the outcome variable (Greenland 1987). Unlike in fixed effects MA model this model does not set $\theta_i, \ldots, \theta_k$ to equal to $\theta$. Moreover, this model allows for some variation accounted for by the predictor variables but not by the random effects.

Stuck et al introduced a fixed effects MR model which utilises a weighted logistic regression approach (Stuck, Siu et al. 1993), to which a patient level or study level covariate could be fitted. Studies are assigned weights based on the number of patients who have and do not have an event. This model requires fitting $2k$ number of cases per study, where $k$ represents the number of arms in the study. The FE model is most appropriate when all the variations can be explained using sampling error. Further, fixed effect MR method has been used by several other authors (Thompson, S. G. 1993) (L'Abbe, Detsky and O'Rourke 1987). In the presence of heterogeneity fixed effects MR is said to produce spurious results (Higgins, J. and Thompson 2004).

2.7.3 **Random effects meta-regression**

Generally, under the random effects MR model, the underlying treatment effect is regressed on a covariate. This model is most suitable when only a part of the variation is explained by the predictive covariates, and a random effect term is used to explain the residual variation. The residual variation is accounted for by
including a random effect to the regression and hence the name ‘random effects’. Since unexplained residual variation is always available in the model, it is argued that a random effects component should always be included in the model.

Berkey (Berkey, Hoaglin et al. 1995) proposed an alternative random effects MR model in which estimation is based on an iterative process. However, this model is criticised as being prone to small biases. Thompson et al introduced a moment based approach of estimation when only one covariate is included in the model (Thompson, S. G. and Sharp 1999). They also compared different approaches of MR and concluded that in many situations the difference in the results between the models is very small. Friedrich et al (Friedrich and Knapp 2011) developed a new method to construct confidence interval for regression coefficients using a principle based on generalised inference.

Stram (Stram 1996) developed a general random effects MR model framework, from which most of other meta analytic models can be viewed as a special case. Lambert and Abrams developed a multilevel model, which could implement weighted random effects MR model (Lambert and Abrams 1995).

In situations where the effectiveness of the treatment depends on the severity of the disease, studies may appear heterogeneous because of the differences in the baseline risk. The baseline risk is usually calculated using the risk of an event in the control group. The baseline risk has been included in the MR model in an effort to explain the heterogeneity between treatment effects (Thompson, S. G., Smith and Sharp 1997) (Arends, Hoes et al. 2000). Inherent
correlation between the observed ORs and the observed baseline risk cause the results to be biased (Sharp, Thompson and Altman 1996). Conducting the MR with large number of studies or studies with larger number of subjects may circumvent this issue. Knapp and Hartung developed (Knapp and Hartung 2003) new test statistics based on the improved estimator of the variance of the parameter estimates, to replace the commonly used test on parameters in random effects MR with one covariate. Because methods designed to explore the heterogeneity are prone to misleading false-positive results, the permutation test proposed by Higgins et al (Higgins, J. and Thompson 2004), to assess the statistical significance of the regression coefficients, is recommended to use before claiming any significant relationship.

2.7.4 **Limitations of meta-regression**

MR is not without its own limitations or restrictions. The main criticism is that MR is prone to ecological bias. The discrepancies between the observed and the underlying association between the treatment effects and covariates, lead the MR to have ecological bias (Berlin, Santanna et al. 2002). The cause of these discrepancies is the use of aggregated data of patient level characteristics. The association of an individual patient characteristic to the treatment effect, is not as the same as the association of the average patient characteristic to the treatment effect (Simmonds and Higgins 2007).

Berlin (Berlin and Antman 1994) has discussed the pitfalls and limitations of MR approach. MR has also been proven to have low statistical power in detecting treatment covariate interactions, particularly the variation in the covariates are
relatively small across studies. Moreover, the degree of freedom available is relatively small when the number of studies included in the MR is low. The effects of individual covariates are not distinguished when the covariates tend to be highly collinear.

The use of IPD in the context of MR holds many benefits including the opportunity to re-analyse the treatment effects, to include explanatory covariates. The advantages of using IDP meta-analysis have earned it the recognition as ‘The Gold Standard’ for MA. Concerns have been raised as to whether the benefits outweigh the costs and time involved in IPD analysis. Moreover, the effort was never justified over the results.

2.8 Summary

The aim of this thesis is to develop a framework of methodologies to design future trials based on existing evidence using evidence synthesis methods. Chapter 2 complies with the early aims of the thesis by reviewing the relevant literature covering the fundamental issues concerning the basic evidence synthesis methods including MA, IC meta-analysis, MTC meta-analysis and MR. The following Chapter 3 further reviews the literature and discusses various aspects and issues pertaining to trial designing.
CHAPTER 3

Literature review on designing new trials

3.1 Randomised controlled trials

Randomised controlled trials (RCTs) are the accepted gold standard in evaluating new health care interventions in EBM (Sackett, Rosenberg et al. 1996). The primary objective of conducting a RCT is to compare the efficacy of two or more treatment regimes. Specific objectives may be to compare the superiority of one treatment over the other, the non inferiority of one treatment with respect to another, or perhaps to ascertain the equivalence of two treatments (O'Hagan, Stevens and Campbell 2005).

Designing a RCT needs careful consideration particularly in selecting the target patient group, the precise therapeutic question and the number of subjects required. Trials that include too few subjects than required lack statistical power to detect small but potentially beneficial therapeutic effects. Conversely, using too many subjects than required in a trial is a waste of resources. Moreover, it is unethical because of the allocation of two many people than required to a treatment which might have been already proven to be inferior.

3.2 Sample size calculation

An important aspect requiring careful consideration in designing RCTs, is the determination of required number of subjects to include in the trial. The
importance of determining an accurate sample size for a clinical trial has widely been acknowledged (Charles, P., Giraudeau et al. 2009). In accordance with the CONSORT statement, the process of sample size calculation needs to be adequately reported and justified in published reports (Moher, D., Schulz et al. 2001).

A clinically important treatment effect can be observed, if the number of subjects in the clinical trial exceeds a certain limit. Too many or less subjects in a clinical trial yield oversized or underpowered results and would not achieve the balance between the resources and the subjects. Underpowered trials may fail to detect a clinically important significant treatment effect whilst oversized trials are a waste of both resources and time. Whenever possible both underpowered and oversized designs should be avoided.

Charles et al in their review identified the use of erroneous methods and inaccurate assumptions in calculating sample sizes of clinical trials reported in medical journals even with high impact factor (Charles, Pierre, Giraudeau et al. 2009). Another issue raised in this review is the frequent absence in reporting the parameters required to calculate sample size in clinical trials.

It is apparent from above evidence of the involvement of elements of discrepancy and misconception in the process of calculating sample size. Despite the importance of having the right size of the sample, methods available to calculate the sample size of RCTs are quite simplistic (Sheppard 1999). In RCT settings, calculation of the sample size considers the minimum treatment difference, which is clinically worthwhile. These values are often
taken from previous relevant trials. Expert opinion is sought, in the absence of previous relevant trials.

Despite receiving widespread criticism (Sutton, A. J., Cooper and Jones 2009) of trials being underpowered, sample size calculations are often reverse-engineered, where the treatment difference assumed in the power calculation is derived having considered the economic and feasibility restrictions on the trials (Schulz and Grimes 2005).

No formal consideration of existing evidence is given in sample size calculation as well. Surprisingly only a little literature is available on informing the sample size of a new trial based on previous trials. A methodology has been recently developed recommending to consider the statistical power of future meta-analyses including the existing trial, in designing new clinical trials (Sutton, A. J., Cooper et al. 2007).

The Importance of sample size planning is acknowledged in estimating accurate parameters too (Maxwell, Kelley and Rausch 2008). Gould et al assess the impact of between study variability on the likelihood of determining true treatment effects. They observed a decrease in statistical power of the existing study to compensate for the added between study variability and suggested a substantially large sample size is required to achieve at least 90% power of detecting a true treatment effects. They developed a methodology to determine the sample size, which compensates for the effect of between-study variability (Gould, Koglin et al. 2009).
3.3 **Statistical power**

The scope of any trial producing experimental data should be sufficient to believe that it will detect the hypothetical effects over the experimental error (Feiveson 2002). In other words, the trial should be adequately powered to be able to detect even the smallest effect that is believed as statistically significant. Determining the power of the statistical test is integral in planning any scientific research. Calculating the statistical power is also important in assessing the ability of the hypothesis to detect any effects within the sample that is believed to be in the population.

The statistical power refers to the probability of rejecting the null hypothesis when the alternative hypothesis is true (Hallahan and Rosenthal 1996). It is the probability of not making a type II error. The power and the type II error are inversely related. To make sense of the probability of the event of 'rejecting the null hypotheses' has to be random.

Statistical power of a test is an integral component in designing a new clinical trial. Once the hypothesis to be tested, the expected minimum clinically relevance effect size and the level of significance are decided upon then the sample size required in the trial could be determined by considering the required power (Sedlmeier and Gigerenzer 1989).

The sample size of a trial is expected to be just large enough to achieve a certain level of power, conventionally 80%. The expected treatment effect is based on the judgment concerning what is the clinically relevance minimum effect size required for effective patient management or the judgment.
concerning what is the anticipated effect size of a new treatment. The choice of clinically important minimum treatment different is not straightforward and involves an element of arbitrariness.

In the context of simulation in which a unique experiment can be run repetitively with collecting new data and testing a new hypothesis at each run, the proportion of the rejections of the null hypothesis is taken as an estimate of the power (Feiveson 2002).

Calculating statistical power of tests related to MA is as important as tests in primary research. In the context of MA, methods to calculate statistical power have not been discussed widely. Hedges et al discussed methods to calculate statistical power of tests involving in MA (Hedges, L. V. and Pigott 2001), including tests on the mean effect sizes, tests of the heterogeneity of effects size parameters and tests for contrasts among effect sizes. The results suggested that the power of meta-analyses is not necessarily very high, as one would expect. The test of heterogeneity in particular is found to be of low statistical power.

In a review of MA carried out by Arnqvist et al (Arnqvist and Wooster 1995) suggested that the use of MA methods drastically reduces the probability of making a type II error, hence increase the power. However, this review is criticised (Peterman 1995) on the basis of not producing any supporting numerical evidence but simply claiming that MA increases statistical power.

In the context of MA, a prior power analysis is required to determine the likelihood of achieving a statistically significance result given the size of the
study, the sample size within a study, the level of significance and the anticipated overall effect size (Hedges, L. V. and Pigott 2001). These parameters are typically unknown before conducting an analysis and estimating involves either scientific guessing or expert opinions. In random effects MA, power analysis becomes slightly complicated due to the involvement of estimating the between studies variation.

Hedges et al have demonstrated that the power of the fixed effects test in MA is rather low. They have further demonstrated that in the presence of heterogeneity of effects the power of mixed effect test is even lower (Hedges, L.V. and Pigott 2004).

It has been proven that incorporating power analysis into MA settings may reduce the risk of misleading conclusions being made (Cohn and Becker 2003). Based on the outcome of the power analysis recommendations could be made about changing the protocol of the MA.

O'Hagan and colleagues introduced an alternative notion to power called 'assurance', which is more relevant in the context of decision-making. They argued that the concept of power is conditional whereas the assurance is not. The power is actually the probability of rejecting the null hypothesis conditional to pre-specified value of the effect estimate. Therefore, the power does not quantify the probability of the trial ending with the null hypothesis being rejected. Conversely, the assurance is specified as being the unconditional probability that the trial yielding a specific outcome (O'Hagan, Stevens and Campbell 2005).
The concept of power has proven to be paramount in developing methodologies in clinical trials. The familiarity of this concept by many users ensures its continued use.

3.4 **Shortcomings of current practices in designing RCTs**

RCTs are regarded as the most reliable form of scientific evidence in healthcare regarding the effectiveness of interventions. RCTs play a major role in making healthcare decisions and setting up guidelines. Therefore, RCTs received substantial attention in biomedical reporting. No other category of biomedical reporting has received such sustained attention apart from RCTs, reflecting their importance in informing healthcare decisions (Clarke and Chalmers 1998).

The shortcomings and deficiencies of designing aspects of RCTs have been previously identified (Hemminki 1982). They particularly looked at the deficiencies in describing the material and methods as well as in analysing and presenting of results. Clarke and colleagues gave a systematic scrutiny of the quality of the introduction and the discussion sections of RCTs (Clarke and Chalmers 1998).

Researchers often make flawed sample size calculations leading to generate underpowered trials (Brasher and Brant 2007). To address issues related to design, relevant existing research evidence should be used to inform the design of new RCTs. Many design elements of RCTs can be informed by existing
research evidence including the hypothesis, the sample size calculation and the primary outcomes.

Currently the use of related existing evidence in informing the design of new RCTs is not commonplace. Although, to a lesser extent, existing research results are used to inform the design of new RCTs, Cooper and colleagues doubt the transparency in which how they had been used (Cooper, N. J., Jones and Sutton 2005).

A tendency of citing only a selective set of relevant previous research has been identified. Previous trials with dramatic results are more likely to be cited than those found ordinary results (Campbell 1990). The authors of a particular country selectively cited literature generated from within that particular country (Gotzsche 1987) (Gibbs 1995).

The piecemeal reporting of research delivers conflicting messages that often confuses the public. This practice further causes individual trials to be looked at in isolation from the whole body of relevant research literature (Clarke and Chalmers 1998). No trial is an island in its own right but every trial is a part of the continent (Clarke and Chalmers 1998). In fact, the nature and the amount of future research should depend on the nature and the amount of existing evidence, not simply because a MA becomes inconclusive (Roloff, Higgins and Sutton 2012).

The practice of not citing existing evidence in designing new trials is concluded to be a research malpractice (Smith, Alison and Goodman 1997) and often
happens due to the absence of universally accepted methodology and stringent guidelines.

### 3.5 Designing new trials based on existing evidence

The fundamental principle in clinical research is that a trial should address a research question needs answering and it should be designed to produce a meaningful answer to the concerned question (Robinson and Goodman 2011). The knowledge and the understanding of relevant preceding research are essential in accurately identifying and addressing these questions (Clarke and Chalmers 1998).

Healthcare decision makers are overwhelmed by the accumulation of new research evidence (Clarke 2004). Not all the evidence is of good quality that could directly inform healthcare decisions. Thus, healthcare decision makers including the public have to identify the relevant good quality research amidst huge amount of information before making well-informed decisions.

Much of the problems related to said information overloading could be eased if new research findings are discussed in the totality of relevant and already existing past research. Systematic reviews and meta-analyses in particular provide a mechanism to summarise and synthesise previous similar evidence that in turn aid to inform healthcare decisions.
It is essential that researchers have adequate knowledge about previous research findings to ensure their trial design is the most appropriate for addressing the question concerned. Researchers have a responsibility towards the participants, that the trial design is optimal. To justify the trial design ethically as well as scientifically, it has to be designed in the light of scientifically defensible assessment of relevant previous research. Therefore, trial has to begin by referring to appropriate systematic reviews to provide a justification to new trial design.

Every RCT contributes to the accumulation of evidence and omission of any RCT causes to underestimate that evidence. Access to previous cumulative evidence in designing new trials is required not only to compute accurate inferences but also to indentify gaps in the literature (Robinson and Goodman 2011). Importantly, uncited evidence does not influence the inferences.

In 1996, the original Consolidated Standards of Reporting Trials (CONSORT) statement put forward a set of evidence-based recommendations on reporting trial findings (Begg, Cho et al. 1996). This statement facilitates the completeness and transparency of reporting of RCTs. It further recommends that the discussion section in a new trial should place the new trial in the context of an up to date systematic review to aid users a more convenient assessment of the quality of the RCT.

The CONSORT statement further recommends authors to discuss individual trial results in the light of totality of available results (Moher, D. , Schulz et al. 2001). This requires authors of RCTs to conduct an up-to-date systematic
review within or alongside the trial and to incorporate or refer it in the discussion section of the trial report.

Several major healthcare journals including the Annals of Internal Medicine, the BMJ, the JAMA, The Lancet, and the New England Journal of Medicine took the recommendations on board. Subsequently, they changed and/or upgraded their policy on reporting RCTs. The BMJ took the groundbreaking step requiring reports to include a summary of what is already known of a subject and of what new evidence adds into the existing evidence.

Clarke and colleagues assessed the extent of major healthcare journals of adherence to the CONSORT recommendations in 1997 (Clarke and Chalmers 1998), 2002 (Clarke, Alderson and Chalmers 2002) and 2005 (Young and Horton 2005) respectively.

In 1997, Clarke and Chalmers examined 26 published RCT reports from the JAMA. One of 26 reports claimed of being the first to address a particular question. The remaining 25 reports are not the very first to address the research question concerned. Surprisingly, only 2 of 25 reports have placed results of the trial in the context of a systematic review (Clarke and Chalmers 1998).

Clarke and colleagues conducted a new review again in 2001, attempting to detect any improvement since 1997 in respect of setting trial results in the context of an up-to-date systematic review (Clarke, Alderson and Chalmers 2002). They examined 33 RCT reports published in five major healthcare journals. The findings are discouraging as only three reports have attempted citing relevant systematic reviews. However, they did not attempt to integrate
results in to a systematic review. No improvement is detected about placing clinical trial results in the context of updated systematic review of relevant studies, in reviews carried out in the space of 4 years.

In 2005, Clarke and colleagues reviewed (Clarke, Hopewell and Chalmers 2007) both introduction and discussion sections of controlled trial reports published in major healthcare journals. They looked at introduction sections to assess the extent to which trials justifying their relevance by referring to a relevant systematic review. They found that only 5 of 18 reports cited a systematic review in the introduction section. None of the 15 reports (not the first trial addressing the research question) sets the trial results in the context of an updated systematic review. Discussion sections of only 5 reports referred to relevant systematic reviews.

Young and Horton (Young and Horton 2005) argued that trials are not published in a way that they justify their existence and relevance. They further highlighted the failure and negligence of journals publishing the unjustifiable research. In recognition, the Lancet agreed to update its policy in this regard. From August 2005, authors of clinical trials submitting to the Lancet need to produce a summary of previous research findings addressing the same research question. In addition to that, the relation of existing and new evidence should be established by referring to a MA or a systematic review.

Some European funding agencies and Centres for Medicare and Medicaid Services now require authors to find existing evidence when designing and reporting RCTs (Robinson and Goodman 2011). A policy which requires
authors to conduct a structured search of relevant trials or at least a description of the search strategy, could be enforced by the International Committee of Medical Journal Editors, just as trial registration is required (Robinson and Goodman 2011).

The use of previous evidence in the justification of trial design is increasingly recognised and now it has become mandatory for applicants to the Medical Research Council (MRC) to refer to a systematic review of evidence to justify the need of the proposed trial (Cooper, N. J., Jones and Sutton 2005).

A recent review (Robinson and Goodman 2011) assessed the extent of citing existing trials in new trial reports and suggested that only less than 25% of existing evidence is cited. A median of 2 preceding trials are cited in reports of new trials. This review is a complete systematic assessment of over four decades covering a full range of health disciplines.

A trial designed without accounting for preceding evidence addressing the same therapeutic question is not only unethical and waste of resources but also poses unnecessary risk to patients being randomised (Young and Horton 2005).

As mentioned before, Sutton and colleagues developed a methodology of designing new trials based on evidence synthesis models (Sutton, A. J., Cooper et al. 2007). Chen and Pei (Chen and Pei 2009) took the concept on board and applied it in designing the next trial in their study.

Based on the recommendations of Sutton et al (Sutton, A. J., Cooper et al. 2007) of the minimal power increase in Bayesian MA settings following the
inclusion of even a larger trial in the presence of heterogeneity, Moran et al (Moran, Graham et al. 2010) excluded a large trial not in consistent with the rest of the trials.

Ferreira et al highlighted the problem of the conventional approach of justifying a further trial as being not considering the potential impact of the new trial on existing evidence. They further developed a method to identify the possible impact of a new trial on the updated MA based on extended funnel plots using benefit-harm trade off models (Ferreira, Herbert et al. 2012).

Roloff et al (Roloff, Higgins and Sutton 2012) argued that in the presence of existing evidence, use of the conditional power is more appropriate than the unconditional power to determine the required additional information size, because some information is already exist that will contribute towards the MA. Based on this argument, they went onto propose a very similar method in spirit to that by Sutton et al (Sutton, A. J., Cooper et al. 2007) but based on conditional power, which avoids the need for simulations, but account for existing evidence to determine the information size required in a further trial.

In an attempt to extend the sample size calculation approach of Sutton et al (Sutton, A. J., Cooper et al. 2007) towards MR, Rotondi et al developed new trials with the goals of examining the modifying effect of study level covariates on observed treatment effect in the light of current available evidence (Rotondi, Donner and Koval 2012).

Universally accepted methodology to inform the design of a new trial based on existing evidence is currently in high demand. It would certainly prove to be a
timely investment. The adherence to such a methodology facilitates designing well-powered clinical trials and prevents designing unethical and unjustified trials.

### 3.6 Summary

This chapter reviews the literature of designing new RCTs with special attention to methodological aspects of sample size calculations. A brief review on statistical power in evidence synthesis methods is undertaken as well. This chapter further discusses the shortcomings of current practices of designing new trials.

This review uncovers the main reason for the non-compliance of updated CONSORT policy guidelines as being the absence of a universally accepted methodology underpinning the use of existing evidence in designing new trials. As the first step towards the development of the methodology, the next chapter conducts a survey to establish the current practices of using previous evidence in designing new RCTs.
CHAPTER 4

Assessing the current practice of using existing evidence in designing RCTs

4.1 Introduction

This chapter provides a brief account of the survey conducted to assess the current practice of using existing evidence when new RCTs are designed. Section 4.2 describes the motivation of the survey followed by the objectives of the survey in section 4.3. The methods employed in the process of data collection and analysis are described in section 4.4. The results are presented in section 4.5. Sections 4.6 and 4.7 present the discussion and the conclusions of this chapter respectively.

4.2 Motivation

The primary objective of this thesis is to develop a methodological framework to inform the design of new RCTs based on evidence synthesis methods. The literature review conducted in Chapters 2 and 3 above has revealed the lack of use, despite the importance, of existing evidence in informing new trial designs (Cooper, N. J., Jones and Sutton 2005). It is however vital to explore the extent to which existing evidence is used to inform the design of RCTs before embark on developing the proposed methodological framework. Further, it is expected that this survey would uncover the limitations and drawbacks in current
practices of using existing evidence and aid in developing methodologies to overcome such limitations and to improve the remaining drawbacks.

### 4.3 Objectives

The central objective of this chapter is to establish the extent of using systematic reviews and especially meta-analyses, which is regarded as the highest level of evidence, in designing new clinical trials.

At the outset, the aim is to determine the key designing/reporting aspects from which a RCT is generally comprised of. Another secondary aim is to sensibly classifying the existing evidence cited in RCTs into meaningful categories. The objective here is to uncover the type of existing evidence used to inform each designing/reporting aspect of a RCT.

This survey is conducted by drawing research articles reporting RCTs and by exploring the type of existing evidence used to influence each designing/reporting aspect. The methods employed in collecting and analysing data are briefly discussed in upcoming section 4.4.

### 4.4 Methods

The methods employed in this survey are split into two sections, i.e. methods of data collection and data analysis. Further, this section addresses two secondary aims of this survey as well, including the determination of key designing/reporting aspects of a RCT and classification of citations into meaningful categories.
The Lancet, one of the major healthcare journals, has updated its policy regarding the reporting of RCTs based on the CONSORT recommendations. Therefore, to begin with, a sample of seven articles $A^1$-$A^7$ reporting RCTs was drawn from the volume 369 (2007) of the Lancet. This is not a random sample. These seven articles were drawn as they appear in the volume 369 of the Lancet.

All articles in the sample are ensured to be reporting RCT results, since the main focus is on designing aspects of RCTs. On inspection, two articles were excluded from the sample: The first $A^4$ was removed on the basis of reporting a very similar topic to another report, and the second $A^7$ is excluded because of reporting an extended follow-up of a previously reported RCT (This does not appear to be reporting an original RCT). Hence, the original sample has shortened to a list of five articles. The list of all seven articles selected initially is given in the appendix A1.

The remaining five articles were examined with special attention to identifying the extent to which existing research had been cited as informing the designing/reporting aspects of each RCT. After careful consideration, three major design aspects have been defined to include the hypothesis, the methods used (specifically sample size, study design and analysis methods) and reporting the primary, secondary and ‘other’ results (including content in the discussion and conclusions section of the paper).

This is followed by defining six different study categories including RCT, observational study, review of trials, systematic review of trials, meta-analysis of
trials and ‘other’. At the outset, both RCTs and observational studies were grouped together under the study design called ‘study’. As the data collection process progresses this category was split into two subgroups namely RCTs and observational studies, because a clear distinction was found between them.

Following the determination of key designing/reporting aspects of RCTs and classification of study categories to which majority of citations belong, the section 4.4.1 below describes the data collection process.

### 4.4.1 Data collection

The data collection process involves considering each citation in turn to evaluate the study category it belongs to and to determine the designing/reporting aspect being informed. For the convenience of data extraction, a grid is designed with the designing/reporting aspects listed as rows and study categories listed as columns.

*Table 4-1: A specimen of the data extraction grid.*

<table>
<thead>
<tr>
<th></th>
<th>Study Trials</th>
<th>Observation Study</th>
<th>Review of trials</th>
<th>Systematic Review of trials</th>
<th>Meta-analysis of trials</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td>Primary Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion And Conclusions</td>
<td>Secondary results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A specimen of the data extraction grid is shown above in table 4-1.
4.4.1.1 **Basic level**

On separate grids, for each of the RCT report in the sample, a brief description was made with respect to how each citation influences the relevant design/reporting aspects in each respective cell of this grid. To make the data collection process much easier and clearer, each column in the grid is further split into three sub columns to accommodate reference numbers from the report, the study category and a brief description of how it is informed. This level of data collection is referred to as the ‘Basic level’.

During the process of extracting data of five RCT reports, it became evident that this process is very time consuming particularly to obtain and examine the original articles of all potentially relevant citations. Therefore, when extracting information from a further sample of five trial reports\(^{A8-A12}\) drawn from volume 371 (2008) of The Lancet (listed in appendix A2) the full text of the cited article was only examined if having read, first, the title, and then, if necessary, the abstract, the design of the study could not be ascertained beyond reasonable doubt.

In the primary analysis, sample specific summaries as well as combined summaries by combining the two samples are produced. The next section 4.4.2 gives a brief account of the process of data analysis.
4.4.2 **Statistical analysis**

The data derived from the basic level of this survey is analysed at three different levels, namely the Yes/No level, count level and summary level. A brief account of each level is given below in turn.

4.4.2.1 **Yes/No level**

At this level, the information available in each cell of the data collection grid at the basic level is transformed to a dichotomous response, i.e. Y or N to indicate whether a specific designing/reporting aspect is informed by a particular study category. This level is referred as Y/N level. A cell is marked with Y to indicate ‘yes’ or N to indicate ‘no’ depending on the availability of information in a particular cell at the basic level. This level helps to determine whether a particular designing/reporting aspect has been informed by a specific study category in a given trial.

4.4.2.2 **Count level**

As the name suggests, the count level specifies the number of instances in which each designing/reporting aspect is informed by the same study category in a given trial.

4.4.2.3 **Summary level**

In addition to aforementioned two levels, the ‘summary level’ of data was created by condensing the data available in the ‘basic level’, across all ten RCT reports in both samples.
This section discusses the procedures of data collection and analysis used in this survey. Section 4.5 below presents the results of this survey.

4.5 **Results**

In this survey, two samples with five articles each have been chosen to extract the data from. Two samples are disparate only with the procedure used to extract the data from them. Hence, in the analysis, two samples are combined to produce combined results. Further, this section produces individual sample specific results as well.

Table 4.2 below presents the results of both samples combined. The number in each cell signifies the number of articles, out of 10, used the study category listed in columns to influence corresponding design aspects listed in rows.

**Table 4-2: Summary results of the combined sample.**

<table>
<thead>
<tr>
<th></th>
<th>Study Trials</th>
<th>Observational study</th>
<th>Review of trials</th>
<th>Systematic Review of trials</th>
<th>Meta-analysis of trials</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Sample size</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Design</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Analysis methods</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Results Discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>And Conclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Results</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Secondary results</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>22</td>
<td>12</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>21</td>
</tr>
</tbody>
</table>

The results shown in table 4-2 above highlight the use of all types of study categories in defining and in support of the hypothesis of a new RCT. In other
words, hypothesis is the only design aspect, which has been influenced by all study categories. Moreover, eight of the ten articles used trials to define or in support of the hypothesis, suggesting trials are the most frequently used study category to inform the hypothesis of a new RCT.

Several design aspects including the sample size, design, analysis methods and secondary results have never been informed by study categories such as reviews, systematic reviews or meta-analysis of trials. None of the study categories except ‘other’ has been used to influence the designing aspect of analysis methods. This is when a citation of a WHO guideline to define anaemia, in the 5th article in the first sample was categorised as ‘other’ as it does not fit in to other study categories.

The least used study category to inform the design of new trials is the ‘review of trials’. It has only been used once to define the hypothesis. Moreover, the study type ‘systematic review of trials’ have only been used in influencing the hypothesis and in support of ‘other results’.

The ‘trials’ and ‘observational studies’ happen to be the most commonly used study categories and they have been used to inform all design aspects except the ‘analysis methods’. Every designing aspect has been influenced by the study category ‘other’ and is equally as influential as trials. As the name suggests the study category ‘other’ is miscellaneous in nature and primarily used to group study categories those do not fit in elsewhere. Hence, this category has null effect in the analysis.
The study category of prime concern, i.e. the MA of trials is found hardly used in planning and designing new RCTs. It has only been used in three separate occasions in this survey, i.e. twice to inform the hypothesis and once to clarify the primary results. None of remaining designing aspects has been influenced by a MA of trials.

Table 4-3 below provides a different perspective to the summary results of the survey. It presents the results of combined samples. The aim here is to detect the highest level of evidence used in informing each designing/reporting aspect by considering each article in turn.

**Table 4-3 : Summary results classified by each article.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Article</th>
<th>Hypothesis</th>
<th>Methods</th>
<th>Results, Discussion and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sample size</td>
<td>Design</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>T MAT O</td>
<td>T</td>
<td>T MAT O</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>T MAT O</td>
<td>T</td>
<td>T MAT O</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>T OS O</td>
<td>T</td>
<td>T OS O</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>T MAT O</td>
<td>T</td>
<td>T MAT O</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>T OS SRT O</td>
<td>T</td>
<td>T OS SRT O</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>SRT O</td>
<td>OS</td>
<td>OS O</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>T O</td>
<td>T O</td>
<td>T O</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>OS SRT O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>T OS O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>T OS SRT O</td>
<td>T</td>
<td>T O</td>
</tr>
</tbody>
</table>
The abbreviations used in the table 4-3 are listed below.

T – Trial  
RT – Review of Trials

O – Other  
OS – Observational studies

MAT – Meta-Analysis of Trials  
SRT – Systematic Review of Trials

The Trials (T) and observational studies (OS) are used to inform every designing aspect except the ‘analysis methods’. Trials are by far the most common type of existing evidence used. Meta-analysis of trials (MAT) which is regarded as the highest level of evidence is used only in three occasions, i.e. in two articles (Article 1 and 2) of the sample one to inform the hypothesis and in article 4 of the sample one in affirming of the primary results. Noticeably, none of the articles in the second sample used a MAT.

Systematic review of trials (SRT) has not been used in the first sample, as is MAT in the second sample. However, in the second sample, SRT is used, predominantly in aid of defining the hypothesis, in four separate instances in three different articles.

In article 5 of both the first and second samples, the hypothesis has been influenced by four distinct study categories. These two are the only instances where four different study categories collectively inform a specific designing aspect. However, MAT is not involved in any of these instances, although SRT and RT have been involved.
In the second sample, the designing aspect called 'sample size' has been influenced in all five articles, whilst the designing aspect called 'design' has been informed in four articles too. This is in contrast with the first sample in which the designing aspects called 'sample size' and 'design' have been informed only once. Markedly, the design aspect called 'analysis methods' is influenced only once in both samples. A discussion of the findings of this chapter is given in section 4.6 below.
4.6 Discussion

This chapter provides a brief account of the survey conducted using two sets of RCT sample reports aiming to assess the extent of the use of existing research evidence to inform the planning and designing of new RCTs. This chapter further highlights the issues preventing relevant existing evidence being cited when designing and reporting trials.

This survey has uncovered common practices of citing existing evidence when reporting new trials. A drug or a treatment often belongs to a particular family and this survey reveals a tendency of citing evidence related to other drugs of the same family in the absence of relevant evidence of the particular drug in concern. For example, in article 1 of the first sample, Exemestane is the drug administered to treat the breast cancer, which belongs to the family of Aromatase inhibitors. Both Anastrozole and Letrozoles belong to the same family as well. Although the trial in concern compares Tomoxifen and Exemenstane the authors however cited a few articles reporting of trials those compare Tamoxifen with other drugs of the Aromotaze inhibitor family. This is in the absence of direct evidence comparing Exemestane and Tamoxifen. This common practice is found to have been used throughout the trial reports included in the survey and concluding that citations are not always directly relevant to the trial in concern.

The same existing evidence is used to inform both the hypothesis and results, often when the new trial is a replicate of a previous trial and that the said previous trial informs both the hypothesis and results of the new trial. The article 3 of the first sample A5 claimed it to be the very first RCT of HIV incidence and
reported that only observational studies were previously addressing this clinical question. This article cites a MA of observation studies as informing the hypothesis and it is the only instance where a meta-analysis of observational studies has been cited in this survey.

Despite the involvement of a certain degree of ambiguity in classifying citations as informing 'design' and 'analysis methods' (subsections of the designing aspect called 'methods') a consistency is maintained throughout to avoid unnecessary conflict in results.

Trials of similar kind are the most frequently cited evidence type when reporting new trials. Following the CONSORT recommendations, most of major medical journals now require authors to conduct a SR or preferably a MA to justify the need of their new trial. However, this survey has not found any instance where the authors attempted to conduct either a SR or a MA of trials addressing the same therapeutic questions.

Since this survey involves only ten trial reports, we believe that it is inadequate to formulate conclusive opinions regarding the use of previous evidence in designing new trials. Moreover, the samples are drawn only from the Lancet only. It is however recommended to draw a relatively large sample from a wider range of journals before constructive opinions could be made. Section 4.7 below draws conclusions of this chapter.
4.7 Conclusions

Despite The Lancet’s’ policy update regarding the reporting requirement of clinical trials, (Young and Horton 2005) two RCT samples drawn from The Lancet in 2007 A1-A7 and 2008 A8-A12 respectively do not provide any evidence of the adherence to the updated policy requirement by the researchers. None of above articles includes a short summary of what is already known of the clinical area in concern and a discussion of the impact of new trial findings on existing knowledge of the clinical question.

This survey concludes that trials and observational studies are the most frequently used types of previous evidence to inform the trial design, followed by the study category called ‘other’ which includes miscellaneous items such as proceedings of conferences, guidelines of certain procedures etc. Moreover, the synthesised evidence such as MA, SR of trials and reviews of trials are rarely used in support of designing and reporting of new trials.

Based on the findings of this survey, none of the trials attempted to discuss their findings in the light of the totality of existing evidence. It is also noteworthy the researchers tendency of often citing previous evidence that produced similar results, as we found only one instance which cites a contrasting previous findings.

Importantly, existing evidence have been found hardly used to support the sample size calculation of a new trial. Despite being an integral design aspect, calculation of sample size of a new trial have been frequently informed either by expert opinion or by other non-research based methods. The use of non-
scientific based methods perhaps could lead to flawed calculations and could well result in under-powered trials being designed. On the other hand, it would be a waste of resources if calculations suggest randomising more than adequate patients. It is conceived that the use of a variety of non-scientific based sources to inform the sample size calculation is ascribable to the lack of widely accepted standard methodology.

We believe that a scientific approach guiding the calculation of exact number of subjects required in a trial to achieve a certain power is very much in demand. The findings of this survey have reassured the need of the methodological framework proposed in this thesis.

The forthcoming chapter employs frequentist MA methods to develop the methodology of designing new RCTs, which would form the initial component of the methodological framework developed during the course of this thesis.
CHAPTER 5

Designing future trials based on existing evidence using frequentist meta-analysis methods

5.1 Introduction

This chapter develops and implements a methodology to design new trials based on existing evidence using frequentist MA models, which eventually will contribute to the framework of methodologies developed in phases throughout this thesis. Both the fixed and the random effects frequentist MA models are employed to develop the methodology. Following this introduction, section 5.2 describes the motivation and section 5.3 sets out the objectives of this chapter. Section 5.4 briefly introduces MA. A description of the existing evidence, which forms the basis of the design of a new trial is given in section 5.5. The methods employed to develop the methodology are outlined in section 5.6. Sections 5.7 and 5.8 develop the methodology to design new trials using FE and RE meta-analyses respectively. The power of the new trial in isolation is explored in section 5.9. The section 5.10 compares the power of the new trial in isolation with that of the existing MA including the new trial. Results are critically evaluated in section 5.11 followed by the discussion in section 5.12. Section 5.13 concludes this chapter.
5.2 **Motivation**

The Chapter 4 of this thesis assesses the extent to which existing evidence is used to inform trial design at present and suggests that systematic reviews and meta-analyses are hardly used in practice to inform the design of a trial. The lack of use may be attributable to non-availability of a reliable methodology, underpinning the use of existing evidence to inform the trial design.

Sutton et al (Sutton, A. J., Cooper et al. 2007) introduced a simulation based methodology of using evidence synthesis models to inform the design of a new trial. They employed Bayesian random effects MA principles to develop the methodology of designing new trials. This chapter extends the methodology introduced by Sutton et al by employing the frequentist approach of random effects MA.

The mean of a RE distribution alone does not adequately describe the distribution of effect sizes estimated by the RE model. In the discussion section Sutton et al further pointed out the importance of investigating the ways of basing inferences on the spread of the distribution of effect sizes (as well as the mean) through the predictive distribution of effects.

5.3 **Objectives**

The primary objective of this chapter is to develop and implement a methodology to design future trials based on existing evidence using frequentist MA principles. The longer-term aim is to encourage researchers to make formal considerations of future MA at the planning and design stage of individual trials.
At the outset, the methodology is developed to be able to estimate the sample size of two-arm RCTs (both arms with equal size) with the outcome being measured on a binary scale and analysed on the OR scale.

Both the FE and RE approaches of MA are explored to estimate the sample size required for a new trial to achieve a certain level of power of the updated MA including the new trial. In addition, the power of the new trial in isolation is also computed to explore the power implications before and after the inclusion of new trial into the MA.

Further, it is aimed to propose a variety of prediction models to accommodate different predictive distributions from which an effect size of a new trial is drawn. The aim is to explore the sensitivity of the power of the new trial alone and the updated MA including the new trial to the model adopted.

Having set out the key objectives of this chapter in this section the upcoming section 5.4 briefly introduces MA which is the fundamental methodology used in this chapter to base the design of new trials.

5.4 Introduction to meta-analysis

As mentioned in section 2.3, MA provides the pinnacle of evidence regarding the effectiveness of interventions. It is the statistical procedure that integrates the results of several combinable studies (Huque 1988). When the effect size is consistent from one study to next, MA yields the common effect size. On the contrary, when the effect sizes do not consistent over studies, but vary from study to study, it facilitates the investigation into the reasons for the variation
(Borenstein, M., Hedges et al. 2005). There are two commonly used MA
models available, namely the fixed effects and the random effects models. The
two models employ distinct approaches of handling the between study variation.
Those two models have been discussed at great length in section 2.4.3 and
2.4.4 in Chapter 2 in the literature review.

The methodological framework developed here uses existing evidence to form
the basis of the design of new trials. Therefore, a set of similar trials addressing
the same therapeutic question is required to perform an initial synthesis. The
following section 5.5 describes the existing evidence.

5.5 Existing evidence

Before embark on to design new trials based on evidence synthesis methods,
the prerequisite is a series of similar trials suitable to be meta-analysed. To
enable a comprehensive comparison of the results of the extensions of methods
developed in this study and the original work by Sutton et al, this chapter uses
the same set of existing evidence used by Sutton et al. The trials included in
the systematic review of antibiotics use for common cold in the Cochrane
database of systematic review are used to illustrate the approach (Arrol and
Kenealy 1999). Table 5-1 below shows the number of events and the total
subjects in both antibiotics and placebo groups in each trial.
Chapter 5  Designing new trials using meta analysis methods

Table 5-1: The list of trials (comparing antibiotics and placebo in treating common cold) constitute the existing evidence to base the design of the new trial.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Year</th>
<th>Antibiotics</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Events</td>
<td>Total</td>
</tr>
<tr>
<td>Herne</td>
<td>1980</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td>Hoaglund</td>
<td>1950</td>
<td>39</td>
<td>154</td>
</tr>
<tr>
<td>Kaiser</td>
<td>1996</td>
<td>97</td>
<td>146</td>
</tr>
<tr>
<td>Lexomboon</td>
<td>1971</td>
<td>8</td>
<td>174</td>
</tr>
<tr>
<td>McKerrow</td>
<td>1961</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Taylor</td>
<td>1977</td>
<td>12</td>
<td>129</td>
</tr>
</tbody>
</table>

5.5.1 Outcome of interest

The outcome of each trial is measured on binary scale using OR and the analysis is carried out on log OR scale. This is because the log OR holds statistical properties that are more closely approximated by the normal distribution.

5.5.2 Definition of an event

An event in this illustrative example is defined as a patient with cold symptoms persisting beyond 7 days.

5.5.3 Estimating the events rate in the control group

This methodology requires the probability of a trial participant in the control group of the new trial would have an event \( P_{\text{control,new}} \) to be known in advance and is usually estimated using the expert opinion, the external data such as data collected routinely on the population of interest or using existing trials.
In this approach, the event rate in the control group \( (P_{\text{control.new}}) \) is computed by taking the mean of the events in the control group of the existing evidence listed in table 5-1 above, i.e. 0.352, in the absence of expert opinion. However, exact value of \( P_{\text{control.new}} \) is often not known as the estimation involves an element of uncertainty. To reflect the uncertainty in the estimation, \( P_{\text{control.new}} \) can be expressed in the form of a distribution. However, for simplicity, \( P_{\text{control.new}} \) is assumed to be known without error.

Having collected existing evidence, the methods employed to design a new trial are discussed in upcoming section 5.6.

5.6 Methods

5.6.1 Null hypothesis tested and power in the updated meta-analysis

In this chapter, the interest is in the power of the updated MA including the new trial. The power of is based on the null hypothesis that the efficacy of two treatments is not different, i.e. two treatments are equally effective \( (H_0: OR=1) \). The decision as to whether to reject or do not reject the null hypothesis is made at 5% level of significance.

5.6.2 Procedure

The methodology to design new trials adopts a simulation based approach (Feiveson 2002) in contrast to a closed form solution. Models with RE
components often are not solvable using closed form solutions and the
simulation approach becomes the most appealing solution. This simulation
based methodology of designing new trials using results of meta-analyses was
first introduced and implemented by Sutton and the colleagues (Sutton, A. J.,
Cooper et al. 2007) to design new trials. The method estimates the sample size
of a new trial needed to be able to achieve a certain power in the updated MA
and vice versa. The sequence of steps of the methodology to compute power is
outlined below (Sutton, A. J., Cooper et al. 2007).

**Step 1**
The predictive distribution of an effect size (log OR) of a new trial ($\theta_{new}$) is
specified using the results of the initial MA of existing trials.

**Step 2**
An observation is then sampled from the predictive distribution, which
represents an effect size of the new trial ($\hat{\theta}_{new}$).

**Step 3**
The effect size of the new trial ($\hat{\theta}_{new}$) drawn in step 2 and the control group
event rate ($P_{treatment,new}$) calculated in section 5.5.3 above are then substituted
to equation 5.1 to derive the probability of a trial participant in the treatment
group of the new trial would have an event ($P_{treatment,new}$).

$$P_{treatment,new} = \frac{\left(\frac{P_{control,new}}{1 - P_{control,new}}\right) \times e^{\hat{\theta}_{new}}}{1 + \left(\frac{P_{control,new}}{1 - P_{control,new}}\right) \times e^{\hat{\theta}_{new}}}$$  (5.1)
**Step-by-step guide to derive the equation 5.1**

Definition of log odds ratio

\[ \hat{\theta}_{new} = \ln \left( \frac{Odds \ in \ treatment \ group}{Odds \ in \ control \ group} \right) \]

\[ \exp (\hat{\theta}_{new}) = \frac{Odds \ in \ treatment \ group}{Odds \ in \ control \ group} \]

Applying the definition of odds into both denominator and numerator

\[ \exp (\hat{\theta}_{new}) = \frac{\frac{Events \ in \ treatment \ group}{Non \ events \ in \ treatment \ group}}{\frac{Events \ in \ control \ group}{Non \ events \ in \ control \ group}} \]

Defining the non events as the total subjects minus the events in each group

\[ \exp (\hat{\theta}_{new}) = \frac{\frac{Events \ in \ treatment \ group}{Total \ subjects \ in \ treatment \ group - Events \ in \ treatment \ group}}{\frac{Events \ in \ control \ group}{Total \ subjects \ in \ control \ group - Events \ in \ control \ group}} \]

Dividing both numerator by total subjects in treatment group and denominator by the total subjects in control group

\[ \exp (\hat{\theta}_{new}) = \frac{\frac{Event \ rate \ the \ treatment \ group}{1 - Event \ rate \ in \ the \ treatment \ group}}{\frac{Event \ rate \ in \ the \ control \ group}{1 - Event \ rate \ in \ the \ control \ group}} \]

Substituting \( P_{treatment,new} \) and \( P_{control,new} \) for \( Event \ rate \ the \ treatment \ group \) and \( Event \ rate \ in \ the \ control \ group \) respectively

\[ \exp (\hat{\theta}_{new}) = \frac{\frac{P_{treatment,new}}{1 - P_{treatment,new}}}{\frac{P_{control,new}}{1 - P_{control,new}}} \]

Event rate in the treatment group of the new trial is given by

\[ P_{treatment,new} = \frac{\left( \frac{P_{control,new}}{1 - P_{control,new}} \right) \times e^{\hat{\theta}_{new}}}{1 + \left( \frac{P_{control,new}}{1 - P_{control,new}} \right) \times e^{\hat{\theta}_{new}}} \]
**Step 4**

The size of the treatment \( n_{\text{treatment,new}} \) and control arm \( n_{\text{control,new}} \) of the new trial is then specified.

**Step 5**

The data representing the new trial is stochastically simulated, i.e. the simulation of the number of events in the treatment \( r_{\text{treatment,new}} \) and the control \( r_{\text{control,new}} \) groups using the binomial distribution as below in equation 5.2 and 5.3 respectively.

\[
r_{\text{treatment,new}} \sim \text{Binomial}(P_{\text{treatment,new}}, n_{\text{treatment,new}}) \quad (5.2)
\]

\[
r_{\text{control,new}} \sim \text{Binomial}(P_{\text{control,new}}, n_{\text{control,new}}) \quad (5.3)
\]

The 2x2 table representing the results of the new trial is then constructed. For simplicity, this procedure is restricted to the case of equal number of subjects randomised into both treatment and control groups.

**Step 6**

The new simulated trial is added to the existing evidence and the updated list of trials is meta-analysed.

**Step 7**

The hypothesis test on which the power is based is then considered (section 5.6.1) and the decision to reject or to retain the null hypothesis in favour of the alternative hypothesis is made.
**Step 8**
A large number of simulations (5000 in this case) is carried out by iterating the process (steps 1-7) and the power is computed as the proportion of simulations in which the null hypothesis is rejected.

**Step 9**
This process (steps 1-8) is repeatedly performed by varying the sample size in each repetition, until the desired level of power is achieved. Sample sizes of 100, 300, 500, 800, 1000, 2000, 3000, 4000 and 5000 in each arm are considered throughout this thesis.

This section discusses the methodology employed to design a new trial. Having discussed the methods and defined the existing evidence in preceding sections, the following section 5.7 embark on to develop the methodology to design new trials using frequentist MA principles.
5.7 Designing a new trial based on existing evidence using a fixed effects meta-analysis model

5.7.1 Fixed effect meta-analysis model

The general form of the fixed effect model can be expressed as

$$Y_i = \mu + e_i$$  (5.4)

The observed effect size estimate of the $i^{th}$ trial is denoted by $Y_i$, where $i$ indexes the $k$ trials included in the MA, $\mu$ is the underlying true treatment effect size, and $e_i$ is the sampling error in the estimation of effect size of the $i^{th}$ trial.

This section describes the fixed effects MA model. The forthcoming section 5.7.2 develops the methodology to design a new trial based on fixed effects MA principles.

5.7.2 Initial fixed effects meta-analysis using existing evidence

As stated above in section 5.6, this approach requires conducting an initial MA of existing evidence (step 1) to derive the inference of the predictive distribution of the new trial. Therefore, an initial MA (fixed effects) is carried out using existing evidence produced in table 5-1, and the resulting forest plot is presented in figure 5-1 below.
The results of the above fixed effects MA suggest the need of further trials addressing the same therapeutic question because the MA is not conclusive. The MA would have been conclusive if the confidence interval does not contain the null value. The statistical test of the pooled OR is non-significant at 5% level (p=0.15) as indicated by the confidence interval (0.59-1.08) of the pooled OR which includes the null value (OR=1).
5.7.3 **Specification of the predictive distribution of a new trial**

5.7.3.1 **The mean of the predictive distribution of the new trial**

The pooled OR resulting from the above fixed effects MA is 0.80 which corresponds to a log OR ($\hat{\mu}$) of -0.219. This estimate is regarded as the mean of the predictive distribution of the new trial (step 1).

5.7.3.2 **The variance of the predictive distribution of the new trial**

The standard error of the pooled log OR ($SE(\hat{\mu})$) of the resulting fixed effects MA is 0.153. This estimate specifies the standard error of the predictive distribution from which an effect size of the new trial is drawn (step 1).

5.7.4 **Prediction model 1 – Designing a new trial based on fixed effects meta-analysis method**

5.7.4.1 **Rationale**

The pooled log OR ($\hat{\mu}$) resulting from the MA of existing trials serves as an estimate of the underlying true effect size as well as an unbiased estimate of an effect to be seen in a future trial (Sutton, A. J., Cooper et al. 2007). Under the assumption that the pooled log OR ($\hat{\mu}$) is normally distributed, the distribution of the underlying effect of a new trial($\theta_{new}$) can be specified as in equation 5.5.

$$\theta_{new} \sim Normal (\hat{\mu}, var(\hat{\mu}))$$ (5.5)
Chapter 5

Designing new trials using meta analysis methods

It is further assumed that the new trial is essentially a replicate of trials included in the existing MA. The only variability between the observed and the underlying effect size is that resulting from the sampling variation.

5.7.4.2 Methods

The event rate in the control group \( P_{\text{control.new}} \) is estimated as illustrated in section 5.5.3. The event rate in the treatment group \( P_{\text{treatment.new}} \) is subsequently derived by substituting into equation 5.1 (step 2).

The sample size of the new trial \( n_{\text{treatment.new}} \) and \( n_{\text{control.new}} \) is then specified and the number of subjects with events in the treatment \( r_{\text{treatment.new}} \) and the control group \( r_{\text{control.new}} \) are then binomially simulated. The number of subjects with non-events is computed by subtracting the number of subjects with events from the total sample size. Once the subjects with events and non-events of the treatment and the control groups are specified, it constitutes the 2 x 2 table of the new trial outcomes (step 3).

The new trial is added into the existing trials and a new MA is performed using all trials, including the new trial (step 4). The p-value of the mean test of the updated MA is determined to reject or retain the null hypothesis which states that the treatment and control are equally effective (step 5). This process is repeated for a large number of times (5000 times in this case) and the power is determined by calculating the proportion of the rejections of the null hypothesis (step 6). This entire process is repeated by specifying different sample sizes of the new trial and the power of the updated MA is reported, for each sample size (step 7).
5.7.4.3 Results – Power of the updated fixed effects meta-analysis

The power of the updated fixed effects MA (including the new trial) is presented in table 5-2 below. The sample size column represents the number of patients randomised to a single arm. The overall sample size is twice as much as the sample size listed in the second column of the table 5-2.

Table 5-2: The table lists the power of the updated fixed effects MA against the sample size of the new trial in each arm.

<table>
<thead>
<tr>
<th>Sample size (one arm)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>24.68</td>
</tr>
<tr>
<td>300</td>
<td>47.06</td>
</tr>
<tr>
<td>500</td>
<td>58.24</td>
</tr>
<tr>
<td>800</td>
<td>64.98</td>
</tr>
<tr>
<td>1000</td>
<td>69.88</td>
</tr>
<tr>
<td>2000</td>
<td>76.64</td>
</tr>
<tr>
<td>3000</td>
<td>80.76</td>
</tr>
<tr>
<td>4000</td>
<td>82.80</td>
</tr>
<tr>
<td>5000</td>
<td>84.34</td>
</tr>
</tbody>
</table>

The results are graphically presented below in figure 5-2.

Figure 5-2: The power curve of the updated fixed effects meta-analysis against the sample size of the new trial in each arm.

Prediction model 1: Fixed effects meta-analysis

The power increases as the sample size increases. The power is relatively high at even smaller sample sizes. The rate at which the power increases with increased sample size declines noticeably after the sample size in each arm reaches 800.
The power implications of the updated MA including a new trial designed based on fixed effects MA principles is described in this section. Upcoming section 5.8 develops and implements the methodology to design new trials based on the random effects MA model and attempts to establish the relationship between the power of the MA and the sample size of the new trial.

5.8 **Designing a new trial based on existing evidence using the random effects meta-analysis model**

5.8.1 **Random effects meta-analysis model**

The general form of the RE model is expressed as

$$Y_i = \theta_i + e_i$$

(5.6)

The observed effect size estimate of the $i^{th}$ study is denoted by $Y_i$, which contains an element of within study error $e_i$. The true underlying treatment effect specific to $i^{th}$ study is denoted by $\theta_i$, where $i$ indexes the $k$ studies.

Within study error $e_i$ is normally distributed with 0 mean and $\sigma_i^2$ variance.

$$e_i \sim N(0, \sigma_i^2)$$

(5.7)

The underlying true overall treatment effect is denoted by $\mu$ and is related to the underlying treatment effect specific to the $i^{th}$ study $\theta_i$ as below.
Therefore

\[ Y_i = \mu + s_i + e_i \]  \hspace{1cm} (5.9)

where \( s_i \) is the heterogeneity element and is distributed as

\[ s_i \sim N(0, \tau^2) \]  \hspace{1cm} (5.10)

The underlying treatment effects in all \( k \) studies \((\theta_1, \ldots, \theta_k)\) are assumed to be a random sample from a normal distribution with mean \( \mu \) and variance \( \tau^2 \).

\[ \theta_i \sim N(\mu, \tau^2) \]  \hspace{1cm} (5.11)

The random effects model of MA is specified in this section. A variety of prediction models is designed in the upcoming sections 5.8.5 - 5.8.8 to be able to examine the fluctuation of power for the changes to the sources of variance of the predictive distribution of the new trial.

### 5.8.2 Initial random effects meta-analysis of existing evidence

This section focuses on the development and implementation of a variety of prediction models to design new trials based on random effects MA principles. This methodology requires an initial MA to specify the predictive distribution of the effect size of a new trial. The results of the initial random effects MA of the trials listed in table 5-1 are shown below in the forest plot in figure 5-3.
The results of the MA highlight the need of more trials addressing the same therapeutic question on the basis that the MA is not conclusive. The resulting pooled OR (0.77) is statistically non-significant at 5% level (p=0.255).

The $I^2$ statistic indicates that 38% of the total variation is caused by the heterogeneity. However this is statistically non-significant at 5% level (p=0.152) and thus failing to provide evidence against the homogeneity of existing trials.

### 5.8.3 Random effects prediction models

In a meta-analytical framework, inferences obtained from a random effects MA can be more reliably used to predict the effect size of a new trial ($\theta_{\text{new}}$) than those from the FE model. The predictive distribution of a new trial is the most
relevant inference drawn from the random effects MA (Higgins, J. P. T., Thompson and Spiegelhalter 2008) for our purposes since it predicts the distribution of an individual future observation. The prediction interval holds the same relationship to a new observation that the confidence interval holds to an unobservable parameter estimate.

### 5.8.4 Specification of the predictive distribution of a new trial

An effect size of a future trial \( \theta_{\text{new}} \) is assumed to be drawn from a normal distribution as below.

\[
\theta_{\text{new}} \sim \text{Normal} (\text{mean}, \text{var}) \tag{5.12}
\]

#### 5.8.4.1 The mean of the predictive distribution of a new trial

The predictive distribution of the new trial of every random effects prediction model developed in upcoming sections (prediction model 2-5) shares a common mean, i.e. the pooled log OR (\( \hat{\mu} \)) of the random effects MA of existing trials. In this illustrative example pooled log OR is estimated to be \( \hat{\mu} = -0.2588 \).

#### 5.8.4.2 The variance of the predictive distribution of the new trial

The variance of the predictive distribution distinguishes the prediction models. The variance is set to take a different value in each prediction model. The standard error of the pooled log OR of the existing MA (\( SE(\hat{\mu}) = 0.227 \)) and the estimated between study variance (\( \hat{\tau}^2 = 0.1075 \)) are used either alone or
collectively to define the variance of the predictive distribution of a new trial in upcoming RE prediction models.

5.8.4.3 Hypothetical models

The RE prediction models those accounted for either of the additive component of the variances in the predictive distribution are rather hypothetical. Nevertheless, these models are expected to provide an insight into the change in the behaviour of the power of the updated MA when the variance of the predictive distribution is altered. Prediction models 2 and 3 developed below are hypothetical and not found in practice.

The prediction model based on the fixed effects MA model (Prediction model 1) was discussed earlier in section 5.7.4 of this chapter. Prediction models 2-5 described in the upcoming sections 5.8.5 - 5.8.8 employ random effects MA principles.
5.8.5 Prediction model 2 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution is specified only using within study variation)

5.8.5.1 Rationale

In this model, the predictive distribution of a new trial is specified to include only the within study variance (\(\text{var}(\hat{\mu})\)) obtained from the existing MA. The model resembles a FE model in that perspective. The model can be written as

\[
\theta_{\text{new}} \sim \text{Normal}(\hat{\mu}, \text{var}(\hat{\mu}))
\]  

(5.13)

5.8.6 Prediction model 3 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution is specified using the between study variations)

5.8.6.1 Rationale

The predictive distribution of an effect size (\(\theta_{\text{new}}\)) of a new trial in this model is designed to contain only the between study component (\(\hat{\tau}^2\)) of the variance (which is derived from the existing random effects MA). Hence, the model takes the form of

\[
\theta_{\text{new}} \sim \text{Normal}(\hat{\mu}, \hat{\tau}^2)
\]  

(5.14)
5.8.7 Prediction model 4 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution includes both within and between study variations - Normal distribution approach)

5.8.7.1 Rationale

The random effects MA model is frequently employed in the presence of and to identify unexplained heterogeneity (Thompson, S.G. 1994). The prediction model developed in this section assumes the variance of the predictive distribution of the effect size of the new trial ($\theta_{new}$) is composed of both within ($\text{var}(\hat{\mu})$) and between study ($\hat{\tau}^2$) components (derived from the results of the existing MA).

The model is written as

$$\theta_{new} \sim \text{Normal} (\hat{\mu} , \text{var}(\hat{\mu}) + \hat{\tau}^2)$$

(5.15)
5.8.8 Prediction model 5 - Designing a new trial using frequentist random effects meta-analysis (the variance of the predictive distribution of a new effect size includes both within and between study variance - t-distribution approach)

5.8.8.1 Rationale

The predictive distribution of this model is based on the prediction interval method derived based on the t-distribution (Higgins, J. P. T., Thompson and Spiegelhalter 2008) which is outlined below.

If $\tau^2$ is known, the mean effect estimate ($\hat{\mu}$) is normally distributed as

$$\hat{\mu} \sim Normal(\mu, SE(\hat{\mu})^2)$$ (5.16)

The predicted effect ($\theta_{new}$) of a new trial, which is assumed to be sufficiently similar to existing trials and suitable to be included in the analysis is given by

$$\theta_{new} \sim Normal (\mu, \tau^2)$$ (5.17)

Assuming the independence of $\theta_{new}$ and $\hat{\mu}$, given $\mu$

$$\theta_{new} - \hat{\mu} \sim Normal (0, \tau^2 + SE(\hat{\mu})^2)$$ (5.18)

Hence

$$\theta_{new} \sim Normal (\hat{\mu}, \tau^2 + SE(\hat{\mu})^2)$$ (5.19)

The value of $\tau^2$ is not known exactly and $\hat{\tau}^2$ is used in place of it. The uncertainty associated with estimation of $\tau^2$ is greater when the number of
studies in the MA is smaller. Both components of the variance of $\theta_{new}$ are affected by the imprecision in $\tau^2$.

Therefore

$$\theta_{new} \sim Normal ( \mu, \hat{\tau}^2 + SE(\hat{\mu})^2 )$$ \hspace{1cm} (5.20)

Taking a t-distribution with $k-2$ degrees of freedom to present this uncertainty

$$\frac{\theta_{new} - \mu}{\sqrt{\hat{\tau}^2 + SE(\hat{\mu})^2}} \sim t_{k-2}$$ \hspace{1cm} (5.21)

The 100(1 - $\alpha/2$)% prediction interval for the effect size of a new trial is hence

$$\hat{\mu} \pm \alpha_{k-2} \sqrt{\hat{\tau}^2 + SE(\hat{\mu})^2}$$ \hspace{1cm} (5.22)

where $\alpha_{k-2}$ is the 100(1 - $\alpha/2$) % percentile of the t-distribution with $k-2$ degrees of freedom.

### 5.8.9 Methods

An effect size is drawn from the predictive distribution specified in the rationale of each respective random effect prediction model. A new trial is designed and power of the updated MA is computed as explained in section 5.6.1 above. The only difference here is that the conduct of the updated MA using RE methods as opposed to FE methods used in section 5.7.2.
5.8.10 **Results of the power of the updated meta-analysis of prediction model 2-5**

Table 5-3 below shows the power of the updated MA including the new trial against the sample size of the new trial of all prediction models including the FE model.

*Table 5-3: The combined results – The power of the updated MA of prediction models 1-5 for sample sizes ranging from 0-5000.*

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed Effects</td>
<td>Random effects Prediction distribution includes within study variation only</td>
<td>Random effects Prediction distribution includes between Study variation only</td>
<td>Random effects Prediction distribution includes both between and within study Variations Normal distribution</td>
<td>Random effects Prediction interval is based on the t-distribution</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>47.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>58.24</td>
<td>21.08</td>
<td>19.02</td>
<td>16.84</td>
<td>15.98</td>
</tr>
<tr>
<td>800</td>
<td>64.98</td>
<td>34.44</td>
<td>31.34</td>
<td>28.40</td>
<td>25.14</td>
</tr>
<tr>
<td>1000</td>
<td>69.88</td>
<td>38.36</td>
<td>34.92</td>
<td>29.76</td>
<td>28.40</td>
</tr>
<tr>
<td>2000</td>
<td>76.64</td>
<td>45.76</td>
<td>38.18</td>
<td>35.64</td>
<td>33.14</td>
</tr>
<tr>
<td>3000</td>
<td>80.76</td>
<td>46.82</td>
<td>41.36</td>
<td>37.04</td>
<td>34.64</td>
</tr>
<tr>
<td>4000</td>
<td>82.80</td>
<td>49.68</td>
<td>42.52</td>
<td>37.40</td>
<td>34.04</td>
</tr>
<tr>
<td>5000</td>
<td>84.34</td>
<td>50.08</td>
<td>42.18</td>
<td>37.80</td>
<td>35.50</td>
</tr>
</tbody>
</table>

Individual power curves of all RE based prediction model are provided below in figure 5-4.
**Figure 5-4**: Description of the results of the power of the updated meta-analysis of prediction models 2-5 in which the sample size ranging from 0-5000.

<table>
<thead>
<tr>
<th>Prediction model 2</th>
<th>Description of the results of the power of the updated meta-analysis of prediction models 2-5 in which the sample size ranging from 0-5000.</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td>The power is zero until the sample size increases above 300. Power then increases rapidly as the sample size increases until sample size reaches about 1000. The rate at which power increases is declining as the sample size increases.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 3</th>
<th>Description of the results of the power of the updated meta-analysis of prediction models 2-5 in which the sample size ranging from 0-5000.</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image2.png" alt="Graph" /></td>
<td>The power remains zero until the sample size reaches above 300. The power increases at a higher rate with increased sample size until sample size reaches 1000. The rate at which power increases is declining for sample sizes above 1000. The power flattens off at around 42% for sample sizes above 3000.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 4</th>
<th>Description of the results of the power of the updated meta-analysis of prediction models 2-5 in which the sample size ranging from 0-5000.</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Graph" /></td>
<td>The updated MA produces lower power at each comparative sample size, compared to that of prediction models include either within or between study variations. The power remains at zero even at 300 patients in each arm. The power increases as the sample size increases at a modest rate until the sample size reaches about 2000 and then reaches a plateau.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 5</th>
<th>Description of the results of the power of the updated meta-analysis of prediction models 2-5 in which the sample size ranging from 0-5000.</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Graph" /></td>
<td>The power remains zero until the sample size increases slightly above 300 and steadily progresses as the sample size increases between 500 and 2000 and flattens off. The power dips slightly, which may be attributable to the sampling variation in the simulation, at 4000. This model produces the lowest power, for a given sample size of the new trial compared to other models.</td>
</tr>
</tbody>
</table>
The power of the updated MA in all random effects based models remains non-zero until the sample size of the new trial reaches 500. Moreover, the power at sample size 500 is larger enough to suggest that the power becomes non-zero when the sample size is within 300 and 500. Section 5.8.11 below explores the power of the updated MA focusing on the specific range, i.e. 300-500.

**5.8.11 Sample size of the new trial in between 300 and 500**

According to above findings, the variance of the predictive distribution of the new trial is found to be dictating the power of the updated MA including the new trial. However, it does not highlight the sample size at which the power begins non-zero in each prediction model. This section conducts the same analysis but focuses on a specific sample size range, i.e. between 300 and 500. Table 5-4 below shows the power of updated MA including the new trial of all RE based prediction models. For completeness, the results of the prediction model based on FE method are also incorporated.
Table 5-4: The combined results – Power of the updated MA of prediction models 1-5 for sample sizes ranging 300-500.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Prediction distribution includes within study variation only</th>
<th>Prediction distribution includes between Study variation only</th>
<th>Prediction distribution includes both between and within study variations (normal distribution)</th>
<th>Prediction interval is based on the t-distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>320</td>
<td>49.40</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>340</td>
<td>50.40</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>360</td>
<td>50.30</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>380</td>
<td>51.46</td>
<td>0.98</td>
<td>0.76</td>
<td>0.68</td>
</tr>
<tr>
<td>400</td>
<td>53.34</td>
<td>5.54</td>
<td>4.84</td>
<td>3.98</td>
</tr>
<tr>
<td>420</td>
<td>53.86</td>
<td>10.94</td>
<td>8.46</td>
<td>7.68</td>
</tr>
<tr>
<td>440</td>
<td>55.80</td>
<td>14.30</td>
<td>12.32</td>
<td>12.40</td>
</tr>
<tr>
<td>460</td>
<td>55.30</td>
<td>18.44</td>
<td>15.28</td>
<td>13.92</td>
</tr>
<tr>
<td>480</td>
<td>56.12</td>
<td>18.46</td>
<td>17.32</td>
<td>16.38</td>
</tr>
<tr>
<td>500</td>
<td>58.24</td>
<td>21.08</td>
<td>19.02</td>
<td>16.84</td>
</tr>
</tbody>
</table>

Figure 5-5 below depicts the power of the updated MA for various sample sizes of the new trial of prediction model 2-5.
Chapter 5  Designing new trials using meta analysis methods

**Figure 5-5**: Description of the results of the power of the updated MA of prediction models 2-5 in which the sample size ranging from 300-500.

<table>
<thead>
<tr>
<th>Prediction model 2</th>
<th>The power becomes positive when the sample size reaches 380. An unusual deviation of power has been identified when sample size is 480. This unexpected deviation could well be attributable to the sampling variation. A new analysis with increased number of simulations is required to confirm the deviation is caused by the sampling variation in the simulation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction model 3</td>
<td>The power becomes positive when the sample size reaches about 360. The power progresses steadily as the sample size increases.</td>
</tr>
<tr>
<td>Prediction model 4</td>
<td>The power remains zero until the sample size reaches up to 380. A steady progress in power is observed all the way, until the sample size reaches about 500, except when the sample size is about 460 where a slight dip in power is detected. This could be attributable to the sampling variation and performing 10,000 simulations should produce more accurate estimates as produced in the context of the prediction model 2 (figure 5-6).</td>
</tr>
<tr>
<td>Prediction model 5</td>
<td>The power turns non-zero when the sample size reaches about 380. The power progresses steadily as the sample size increases, apart from when the sample size is about 480. The increase of power at this particular sample size is inconsistent with the rest of the points. This could well be attributable to the sampling variation as illustrated in figure 5-6.</td>
</tr>
</tbody>
</table>
Figure 5-6 below shows the power curve of prediction model 2 in which each power estimate is based on 10,000 simulations.

**Figure 5-6**: The power curve of prediction model 2 based on 10,000 simulations.

With 10,000 simulations, the power of the updated MA increases smoothly as the sample size increases. As expected, results of this analysis do not indicate any deviation as is in the analysis with 5000 simulations. This confirms that the deviation observed in the analysis with 5000 simulations is due to the sampling variation in the simulation.

The summary plot below in figure 5-7 combines all individual power curves based on all prediction models developed in this chapter.

**Figure 5-7**: The combined power curves – Power of updated meta-analysis including the new trial, of all prediction models.
The power of the updated MA based on the fixed effect model (Prediction model 1) is substantially higher than all the RE models (Prediction models 2-5) at all sample sizes considered. The power of the updated MA which includes the new trial designed based on the predictive distribution accounted for both within and between study variance (Prediction model 4) is always smaller than that of the models which accounted for either within (Prediction model 2) or between study variations (Prediction model 3). These results suggest, as would be expected, that the more variability the predictive distribution of the new trial accounted for the less power the updated MA produces when the new trial is added to the existing MA.

The power of the updated MA in which the new trial is designed using the prediction interval based on the t-distribution (Prediction model 5) is lower than that of other random effect models. The power of this model is not far from that of prediction model 4, which is based on the normal distribution and accounted for both the within and the between study variations. The only difference between these two models is the underlying distributional assumptions about the predictive distribution of the effect size of the new trial. In theory, when the sample size is larger enough to justify the normal approximation of the t-distribution, the power should not be different.

However these results do not appear consistent with those obtained by Sutton et al (Sutton, A. J., Cooper et al. 2007) using a Bayesian approach to analyse the RE models. The power of the Bayesian RE model by Sutton et al is very close to zero. Although the power of updated MA based on RE prediction models in this study reaches a maximum of 40%, these two results are not quite
comparable. The results of Sutton et al are readily comparable with that in Chapter 9 of this thesis, which employs a Bayesian approach to design future trials.

In preceding sections, a prediction model based on the fixed effects MA model and four prediction models based on the random effects MA model have been developed to explore the power implications of the updated MA including the new trial. An investigation into the behaviour of the power of the new trial alone could help to understand the impact on the power following the inclusion of the new trial into the existing MA. Section 5.9 below explores the power of the new trial on its own (without inserting it into the existing MA).

5.9 Power of the new trial in isolation

5.9.1 Objectives

The objective of this section is to determine the power of the new trial in isolation. The power of a new trial alone should always be considered, and will often be regarded (at least by the trial investigators) as more important than the power of the subsequent MA. The power of a new trial dictates the number of patients that should be included in a trial, given the expected treatment effect and the significance level.

Thus far, the power of the updated MA including the new trial has been considered. To understand the impact of the inclusion of the new trial to the MA it is imperative to know the power of the new trials alone. The power of
individual trials designed based on prediction models described in sections 5.7.4, 5.8.5 – 5.8.8 is calculated to examine how the differences in the variation in the predictive distribution of a new trial influence the power of individual new trials.

5.9.2 Methods

5.9.2.1 Null hypothesis and power

The power of the new trial in isolation is calculated based on the null hypothesis that no difference in the efficacy of two treatments compared in the trials i.e. OR=1, at 5% level of significance.

5.9.2.2 Procedure

The process of powering the new trial in isolation entails designing a trial (steps 1-3 described in the methods section, i.e. section 5.6) and determining the power of the new trial based on the null hypothesis specified in section 5.9.2.1 above using the simulation based approach used before.

The power of the new trial developed based on prediction models 1-5 described in section 5.7.4, 5.8.5 – 5.8.8 is examined, and the results are presented in upcoming section 5.9.3.
5.9.3 Results

The prediction models developed in preceding sections 5.7.1, 5.8.3 – 5.8.7 are disparate in terms of the variability they account for in the prediction distribution of the new trial. Due to dissimilarities in the variance of predictive distributions, the power of the individual trials designed based on these prediction models are expected to vary. Table 5-5 below shows the power of individual trials based on distinct prediction models, against the sample size in each arm of the new trial.

**Table 5-5**: The power of the new trials in isolation designed based on respective prediction models against the sample size of the new trial.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed effects</td>
<td>Random effects</td>
<td>Random effects</td>
<td>Random effects</td>
<td>Random effects</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>13.40</td>
<td>19.30</td>
<td>25.52</td>
<td>28.64</td>
<td>34.24</td>
</tr>
<tr>
<td>200</td>
<td>21.56</td>
<td>32.26</td>
<td>38.60</td>
<td>41.08</td>
<td>47.38</td>
</tr>
<tr>
<td>300</td>
<td>30.50</td>
<td>39.84</td>
<td>46.50</td>
<td>51.44</td>
<td>53.24</td>
</tr>
<tr>
<td>500</td>
<td>41.96</td>
<td>51.90</td>
<td>55.84</td>
<td>61.74</td>
<td>63.92</td>
</tr>
<tr>
<td>800</td>
<td>54.72</td>
<td>60.90</td>
<td>64.82</td>
<td>68.48</td>
<td>70.64</td>
</tr>
<tr>
<td>1000</td>
<td>57.32</td>
<td>64.70</td>
<td>68.02</td>
<td>71.04</td>
<td>73.14</td>
</tr>
<tr>
<td>2000</td>
<td>71.22</td>
<td>75.24</td>
<td>75.76</td>
<td>79.82</td>
<td>81.48</td>
</tr>
<tr>
<td>3000</td>
<td>77.74</td>
<td>80.04</td>
<td>81.42</td>
<td>83.18</td>
<td>84.36</td>
</tr>
<tr>
<td>4000</td>
<td>81.60</td>
<td>82.36</td>
<td>84.16</td>
<td>85.64</td>
<td>86.24</td>
</tr>
<tr>
<td>5000</td>
<td>83.08</td>
<td>85.08</td>
<td>84.90</td>
<td>86.06</td>
<td>88.34</td>
</tr>
</tbody>
</table>

The corresponding summary plot is presented in figure 5-8 below.
Chapter 5  Designing new trials using meta analysis methods

Figure 5-8: The combined power curve – Power in isolation of the new trial designed based on prediction models 1-5.

The power of each individual trial increases as the sample size of the new trial increases. As expected, RE power curves reach a plateau when the sample size reaches about 5000 and FE Power curve asymptotically reach 100%. The new trial designed based on the fixed effects MA model (prediction model 1) produces the lowest power. The new trial designed based on the RE model where the prediction interval of the new effect size is derived using the t-distribution (prediction model 5) produces the highest power, as far as the power of the new trial alone is concerned.

The comparison of models those account for only the within study variation, indicates that the RE model produces relatively large power over the FE model, for all sample sizes considered. A greater power is observed in the RE model
which accounts for the between study variation alone over the model which accounts for the within study variation alone. As expected in theory, the two prediction models those account for both the within and the between study variations (Prediction model 4 and 5), produce almost equal power, in each sample size considered.

Thus far, the power of the updated MA including the new trial based on a variety of prediction models has been considered. In addition to that, the power of individual trials designed based on the same prediction models are considered. Upcoming section 5-10, attempts to integrate all these results.

5.10 **Comparison – The power of the new trial alone and that of the updated meta-analysis following the inclusion of the new trial**

It is aimed to determine the extent to which the power of the new trial is changed, following the inclusion of it into existing MA. Table 5-6 below lists the power of the new trial alone together with that of the updated MA for each prediction model.
Table 5-6: The comparison of power of the new trial alone and the power of the updated meta-analysis, of all prediction models, against the sample size of the new trial.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed effects</td>
<td>Random effects</td>
<td>Predictive distribution includes only within study variation</td>
<td>Predictive distribution includes only between study variation</td>
<td>Predictive distribution includes both within and between study variations</td>
</tr>
<tr>
<td>New trial</td>
<td>Updated MA</td>
<td>New trial</td>
<td>Updated MA</td>
<td>New trial</td>
<td>Updated MA</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>13.40</td>
<td>24.68</td>
<td>19.30</td>
<td>0</td>
<td>25.52</td>
</tr>
<tr>
<td>200</td>
<td>21.56</td>
<td>40.38</td>
<td>32.26</td>
<td>0</td>
<td>38.60</td>
</tr>
<tr>
<td>300</td>
<td>30.50</td>
<td>47.06</td>
<td>39.84</td>
<td>0</td>
<td>46.50</td>
</tr>
<tr>
<td>500</td>
<td>41.96</td>
<td>56.94</td>
<td>51.90</td>
<td>21.40</td>
<td>55.84</td>
</tr>
<tr>
<td>800</td>
<td>54.72</td>
<td>64.98</td>
<td>60.90</td>
<td>34.44</td>
<td>64.82</td>
</tr>
<tr>
<td>1000</td>
<td>57.32</td>
<td>69.88</td>
<td>64.70</td>
<td>38.36</td>
<td>68.02</td>
</tr>
<tr>
<td>2000</td>
<td>71.22</td>
<td>76.64</td>
<td>75.24</td>
<td>45.76</td>
<td>75.76</td>
</tr>
<tr>
<td>3000</td>
<td>77.74</td>
<td>80.76</td>
<td>80.04</td>
<td>46.82</td>
<td>81.42</td>
</tr>
<tr>
<td>4000</td>
<td>81.60</td>
<td>82.80</td>
<td>82.36</td>
<td>49.68</td>
<td>84.16</td>
</tr>
<tr>
<td>5000</td>
<td>83.08</td>
<td>84.34</td>
<td>85.08</td>
<td>50.08</td>
<td>84.90</td>
</tr>
</tbody>
</table>

From the table above it is apparent that the updated MA based on the FE model produces more power than the new trial alone at each sample size. In contrast, in all the RE prediction models the isolated power of the new trial diminishes following being included into existing MA. The more variability the prediction distribution of the effect size accounts for, the higher the power of the new trial alone. The exact reverse pattern to this is observed in the power of the updated
MA, in which the models accounted for more variability produce relatively less power.

### 5.11 Critical evaluation of the results

To enable a critical evaluation of the power of prediction models based on the FE and RE models, important summary estimates and other relevant information are presented in figure 5-9 below.

*Figure 5-9: A comparison of the output from initial fixed and random effects meta-analyses.*

Table 5-7 below combines the summary estimates of both fixed and random effects methods for the convenience of illustration.
Table 5-7: Table summarising the estimates obtained from initial fixed and random effects meta-analyses.

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>Log OR</td>
</tr>
<tr>
<td>Pooled mean effect</td>
<td>0.803</td>
<td>-0.219</td>
</tr>
<tr>
<td>Standard Error of the pooled mean effect</td>
<td>0.153</td>
<td>0.227</td>
</tr>
<tr>
<td>Lower Confidence Limit</td>
<td>0.595</td>
<td>0.494</td>
</tr>
<tr>
<td>Upper Confidence Limit</td>
<td>1.083</td>
<td>1.206</td>
</tr>
<tr>
<td>Heterogeneity estimate</td>
<td></td>
<td>0.1075</td>
</tr>
<tr>
<td>Variance of $i^{th}$ study</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2 + \tau^2$</td>
</tr>
<tr>
<td>Weight of $i^{th}$ study</td>
<td>$\frac{1}{\sigma_i^2}$</td>
<td>$\frac{1}{\sigma_i^2 + \tau^2}$</td>
</tr>
</tbody>
</table>

5.11.1 **Power of the new trial in isolation of the fixed effects model and the random effects model accounted for only within study variation**

The power of the new trial alone from the FE model (Prediction model 1) is smaller than that from the RE model which accounted for only the within study variation (Prediction model 2). The within study variation in the predictive distribution of a new effect size in each prediction model is deemed to be the standard error of the pooled mean effect size of the existing MA.

Although these two models accounted only for within study variation, the magnitude of the within study variation in the predictive distribution of the new trial in them are not identical. In the FE context this value is estimated as $0.153^2$ and in the RE context it is about $0.227^2$. 

105
In addition to that, the pooled log ORs are different in the initial fixed (0.803) and the random effects (0.772) meta-analyses. It is these pooled log ORs of the existing MA, which are considered the mean of the predictive distribution of a new trial. In this particular dataset, the resulting pooled OR of the fixed effects model is situated much closer to the null value (OR=1 or log OR=0) than the RE counterpart. The predictive distribution of a new effect size of the RE model accounted for more variability than in the FE model. Hence, an effect size drawn from the RE predictive distribution has more chance of being selected from the far end of the distribution (further from the pooled mean effect) and consequently being significant. This is because the confidence interval is less likely to include the null value. This causes the RE model to produce relatively high power compared to that of the FE counterpart as far as the power of the new trial in isolation is concerned.

The scatter plot in figure 5-10 below shows the association between the distance of an effect size (log OR) drawn from the predictive distribution based on the fixed effects MA model (prediction model 1) from the null value and the resulting p-values following the new trial is powered in isolation. Each point in this plot represents a new effect size drawn from the FE predictive distribution.
Each point in the above plot represents a new trial with 250 subjects in each arm. The red horizontal line demarcates the conventional p-value of 0.05. The further the effect size lies from the null value the more likely the p-value becomes less than 0.05, which indicates its significance.

In the context of the RE, as the variance of the predictive distribution increases, the new trial in isolation produces more power. Predictive distributions of all the RE models (prediction model 2-5) share a common mean. They are different only with respect of the variance in the predictive distribution. Increased variability allows effect sizes to be drawn from a wider range rather than from the more condensed ranges, and to locate further from the mean. Importantly this means further from the null value. The confidence interval of the effect size
of the new trial is therefore less likely to cross the null value (OR=1 or log OR=0) and hence the sampled value will positively contribute towards power.

The scatter plot below in figure 5-11 illustrates the above argument. The p-value of the effect size of the new trial is more likely to be below 0.05 if the effect size is further from the null value (log OR=0).

**Figure 5-11**: The p-value against the distance of the log OR from the null value in the random effects model.

Each scatter in this plot represents a new effect size drawn from the RE predictive distribution that accounted for both within and between study variations. Similar to the FE context, effect sizes below the horizontal red line at p-value 0.05 indicates the statistically significant effect sizes. In comparison to the FE context illustrated in figure 5-10, in the RE context a higher proportion of effect sizes are drawn further from the null value. It is noticeable that effect sizes further from zero are very likely to be significant and contribute towards
the power. The increased variability in the predictive distribution in RE prediction models allows effect sizes being drawn further from the null value. These effect sizes turn out to be statistically significant and increases the power.

### 5.11.2 Increased power in the fixed effects updated meta-analysis following the inclusion of the new trial to the existing meta-analysis

When the new trial designed based on the fixed effects MA model is included into the existing MA, the power of the updated fixed effects MA is shown to have been increased (prediction model 1). The variance of the pooled OR of the existing MA (with n number of trials) before the insertion of the new trial \(V_n\) is given in equation 5.23 as below.

\[
V_n = SE_n^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \ldots + \frac{1}{\sigma_n^2}}
\]  
(5.23)

The confidence interval \((CI_n)\) of the pooled mean OR of the existing MA is given in equation 5.24 as below.

\[
CI_n = \hat{\mu}_n \pm 1.96 \times SE_n(\hat{\mu}_n)
\]  
(5.24)

Following the insertion of the new trial to the existing MA, the variance of the pooled mean OR \(V_{n+1}\) is given in equation 5.25 as follows.

\[
V_{n+1} = SE_{n+1}^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \ldots + \frac{1}{\sigma_n^2} + \frac{1}{\sigma_{n+1}^2}}
\]  
(5.25)
The confidence interval \( (CI_{n+1}) \) of the pooled mean OR of the updated MA after adding the new trial is given in the equation 5.26.

\[
CI_{n+1} = \hat{\mu}_{n+1} \pm 1.96 \times SE_{n+1}(\hat{\mu}_{n+1}) \tag{5.26}
\]

The within study variation of the new trial \( (\sigma^2_{n+1}) \) is always positive and cause the denominator of equation 5.25 to increase following the insertion of the new trial. Consequently the variance \( (V_{n+1}) \) is decreased, indicating the reduction in the variance and the standard error of the pooled OR of the updated MA (i.e. \( V_{n+1} < V_n \)). Thus the confidence interval \( (CI_{n+1}) \) of the pooled OR of the updated MA becomes narrower than that \( (CI_n) \) of the existing MA (i.e. \( CI_{n+1} < CI_n \)). Narrower confidence intervals are less likely to include the null value and they hence are accounted with increased power of the updated MA.

### 5.11.3 Decreased power in the random effects updated meta-analysis following the inclusion of the new trial into the existing meta-analysis

In contrast to the FE model, the power of the updated random effects MA models are shown to have been decreased, compared to that produced by the new trial alone. The variance of the pooled OR of the existing random effect MA \( (V_n) \) is given in equation 5.27. \( \tau^2 \) denotes the heterogeneity estimate of the existing MA.

\[
V_n = SE^2_n = \frac{1}{\frac{1}{\tau_1^2 + \sigma_1^2} + \frac{1}{\tau_2^2 + \sigma_2^2} + \cdots + \frac{1}{\tau_n^2 + \sigma_n^2}} \tag{5.27}
\]
Following the inclusion of the new trial, the variance of the updated MA ($V_{n+1}$) is given in the equation 5.28. $\tau_2^2$ denotes the heterogeneity estimate of the updated MA.

$$V_{n+1} = SE_{n+1}^2 = \frac{1}{\frac{1}{\tau_2^2 + \sigma_1^2} + \frac{1}{\tau_2^2 + \sigma_2^2} + \cdots + \frac{1}{\tau_2^2 + \sigma_n^2} + \frac{1}{\tau_2^2 + \sigma_{n+1}^2}}$$

(5.28)

Insertion of a new trial, which is drawn from a predictive distribution with a wider variance (which account for both within and between study variance) to the existing MA could cause the heterogeneity of the updated MA to increase, i.e. ($\tau_2^2 > \tau_1^2$). Increased $\tau_2^2$ causes the denominator of the equation 5.28 to decrease, given $\sigma_i^2$ remains constant. The reduced denominator makes the magnitude of $V_{n+1}$ become larger than $V_n$. This in turn affects the confidence interval of the pooled OR of the updated MA ($CI_{n+1}$) and cause to make it wider than that of existing MA ($CI_n$). Wider confidence intervals are more likely to include the null value (OR=1 or log OR = 0) and lead to diminish the power of the random effects updated MA.

### 5.11.4 Zero power of the updated random effects meta-analysis for smaller sample sizes of the new trial

In all the RE prediction models, the power of the update MA remains zero for smaller sample sizes of the new trial (figure 5-4). Sample size is inversely related to the standard error of the mean and hence the power of the test (Maxwell, Kelley and Rausch 2008). The inverse variance method of MA, which is the method employed in this thesis, assigns a lower weight to those trials with higher variances and vice versa (Borenstein, M., Hedges et al. 2009).
Therefore, a new trial with smaller sample size (larger variance \( \sigma^2 \)) is assigned a relatively smaller weight \((w_{n+1})\). Equation 5.29 below presents the calculation of the pooled effect estimate \( (\hat{\mu}_{n+1}) \) of the updated MA using the inverse variance weighted method.

\[
\hat{\mu}_{n+1} = \frac{w_1 \hat{\theta}_2 + w_2 \hat{\theta}_2 + \cdots + w_n \hat{\theta}_n + w_{n+1} \hat{\theta}_{n+1}}{w_1 + w_2 + \cdots + w_n + w_{n+1}}
\]  

(5.29)

The lower the weight assigned to the new trial \((w_{n+1})\) the less its influence on determining the pooled effect size \( (\hat{\mu}_{n+1}) \). The location of the pooled effect size is hence not much affected by including a small new trial. The variance of the pooled OR of the updated MA is determined using the formulae given in equation 5.30.

\[
Var(\hat{\mu}_{n+1}) = \frac{1}{w_1 + w_2 + \cdots + w_n + w_{n+1}}
\]  

(5.30)

As the weight assigned to the new trial \( (w_{n+1}) \) is decreased the variance of the pooled effect size is increased \((Var(\hat{\mu}_{n+1}))\) and leading to a wider confidence interval and hence lower power. The reason for the power of the updated MA to remain zero for smaller sample sizes of the new trial could thus be due to the weight assigned being not sufficient to influence the location of the pooled effect size until the sample size increases above a certain value.
5.11.5 **Increased power in the updated meta-analysis following the inclusion of the new trial based on the prediction model accounted both within and between study variation**

The greater the weight assigned to the new trial the more influence it has on determining the pooled effect size (equation 5.29) and causing the pooled OR to be dragged toward the location of the effect size of the new trial. In the RE model which employs both within and between study variances, the wider prediction distribution makes it possible to draw effect sizes far from the null value (OR=1 or log OR=0). The fact that the new trial is far from the null value coupled with the narrower confidence interval makes the confidence interval less likely to include the null value and hence increased power.

5.11.6 **Larger sample sizes of the new trial**

Sample sizes of the trials included in the existing MA are not particularly large, so a new trial with a large sample size attracts a greater weight and the new trial begins to dominate the updated MA and has a greater influence on both the pooled mean effect size and corresponding confidence interval. The greater the weight assigned to the new trial the less the variance of the pooled OR. This in turn produces a narrower confidence interval of the pooled OR and leads to increase the power of the updated MA.

This section critically evaluates the results obtained from the analysis conducted throughout this chapter. Following section 5.12 discusses the results
of this chapter in the light of similar results and discusses some specific issues related to the work.
5.12 **Discussion**

This chapter adopts a frequentist approach to develop a framework of methodologies to design a new trial based on previous evidence using evidence synthesis models. Although most aspects of trial design could be influenced by previous evidence, this thesis primarily concerns about calculating the sample size of RCTs. The use of MA has been increasingly recognised to inform healthcare decisions and policies. In order to produce reliable results, the updated MA including the new trial needs to be adequately powered. The distinct feature of this methodology is that it takes into account not merely the power of the new trial alone but also the power of the updated MA too, when designing a new trial. Moreover, the power of the new trial in isolation has been explored to discover the extent to which the addition of a new trial affects the power of the existing MA.

5.12.1 **Key findings of this chapter**

This thesis extends the methodology developed by Sutton et al, using a variety of prediction models based on frequentist random effects MA. To inform the design of the new trial Sutton et al employed a Bayesian random effects MA approach. In contrast, this chapter devises a variety of prediction models to inform the design of a new trial based on frequentist RE principles. This chapter employs the same existing evidence used by Sutton et al to base the new trial design to enable a comprehensive comparability.

When the new trial is designed using the fixed effects MA principles both this study and Sutton et al obtained approximately equal power estimates in the
updated MA (section 5.7.4). Chapter 9 of this thesis employs a Bayesian MA approach to develop the methodology of designing new trials, which enables a comparison of the Bayesian results obtained in this chapter with those by Sutton et al using Bayesian RE model.

Five prediction models have been designed based on frequentist principles of MA and each model adopts a different prediction distribution to draw effect sizes of new trials. The variance of each predictive distribution is set to be different from each other to examine the impact of the changes to the variance on power. When powering the trials in isolation, the FE model reported the lowest power and the RE model that accounted for both the within and the between study variation based on the t-distribution reported the highest power. The more variance accounted in the predictive distribution the more power the trial in isolation produces.

The power of the updated MA based on the FE model produces more power than that of the new trial designed using the FE MA principles. This result is in consistent with that of Sutton et al. The power of the updated MA based on all RE models diminishes following the inclusion of the existing MA to the new trial. Only difference between the FE and RE models are the inclusion of the heterogeneity element in to the analysis. Therefore, the reason for the contrasting trend in power of FE and RE prediction models following the inclusion of new trial is the introduction of the heterogeneity element in to the analysis.
Although the RE models those accounted for either the within or between study variance in the predictive distribution of the new trial might not relevant in practice, they provide an insight into the extent of the impact of each variable component of the existing MA on the power of the new trial and updated MA.

### 5.12.2 Related research

The methodology to design new trials based on existing evidence was first introduced by Sutton et al (Sutton, A. J., Cooper et al. 2007), in which they recommended to consider the future MA including the new trial at the planning and design stages of a new trial.

The most recent and possibly the most relevant related work to the MA aspect of present work is proposed by Roloff and Higgings (Roloff and Higgins 2011). They implement the adaptive clinical trial design methods proposed by Bauer and Kohne (Bauer and Kohne 1994) to design the next study, which will then be added into existing MA. They consider the meta-analytic result in a similar vein in this thesis, as a basis for designing a new trial. Moreover, they emphasized the value of using conditional power as opposed to unconditional power in computing the additional information size required in a new trial, in the presence of existing evidence.

Their findings are very similar to that found in this thesis, where both conclude that in the presence of heterogeneity one additional study might not be sufficient to yield a statistical significant result in the updated MA. However, our approach does not require the specification of clinically meaningful minimum effect size to act as the alternative hypothesis, which is not trivial to find. Beside,
simulation based approach adopted here does not require approximations as required in their approach that leads naturally to algebraic formulae (Roloff, Higgins and Sutton 2012).

Thorlund and colleagues also proposed an approach to determine the sample size of a future trial in the settings of prospective MA (Thorlund, Anema and Mills 2010). They did not consider the power but simply focused on the statistical significance of the subsequent MA. They illustrated the utility of the Trial Sequential Analysis approach to determine additional number of subjects needs to be randomised before the prospective MA becomes conclusive.

Hinchliffe et al (Hinchliffe, Crowther et al. 2012) also explored the power implications of the updated MA to inform the design of future trials. However, their work is content specific and different to the present work, as the study focuses only on diagnostic accuracy test studies whilst the present work is applicable to any clinical trial context in general.

5.12.3 Limitations of the methodology

Moreover, the methodology adopted is limited just to the case of balanced trials in terms of the number of subjects in each arm. The methodology could be enhanced to incorporate the design of unbalanced trials as well. In addition to that, design of trials with multiple arms could be explored as future work, as this chapter limits to design trials with two arms.
5.12.4 **Potential Future work**

This methodology requires specifying the estimate of the event rate in the control group of the new trial. In the present work, the mean event rate in the control group of existing trials is used, assuming that this parameter is exactly known without uncertainty. An aspect of improvement to this methodology could be to discover a reliable way to incorporate the uncertainty in the estimation of the event rate in the control group, in a frequentist platform.

To make this methodology usable in practice, it is essential to develop purpose-built user-friendly software, which is capable of fitting approximate power curves with minimum number of points. This might facilitates the exploration of alternative design strategies by way of a sensitivity analysis to choose the best possible design strategy.

However, results obtained in this chapter should be interpreted with care, as they are dataset-specific. This methodology needs to be applied to a variety of datasets before any general conclusions are made confirming the results obtained here are not atypical.

This section provides a general discussion to this chapter by discussing the results and recommending future related work. Upcoming section 5.13 concludes this chapter.
5.13 Conclusions

In certain areas where some evidence already exists, the importance of MA is increasingly recognised as a quantitative and explicit framework to inform the design of new trials. Despite the importance and widespread recognition, MA is practically underutilised as a basis to inform the trial design (Cooper, N. J., Jones and Sutton 2005). The lack of practical embracement is thought to be not because of a philosophical objection to the idea but due to unavailability of a well developed framework of methodology and facilitating software.

The power of the updated MA is very much dependent on the MA model employed to combine the trials. The appropriateness of the choice of model is hindered by the existence of publication bias in trials, as the presence of publication bias would produce misleading results. Moreover, to avoid misleading results it is recommended to assess the validity of the existing MA.

This thesis considers the context of a set of binary data based on OR scales to develop the methodology to design RCTs. The methodology could be extended to use with other outcome indices and importantly to use data measured on a continuous scale.

Since meta-analytic principles are increasingly used in other areas such as epidemiology, this approach of designing trials based on existing evidence could be generalisable to underpin the design of other (epidemiological) studies.

If the interest is focused on the results of the new trial alone, then it is recommended to employ the random effects MA principles (both within and
between variance) to base the design on. To achieve the optimum power the method that employs the t-distribution to approximate the predictive distribution should be used.

If the focus is on the results of the updated MA, to achieve a greater power, the new trial should be designed and updated using the fixed effects MA principles. However, the widespread reluctance to accept the fixed effects MA method as a standard practice might prevent widespread acceptance of this methodology. When the new trial is designed based on the random effects MA methods, the updated MA provides a modest power.

In fact, the recommended model to use in practice relies on available size of the sample as well. Since the power of the updated MA remains zero until the sample size increases up to a certain level, the model based on random effect model would not be the best option, given the size of the available sample is small. The FE method proves to be the best option in this context.

As a concluding remark, it should be acknowledged that the methodology outlined here is rather a logical approach, if adhered to, which would optimize the use of available scarce resources by preventing underpowered trials being designed.
CHAPTER 6

Designing future trials based on existing evidence using indirect comparison meta-analysis methods

6.1 Introduction

This chapter develops a methodology to design future trials based on existing evidence using indirect comparison MA methods. This methodology is eventually integrated into the broader framework of methodology developed to design future trials, which is the primary objective of this thesis. Following this introduction the motivation to the work in this chapter is explained in section 6.2. Specific objectives of this chapter are outlined in section 6.3. Section 6.4 briefly introduces the indirect comparison MA methods. Section 6.5 discusses about the gathering of existing evidence on which the initial meta-analyses are based to derive the indirect estimates of the new head-to-head trials. The methods employed in this chapter are detailed in section 6.6. Section 6.7 develops the methodology to design new head-to-head trials. Within that sub section 6.7.1 employs the FE methods and section 6.7.2 employs the RE method to specify the predictive distribution of the new head-to-head trial. A critical evaluation of the results is given in section 6.8. The discussion of this chapter is given in section 6.9. Section 6.10 concludes the chapter.
6.2 Motivation

Rapid advances in health technology lead to an increase in the number of treatments available for a given clinical condition. To receive the regulatory approval and for economic evaluation, a new drug should demonstrate its efficacy over only placebo in a controlled trial (Signorovitch, Wu et al. 2010). It is not mandatory for many of these competing treatments to be compared head-to-head in a RCT setting, which most reliably measures the relative efficacy between them. In the absence of head-to-head trials between competing treatments, indirect evidence could possibly provide the most reliable measures of the relative efficacy. Empirical evidence has shown that the results of indirect comparisons are often not significantly different from those of direct comparisons (Song, F., Altman et al. 2003).

Although indirect evidence is employed to evaluate competing treatments, the potential of indirect evidence as a basis for designing new head-to-head trials has not been considered elsewhere.

6.3 Objectives

The primary objective of this chapter is to develop a methodology to design new head-to-head trials using indirect comparison MA techniques, based on existing evidence. This methodology is particularly useful in situations where direct comparisons between two competing treatments are limited or not available at all.
Chapter 6  Designing new trials using indirect comparison methods

The aim here is to design new head-to-head trials using a series of prediction models with distinct predictive distributions from which an effect size of the new trial is drawn.

The section below briefly introduces the indirect comparison MA.

6.4  Introduction to indirect comparison meta-analysis

Section 2.5 of the literature review in Chapter 2 consists of an extensive review of the indirect comparison MA methods. Two main approaches have been identified in the literature for conducting an indirect comparisons, namely the naïve and the adjusted methods (Song, F. Altman, D. G. Glenny, A. Eastwood, A. Deeks, J. 2007). These two methods are discussed briefly in sections 6.4.1 and 6.4.2 below.

6.4.1  Naïve method of indirect comparison

The Naïve approach compares one treatment from one trial with the second treatment from a different trial. Such indirect comparisons may lose the power of the randomisation. Therefore, the comparison may suffer the biases that observational studies often suffer, e.g. confounding. The comparison of different arms from different trials could introduce selection bias and subsequently lead to generation of inaccurate estimates and invalid policy decisions. (Glenny, Altman et al. 2005)
6.4.2 **Adjusted method of indirect comparison**  
**(Bucher’s method)**

Different patient characteristics and other prognostic factors could cause discrepancies between the relative efficacies of treatments compared indirectly. To overcome the drawbacks of the naïve method, Bucher et al suggested a method, which involves adjusting the relative efficacies of two treatments ($T_1$ and $T_2$) with a common comparator (C) to determine the relative efficacy between two treatments ($T_1$ and $T_2$) (Bucher, Guyatt et al. 1997).

This chapter employs Bucher s’ method of adjusted indirect comparisons to develop the methodology of designing new trials. For illustration, the outcome is measured on OR scale of binary data although it is readily modifiable to accommodate other outcome indices such as risk ratio and risk differences.

Bucher introduced two formulae to compute the indirect log OR (equation 6.1) and its variance (equation 6.2). The adjusted indirect log OR between treatment $T_1$ and $T_2$ ($log^{ind} OR_{T1T2}$) is estimated by equation 6.1 below.

$$log^{ind} OR_{T1T2} = log OR_{CT2} - log OR_{CT1}$$

(6.1)

The variance of the indirect log OR between treatments $T_1$ and $T_2$ is given by equation 6.2 below.

$$Var[log^{ind} OR_{T1T2}] = Var[log OR_{T1C}] + Var[log OR_{T2C}]$$

(6.2)

where $log OR_{T1C}$ denotes the log OR between the treatments $T_1$ and C and $log OR_{T2C}$ represents the log OR between the treatment $T_2$ and C.
6.4.3 Underlying assumptions

The validity of both the results and the methods of an IC relies on two main assumptions being fulfilled: the homogeneity and the similarity assumptions (Glenny, Altman et al. 2005). These two assumptions are discussed briefly in upcoming sections 6.4.3.1 and 6.4.3.2 respectively.

6.4.3.1 Homogeneity assumption

The homogeneity assumption in IC contexts is the same assumption as in standard MA. In the FE model, all trials are assumed to be estimating one underlying true effect size. In the RE model, different underlying effect sizes are assumed to be a random sample from effect size population, distributed around a typical value (DerSimonian, Rebecca and Laird 1986).

When multiple treatments are involved, in IC context, the homogeneity assumption of standard MA should be met. That is, the two sets of trials comparing two treatments with a common comparator should be homogeneous enough to be combined in a MA.

6.4.3.2 Similarity assumption

The similarity assumption in the settings of adjusted IC means that the relative efficacy of the treatment is consistent in patients included in different trials. In other words, the results of the $T_1$ vs C trial would have been observed in the $T_2$ vs C trial if treatment $T_2$ was replaced with treatment $T_1$, and vice versa. Further, the two sets of trials involved in indirect comparisons are assumed interchangeable. (Song, F., Glenny and Altman 2000)
Having outlined the underlying assumptions concerning the IC methods in this section, upcoming section 6.5 describes the nature of the existing evidence used and the process of gathering them.

6.5 **Existing evidence**

Since the methodological framework being developed throughout this thesis to design new trials is based on evidence synthesis principles, a prerequisite is a set of meta-analysable trials to constitute the existing evidence. Two groups of trials comparing treatments $T_1$ vs $C$ and $T_2$ vs $C$, where $C$ being the common comparator are required to conduct two separate initial meta-analyses to estimate the relative efficacies of the treatment $T_1$ and $C$ ($\log OR_{T_1C}$) and of the treatment $T_2$ and $C$ ($\log OR_{T_2C}$).

Salanti and colleagues (Salanti, G., Kavvoura and Ioannidis 2008) cite an article reporting a potential set of meta-analysable trial results (Otoul, Arrigo et al. 2005). In this paper Otoul and the colleagues (Otoul, Arrigo et al. 2005) conducted an indirect comparison of Levetiracetam (LEV) with other second generation antiepileptic drugs as add-on therapy in treating partial epilepsy. Topiramate (TPM) is one of the six antiepileptic drugs to which LEV is indirectly compared. LEV and TPM add-on therapy are thus identified as the two treatments to be compared head-to-head in an indirect settings. Both LEV and TPM have already been compared with placebo (PLC) in a few trials. For illustrative purposes, LEV, TPM and PLC is treated as $T_1$, $T_2$ and $C$ respectively.
6.5.1 **Trials comparing LEV add-on therapy and PLC**

6.5.1.1 **Initial list of trials comparing LEV and PLC**

In 2005, Otoul et al (Otoul, Arrigo et al. 2005) conducted a MA of trials estimating the efficacy and the safety of LEV as add-on therapy (in combination with other prescribed medications) compared to PLC in treating patients with partial epilepsy. The response rate and the withdrawal rate (as a measure of safety) were evaluated as outcome measures. The trials included in the MA are tabulated below in table 6-1.

*Table 6-1: Initial list of trials comparing LEV add-on therapy and PLC in treating partial epilepsy extracted from Otoul et al.*

<table>
<thead>
<tr>
<th>Trial</th>
<th>PLC</th>
<th>LEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorvon et al (Shorvon, Löwenthal et al. 2000)</td>
<td>7</td>
<td>112</td>
</tr>
<tr>
<td>Cereghino et al (Cereghino, Biton et al. 2000)</td>
<td>7</td>
<td>95</td>
</tr>
<tr>
<td>Ben-Menachem et al (Ben-Menachem, E. and Falter 2000)</td>
<td>15</td>
<td>105</td>
</tr>
</tbody>
</table>

As this MA was performed in 2000, it is very likely that new trials addressing the same therapeutic questions have been published since. A literature search was undertaken to identify similar new trials published since year 2000, addressing the same therapeutic questions, to update the list of trials included in the MA.

6.5.1.2 **Updated list of trials comparing LEV and PLC**

To identify new relevant trials published since 2000, an extensive search was carried out in the clinical trials section of the Cochrane library, Pubmed and
Medline databases using keywords ‘levetiracetam’ ‘placebo’ ‘epilepsy’ and ‘randomised’ in any field.

As expected, the Pubmed and Medline databases produce identical results. The search results from the Cochrane database of RCTs, produced a list of 10 records. Out of ten reports, one was excluded from the analysis as it is not fully reported as yet; only an abstract is available (Tonner 2008). Another report is excluded on the basis of not being a randomised trial (Ben-Menachem, E., Edrich et al. 2003). The third report is removed from the analysis because it is a systematic review but not a randomised trial (Cramer 2003).

Three more trials are excluded from the analysis due to mismatches in population. Two of them report of treating idiopathic generalised epilepsy but not partial epilepsy, which is the patient population of concern here (Noachtar 2008) (Berkovic 2007). The third trial left out on the population mismatch basis had recruited patients with pharmaco-resistant focal epilepsy (Wang-Tilz, Stefan et al. 2005).

One more trial is left out from the analysis because it does not consider LEV add-on therapy, which is the treatment regime of interest in this analysis. This article reports the efficacy of LEV administered intravenous in contrast to orally administered LEV (Ramael 2006).

Having excluded seven items from the list of 10 reports, the search results reduces to three reports. These are integrated into the list of trials to form the existing evidence. One of these three eligible trials is dedicated to paediatric population (Verdru 2005). Since the inclusion criteria of the original MA in Otoul
et al has not put in place any age limitations, inclusion of this particular trial would not affect the consistency of the MA. In fact, a significant number of child patients are included in the trials of the original MA comparing LEV and PLC.

Second of the three trials found suitable to be added, compares the efficacy of two LEV doses in treating partial epilepsy (Boon 2002). A few of other trials included in existing MA also compare multiple LEV doses; all LEV doses are combined in the analysis in computing the efficacy of LEV against PLC. Therefore, this particular trial meets the inclusion criteria to be included into the existing MA.

Although the third of the three trials is set in a Taiwanese population (Tsai, Yen et al. 2006), it is eligible to be included in to existing MA as it meets the inclusion criteria.

This section uncovers three RCTs as suitable to be included in to the existing dataset listed in table 6-1. The table 6-2 below shows the number of events (patients with > 50% response rate) and the total sample size in each of six trials comparing LEV add-on therapy with PLC in treating partial epilepsy.
Table 6-2: The updated list of trials comparing LEV add-on therapy with PLC to constitute existing evidence to base the analysis on.

<table>
<thead>
<tr>
<th>Trial</th>
<th>PLC</th>
<th>LEV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shorvon et al (Shorvon, Löwenthal et al. 2000)</td>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td>Cereghino et al (Cereghino, Biton et al. 2000)</td>
<td>7</td>
<td>76</td>
</tr>
<tr>
<td>Ben-Menachem et al (Ben-Menachem, E. and Falter 2000)</td>
<td>15</td>
<td>71</td>
</tr>
<tr>
<td>Tsai et al (Tsai, Yen et al. 2006)</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Verdu et al (Verdu 2005)</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>Boon et al (Boon 2002)</td>
<td>21</td>
<td>108</td>
</tr>
</tbody>
</table>

This section describes the process of obtaining an up to date set of trials comparing LEV add-on therapy with PLC. Section 6.5.2 below describes the approach of obtaining an up to date list of meta-analysable trials comparing TPM add-on therapy and PLC.

6.5.2 Trials comparing TPM add-on therapy and PLC

The aim of this section is to identify relevant trials published since 2000 to be included into the existing evidence, comparing TPM add-on therapy with PLC. A search in the Cochrane database of systematic reviews has uncovered a systematic review conducted in 2008 which includes up-to-date list of trials comparing TPM add on therapy and PLC for patients with partial epilepsy (Jette 2008). This review consists of 10 trials comparing response rate of patients treated with TPM add on therapy against PLC. No further relevant trials published since 2008 could be found in the literature. The list of trials included in the Cochrane review is given in table 6-3 below.
Table 6-3: The updated list of trials comparing TPM add-on therapy with PLC in treating partial epilepsy that constitutes existing evidence to base the analysis on extracted from Jette 2008.

<table>
<thead>
<tr>
<th>Trial</th>
<th>PLC</th>
<th>TPM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Total</td>
</tr>
<tr>
<td>Ben-Menachem (Ben-Menachem, E., Sharief et al. 1996)</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Elterman (Elterman, Glauser et al. 1999)</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>Faught (Faught, Wilder et al. 1996)</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>Guberman (Guberman, Neto and Gassmann 2002)</td>
<td>22</td>
<td>92</td>
</tr>
<tr>
<td>Korean (Group 1999)</td>
<td>11</td>
<td>85</td>
</tr>
<tr>
<td>Privitera (Privitera, Fincham et al. 1996)</td>
<td>4</td>
<td>47</td>
</tr>
<tr>
<td>Rosenfeld 1996</td>
<td>8</td>
<td>42</td>
</tr>
<tr>
<td>Sharief (Sharief, Viteri et al. 1996)</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Tassinari (Tassinari, Michelucci et al. 1996)</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>Yen (Yen, Yu et al. 2000)</td>
<td>3</td>
<td>23</td>
</tr>
</tbody>
</table>

6.5.3 Definition of an event

The outcome of interest in all the trials is the response rate. In here, an event is defined as the reduction in seizure frequency (at least 50 percent) compared that at the baseline. The responder rate is one of the secondary outcomes measured in each individual trial listed in table 6-2 and 6-3.

6.5.4 Estimating the events rate in the control group

The methodological framework developed in this thesis requires the control group event rate \( P_{T2, new} \) to be known beforehand. In the absence of expert opinion \( P_{T2, new} \) is estimated from existing evidence.
For convenience, in the settings of indirect comparisons, $T_1$ is set to represent the treatment whereas $T_2$ is deemed the control. The event rate in the $T_2$ group (control group in this setting) of the new trial ($P_{T2,new}$) is determined by taking the mean of the event rate in trials comparing $T_2$ with $C$ listed in the existing evidence (table 6-2 and 6-3).

In the running example, LEV and TPM represent $T_1$ and $T_2$ respectively and the control group event rate ($P_{T2,new}$) is estimated to be 0.4477.

This section describes the way in which an up-to-date list of trials is obtained to constitute existing evidence to base the initial analysis necessary to the development of the methodology. The forthcoming section 6.6 explains how the existing evidence is utilised in developing the methodology to design new trials.

**6.6 Methods**

**6.6.1 Null hypothesis tested and power in the context of indirect comparisons**

In the context of indirect comparisons, the focus is restricted to the power of the new head-to-head trial in isolation (In contrast to the power of the updated MA considered elsewhere in this thesis). The use of indirect evidence in place of direct evidence is justified by the lack of availability of direct evidence. Hence, the head-to-head trial designed using indirect evidence could not be included into a MA of head-to-head trials to discover the power of the updated MA. This
limits the focus of this chapter only to discover the power implications of the new head-to-head trials alone.

The null hypothesis in the indirect comparison MA states that the efficacy of two competing treatments is not different, i.e. two competing treatments are equally effective ($H_0: OR=1$). The decision as to whether to reject or do not reject the null hypothesis is based on 5% level of significance.

### 6.6.2 Procedure

Throughout this thesis a unique methodology is employed to design new trials, which is first introduced by Sutton and the colleagues (Sutton, A. J., Cooper et al. 2007). However, when addressing different situations a modification to this methodology is necessary. This general methodology is tailored to address the context of indirect comparisons and is specified in nine steps below.

**Step 1**

Bucher's adjusted indirect estimate of the log OR comparing treatment $T_1$ and $T_2$ is calculated using the formulae given in equation 6.1. The variance of the indirect estimate of log OR between treatment $T_1$ and $T_2$ is calculated using the formulae given in equation 6.2. Bucher's adjusted indirect estimates are then employed to specify the predictive distribution of a new head-to-head trial ($\theta_{T1T2(new)}$) comparing treatments $T_1$ and $T_2$. 
Step 2

An effect size perceived to be from the new head-to-head trial ($\hat{\theta}_{T1T2(new)}$) is drawn from the predictive distribution specified in step 1. This drawn effect size represents the relative efficacy between $T_1$ and $T_2$ in the new head-to-head trial.

Step 3

The effect size ($\hat{\theta}_{T1T2(new)}$) drawn in step 2 and the value of $P_{T2.new}$ computed in section 6.5.4 above are then substituted to the equation 5.1 in section 5.6.2 of Chapter 5, to derive the event rate in the $T_1$ group of the new trial ($P_{T1.new}$).

Step 4

The number of subjects in the $T_1$ group ($n_{T1.new}$) and the $T_2$ group ($n_{T2.new}$) of the new trial is (sample size of the new head-to-head trial) then specified.

Step 5

The number of events in the $T_1$ group ($r_{T1.new}$) and the $T_2$ group ($r_{T2.new}$) are binomially simulated as specified in equation 5.2 and 5.3 of section 5.6.2 in Chapter 5. Subsequently, the 2x2 table representing the outcome of the new trial is constructed.

Step 6

The effect size (log OR and standard error) of the new head-to-head trial is then computed.
This is mathematically performed by conducting a meta-analysis of the new head-to-head trial in isolation.

**Step 7**

The decision as to whether to reject or not to reject the null hypothesis is then made based on the criteria specified in section 6.6.1 above.

**Step 8-9**

Step 8 and 9 are identical to corresponding steps specified in section 5.6.2 of Chapter 5 apart from power of the new trial in isolation is computed in here in contrast to the power of the updated MA considered in Chapter 5.

This section explains the methodology employed to compute the power of the new head-to-head trial designed using indirect comparison MA methods.

### 6.7 Designing a new head to head trial using indirect comparison principles

Several prediction models based on indirect comparison MA principles are employed to investigate the impact of the change in the within and the between study variations of the predictive distribution on statistical power of the new trial alone. Both fixed and the random effects MA methods underlie these prediction models. The FE method of indirect comparison MA underlies the prediction model 6, where as the RE indirect comparison method of MA underlies the prediction model 7-10.
6.7.1 **Designing a new trial based on existing evidence using indirect comparison meta-analysis methods (fixed effects method)**

6.7.1.1 **Initial fixed effects meta-analysis**

The existing evidence compiled in sections 6.5.1 and 6.5.2 is used to conduct two initial FE meta-analyses. The resulting summary effects are used to derive the indirect estimates of the relative efficacy of the competing treatments from which the parameters of the predictive distribution of an effect size of new head-to-head trial. Sections 6.7.1.1.1 and 6.7.1.1.2 below present the results of the two initial FE meta-analyses of trials comparing PLC vs LEV and PLC vs TPM respectively.

6.7.1.1.1 **Initial fixed effects meta-analysis of trials comparing LEV add-on therapy with PLC**

The list of trials shown in table 6-2 is subjected to a fixed effects MA and the resulting forest plot is given in figure 6-1.
Figure 6-1: Resulting forest plot of the fixed effects MA of trials comparing PLC and LEV add-on therapy.

The initial fixed effects MA of six trials comparing PLC and LEV produced a pooled OR \((OR_{CT1})\) of 0.24 with a confidence interval of 0.18-0.31. (The corresponding log OR \((\log OR_{CT1})\) is -1.4457 with standard error \((se(\log OR_{CT1}))\) of 0.1423). The next section conducts the fixed effects MA of trials comparing PLC with TPM add-on therapy.

6.7.1.1.2 Initial fixed effects meta-analysis of trials comparing PLC with TPM add-on therapy

The resulting forest plot of the fixed effects MA of trials comparing TPM adjunctive therapy and PLC is shown in figure 6-2.
Figure 6-2: Resulting forest plot of the fixed effects MA of trials comparing PLC and TPM add-on therapy.

<table>
<thead>
<tr>
<th>Study</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben-Menachem</td>
<td>0.02 (0.00, 0.42)</td>
<td>5.82</td>
</tr>
<tr>
<td>Elterman</td>
<td>0.39 (0.15, 1.02)</td>
<td>6.35</td>
</tr>
<tr>
<td>Faught</td>
<td>0.33 (0.14, 0.76)</td>
<td>10.47</td>
</tr>
<tr>
<td>Guberman</td>
<td>0.38 (0.22, 0.68)</td>
<td>19.43</td>
</tr>
<tr>
<td>Korean</td>
<td>0.15 (0.07, 0.31)</td>
<td>18.15</td>
</tr>
<tr>
<td>Privitera</td>
<td>0.14 (0.05, 0.40)</td>
<td>12.45</td>
</tr>
<tr>
<td>Rosenfeld</td>
<td>0.22 (0.10, 0.51)</td>
<td>13.27</td>
</tr>
<tr>
<td>Sharief</td>
<td>0.17 (0.03, 0.92)</td>
<td>3.55</td>
</tr>
<tr>
<td>Tassinari</td>
<td>0.13 (0.03, 0.51)</td>
<td>5.97</td>
</tr>
<tr>
<td>Yen</td>
<td>0.16 (0.04, 0.71)</td>
<td>4.54</td>
</tr>
<tr>
<td>Overall (I-squared = 16.6%, p = 0.290)</td>
<td>0.23 (0.17, 0.31)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Initial MA of ten trials comparing PLC and TPM add-on therapy produced an OR ($OR_{CT2}^F$) of 0.23 with the confidence interval range from 0.17 to 0.31. The corresponding log OR ($\log OR_{CT2}^F$) is -1.4745 with the standard error ($se(\log OR_{CT2}^F)$) of 0.1502.

Following the estimation of the individual summary estimates of each treatment with the common comparator, upcoming section 6.7.1.2 combines them to derive the indirect estimates of the relative efficacy of the competing treatments.
6.7.1.2 Specify the predictive distribution of a new head-to-head trial

6.7.1.2.1 The mean of the predictive distribution of the new head-to-head trial

This methodology derives the mean of the predictive distribution from the indirect pooled log OR of existing indirect evidence. The FE indirect estimates of the relative efficacy between LEV and TPM are calculated using the Bucher's formulae specified in equation 6.1.

\[ \log^{ind} OR_{T1T2}^F = -1.4745 \cdot (-1.4457) = -0.0288 \]

6.7.1.2.2 The variance of the predictive distribution of the new head-to-head trial

In the FE context, the variance of the predictive distribution of a new head-to-head trial is derived from the variance of the indirect pooled log OR of the indirect evidence, which is calculated as specified in equation 6.2 for the running example.

\[ Var[\log^{ind} OR_{T1T2}^F] = (0.1423)^2 + (0.1502)^2 = 0.0428 \]

The standard error of the indirect relative efficacy of treatment T1 and T2 is specified as below.

\[ se[\log^{ind} OR_{T1T2}^F] = 0.2069 \]
The prediction model 6 in upcoming section 6.7.1.3 develops the methodology to design a new head-to-head trial using indirect evidence via a fixed effects method of MA.

6.7.1.3 Prediction model 6: Designing a new head-to-head trial using fixed effects indirect comparisons meta-analysis method

6.7.1.3.1 Rationale

The mean and the variance of the predictive distribution of the new head-to-head ($\theta_{T1T2new}$) trial are specified using indirect estimates calculated in section 6.7.1.2 above. This prediction model assumes the presence of only the within study variation in the predictive distribution of the new trial. This prediction model could be specified as in equation 6.3 below (step 1).

$$\theta_{T1T2new} \sim Normal\left(\log^{ind}OR_{T1T2}, Var[\log^{ind}OR_{T1T2}]\right)$$ (6.3)

6.7.1.3.2 Method

The simulation process begins with sampling an effect size from the predictive distribution specified in equation 6.3 above (step 2). Initially a sample of 100 subjects is selected for each arm (step 3). The number of events and non-events in both $T_1$ and $T_2$ groups are then binomially simulated (step 4) and the 2x2 table represent the results of the new head-to-head trial is created.
The power of the new head-to-head trial alone is computed using the methodology as specified in step 6. The process is repeated by varying sample sizes of 200, 300, 500, 800, 1000, 2000, 3000, 4000 and 5000 in each arm (step 7). The power of the new trial for specified sample sizes is computed and presented in the upcoming section 6.7.1.3.3.

### 6.7.1.3.3 Results

The power of the individual head-to-head trial in isolation, when the sample size of the new trial varies between 0-5000 in each arm, is calculated and presented in table 6-4 below.

**Table 6-4**: The power of the new head-to-head trial in isolation for various sample sizes of the new trial designed using the prediction model 6 (The variance of the predictive distribution accounted for only the within study variation).

<table>
<thead>
<tr>
<th>Sample size (One arm)</th>
<th>Power of the new head-to-head trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>12.24</td>
</tr>
<tr>
<td>200</td>
<td>17.44</td>
</tr>
<tr>
<td>300</td>
<td>23.20</td>
</tr>
<tr>
<td>500</td>
<td>31.06</td>
</tr>
<tr>
<td>800</td>
<td>39.60</td>
</tr>
<tr>
<td>1000</td>
<td>43.42</td>
</tr>
<tr>
<td>2000</td>
<td>56.74</td>
</tr>
<tr>
<td>3000</td>
<td>63.20</td>
</tr>
<tr>
<td>4000</td>
<td>67.26</td>
</tr>
<tr>
<td>5000</td>
<td>70.22</td>
</tr>
</tbody>
</table>

The power curve 6-3 below illustrates the impact of the changes to the size of the trial on its power.
Chapter 6  Designing new trials using indirect comparison methods

**Figure 6-3**: The power curve of the new head-to-head trial in isolation for various sample size of the new trial designed using the prediction model 6 (The variance of the predictive distribution of a new effect size is accounted for only the within study variation).

<table>
<thead>
<tr>
<th>Prediction model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Graph showing the power curve of the new head-to-head trial" /></td>
</tr>
<tr>
<td>The power of the new head-to-head trial designed based on the prediction model 6 rises as the sample size increases. It does not reach a plateau even at very high sample size of the new trial, although the rate at which the power increases is falling as the sample size increases.</td>
</tr>
</tbody>
</table>

The power of a new head-to-head trial designed using the fixed effects MA methods based on indirect evidence is explored in this section. Upcoming section 6.7.2 develops the methodology to design new trials based on indirect evidence using RE method of indirect comparison MA.

### 6.7.2 Designing a new head-to-head trial based on indirect evidence using random effects meta-analysis model

#### 6.7.2.1 Initial random effects meta-analysis

Since this methodology to design new head-to-head trial using indirect evidence is based on evidence synthesis methods, two initial meta-analyses are required to derive the indirect relative effect size and corresponding variance to inform the predictive distribution of an effect size of the new head-to-head trial.
6.7.2.1.1 **Initial random effects meta-analysis of trials comparing PLC with LEV add-on therapy**

The random effects MA is performed using the updated list of trials between PLC and LEV add-on therapy listed in table 6-2. The resulting forest plot is presented in figure 6-4 below.

Figure 6-4: The resulting forest plot of the initial random effects MA of trials comparing PLC with LEV-add on therapy.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorvon</td>
<td>0.17 (0.08, 0.39)</td>
<td>12.03</td>
</tr>
<tr>
<td>Cereghino</td>
<td>0.13 (0.06, 0.29)</td>
<td>12.06</td>
</tr>
<tr>
<td>Ben-Menachem</td>
<td>0.26 (0.14, 0.48)</td>
<td>20.29</td>
</tr>
<tr>
<td>Tsai</td>
<td>0.15 (0.05, 0.46)</td>
<td>6.91</td>
</tr>
<tr>
<td>Verdu</td>
<td>0.30 (0.16, 0.57)</td>
<td>19.47</td>
</tr>
<tr>
<td>Boon</td>
<td>0.32 (0.19, 0.54)</td>
<td>29.25</td>
</tr>
<tr>
<td>Overall</td>
<td>0.24 (0.18, 0.32)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

The pooled OR (\(OR_{CT1}^R\)) of PLC compared to LEV is estimated to be 0.24 with confidence interval ranging from 0.18 to 0.32. The corresponding log OR (\(\log OR_{CT1}^R\)) is -1.4258 with a standard error (\(se(\log OR_{CT1}^R)\)) of 0.1488. The \(I^2\) statistic suggests that only a 6% of the total variability is attributable to the between study variation; this is non-significant at 5% (p-value 0.378). Moreover, the heterogeneity (\(\tau^2_{CT1}\)) of these trials is estimated to be 0.0082 at 5% (p-value =0.378) at 5%. The results presented here are only slightly different to that of
Chapter 6  Designing new trials using indirect comparison methods

the fixed effect equivalent (section 6.7.1.1.1). In fact, the OR is almost equal when rounded at two decimal places.

As expected, the resulting confidence intervals of the pooled OR in the random effects model are slightly wider (0.18 - 0.32) than the FE counterpart (0.18 - 0.31). In addition to that, both the pooled log OR (-1.4457) and its standard error (0.1423) in the FE model are slightly smaller than the RE equivalents (-1.4258 and 0.1488).

This section conducts initial random effects MA of trials between PLC and LEV. The following section conducts the initial random effects MA of trials comparing PLC with TPM add-on therapy.

6.7.2.1.2 **Initial random effects meta-analysis of trials comparing PLC with TPM add-on therapy**

This section is aimed to estimate the pooled relative effect sizes and corresponding variances of trials comparing PLC and TPM add-on therapy. The outcome is measured on OR scale. Figure 6-5 below presents the forest plot resulting from the initial MA.
Figure 6-5: The resulting forest plot of the initial random effects MA of trials comparing PLC with TPM add-on therapy

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben-Menachem</td>
<td>0.02 (0.00, 0.42)</td>
<td>1.37</td>
</tr>
<tr>
<td>Elterman</td>
<td>0.39 (0.15, 1.02)</td>
<td>10.51</td>
</tr>
<tr>
<td>Faught</td>
<td>0.33 (0.14, 0.76)</td>
<td>13.17</td>
</tr>
<tr>
<td>Guberman</td>
<td>0.38 (0.22, 0.68)</td>
<td>23.01</td>
</tr>
<tr>
<td>Korean</td>
<td>0.15 (0.07, 0.31)</td>
<td>15.37</td>
</tr>
<tr>
<td>Privitera</td>
<td>0.14 (0.05, 0.40)</td>
<td>8.69</td>
</tr>
<tr>
<td>Rosenfeld</td>
<td>0.22 (0.10, 0.51)</td>
<td>13.42</td>
</tr>
<tr>
<td>Sharief</td>
<td>0.17 (0.03, 0.92)</td>
<td>3.89</td>
</tr>
<tr>
<td>Tassinari</td>
<td>0.13 (0.03, 0.51)</td>
<td>5.52</td>
</tr>
<tr>
<td>Yen</td>
<td>0.16 (0.04, 0.71)</td>
<td>5.04</td>
</tr>
<tr>
<td>Overall (I-squared = 16.6%, p = 0.290)</td>
<td>0.23 (0.16, 0.33)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

The between study variation ($\tau^2$) is estimated to be 0.0496 and found to be non-significant at 5% (p-value of 0.290). Moreover, the $I^2$ statistic suggests that 16.6% (p-value = 0.290) of the total dispersion is accounted for by the between study variation. The pooled OR of PLC compared to TPM add-on therapy ($OR_{CT2}$) is estimated to be 0.23 with confidence intervals ranging from 0.16 to 0.33. The corresponding log OR ($log OR_{CT2}$) is -1.4636 with a standard error (se[log $OR_{CT2}$]) of 0.1749.

In comparison to the corresponding initial fixed effects MA conducted in section 6.7.1.1.2, both analyses produce an equal OR (0.23) at two decimal places. Moreover, both the pooled log OR (-1.4636) and its standard error (0.1749) in
the RE model are slightly larger than that obtained in the FE analysis (-1.4745 and 0.1502).

This section conducts two initial RE meta-analyses and upcoming section employs Bucher's method to estimate the indirect inferences required to derive the predictive distribution of the new head-to-head trial.

6.7.2.2 Specifying the predictive distribution of a new head-to-head trial

6.7.2.2.1 Mean of the predictive distribution of a new head-to-head trial

The mean of the predictive distribution of all random effect based prediction models is assumed to be the indirect pooled effect estimate derived from the indirect evidence. In the running example, the indirect estimate of the log OR (random effects) between LEV and TPM is calculated using the Bucher's method specified in equation 6.1.

\[
\log^{ind}OR_{T1T2} = -1.4636 - (-1.4258) = -0.0378
\]

6.7.2.2.2 Variance of the predictive distribution of a new head-to-head trial

Each RE based prediction model specifies a different variance in the predictive distribution from which an effect size of a new head-to-head trial is drawn. The variance is specified, either alone or in combination, using the variance of the
indirect pooled relative efficacy ($\text{Var}[\text{log}^{\text{ind}} \text{OR}^R_{T1T2}]$) and indirect heterogeneity estimate ($\hat{\tau}^2_{T1T2}$) between competing treatments.

**Computing the variance of indirect log OR**

The variance and standard error of the indirect log OR estimate in the running example is calculated using equation 6.2.

$$\text{Var}[\text{log}^{\text{ind}} \text{OR}^R_{T1T2}] = (0.1749)^2 + (0.1488)^2 = 0.0527$$

$$\text{se}[\text{log}^{\text{ind}} \text{OR}^R_{T1T2}] = 0.2296$$

**Computing the indirect estimate of heterogeneity between LEV and TPM trials**

For simplicity majority of this chapter (except in prediction model 10) assumes that the mean heterogeneity in two groups of trials can be used as an estimate of the heterogeneity in trials between competing treatments. The between study variations in head-to-head trials comparing LEV and TPM is ($\hat{\tau}^2_{T1T2}$) is computed as specified in equation 6.4 below.

$$\hat{\tau}^2_{T1T2} = \frac{\hat{\tau}^2_{CT1} + \hat{\tau}^2_{CT2}}{2}$$  \hspace{1cm} (6.4)

where $\hat{\tau}^2_{CT1}$ and $\hat{\tau}^2_{CT2}$ denote the estimated between study variance in trials comparing treatment $T_1$ and $T_2$ with the common comparator respectively.

In this illustrative example, magnitude of $\hat{\tau}^2_{T1T2}$ is computed using equation 6.4 above.
\[ \hat{\tau}^2_{T1T2} = \left[ \frac{0.0082 + 0.0496}{2} \right] = 0.0289 \]

\[ \hat{\tau}_{T1T2} = 0.17 \]

The between study variance and the standard error of trials comparing LEV and TPM is thus computed to be 0.0289 and 0.17 respectively.

**Hypothetical models**

As described in detail in section 5.8.4.3 of Chapter 5, RE based prediction models accounted for only one component of the variance, i.e. either within (prediction model 7) or between study variance (prediction model 8) in the predictive distribution of the new head-to-head trial is not found in practice and are hypothetical. Nevertheless, their presence gives an insight into the fluctuation of the power of the new head-to-head trial to the changes in the variability in the predictive distribution of the new trial.

This section conducts two initial RE meta-analyses and specify the predictive distribution of the new head-to-head trial. Upcoming sections 6.7.2.3 - 6.7.2.6 describe the process of developing the methodology based on random effects MA principles using indirect evidence.
6.7.2.3 **Prediction model 7 : Designing a new head-to-head trial using random effects indirect comparisons meta-analysis methods (predictive distribution accounted for only the within study variation)**

6.7.2.3.1 **Rationale**

The effect size of the new head-to-head trial is drawn from a normal distribution with mean equivalent to the RE indirect pooled log OR between $T_1$ and $T_2$ ($\log^{\text{ind}} OR_{T1T2}^R$) and the variance equivalent to the RE variance of the indirect pooled log OR between $T_1$ and $T_2$ ($\text{Var}[\log^{\text{ind}} OR_{T1T2}^R]$) as specified in equation 6.5 below.

$$\theta_{T1T2\text{new}} \sim \text{Normal}(\log^{\text{ind}} OR_{T1T2}^R, \text{Var}[\log^{\text{ind}} OR_{T1T2}^R]) \quad (6.5)$$

6.7.2.4 **Prediction model 8 : Designing a new head-to-head trial using random effects indirect comparisons meta-analysis methods (The variance of the predictive distribution accounted for only the between study variation)**

6.7.2.4.1 **Rationale**

This model assumes that only the between study variation is included in the variance of the predictive distribution of an effect size of a new head-to-head trial. The within study variation component is excluded from the variance of the
predictive distribution, on purpose. An effect size of the new head-to-head trial is drawn from a normal distribution with the mean and the variance specified as below in equation 6.6.

\[ \theta_{T1T2_{new}} \sim \text{Normal} (\log^{\text{ind}} OR^R_{T1T2}, \hat{\tau}^2_{T1T2}) \]  

(6.6)

Since the heterogeneity parameter is unknown, for simplicity the uncertainty in the estimation of \( \tau^2_{T1T2} \) is ignored and the estimator (\( \hat{\tau}^2_{T1T2} \)) is used in place.

6.7.2.5 **Prediction model 9 : Designing a new head-to-head trial based on random effects indirect comparisons meta-analysis methods (The predictive distribution accounted for both within and between study variance: mean heterogeneity approach)**

6.7.2.5.1 **Rationale**

This prediction model accounted for both the within and between study components of variations in the predictive distribution of a new effect size. In fact, it can be viewed as a model composited of both prediction model 7 and 8. The within study variation accounted for in the prediction model 7 and the between study variation accounted for in the prediction model 8, are amalgamated in estimating the variance of the predictive distribution of the new head-to-head trial.

The effect size of the new head-to-head trial is drawn from a normal distribution with mean and variance are specified as below in equation 6.7.
\[ \theta_{T_1T_2^{\text{new}}} \sim \text{Normal} \left( \log^{ind} OR_{T_1T_2}^R, \text{Var}[\log^{ind} OR_{T_1T_2}^R] + \hat{\tau}^2_{T_1T_2} \right) \] (6.7)

For simplicity, this prediction model also disregards the uncertainty in the estimation of the heterogeneity parameter.

### 6.7.2.6 Prediction model 10: Designing a new head-to-head trial using random effects indirect comparisons meta-analysis methods (Both within and between study variation are assumed in the predictive distribution of a new effects size: The residual heterogeneity approach)

#### 6.7.2.6.1 Rationale

This prediction model intends to obtain a unique heterogeneity estimate (\( \hat{\tau}^2_{T_1T_2} \)) to specify the between study variance element of the predictive distribution of a new head-to-head trial by combining the two sets of indirect evidence. This is in contrast to the practice of using the mean of the two heterogeneity estimates of two set of trials employed in prediction model 8 and 9 developed before.

The two sets of trials comparing treatments T_1 and T_2 with the common comparator are merged to estimate the common heterogeneity estimate. Using an indicator variable, the data pertaining to treatments T_1 and T_2 are distinguished. The random effects MR model is fitted to estimate the residual heterogeneity statistic. This model produces the residual heterogeneity, which is the between study variation which is not explained by the covariates (Harbord...
2008). The MR model is an extension to the random effects MA model, which includes a covariate. The indicator variable features as the covariate in this context.

The general form of the MR model is given in equation 6.8 below.

\[ y_i = x_i \beta + \epsilon_i + u_i \]  \hspace{1cm} (6.8)

where \( y_i \) represents the outcome measure of the \( i^{th} \) trial and \( x_i \) represents the value of the covariate attributable to the \( i^{th} \) trial. \( \beta \) denotes the regression coefficients. In this model, \( \beta \) is a \( k \times 1 \) vector of coefficients and \( x_i \) is a \( 1 \times k \) vector of covariates in the \( i^{th} \) study. \( \epsilon_i \) and \( u_i \) are distributed as below.

\[ \epsilon_i \sim Normal(0, \sigma^2) \]  \hspace{1cm} (6.9)

\[ u_i \sim Normal(0, \tau^2) \]  \hspace{1cm} (6.10)

Once the residual heterogeneity (\( \hat{\tau}^2_{T1T2} \)) is estimated, it could along with the standard error of the indirect log OR (\( Var[\log^{ind} OR_{T1T2}^R] \)) be used to inform the variance of the predictive distribution of a new head-to-head trial.

\[ \theta_{T1T2\text{new}} \sim Normal(\log^{ind} OR_{T1T2}^R, Var[\log^{ind} OR_{T1T2}^R] + \hat{\tau}^2_{T1T2}) \]  \hspace{1cm} (6.11)

Equation 6.11 above specifies the predictive distribution of an effect size of the new head-to-head trial.
6.7.2.7 Methods

An effect size is drawn from the predictive distribution specified in each prediction model. A new head-to-head trial is designed using the procedure described in section 6.6.2. The power of the new head-to-head trial in isolation is recorded.

6.7.2.8 Power results of prediction models 6-10

To enable a comprehensive comparison between the power results of the prediction models, all the results of prediction model 6-10 are summarised in table 6-5. This table contains the power of the new head-to-head trial and corresponding sample sizes of the new head-to-head trial.
Table 6-5: The table combining the results of all prediction models based on indirect comparison meta-analysis.

<table>
<thead>
<tr>
<th>Sample Size (one arm)</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed effects</td>
<td>Random effects - Within Study Variation only</td>
<td>Random effects - Between Study Variation only</td>
<td>Random effects Between and Within study variations (Mean heterogeneity approach)</td>
<td>Random effects Between and within study variations (Residual heterogeneity approach)</td>
</tr>
<tr>
<td>100</td>
<td>12.24</td>
<td>13.74</td>
<td>10.16</td>
<td>25.58</td>
<td>27.34</td>
</tr>
<tr>
<td>200</td>
<td>17.44</td>
<td>19.80</td>
<td>13.96</td>
<td>37.28</td>
<td>39.16</td>
</tr>
<tr>
<td>300</td>
<td>23.20</td>
<td>26.24</td>
<td>18.76</td>
<td>46.30</td>
<td>48.40</td>
</tr>
<tr>
<td>500</td>
<td>31.06</td>
<td>34.44</td>
<td>25.14</td>
<td>54.96</td>
<td>56.88</td>
</tr>
<tr>
<td>800</td>
<td>39.60</td>
<td>43.82</td>
<td>32.28</td>
<td>63.14</td>
<td>64.72</td>
</tr>
<tr>
<td>1000</td>
<td>43.42</td>
<td>47.68</td>
<td>36.04</td>
<td>66.82</td>
<td>68.38</td>
</tr>
<tr>
<td>2000</td>
<td>56.74</td>
<td>60.66</td>
<td>50.22</td>
<td>76.00</td>
<td>77.48</td>
</tr>
<tr>
<td>3000</td>
<td>63.20</td>
<td>66.70</td>
<td>57.38</td>
<td>79.86</td>
<td>81.16</td>
</tr>
<tr>
<td>4000</td>
<td>67.26</td>
<td>70.38</td>
<td>62.42</td>
<td>82.86</td>
<td>83.72</td>
</tr>
<tr>
<td>5000</td>
<td>70.22</td>
<td>72.78</td>
<td>65.54</td>
<td>83.16</td>
<td>84.18</td>
</tr>
</tbody>
</table>

Figure 6-6 below shows individual power curves based on random effects IC methods.
**Figure 6-6**: Individual power curves of prediction models 7-10.

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prediction model 7</strong> - Within study variance only</td>
<td>The power of the new head-to-head trial increases gradually with the increased sample size. The rate of growth of power when the sample size increases beyond 2000 is falling compared to when it is below 2000. The power does not reach a plateau even at higher sample sizes of the new head-to-head trial.</td>
<td><img src="image1" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Prediction model 8</strong> - Heterogeneity only</td>
<td>The graph indicates a gradual increase in power of the new head-to-head trial with increasing sample sizes. Moreover, the power does not flatten off even at a sample size as high as 5000 in one arm. However, the rate at which power increases is falling as the sample size increases.</td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Prediction model 9</strong> - With study variance and average heterogeneity</td>
<td>The power of the new head-to-head trial designed based on prediction model 9, is relatively higher even at lower sample sizes, when compared to that of other comparative prediction models, which accounted for either within or between study variation. The power begins to flatten off around mid 80% as the sample size increases above 3000. The rate of power increases is falling noticeably with increased sample size greater than 2000.</td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Prediction model 10</strong> - Within study variance and residual heterogeneity</td>
<td>As expected, with increased sample sizes the power of the new head-to-head trial increases as well. However, the rate at which the power increases is falling as the sample size increases. For sample sizes of the new trial over 2000 in each arm, the rate at which the power increases is markedly falling. The power is flattening off around 84% and reaching a plateau at higher sample sizes.</td>
<td><img src="image4" alt="Graph" /></td>
</tr>
</tbody>
</table>
The combined power curve below in figure 6-7 reflects the impact of different prediction models on power of the new head-to-head trial.

**Figure 6-7**: The combined power curves of prediction model 6-10.

Two prediction models accounted for both the within and the between study variances (Prediction model 9 and 10) produce almost equal power at all specified sample sizes. Moreover, the new head-to-head trials based on these two models produced the highest power at all sample sizes considered, in comparison to power of other prediction models. The prediction model 8 which accounted for only the between study variances produces the lowest power at all sample sizes specified. The RE model which accounted for only the within study variation produces a slightly higher power than that of the FE model at all sample sizes.
6.8 **Critical evaluation of the results**

To facilitate a critical evaluation of the results, following table 6-6 is constructed to showcase vital point estimates derived in the analysis, namely the pooled log ORs, corresponding standard errors and variances, along with the heterogeneity estimates in both FE and RE scenarios.

*Table 6-6: Summary statistics of indirect comparison MA methods.*

<table>
<thead>
<tr>
<th>Indirect Estimates</th>
<th>Fixed effects</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled log OR</td>
<td>-0.029</td>
<td>Pooled log OR</td>
</tr>
<tr>
<td>Variance of the pooled log OR</td>
<td>0.04296</td>
<td>Variance of the pooled OR</td>
</tr>
<tr>
<td>Standard error of the pooled log OR</td>
<td>0.2072</td>
<td>Standard error of the pooled log OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Between study standard error – Mean heterogeneity approach</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Between study standard error – Residual heterogeneity approach</td>
</tr>
</tbody>
</table>

Predictive distributions of models based on RE principles are set to have a common mean and distinct variances. Therefore, it is apparent that the reason for different models to produce different power is the different variances each model accounted for.

The magnitude of the heterogeneity ($\tau$) estimated by the mean of the heterogeneity estimates obtained from two initial meta-analyses (0.1701), is only slightly lower from the residual heterogeneity estimated after allowing for dummy covariates (0.1921). Therefore, the power of the new trial alone in the prediction model 9 is only slightly lower than that of prediction model 10, at all sample sizes considered. Given all other factors remain equal in the prediction
model 9 and 10; the cause of the difference in power is only attributable to the
difference in the accounted variability. The model, which accounted for a little
more variability yielded a little more power.

In the context of the RE, the magnitude of the within study variance \((0.2299^2)\) is
larger than the between study variation \((0.1701^2)\). This has caused the power of
the new trial which accounted only for the within study variation (prediction
model 7) to be higher than that which accounted only for the between study
variation (prediction model 8).

The power of the new trial derived from the predictive distribution which
accounted for both the within and the between study variations (prediction
model 9) is higher than that of individual trials those accounted either the within
(prediction model 7) or between study variation (prediction model 8). The
indirect estimate of the within study variation \((0.2299^2)\) is larger than the
between study variation \((0.1701^2)\) counterpart. A new trial based on an effect
size sampled from a wider prediction interval is likely to produce higher power
than that with narrower prediction interval.

In a wider predictive distribution, a new trial is highly likely to be drawn from the
extreme end of the distribution (tails) and the confidence interval around it
therefore is less likely to include the null value. This causes the power of the
trials derived from wider predictive distributions to become higher.

When compared, the FE model (prediction model 6) yielded less power than the
RE model (prediction model 7) accounted for only the within study variation.
The explanation of this phenomenon is of two fold. The mean of the RE
predictive distribution (-0.038) locates further from the null value than the FE counterpart (-0.029). In addition to that, the dispersion in the RE model (0.2299) is wider than in the FE (0.2072) model as well. These two factors in combination cause the power of the RE model to yield more power than the FE equivalent.
6.9 Discussion

This chapter contributes to the framework of methodology for designing new trials using existing evidence, by developing and implementing a methodology to design new head-to-head trials based on indirect comparison MA methods. The methodology facilitates determination of the required sample size of a new head-to-head trial to achieve the power of the new trial in isolation. Adherence to this methodology would be beneficial in planning and designing new trials between two competing treatments those have not been or rarely compared head-to-head before, but where each has been compared to a common comparator.

The results reflect that the power of the new head-to-head trial is heavily reliant on the variance of the predictive distribution of the new trial. The more variability the predictive distribution accounted for the greater the power the new head-to-head trial in isolation. This phenomenon is in accord with the results produced in section 5.9 of Chapter 5, where the power of the new trial in isolation designed based on frequentist MA principles is explored.

The decision as to whether to use indirect evidence should more often be relied on the results of the significance test of the difference between the direct and indirect estimates. If direct and indirect results are significantly different then the use of indirect evidence in place of direct evidence may yield spurious findings. Hence, before combining them, the reasons for the difference should be investigated. If the discrepancies between the direct and indirect estimates are in the same direction, it is not clinically importance. However, inconsistent estimates or estimates of opposite directions do not necessarily imply that the
indirect estimates are wrong or misleading. However, further investigations are necessary to explain the cause of the discrepancy before indirect evidence is used as a surrogate of direct evidence or to combine both (Song, F., Altman et al. 2003).

However, empirical evidence has indicated that in most cases direct evidence is not significantly different from that of indirect evidence. The internal validity of the trials included in IC should be properly examined to reveal the causes of the discrepancies between the direct and indirect evidence. For example, biases in individual trials could propagate in to indirect comparisons. Further to that, the differences in treatments, populations, other patient characteristics, and importantly methodological flaws in trials could cause discrepancies.

Song and colleagues explored the key methodological problems associated with indirect comparisons (Song, F., Loke et al. 2009). One of the major methodological problems is identified as use of an incomplete search strategy or inappropriate inclusion of relevant studies. To overcome this specific problem the updated datasets (tables 6-2 and 6-3) in this study are retrieved in a systematic search in major databases including the Pubmed and the Cochrane database.

Unclear understanding and inadequate specification of assumptions are identified as further methodological problems by Song and colleagues (Song, F., Loke et al. 2009). However, these assumptions are not without practical limitations. For example, the similarity assumption requires the relative efficacy of a treatment to be consistent in patients across all trials. In other words, the
relative efficacy should be generalisable across trials. Often the generalisability of trial results is questionable in the light of restricted inclusion criterion, exclusion of patients, and differences in trial settings (Song, F., Altman et al. 2003).

It is important to acknowledge the fact that results specified in this chapter are entirely dependent on the validity of Bucher’s method of adjusted IC, which is central to the derivation of indirect estimates. Although Bucher’s method is commonly used in evaluating treatments, it has received some criticism too, particularly in relation to the assumption about the consistency of the relative effect across trials (Chou, Fu et al. 2006). Hence, these criticisms carry over to the methodology being developed in this chapter.

However, it is worth noting that IC methods have not yet received widespread acceptance in evaluating healthcare due to conflicting evidence emerged about the validity of indirect comparisons. The results of IC are suggested to be less biased than that of head-to-head trials because of the results of the simulations showed that in certain circumstances adjusted indirect comparisons may counterbalance the bias (Song, F., Harvey and Lilford 2008). In contrast, another study (Chou, Fu et al. 2006) concluded that the indirect results are unreliable for complex and rapidly evolving interventions. Hence, the potential usefulness of IC methods in evaluation treatments are often overshadowed by the concerns about their results; hence they are currently not an automatic choice.
This methodology could be readily extended to be able to accept other binary outcome measures, including risk ratio. Another aspect to explore in the future is to make this methodology to work with outcomes measured on continuous scale.
6.10 Conclusions

Direct evidence from well designed randomised trials is regarded as the best evidence, although the use of indirect evidence is inevitable due to the increased number of treatment efficacy comparisons, as a result of rapid advances in health technology.

This chapter concludes that the power of the new head-to-head trial is directly associated with the variance of the predictive distribution from which an effect size of a new head-to-head trial is drawn from. The more variability accounted by the predictive distribution, the more likely that the confidence interval around the effect estimate does not include the null value, and hence, more power accrues. Therefore, to improve the power of the head-to-head trial the variance of the predictive distribution should include both between and within study components of variances.

This study introduces two approaches to incorporate between study variance into the predictive distribution (prediction model 9 and 10). Potential users of the approach should employ both models to estimate the between study variance and design the trial using the model which yield the higher between study variances.

The approach to design new head-to-head trial is illustrated using a specific dataset and it is therefore important to understand that the results may be dependent on specific features of the dataset. When direct evidence is available but not sufficient direct and indirect evidence should be combined to obtain more precise estimates of pooled summary effects. This recommendation can
be applied to inform the design of new trials between competing treatments, which is explored in upcoming Chapter 7.
CHAPTER 7

Designing future trials by combining both direct and indirect evidence using Mixed Treatment Comparison (MTC) methods

7.1 Introduction

This chapter contributes to the methodological framework of designing future trials by developing a methodology to combine both direct and indirect evidence. It adopts an approach based on the power of the Mixed Treatment Comparison (MTC) meta-analysis to determine the sample size of a new head-to-head trial. This introduction is followed by the motivation to this particular work in section 7.2 and specific objectives of this chapter are outlined in section 7.3. Section 7.4 briefly introduces the concept of the MTC approach. Section 7.5 describes the existing evidence on which the methodology is based. The methods employed to develop the methodology of designing new head-to-head trial are explained in section 7.6. The methodology to design new head-to-head trials based on MTC meta-analysis is developed in section 7.7, within that subsection 7.7.1 employs the FE model and subsection 7.7.2 employs the RE model. A critical evaluation of the results is provided in section 7.8. The discussion is given in section 7.9 and section 7.10 gives the conclusions drawn from this chapter.
7.2 Motivation

The explosion in the number of systematic reviews and the proliferation of treatment options available to treat a given clinical condition have created the need for the MTC meta-analysis, in which both direct and indirect evidence are combined to evaluate the treatment efficacy of two competing treatments (Caldwell, D.M., Ades and Higgins 2005). (A detailed review of the MTC methods has been given in section 2.6 of Chapter 2.)

In the absence of high quality randomised trials comparing all eligible treatments, investigators often rely on indirect evidence as a basis of making health care policies and informing decisions. Hence, the evidence available in indirect comparisons regarding the efficacy of target direct comparisons should not be disregarded (Lu and Ades 2004). In situations where direct evidence is available but not sufficient indirect evidence could provide supplementary evidence.

Conflicting concerns have been expressed over the validity of indirect comparisons. For example, Yazdanpanah et al have found indirect evidence consistent with the direct evidence (Yazdanpanah, Sissoko et al. 2004) whilst Song et al found discrepancies in the comparison between the direct and indirect evidence (Song, F., Altman et al. 2003). Moreover, indirect comparisons are regarded as not randomised trials but observational studies across trials and therefore they could be subjected to biases like confounding (Higgins, J. P. T., Green and editors 2005). The strength of the indirect evidence is limited and are prone to produce spurious efficacy results when differences in methodology, populations and outcome measures exist in two sets of trials (Bucher, Guyatt et
al. 1997). Whenever possible, direct evidence should take precedence over the indirect evidence, as the basis of forming decisions. In certain situations, direct evidence is not conclusive in its own right but collectively direct and indirect evidence may be conclusive (Caldwell, D.M., Ades and Higgins 2005).

The quantitative combination of direct and indirect evidence could improve both the statistical power of the hypothesis test and the precision of the derived estimates, if no clinically important discrepancies between the two sources exist. It further strengthens the inferences, simply because of increased amount of information (Song, F., Altman et al. 2003). Although recent research urges the value of combining direct and indirect evidence, it does not seem to have been practically embraced yet. Moreover, Salanti et al recommends extending the original idea of Sutton et al (Sutton, A. J., Cooper et al. 2007) to consider the future MA including the new trial when designing a new trials, towards a MTC framework (Salanti, G., Higgins et al. 2007).

The key objectives of this chapter are set out in section 7.3 below.

### 7.3 Objectives

The primary objective of this chapter is to contribute to the methodological framework developed in stages throughout this thesis, with a methodology to design new head-to-head trials based on the MTC framework, which facilitates integrating both direct and indirect evidence.
Moreover, the aim is to plan and determine the sample size of a new head-to-head trial with the expectation that it will be a part of the subsequent MTC meta-analysis.

Since the methodology developed in Chapter 6 employs indirect evidence to form the basis of the design of a new head-to-head trial, the methodology developed here is perceived as an extension to Chapter 6. Building on from prediction models developed in Chapter 6, this chapter examines the implicit variation in power of the MTC meta-analysis, for changes to the variance of the predictive distribution from which an effect size of a new head-to-head trial is drawn.

Having set out the objectives of this chapter, the forthcoming section 7.4 briefly introduces MTC meta-analysis.

### 7.4 Introduction to Mixed Treatment Comparison meta-analysis

The process of incorporating indirect evidence to direct evidence is referred to as the Mixed Treatment Comparison (MTC) (Salanti, G., Kavvoura and Ioannidis 2008). In addition to allowing for appropriately combining both direct and indirect evidence (O'Regan, Ghement et al. 2009), MTC offers further benefits by simultaneously ranking all competing treatments based on a coherent judgment regarding their efficacies (Caldwell, D. M., Welton and Ades 2009).
The focus of this chapter is to combine direct and indirect evidence in a MTC setting. The direct and indirect evidence on relative efficacies of treatments \( T_1 \) and \( T_2 \) are appropriately combined to obtain the pooled relative efficacy between treatment \( T_1 \) and \( T_2 \). Both direct and indirect evidence are assigned weights equal to reciprocal of the variance of the relative efficacies. The inverse variance approach of combining direct and indirect evidence is specified in equation 7.1 below.

\[
\hat{\theta}_{Pooled}^{T_1T_2} = \frac{w^{direct}\hat{\theta}_{direct}^{T_1T_2} + w^{indirect}\hat{\theta}_{indirect}^{T_1T_2}}{w^{direct} + w^{indirect}} \tag{7.1}
\]

where \( w = \frac{1}{se(d_{T_1T_2})^2} \)

As for any other form of MA, the validity of the findings from the methods based on the MTC methods require adherence of underlying assumptions, which are discussed in forthcoming section 7.4.1.

### 7.4.1 Assumptions

Three levels of assumptions apply in MTC meta-analysis (Song, F., Loke et al. 2009) including homogeneity, similarity and consistency assumptions. The homogeneity and the similarity assumptions have been explained in section 6.4.3 of Chapter 6. Since the MTC meta-analysis is deemed a generalisation of the standard MA, both the MTC and the standard MA require similar assumptions (Hasselblad 1998) namely the homogeneity. Moreover, the MTC is regarded as an extension to the IC and hence, there are common assumptions shared between them, such as the similarity assumption. Both IC and MTC
assume the similarity of trials included in the analysis. The consistency assumption, which is briefly described in section 7.4.1.1 below, only pertains to the MTC meta-analysis.

### 7.4.1.1 Consistency assumption

The consistency of evidence is assumed when indirect and direct evidence are quantitatively combined (Song, F., Loke et al. 2009). The consistency of evidence is often assessed by comparing the efficacy estimates from the two sources of evidence. If combined, inconsistent evidence may produce invalid and misleading pooled estimates. Possible causes of the inconsistency should be investigated, including play of chance, invalid methods of indirect comparisons, bias in direct comparisons, and heterogeneity (Song, F., Harvey and Lilford 2008).

Forthcoming section 7.6 describes the process of collecting existing indirect evidence.

### 7.5 Existing evidence

This chapter extends the methodology developed in Chapter 6 and therefore it is vital to use the same existing evidence in both chapters that allows a valid comparison between the results yielded in two methodologies. Two sets of trials comparing LEV and TPM with PLC listed in table 6-2 and 6-3 in Chapter 6 constitute the existing evidence in the MTC meta-analysis framework as well.
7.5.1 **Indirect evidence**

This methodological framework requires existing evidence to form the basis of the analysis to design new trials. Therefore, as required in Chapter 6, two sets of trials comparing two competing treatments with a common comparator ($T_1$ vs $C$ and $T_2$ vs $C$) are a prerequisite to conducting two initial meta-analyses to derive indirect inferences.

The indirect relative efficacy ($\log^{\text{ind}} OR_{T_1T_2}$) of treatments $T_1$ and $T_2$ and its variance ($\text{Var}[\log^{\text{ind}} OR_{T_1T_2}]$) are computed using Bucher's adjusted IC method by substituting the resulting inferences of two initial meta-analyses as specified in equation 6.1 and 6.2 in Chapter 6.

7.5.2 **Direct evidence**

The distinct feature of this methodology is the use of the same predictive distribution to derive both direct and indirect inferences. The predictive distribution of the new head-to-head trial is initially specified using the indirect inferences. A new head-to-head trial is designed by drawing an effect size from this predictive distribution. Despite being originated from indirect evidence, the new head-to-head trial serves as the direct evidence, which will then be combined with the indirect inferences to form the MTC meta-analysis.
7.5.3 Event of interest

7.5.3.1 Definition of an event

Since the same existing evidence is used in IC and MTC contexts, the definition of an event is exactly as same as that defined in section 6.5.3 of Chapter 6.

7.5.3.2 Estimating the event rate in the control group

Knowing the event rate in the control group \( (P^{T2,new}) \) is a prerequisite common to all components of the methodological framework developed in this thesis. As computed in section 6.5.4 of Chapter 6, in the illustrative example, that indirectly compares LEV and TPM, the event rate in the control group is 0.4477.

Following section 7.6 describes the methods employed to develop the methodology of designing new head-to-head trials using the MTC framework.

7.6 Methods

This section introduces the notion of power in the context of MTC meta-analysis. Further it distinguishes the indirect and indirect evidences. The procedure for designing a new head-to-head trial using the MTC meta-analysis is also specified.
### 7.6.1 Null hypothesis tested and the power of the MTC meta-analysis

The interest lies in here is not on the power of the new head-to-head trial in isolation (which was the focus in Chapter 6) but the power of the MTC meta-analysis, which combines both the indirect and direct evidence.

In the MTC context, the power is based on the test of the main effect in which the null hypothesis specifies that treatments T1 and T2 are equally effective ($H_0: OR=1$). The statistical significance of this test is evaluated at 5% level of significance.

### 7.6.2 Procedure

Step-by-step guide to combine direct and indirect evidence and to design a new study based on MTC meta-analysis is given below.

**Step 1 - 5**

The process of designing a new head-to-head trial between treatments T₁ and T₂ from indirect evidence is exactly the same as that introduced in section 6.6.2 of Chapter 6. The new head-to-head trial represents the direct evidence in this particular MTC framework.

**Step 6**

This step combines the direct and indirect evidence using the formulae specified in equation 7.1, in the MTC settings. Both the direct and indirect
evidence are assigned weights equal to the reciprocal of the variances of their pooled effect size (inverse variance method).

**Step 7**

The decision as to whether to reject or not to reject the null hypothesis is decided based on the criteria specified in section 7.6.1 above.

**Step 8 - 9**

Steps 8 and 9 are exactly as same as step 8-9 described in section 5.6.2 of Chapter 5, except in here the power of the MTC meta-analysis is computed as opposed to the power of the updated MA computed in Chapter 5.

This section describes the methods used to compute the power of the MTC meta-analysis.

### 7.7 Designing a new head to head trial using MTC methods

This chapter proposes a series of prediction models based on MTC meta-analysis methods to design new head-to-head trials. Both fixed and random effects MTC methods are used to design prediction models. Section 7.7.1 explores FE methods and section 7.7.2 explores random effects MA to design new head-to-head trials within a MTC framework.
7.7.1 **Designing a new head-to-head trial using fixed effects MTC meta-analysis methods**

Since the same existing evidence is used in both the IC and MTC settings, the initial FE meta-analyses of trials between LEV with PLC and TPM with PLC are identical to those conducted in sections 6.7.1.1.1 and 6.7.1.1.2 of Chapter 6.

7.7.1.1 **Specifying the predictive distribution of a new head-to-head trial**

The computation of the mean and the variance of the predictive distribution of a new head-to-head trial is identical to that specified in section 6.7.1.2 of Chapter 6.

Section 7.7.1.2 below designs the new head-to-head trial using fixed effects MA model within MTC framework.

7.7.1.2 **Prediction model 11: Designing a new head-to-head trial using fixed effects MTC model**

7.7.1.2.1 **Rationale**

This prediction model considers the power of the MTC meta-analysis in a FE setting. The pooled log OR and corresponding variance derived from Bucher’s method specified in section 6.7.1.2 of Chapter 6 represent the indirect evidence. Using these indirect inferences, the predictive distribution of an effect in a new head-to-head trial is then specified as given in formulae 7.2 below.
This prediction model assumes that an effect size of the new head-to-head trial ($\theta_{T1T2new}$) is normally distributed with the mean equivalent to the indirect pooled log OR ($\log^{ind} OR_{T1T2}$) and the variance equivalent to the variance of the indirect pooled log OR ($Var[\log^{ind} OR_{T1T2}]$) of treatment T1 and T2.

$$\theta_{T1T2new} \sim Normal(\log^{ind} OR_{T1T2}, Var[\log^{ind} OR_{T1T2}]) \quad (7.2)$$

### 7.7.1.2.2 Methods

The predictive distribution of the new head-to-head trial is specified as given in equation 7.1 (step 1). An effect size is sampled from the specified predictive distribution assuming the sample size in the T1 and T2 groups (step 2). The event rate in the T1 group is then computed (step 3). The number of subjects in the T1 and T2 groups are then specified (step 4). The number of events and non-events in the new head-to-head trial then binomially simulated. The 2x2 table of the results of the new head-to-head trial is then developed which represents the direct evidence (step 5). This direct evidence is then combined with indirect evidence via a fixed effects MTC meta-analysis (step 6). The p-value of the MTC meta-analysis is determined. The decision as to whether to reject or do not reject the null hypothesis is made (step 7). This process is repeated for a larger number of times. The power of the MTC meta-analysis is determined as the proportion of significant number of p-values (step 8). The process is repeated for varying sample sizes of the new head-to-head trial. The power of the MTC meta-analysis at each sample size is determined (step 9) and reported in section 7.7.1.2.3 below.
### 7.7.1.2.3 Results

The power of the fixed effects MTC meta-analysis at specific sample sizes of the new head-to-head trial is given in table 7-1 below.

**Table 7-1**: The power of the fixed effects MTC meta-analysis for specific sample sizes of the new head-to-head trial designed based on prediction model 11.

<table>
<thead>
<tr>
<th>Sample size of the new head-to-head trial (One arm)</th>
<th>Power of the MTC meta-analysis (Fixed Effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.52</td>
</tr>
<tr>
<td>200</td>
<td>5.86</td>
</tr>
<tr>
<td>300</td>
<td>11.84</td>
</tr>
<tr>
<td>500</td>
<td>23.44</td>
</tr>
<tr>
<td>800</td>
<td>35.10</td>
</tr>
<tr>
<td>1000</td>
<td>39.14</td>
</tr>
<tr>
<td>2000</td>
<td>54.94</td>
</tr>
<tr>
<td>3000</td>
<td>62.18</td>
</tr>
<tr>
<td>4000</td>
<td>66.70</td>
</tr>
<tr>
<td>5000</td>
<td>69.84</td>
</tr>
</tbody>
</table>

The data tabulated above is graphically displayed in figure 7-1 below.

**Figure 7-1**: The power curve of the fixed effects MTC meta-analysis for specific sample sizes of the new head-to-head trial designed using prediction model 11.
This section explores the power implications of the fixed effects MTC settings. The next section designs new head-to-head trials using the random effects MTC settings.

### 7.7.2 Designing new head-to-head trial using random effects MTC meta-analysis methods

The prediction models based on random effects MTC methods employ the random effects method to conduct two initial meta-analyses between competing treatments T1 and T2 with the common comparator. The predictive distribution of a new effect size in these prediction models shares a common mean. However, the variance is set to vary across prediction models. The prediction models 12-15 developed in upcoming sections 7.7.2.2-7.7.2.5 extend the IC based prediction models 6-10 developed in Chapter 6.

#### 7.7.2.1 Specifying the predictive distribution of a new head-to-head trial

The process of specifying and computing the predictive distribution of an effect size of a new head-to-head trial is identical to that undertaken in section 6.7.2.2 of Chapter 6. The inferences of the predictive distribution are derived from the results of two initial RE meta-analyses carried out in section 6.7.2.1 of Chapter 6 between competing treatments LEV and TPM with PLC.
7.7.2.2 Prediction model 12: Designing a new head-to-head trial using random effects MTC methods: The predictive distribution includes only the within study variation

7.7.2.2.1 Rationale

Many aspects of this prediction model are akin to that of prediction model 7, including the conduct of two initial meta-analyses and the derivation of adjusted indirect estimates between LEV and TPM. These indirect estimates specify the parameters of the predictive distribution of the new head-to-head trial which includes only the within study variance.

\[ \theta_{T1T2new} \sim Normal \left( \log^{ind} OR_{T1T2}, \Var \left[ \log^{ind} OR_{T1T2} \right] \right) \]  \hspace{1cm} (7.3)

The predictive distribution of an effect size of a new head-to-head trial \((\theta_{T1T2new})\) is normally distributed as specified in equation 7.3 above.

7.7.2.3 Prediction model 13: Designing a new head-to-head trial using random effects MTC methods: the predictive distribution includes only the between study variation

7.7.2.3.1 Rationale

This prediction model expands the scope of the prediction model 7 in a way that it allows combining both direct and indirect evidence in a MTC structure. The
The variance of the predictive distribution of the new head-to-head trial includes only the heterogeneity. The heterogeneity estimate of the predictive distribution of the new head-to-head trials ($\hat{\tau}_{T1T2}^2$) is estimated by taking the mean of two heterogeneity estimates ($\hat{\tau}_{T1c}^2$ and $\hat{\tau}_{T2c}^2$) of two initial MA comparing T1 and T2 with the common comparator. An illustration of the computation of the heterogeneity estimate is given in section 6.7.2.2.2 of Chapter 6.

Following equation 7.4 specifies the predictive distribution of an effect size of new ($\theta_{T1T2new}$) head-to-head trial.

$$\theta_{T1T2new} \sim \text{Normal} \left( \log^{\text{ind}} \hat{\text{OR}}_{T1T2}^R, \hat{\tau}_{T1T2}^2 \right) \quad (7.4)$$

In this simplest case, the uncertainty in the estimation of the heterogeneity parameter is excluded and only a point estimate is used instead.

### 7.7.2.4 Prediction model 14: Designing a new head-to-head trial using random effects MTC methods: The predictive distribution accounts for both the within and between study variations: The mean heterogeneity approach

#### 7.7.2.4.1 Rationale

This prediction model extends the prediction model 9, in that the direct evidence could also be included in the model. The variance of the predictive distribution
of this prediction model accommodates both elements of the variance, i.e. the within ($Var[\log^{\text{ind}} OR^{R}_{T1T2}]$) and between study variance ($\hat{\tau}^2_{T1T2}$).

This model specifies the predictive distribution of the new head-to-head trial ($\theta_{T1T2\text{new}}$) as given in equation 7.5 below.

$$\theta_{T1T2\text{new}} \sim \text{Normal} \left( \log^{\text{ind}} OR^{R}_{T1T2} , Var[\log^{\text{ind}} OR^{R}_{T1T2}] + \hat{\tau}^2_{T1T2} \right)$$

The heterogeneity component of the predictive distribution is estimated by taking the average of two resulting heterogeneity estimates of initial meta-analyses. This calculation is performed in section 6.7.2.2.2 of Chapter 6.

### 7.7.2.5  Prediction model 15 - Designing a new head-to-head trial using random effects MTC methods: The predictive distribution accounts for both the within and between study variation: The residual heterogeneity approach

#### 7.7.2.5.1  Rationale

This prediction model forms an extension to the prediction model 10 in that it incorporates direct evidence in addition to indirect evidence, in a MTC setting.

The variance of the predictive distribution includes both within ($Var[\log^{\text{ind}} OR^{R}_{T1T2}]$) and between study variances ($\hat{\tau}^2_{T1T2}$). The between study variance is derived from the residual heterogeneity estimated from fitting the random effects MR model to the combined dataset. The mathematical formulae
Chapter 7  
Designing new trials using mixed treatment comparisons methods

pertaining to MR model is specified in section 6.7.2.6.1 of Chapter 6 under prediction model 10. The predictive distribution of the new head-to-head trial \((\theta_{T1T2new})\) is specified in equation 7.6 below.

\[
\theta_{T1T2new} \sim Normal \left( \text{log}^{\text{ind}} OR_{T1T2}^R, \text{Var}[\text{log}^{\text{ind}} OR_{T1T2}^R] + \hat{\tau}_{T1T2}^2 \right)
\]  
(7.6)

The mean and the variance of predictive distributions based on prediction models 14 and 15 are identical, although the process of estimating the heterogeneity separates these two models.

7.7.3 Methods

The methods used in random effects MTC settings are exactly as same as that specified in section 7.7.1.2.2 which specifies the method using the FE method.

7.7.4 Power results of prediction models 11-15

The power of the MTC meta-analysis of prediction models developed in section 7.7.2.2 – 7.7.2.5 above is summarised in table 7-2 below. The power results of the fixed effects MTC methods are also incorporated into the table aiming a detailed comparison between prediction models.
Table 7-2: Table summarising the power of MTC meta-analysis based on prediction models 11 – 15.

<table>
<thead>
<tr>
<th>Sample Size of the new head to head trial (one arm)</th>
<th>Power of the Mixed Treatment Comparison meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model11</td>
</tr>
<tr>
<td></td>
<td>Fixed effects</td>
</tr>
<tr>
<td></td>
<td>Predictive distribution accounts for only within study variation</td>
</tr>
<tr>
<td>100</td>
<td>0.52</td>
</tr>
<tr>
<td>200</td>
<td>5.86</td>
</tr>
<tr>
<td>300</td>
<td>11.84</td>
</tr>
<tr>
<td>500</td>
<td>23.44</td>
</tr>
<tr>
<td>800</td>
<td>35.10</td>
</tr>
<tr>
<td>1000</td>
<td>39.14</td>
</tr>
<tr>
<td>2000</td>
<td>54.94</td>
</tr>
<tr>
<td>3000</td>
<td>62.18</td>
</tr>
<tr>
<td>4000</td>
<td>66.70</td>
</tr>
<tr>
<td>5000</td>
<td>69.84</td>
</tr>
</tbody>
</table>

Figure 7-2 below presents the resulting individual power curves of RE based models, i.e. prediction model 12-15.
**Figure 7-2**: The individual power curves of prediction models 12-15.

### Prediction model 12 - Within study variance only

The power remains zero at least until the sample size increases close to 500. Thereafter, the power gradually increases as the sample size increases. Within the considered range of sample size, the power does not flatten off even at very larger sample sizes.

### Prediction model 13 - Average heterogeneity only

For smaller sample sizes of the new trial, i.e. at least until close to 500, the power remains zero. A rapid increase of power is observed as the sample size increases above 500. The power does not come to a plateau within the sample size range considered. The rate of power increase drops as the sample size increases.

### Prediction model 14 - With study variance and average heterogeneity

As observed in all prediction models based on RE, the power remains zero until the sample size reaches close to 500. Beyond that, a gradual increase of power is observed with increased sample size.

### Prediction model 15 - Within study variance and residual heterogeneity

The power does not become non-zero until the sample size reaches close to 500. Beyond that, the power gradually increases as the sample size rises. Within the sample size range considered, the power does not reach a plateau.
Figure 7-3 below combines all power curves presented in figure 7-2 above.

**Figure 7-3**: Summary plot of the statistical power for specific sample sizes for the prediction model 11-15.

The fixed effects MTC meta-analysis model (Prediction model 11), yields the highest power at all specified sample sizes of the new head-to-head trial. In contrast to the prediction models based on random effects MA model, this model produces power even at low sample sizes of the new head-to-head trial.

All prediction models based on random effects MA report zero power until the sample size of the new head-to-head trial increases to values close to 500. Further, in these models the power of the MTC meta-analysis is very much dependent on the variance of the predictive distribution. The more variance accounted in the predictive distribution the less power yields in the MTC meta-analysis. The RE based prediction models accounted for both elements of
variance, i.e. prediction models 14 and 15, yield relatively lower power compared to that of the RE models accounted for either of the variance components. These two models yield almost similar power although the estimation of heterogeneity employs different techniques.

### 7.8 Critical evaluation of the results

This chapter develops a methodology to explore power implications by combining indirect evidence with direct evidence within a MTC structure. The indirect evidence of the relative efficacy of two competing treatments is derived using Bucher’s methods and they are exactly similar to that produced in Chapter 6. Hence, this evaluation is also based on point estimates of IC presented in table 6-6. In addition to that, for clarity and convenience of illustration table 7-3 below introduces some notations that could be used to refer to various elements of the models.

**Table 7-3:** The notations of various model elements used for the illustration.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect</td>
<td>Direct</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pooled effect estimate</th>
<th>$\hat{\theta}_i^F$</th>
<th>$\hat{\theta}_i^D$</th>
<th>$\hat{\theta}_i^R$</th>
<th>$\hat{\theta}_i^D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight assigned</td>
<td>$w_i^F$</td>
<td>$w_i^D$</td>
<td>$w_i^R$</td>
<td>$w_i^D$</td>
</tr>
<tr>
<td>Standard error of the pooled effect estimate</td>
<td>$\hat{\sigma}_i^F$</td>
<td>$\hat{\sigma}_i^D$</td>
<td>$\hat{\sigma}_i^R$</td>
<td>$\hat{\sigma}_i^D$</td>
</tr>
<tr>
<td>Pooled effect estimate- MTC</td>
<td>$\hat{\mu}_F$</td>
<td>$\hat{\mu}_R$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard error of the pooled effect estimate- MTC</td>
<td>$SE(\hat{\mu}_F)$</td>
<td>$SE(\hat{\mu}_R)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity estimate between direct and indirect evidence- MTC</td>
<td>$\hat{\tau}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The behaviour of the power of MTC meta-analysis is akin to that of the updated MA (frequentist) including the new trial, explored in section 5.8.10 of Chapter 5. Central to all the variations of power between the FE and the RE models is the distance of $\hat{\theta}^F_i$ and $\hat{\theta}^R_i$ from the null value (log OR=0). The power of the MTC is associated with the confidence interval of the pooled effect size ($\mu_F$ and $\mu_R$).

The confidence interval is a direct function of the standard error of the pooled effect size ($SE(\mu_F)$ and $SE(\mu_R)$). Consequently, there is an association between the power and the standard error of the pooled effect estimate.

In this MTC approach, in a given prediction model, the indirect estimates, which defines the predictive distribution are always fixed. The direct evidence derived from the new head-to-head trial varies in each iteration. Most of the results and phenomenon found in this chapter can be explained with the help of following four equations (7.7 - 7.10) presented below.

In the FE model

$$SE(\mu_F) = \frac{1}{\sqrt{w^F_I + w^F_D}}$$  \hspace{1cm} (7.7)

$$SE(\hat{\mu}_F) = \frac{1}{\sqrt{\frac{1}{\sigma^F_I} + \frac{1}{\sigma^F_D}}}$$  \hspace{1cm} (7.8)

In the RE model

$$SE(\hat{\mu}_R) = \frac{1}{\sqrt{w^R_I + w^R_D}}$$  \hspace{1cm} (7.9)
Chapter 7  Designing new trials using mixed treatment comparisons methods

\[ SE(\mu_R) = \frac{1}{\sqrt{\frac{1}{\tau^2} + \frac{1}{\hat{\sigma}_i^R} + \frac{1}{\hat{\sigma}_d^R}}} \] (7.10)

The power of the MTC meta-analysis decreases as the variance in the predictive distribution of the new head-to-head trial increases. The wider the predictive distribution the more chance an effect size to be drawn further from the mean (\(\hat{\theta}_i^R\)). Thus, the heterogeneity (\(\hat{\tau}\)) between the direct (\(\hat{\theta}_d^R\)) and indirect evidence (\(\hat{\theta}_i^R\)) is increased. In equation 7.10, given the value of \(\hat{\sigma}_i^R\) is constant, increased \(\hat{\tau}\) leads to yield increased \(SE(\mu_R)\). A higher standard error produces a wider confidence interval of \(\mu_R\) and subsequently causes to lower the power.

The power of the random effects MTC meta-analysis of all prediction models is zero until the sample size increases up to slightly below 500. New head-to-head trials with smaller sample sizes are associated with higher within study errors (\(\hat{\sigma}_d^R\)). In equation 7.10, an increased \(\hat{\sigma}_d^R\), given both \(\hat{\sigma}_i^R\) and \(\hat{\tau}\) are fixed could cause \(SE(\mu_R)\) to increase and hence the power of the MTC meta-analysis decreases. For smaller sample sizes (in this case below 500), irrespective of whether the effect size is drawn from closer to or further from the mean of the predictive distribution (i.e. regardless of the value of \(\hat{\tau}\), the magnitude of \(\hat{\sigma}_d^R\) is smaller enough to exclusively create large enough magnitudes of \(SE(\mu_R)\) to produce wider confidence intervals of \(\mu_R\), for all possible drawn effect sizes. This could be the reason for the random effects MTC model to produce zero power at smaller sample sizes. The phenomenon of yielding zero power for new trials with smaller sample sizes in the RE models, can be explained as above.
The MTC meta-analysis including the new trial designed using the prediction model based on FE model yields the highest power amongst the rest of the models, at all sample sizes specified. This is because, less variability is accounted for in the predictive distribution of the new head-to-head trial (only the within study variance). Interestingly, the MTC meta-analysis based on FE model produces more power than that of the RE model which accounted for only the within study variance. This is partly caused by the difference in the distance from the mean of the fixed ($\hat{\theta}_f^R$) and random effects ($\hat{\theta}_r^R$) predictive distributions to the null value. The mean of the FE ($\hat{\theta}_f^R$) locates further from the null value, than the random effects ($\hat{\theta}_r^R$) counterpart. Hence the confidence interval of pooled effect size ($\hat{\mu}_f$) is less likely to cross the null value than its RE counterpart, hence resulting in increased power. However, this phenomenon is example specific.

Moreover, the random effects MTC model accounted for the heterogeneity between the direct and indirect evidence. Therefore, if two competing effect sizes drawn from the FE and RE predictive distributions, with identical standard errors ($\hat{\sigma}_f^R = \hat{\sigma}_r^F$ and $\hat{\sigma}_r^R = \hat{\sigma}_r^F$) are considered, the standard error of the pooled effect estimate in MTC meta-analysis based on the RE model, is larger than that in the FE counterpart. This wider standard error and resulting larger confidence interval causes the RE based MTC meta-analysis to produce less power than that based on FE model.
7.9 Discussion

This chapter contributes to the methodological framework by developing the module dedicated to design new head-to-head trial using MTC meta-analysis methods. It is interesting to note that the process of computing power of the updated MA (in Chapter 5) is analogous to the process of computing power based on MTC meta-analysis in this chapter. In addition to that, the pattern in the results is similar in these two contexts. In both contexts, the prediction model based on the FE model produces the highest power amongst all the models. Moreover, the more variability the prediction model accounted for the lesser the power produced by the MTC/updated meta-analysis.

All prediction models designed in this chapter are natural extensions to prediction models designed using IC settings in Chapter 6. Prediction models based on IC principles in Chapter 6 only accounted for the indirect evidence. In contrast, prediction models in this chapter employ indirect evidence to design a new head-to-head trial and combine that with indirect evidence in an MTC setting.

One of the downsides of this methodology is that it could be perceived as a MA of only two trials, although it summarises entire indirect evidence and the direct evidence derived within indirect evidence, pertaining to a specific clinical condition. Having only two trials in the MA affects the accuracy of the heterogeneity estimate and consequently on weights assigned and on summary estimates. However, we argue that the accuracy of the heterogeneity estimation in the context of mixed treatment comparison should be even higher as it combines both direct and indirect evidence.
Another concern in here is the bias caused by breaking the randomisation by combining direct and indirect evidence (Song, F., Altman et al. 2003). However, Lu and Ades argued that when combining direct and indirect evidence, the weighted average should be regarded a valid estimate if all trials in the evidence base are validly estimating a causal effect (Lu and Ades 2006).

Concerns have been expressed over the validity of combined estimates if direct and indirect evidence are sourced from two different populations (Lu and Ades 2006). However, these concerns are not applicable to the situation addressing in this chapter because both direct and indirect evidence are originated from the same source. Moreover, the two groups of trials (LEV vs PLC and TPM vs PLC) used in deriving the relative efficacy between two competing treatments (LEV vs TPM) are treating the same patient population as well. This further supports the validity of the combined estimates obtained in this chapter.

Future research could investigate the development and properties of ensuring consistency between the indirect and direct evidence. Inconsistent evidence could be discarded in the simulation process while performing more iterations to compensate for the discarded iterations to keep the total number of iterations to a constant (in this case 5000 iterations).

The indirect evidence contains more uncertainty than the corresponding direct evidence. The correlation between LEV and TPM groups could artificially reduce the uncertainty (Lu and Ades 2004). In future work, the correlation structure needs to be considered.
A further limitation of the methodology developed in this chapter is that it is restricted to designing trials between competing treatments that had been compared with a common comparator. One important challenge in the future is to extend the methodology to be able to design trials comparing potentially competing treatments those are connected in a network of complex evidence structures. Further to that, this methodology could be extended to be able to design trials with multiple arms, from an evidence structure that consists of a series of two and multiple arm comparisons and which connects a network of treatments.
Chapter 7

Designing new trials using mixed treatment comparisons methods

7.10 Conclusions

The methodology developed in this chapter seems particularly appealing in situations that need to borrow strength from indirect evidence because the direct evidence is limited or not available.

In the context of designing new trials based on direct and indirect evidence, the fixed effects MA should be used provided there is not substantial heterogeneity exists between trials, or in other words, all trials in each group may be regarded as estimating a single underlying effect size.

In the case of limited available sample size of a new study, prediction models based on a RE model are not particularly appealing as the basis for trial design, as the power of the MTC meta-analysis remains zero at least sample size rises up to 500. Despite reporting the lowest power, whenever possible, RE prediction models which accounted for both the within and between study variations, should be employed in designing new head-to-head trial as they attempt to describe the full variation between trials.

Although results observed in this chapter are example-specific in quantitative terms, the trends or the patterns detected in results may well not be in qualitative terms. For example, the pattern of yielding less power in the updated/MTC meta-analysis when both forms of variance are accounted for is generic in both frequentist MA and MTC meta-analysis settings (figure 5-4 of Chapter 5 and figure 7-2 of Chapter 7).
Further to that, the phenomenon of observing zero power at smaller sample sizes of the new head-to-head trial is typical of use of a RE model, as far as the power of the updated/MTC meta-analysis is considered (figure 5-4 of Chapter 5 and figure 7-2 of Chapter 7).

Thus far, this thesis has considered developing methodologies to design new trials employing MA approaches. Chapter 8 upcoming considers developing a methodology of designing new trials using MR approaches.
CHAPTER 8

Designing future trials based on existing evidence using meta-regression methods

8.1 Introduction

This chapter develops a methodology to design future trials based on existing evidence using Meta Regression (MR) principles. The motivation to this section is briefly discussed in section 8.2 followed by the objectives in section 8.3. The theory used in this chapter is described in section 8.4. Section 8.5 discusses the existing evidence used throughout this chapter. Several models based on MA, and meta-regression (MR) are compared in section 8.7. Section 8.8 models the relationship between the effect size and the administered Levetiracetam (LEV) dose, for accumulating evidence available at three distinct time points. Section 8.9 develops the methodology to design a single trial. Within that, section 8.9.1 explores the possibility of designing two-arm trial and section 8.9.2 explores the design of three-arm trials. Designing multiple two-arm trials is considered in section 8.10. A dose wise comparison of power is carried out in section 8.11. The results are critically evaluated in section 8.12. A general discussion to this chapter is given in section 8.13 and section 8.14 concludes this chapter.
8.2 Motivation

The MR approach is generally used in modelling the variation in observed effect sizes of primary study data with study level covariates. The literature in relation to MR has been previously reviewed in section 2.7 of Chapter 2 and found no evidence of previous application of MR methods to inform the design of future trials. Therefore, the proposed approach in this chapter of designing future trials based on MR principle is assured to be novel area of research. Moreover, the potential to incorporate study specific covariates in to future study design may lead it to receive widespread attention, and perhaps extensive use in practice.

Sutton et al (Sutton, A. J., Cooper et al. 2007), who originally proposed the simulation based approach of power calculation in designing new trials, have not designed and implemented the approach via MR although they discussed the feasibility of the approach. However, an adaption of this approach into MR was published in 2012 in which a quantitative insight into the amount of support that a new trial may provide to the hypothesis that the covariate is an effect modifier in an updated MR is presented (Rotondi, Donner and Koval 2012).

This section discusses the motivation of this chapter. The next section outlines the key objectives to be achieved in this chapter.
8.3 Objectives

The fundamental shortcoming in the existing methodology addressed through the development of methods throughout this thesis is the lack of use of the existing evidence in the planning and designing stages of new clinical trials. The principal objective of this chapter is to develop a methodology to design new trials including specific study level covariates, based on existing evidence, using MR methods.

This chapter further aims to identify time points (within the span of time where the existing evidence is accumulated) at which the need to design further trials could have been triggered, by modelling the effect size with specific covariates in a MR framework. In the context of the illustrative example used in this chapter, this investigation is carried out at three distinct time points.

In addition to that, this chapter aims to design two-arm and three-arm trials, based on existing evidence using a MR approach. In similar manner to the methodologies developed in previous chapters, the interest is in the power of the updated MR including the new trial.

Further, this chapter focuses on developing multiple trials (Two trials and five trials in this instance by way of illustration), on adding the new trials into the existing MR, and on exploring the fluctuation of power to changes in sample size.

The upcoming section briefly describes the key theoretical aspects employed in this chapter.
8.4 Theory used

8.4.1 Meta-regression model

Generally, under the MR model, the underlying treatment effect is regressed on a study level covariate. It is assumed that $i$ of a total of $n$ studies provide an estimate, $y_i$, of effect of interest, i.e. the log odds ratio in the running example. Each study provides the standard error of this estimate, $\sigma_i$, which is assumed to be known.

8.4.1.1 Fixed effects meta-regression model

This model is outlined in Hedges and Cooper's (Cooper, H. and Hedges 1994) and could include both discrete and continuous predictor variables. Suppose there are $k$ independent effect size estimates, $y_i$, with sampling variances $\epsilon_i$ and $p$ known predictor variables $X_i, \ldots, X_p$. The fixed effects MR extends fixed effects MA model by replacing the mean, $\theta_i$, with a linear predictor, $x_i\beta$.

The fixed effect MR model for $k$ independent effect size parameters ($\theta_1, \ldots, \theta_k$) can be written as (Sutton, A., Abrams et al. 2000)

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad (8.1)$$

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \quad (8.2)$$

where $i = 1, \ldots, k$ and $x_{i1}, \ldots, x_{ip}$ are the values of the predictor variables $X_1, \ldots, X_p$ for the $i^{th}$ study. $\beta_0, \beta_1, \ldots, \beta_p$ are unknown regression coefficients to
be estimated and are often estimated using weighted least square algorithms. The weights are defined as the reciprocal of the sampling error.

### 8.4.1.1.1 Differences to the fixed effects meta-analysis model

Fixed effects MA assumes all true effects equal to $\theta$ while FE regression let the true value vary between trials, as is also the case in the random effects MA. However, here the covariate predictor variables are responsible for the variation not the random effects. The variation is assumed not random but predictable.

### 8.4.1.1.2 Practical limitations of the fixed effects meta-regression model

The fixed effects MR model is based on the assumption that the total variation is accounted for by the predictors. However, in practice the total variability in true effect sizes is not completely accounted for by the covariates. This poses some practical limitations on the use of fixed effect MA. The amount of variability left unexplained by covariates is named residual heterogeneity. The residual heterogeneity is accounted for in a random effects MR model.

### 8.4.1.2 Random effects meta-regression model

The model present below is a modified version from Raudenbush in (Cooper, H. and Hedges 1994). This model is most suitable when only a part of the variation is explained by the predictive covariates, and a random effect term ($u_i$) is used to explain the residual variation. The residual variation is accounted for by including a random effect to the regression and hence the name 'random effects'. The model is given below (Sutton, A., Abrams et al. 2000),
\[
\theta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + u_i
\]  \hspace{1cm} (8.3)

where, \( \theta_i \) is the true effect size in the ith study.

\( \beta_0 \) is the intercept of the model. \( \beta_1, \ldots, \beta_p \) are the regression coefficients. \( u_i \) is the random effects of study \( i \), which is the difference between the true effect size and the value predicted based on the model.

\( u_i \) is assumed independent and distributed as below with zero mean and variance \( \tau_i^2 \).

\[
u_i \sim N(0, \tau_i^2) \]  \hspace{1cm} (8.4)

When dealing with a single covariate, a simplified version of this model is used as below.

\[
\theta_i = \beta_0 + \beta_1 x_{i1} + u_i
\]  \hspace{1cm} (8.5)

where \( \theta_i \) is the true effect size, \( \beta_0 \) is the intercept of the model and \( \beta_1 \) is the regression coefficient capturing the association between the treatment effect and the study characteristics. \( x_{i1}, \ldots, x_{ip} \) are the values of the predictor variables \( X_1, \ldots, X_p \) for the \( i^{th} \) study.

Since unexplained residual variation is always included explicitly in the model, it is argued that a RE component should always be included in the model.

Therefore, the development of the methodology is limited to the random effects MR principles.
### 8.4.1.3 Weighted regression and residual heterogeneity

In meta-regression analysis trials should be weighted so that trials that are more precise become more influential in the analysis. Generally, the weights assigned are equal to the inverse of sum of within study variance and the residual between study variance. This is however different to that in the random effects MA since random effects MR uses the residual between study variation, which is the heterogeneity not explained by the covariate in the model. Therefore, the weights assigned in random effects MA are somewhat different to that used in random effects MR.

#### 8.4.1.3.1 The weighted Least Square Estimation

In the fixed effects MR, the weights are defined by the reciprocal of the sampling variation. Hence the weight given in the \(i^{th}\) trial is given by

\[
    w_i^F = \frac{1}{\sigma_i^2}
\]  

(8.6)

Optimum weights are assigned in the random effects MR by the inverse of each trials variance. As explained both components of variability are involved in the random effects MR and the weight given in the \(i^{th}\) study is as follows.

\[
    w_i^R = \frac{1}{\sigma_i^2 + \tau^2}
\]  

(8.7)

Both \(\sigma_i^2\) and \(\tau^2\) are not exactly known and are estimated from data at hand. The estimation of regression coefficients requires preliminary estimates of \(\sigma_i^2\) and \(\tau^2\), to calculate corresponding weights. Two types of approaches are commonly used in practice to estimate of the heterogeneity namely the methods of moments and the residual maximum likelihood method.
8.4.1.3.1.1 Estimating the heterogeneity

Two widely used techniques are found in the literature for estimating the heterogeneity parameter and described below.

**Methods of moments estimation**

This is a method based on a non-iterative process. This method at the beginning requires provisional estimates of $\beta$ vector to compute $\tau^2$ and weights. These weights are used in a weighted least square regression to obtain new estimates of the $\beta$ vector.

**Maximum likelihood estimation**

This is an alternative estimation technique, which simultaneously estimates both $\beta$ vector and weights. However, this requires a further simplification that the distribution of the observed effect sizes is normal.

8.4.2 The Hat matrix and leverage

In regression analysis, the hat matrix (H) or projection matrix is studied by many to detect the leverage points that are based on the diagonal elements of the hat matrix. (Chave and Thomson 2003)

In regression

$$\hat{y} = X.\hat{\beta} \quad (8.8)$$

Where $\hat{y}$ is the predicted y values, The weighted least square solution of $\hat{\beta}$
\[
\hat{\beta} = (X'X)^{-1}X'y 
\]  \hspace{1cm} (8.9)

Hence

\[
\hat{y} = X (X'X)^{-1}X'y 
\]  \hspace{1cm} (8.10)

The predicted value of \( y \), i.e. \( \hat{y} \) could be obtained by multiplying \( y \) by \( X (X'X)^{-1}X' \). This element is called the hat matrix (H).

\[
H = X (X'X)^{-1}X' 
\]  \hspace{1cm} (8.11)

The hat matrix has some interesting statistical properties.

\[
\begin{bmatrix}
\hat{y}_1 \\
\vdots \\
\hat{y}_n \\
\end{bmatrix} = 
\begin{bmatrix}
h_{11} & h_{1i} & h_{1n} \\
h_{i1} & h_{ii} & h_{in} \\
h_{n1} & h_{ni} & h_{nn} \\
\end{bmatrix} 
\begin{bmatrix}
y_1 \\
\vdots \\
y_n \\
\end{bmatrix} = 
\begin{bmatrix}
h_{11}y_1 + h_{1i}y_i + h_{1n}y_n \\
h_{i1}y_1 + h_{ii}y_i + h_{in}y_n \\
h_{n1}y_1 + h_{ni}y_i + h_{nn}y_n \\
\end{bmatrix} 
\]  \hspace{1cm} (8.12)

The diagonal elements of the hat matrix are called leverage points. Each diagonal element in the hat matrix determines the influence of each observation (trial in the case of MR) on its own predicted value as apparent from highlighted elements of the matrix (Chatterjee and Hadi 1986).

This thesis attempts to explore the possibility of using existing evidence to form the basis of designing future trials. Therefore as a prerequisite to the development of the methodology, a series of trials addressing the same therapeutic question is necessary to constitute the existing evidence. Section 8.5 below describes the process used to select the existing evidence and the mechanism used to verify the consistency of the existing evidence.
8.5 Existing evidence

8.5.1 Gathering existing evidence

To maintain the consistency and to establish the continuity of the tasks undertaken throughout the course of the thesis, this chapter employs a part of the same existing evidence used in preceding Chapters 6 (IC methods) and 7 (MTC methods) to design future trials. The set of trials listed in table 8-1 below is initially extracted from the review of Levetiracetam add-on for drug-resistant localization related (partial) epilepsy published in the Cochrane Database of Systematic Review of (Chaisewikul, Privitera et al. 2001).

Table 8-1 below lists all trials that constitute the existing evidence on which the analysis is based. These trials compare the efficacy and safety of different doses of LEV against PLC. Notably this dataset contains three trials comparing multiple LEV doses (three arm trials) and two two-arm trials.

Table 8-1: Trials comparing efficacy and safety of different doses of Levetiracetam (LEV) with Placebo (PLC) in treating partial epilepsy, which constitute the existing evidence.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Year</th>
<th>PLC</th>
<th>LEV 1000</th>
<th>LEV 2000</th>
<th>LEV 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Events</td>
<td>Events</td>
<td>Events</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>Total</td>
<td>Total</td>
<td>Total</td>
</tr>
<tr>
<td>Cereghino et al</td>
<td>2000</td>
<td>10</td>
<td>95</td>
<td>31</td>
<td>98</td>
</tr>
<tr>
<td>Shorvon et al</td>
<td>2000</td>
<td>7</td>
<td>112</td>
<td>25</td>
<td>106</td>
</tr>
<tr>
<td>Ben-Menachem et al</td>
<td>2000</td>
<td>15</td>
<td>105</td>
<td>48</td>
<td>183</td>
</tr>
<tr>
<td>Boon et al</td>
<td>2002</td>
<td>21</td>
<td>172</td>
<td>48</td>
<td>183</td>
</tr>
<tr>
<td>Tsai et al</td>
<td>2006</td>
<td>5</td>
<td>47</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>
The previously used (table 6-2) dataset is augmented by incorporating a column representing the dose administered in each trial, which is the study level covariate of concern here. This facilitates the modelling of the relationship of the efficacy and the dose level.

Each individual trial in the dataset is aiming to investigate the efficacy of one or more dose of Levetiracetam (LEV) with placebo (PLC) as an adjunctive therapy in treating partial epilepsy.

Because it has been a year since this dataset was last used in developing methods using IC methods, it is prudent to ensure that the dataset is up-to-date by systematically searching for relevant similar trials published recently, before embark in on the analysis.

### 8.5.2 Updating existing evidence

#### 8.5.2.1 Searching for new trials

The Cochrane Database of Systematic Reviews was searched to identify if any new trials are added in to this specific systematic review. Few changes to the latest issue of the review (issue 1, 2010) were discovered although the conclusions remain unchanged. It was further noticed that the content of the review was last updated in June 2005 despite the cover page suggested it was last updated in 2010. I contacted the principal author of this review requesting the reason for this observed discrepancy, but our email queries were unanswered. I then contacted a co-author of the review but the response gave little clarification of the situation.
Subsequently, a systematic search was conducted with key words ‘levetiracetam’ and ‘adjunctive’ and ‘partial epilepsy’ in academic databases including the ‘Web of Science’, ‘Scopus’ and ‘Cochrane Databases of Systematic Review’, in order to identify relevant new trials recently added to the literature.

The search undertaken in Scopus and Web of Science yielded seven potentially relevant trials. Three trials, focusing on the assessment of the efficacy of LEV as an adjunctive therapy for patients with uncontrolled partial seizure, were excluded because they do not have a placebo arm. (Heo, Lee et al. 2007; Somerville, McLaughlin et al. 2007; Steinhoff, Somerville et al. 2007). Another trial consisting of a child population was also excluded due to a population mismatch. (Jesus Eric, Douglas et al. 2009). The remaining three trials identified have met the inclusion criteria and were subsequently included in the analysis. Two of three trials (Zhou, Zhang et al. 2008; Xiao, Li et al. 2009) compare LEV 3000 and one compares LEV 1000 (Peltola, Coetzee et al. 2009) with PLC.

### 8.5.2.2 Incorporation of new trials into the existing evidence

The new trials identified in all three databases are identical except that the Cochrane Database yielded another relevant trial by Wu et al (Wu, Hong et al. 2009). This increases the count of relevant trials up to four. Table 8-2 below shows the revised list of trials following the inclusion of four more relevant trials into the existing dataset. Notably, all the trials added later are two-arm trials.
Table 8-2: Trials constituting the existing evidence in the meta-regression approach, comparing LEV add-on therapy and PLC in treating partial epilepsy.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Month/ Year of publication</th>
<th>PLC Events</th>
<th>Total</th>
<th>LEV 1000 Events</th>
<th>Total</th>
<th>LEV 2000 Events</th>
<th>Total</th>
<th>LEV 3000 Events</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereghino et al</td>
<td>Jul-00</td>
<td>7</td>
<td>95</td>
<td>36</td>
<td>98</td>
<td>40</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shorvon et al</td>
<td>Sep-00</td>
<td>7</td>
<td>112</td>
<td>22</td>
<td>106</td>
<td>37</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben-Menachem et al</td>
<td>Oct-00</td>
<td>15</td>
<td>105</td>
<td>15</td>
<td>105</td>
<td>71</td>
<td>181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boon et al</td>
<td>01-Jan-02</td>
<td>21</td>
<td>172</td>
<td>48</td>
<td>183</td>
<td>60</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsai et al</td>
<td>Jan-06</td>
<td>5</td>
<td>47</td>
<td>20</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou et al</td>
<td>Feb-08</td>
<td>2</td>
<td>11</td>
<td></td>
<td></td>
<td>8</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peltota et al</td>
<td>Mar-09</td>
<td>23</td>
<td>79</td>
<td>34</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xiao et al</td>
<td>Mar-09</td>
<td>11</td>
<td>28</td>
<td></td>
<td></td>
<td>13</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al</td>
<td>Mar-09</td>
<td>26</td>
<td>100</td>
<td></td>
<td></td>
<td>57</td>
<td>102</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As set out in the objectives, one aim of this chapter is to establish what evidence is available at the designing stage of each new trial to discover the benefits each trial could have had if existing evidence was considered. To this end, further information is obtained for each of the relevant trials including the period of data collection and the date of publication. This relevant information is extracted from each individual article and reported in table 8-3 below.
Table 8-3: The table shows the duration of data collection and other publication related dates of each trial.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Start date of data collection</th>
<th>End date of data collection</th>
<th>Month/Year of Acceptance</th>
<th>Early view publication</th>
<th>Month/Year of Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereghino et al</td>
<td>01-Sep-94</td>
<td>07-Mar-96</td>
<td>17-Mar-00</td>
<td></td>
<td>Jul-00</td>
</tr>
<tr>
<td>Shorvon et al</td>
<td>Not Available</td>
<td>Not Available</td>
<td>12-May-00</td>
<td></td>
<td>Sep-00</td>
</tr>
<tr>
<td>Ben-Menachem et al</td>
<td>Jun-95</td>
<td>Mar-97</td>
<td>25-May-00</td>
<td></td>
<td>Oct-00</td>
</tr>
<tr>
<td>Boon et al</td>
<td>Oct-93</td>
<td>Oct-95</td>
<td>28-Sep-01</td>
<td>01-Jan-02</td>
<td></td>
</tr>
<tr>
<td>Tsai et al</td>
<td>Nov-00</td>
<td>Feb-02</td>
<td>20-Aug-05</td>
<td>10-Jan-06</td>
<td>Jan-06</td>
</tr>
<tr>
<td>Zhou et al</td>
<td>01-Jul-04</td>
<td>03-May-06</td>
<td>07-Aug-07</td>
<td>19-Nov-07</td>
<td>Feb-08</td>
</tr>
<tr>
<td>Peltoia et al</td>
<td>21-Aug-06</td>
<td>30-May-07</td>
<td>04-Aug-08</td>
<td>24-Oct-08</td>
<td>Mar-09</td>
</tr>
<tr>
<td>Xiao et al</td>
<td>Not Available</td>
<td>Not Available</td>
<td>26-Sep-08</td>
<td>29-Jan-09</td>
<td>Mar-09</td>
</tr>
<tr>
<td>Wu et al</td>
<td>Jul-04</td>
<td>May-05</td>
<td>29-May-08</td>
<td>24-Jul-08</td>
<td>Mar-09</td>
</tr>
</tbody>
</table>

Table 8-3 above lists all trials in chronological order starting from the earliest publication. The start and end date of data collection are not available in trials reported by Shorvon and Xiao et al. Zhou et al only reported the duration in which the patients were screened. Consequently, the end date of data collection is estimated by adding 48 weeks to the end date of patient screening.

Before performing the synthesis, it is important to be certain that the pooling would produce sensible results. The outcome reported in each trial should be meaningfully comparable with the rest of the trials. Section 8.5.3 below investigates the consistency of the existing evidence found in this section.
8.5.3 The consistency of the existing evidence included in the analysis

Before proceeding with the analysis, it is essential to ascertain the consistency of the outcome measures reported in each trial. Trials generally collect a wide variety of data and report a wide range of results. In the synthesis, it is necessary to ensure the consistency and the compatibility of outcomes being combined. The synthesis of inconsistent outcomes generates flawed results.

8.5.3.1 Different patient population and treatment periods

In LEV trials, the outcome of interest is 50% responder ratio, which is defined as the proportion of patients received at least 50% recovery compared to the baseline. Due to dissimilarities in objectives and various design aspects of trials, analyses are carried out and results are produced for different patient populations (Per Protocol and Intention to Treat population) and treatment periods (titration and evaluation period, evaluation period only). Using different combinations of patient populations and treatment periods, an array of distinct responder ratios could be calculated.

Table 8-4 lists six distinct population/period combinations which have been used to base the responder ratio computations. These six combinations are listed below and assigned letters A-F, for the convenience of referring to them subsequently.


B: Per Protocol population in Titration and Evaluation Period : PP/TEP
Chapter 8

Designing new trials using meta-regression methods

C: Inferential Intention To Treat population in Evaluation Period : IIT/EP

D: Inferential Intention To Treat population in Titration and Evaluation Period : IIT/TEP

E: Intention To Treat population in Evaluation Period : IIT/EP

F: Intention To Treat population in Titration and Evaluation Period : IIT/TEP

Table 8-4 below represents the population/period combinations on which the responder rate is based on, in each trial.

**Table 8-4: Population/Period combinations used in each trial in computing the responder ratio.**

<table>
<thead>
<tr>
<th>Ref</th>
<th>Population / Period</th>
<th>Ceregheno</th>
<th>Shorvon</th>
<th>Boon</th>
<th>Tsai</th>
<th>Ben-Menachem</th>
<th>Zhou</th>
<th>Peltola</th>
<th>Xiao</th>
<th>Wu</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>PP/EP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>PP/TEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>IITT/EP</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>IITT/TEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>ITT/EP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>ITT/TEP</td>
<td>F</td>
<td></td>
<td>F</td>
<td></td>
<td>F</td>
<td></td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

According to table 8-4 above, four trials each used IITT/EP (C) and ITT/TEP (F) combinations to compute responder ratios. This limits the number of trials being included in any given synthesis to a maximum of four and consequently made five remaining trials being redundant. It was however realised that the Cochrane review of Levetiracetam add-on therapy for drug resistant partial epilepsy (Chaisewikul,Privitera et al. 2001) from which the original existing evidence is retrieved (table 6-1), in some computations used outcome measures which are
not available in original articles. i.e. despite the Shorvon and Ben-Menachem trials basing the analysis on the outcomes measured on IITT/EP (C) combination, this review however computed the responder ratio based on ITT/TEP (F) combination. Thus, the total number of trials that uses ITT/TEP combination to compute responder rate increases up to six, and presented in table 8-5 below.

**Table 8-5**: Revised table following the incorporation of information from the Cochrane Review about population/period combination used to compute responder rate.

<table>
<thead>
<tr>
<th>Combination</th>
<th>Cereghino</th>
<th>Shorvon</th>
<th>Boon</th>
<th>Tsai</th>
<th>Ben-Menachem</th>
<th>Zhou</th>
<th>Peltola</th>
<th>Xiao</th>
<th>Wu</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP/EP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP/TEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IITT/EP</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IITT/TEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT/EP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT/TEP</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

According to table 8-5 above, the ITT/TEP (F) combination has been used to compute responder ratios in six trials. Three of nine trials (Boon et al, Zhou et al and Peltola et al) do not compute responder rate based on ITT/TEP (F) combination. The lead authors of these three trials have been contacted via email enquiring the availability of any unpublished data or results related to the responder rates computed basing ITT/TEP combination. None of the authors responded. In addition to that, the lead author of the Cochrane Review of Levetiracetam add-on for partial epilepsy (Chaisewikul, Privitera et al. 2001) was contacted to check if they possess any relevant unpublished data. These email queries were also unanswered.
Finally, the decision was made to include trials that report responder ratios based on ITT/TEP (F) into the main analysis based on the highest number of trials available. Existing evidence on which the synthesis is based is thus consisted of six trials. These trials compare efficacy and safety of LEV add-on therapy with PLC for patients with partial epilepsy and are given in table 8-6 below.

**Table 8-6**: Trials reporting the responder rate computed using the ITT/TEP combination with number of events and total in each arm.

<table>
<thead>
<tr>
<th>Trial</th>
<th>PLC</th>
<th>LEV 1000</th>
<th>LEV 2000</th>
<th>LEV 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Total</td>
<td>Events</td>
<td>Total</td>
</tr>
<tr>
<td>Cereghino et al</td>
<td>7</td>
<td>95</td>
<td>36</td>
<td>98</td>
</tr>
<tr>
<td>Shorvon et al</td>
<td>7</td>
<td>112</td>
<td>22</td>
<td>106</td>
</tr>
<tr>
<td>Ben-Menachem et al</td>
<td>15</td>
<td>105</td>
<td>22</td>
<td>106</td>
</tr>
<tr>
<td>Tsai et al</td>
<td>5</td>
<td>47</td>
<td>20</td>
<td>46</td>
</tr>
<tr>
<td>Xiao et al</td>
<td>11</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al</td>
<td>26</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The list of trials constituting the existing evidence to be used in the analysis is now finalised. However, it should be noted that the power of the existing meta-regression is expected to be low because it includes only of six trials.

### 8.5.4 Transforming three arm trials into two-arm trials

The majority of this chapter addresses the methods of designing two-arm trials, except section 8.9.2 which explores designing three-arm trials. For simplicity, a three-arm trial is treated as a combination of two two-arm trials. Hence the evidence from three-arm trials in existing evidence in table 8-6 is split into two-arm comparisons. To model the efficacy of specific LEV doses with PLC, trials,
which compare more than one LEV doses with PLC are split as though they constitute two trials of two-arms.

### 8.5.4.1 The correlation in the treatment effects of three-arm trials

The resultant two treatment effects in a three-arm trial are correlated as the PLC arm is common to both treatments. For example, treatment effects of $OR_{LEV2000PLC}$ and $OR_{LEV1000PLC}$ in a three-arm trial of LEV 1000, 2000 and PLC are correlated in this running example. The resulting placebo arms of both two-arm trials are assumed independent one another (as if they are not arising from the same placebo arm). This is to avoid the need to account for the inherited correlation in individual treatment effects of two two-arm trials produced from a three-arm trial.

The recommendations outlined in the Cochrane Handbook of Systematic Reviews of interventions (Higgins, J. P. T., Green and editors 2005) has been adhered to, in converting three-arm trials into two-arm trials. Thus, the number of events and total patients in the PLC group is equally split between two treatment groups to avoid duplication of evidence and inflation of power.

The trial reported by Cereghino et al is a three-arm trial, which compares LEV 1000 and 3000 with PLC. This trial is split to construct two two-arm trials. Each of these two-arm trials has a PLC arm and one of LEV doses (from the three-arm trial) being the second arm. The three-arm trial reported by Shorvon et al is also split into two trials of two-arm as though they compare LEV 1000 and LEV 2000 with PLC individually.
Table 8-7 below shows the number of events and non-events in each arm alongside the corresponding LEV dose in each trial. The three-arm trials in previous table 8-6 are represented by two two-arm trials in below table.

Table 8-7: Existing evidence having split three-arm trials into two-arm trials.

<table>
<thead>
<tr>
<th>Trial</th>
<th>LEV Events</th>
<th>LEV Non events</th>
<th>PLC Events</th>
<th>PLC Non events</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereghino1</td>
<td>36</td>
<td>62</td>
<td>3</td>
<td>44</td>
<td>1000</td>
</tr>
<tr>
<td>Cereghino2</td>
<td>40</td>
<td>61</td>
<td>4</td>
<td>44</td>
<td>3000</td>
</tr>
<tr>
<td>Shorvon1</td>
<td>22</td>
<td>84</td>
<td>3</td>
<td>52</td>
<td>1000</td>
</tr>
<tr>
<td>Shorvon2</td>
<td>37</td>
<td>69</td>
<td>4</td>
<td>53</td>
<td>2000</td>
</tr>
<tr>
<td>Ben-Menachem</td>
<td>71</td>
<td>110</td>
<td>15</td>
<td>90</td>
<td>3000</td>
</tr>
<tr>
<td>Tsai</td>
<td>20</td>
<td>26</td>
<td>5</td>
<td>42</td>
<td>2000</td>
</tr>
<tr>
<td>Xiao</td>
<td>13</td>
<td>15</td>
<td>11</td>
<td>17</td>
<td>3000</td>
</tr>
<tr>
<td>Wu</td>
<td>57</td>
<td>45</td>
<td>26</td>
<td>74</td>
<td>3000</td>
</tr>
</tbody>
</table>

The three-arm trials reported by Cereghino and Shorvon were split into two two-arm trials and are referred as Cereghino1, Cereghino2, Shorvon1 and Shorvon2 in table 8-7 above. The proposed methodologies developed in upcoming sections employ existing evidence given in table 8-7 above to base the analysis on.

8.5.5 Event of interest

8.5.5.1 Definition of an event

The definition of an event is identical to that specified in section 6.5.3 of Chapter 6. This is because this chapter employs part of the existing evidence used in the IC context.
8.5.5.2 Estimating the event rate in the control group

This methodology requires the event rate in the control group ($P_{\text{Control\_new}}$), i.e. PLC group of a new trial to be known beforehand. This is usually estimated by taking the mean of the event rate of the control group of the existing trials. In the illustrative example this value is estimated to be 0.1467.

The next section discusses the development and implementation of the methodology.

8.6 Development of the Methodology

The methodology of developing new trials based on existing evidence using MR principles are developed in phases.

**Phase one:**

In this phase an assessment is carried out about the MR methods (when covariates are ignored) producing identical summary measures to that of MA methods.

**Phase two:**

Phase two carries out the initial synthesis of the existing evidence. It further aims to identify the evidence available to researchers (specifically the authors of trials available in the dataset) and to determine the preliminary synthesis they should have done in the planning stage of their trial before embark on to design and conduct of their respective trials.
Phase three:

This phase discusses the development of the new methodology to design new two-arm trials.

Phase four:

This phase develops the methodology to design new three-arm trials.

Phase five:

Development of the methodology to design multiple new trials is discussed here, particularly the designing of two two-arm trials.

Phase six:

Development of the methodology to design multiple new trials is discussed here, particularly the designing of five two-arm trials.

8.6.1 Null hypothesis tested and power in the context of the meta-regression

The null hypothesis tested in the MR model is that there is no association between the treatment efficacy and the predictors, i.e. the LEV dose in the running example. In other words, the regression coefficient of the association between the observed treatment effects and study level covariates, i.e. LEV dose is zero. ($H_0: \beta_1 = 0$)
In the context of MR, the power is referred to as the probability that the test will correctly reject the above null hypothesis. The null hypothesis will not be rejected if the confidence interval of the above regression coefficient contains zero. The power of the updated MR including the new trial is of interest here.

Each phase is discussed in detail in upcoming sections from 8.7 to 8.10.

8.7 **Phase 1: The consistency of summary measures computed using meta-analysis and meta-regression models**

The aim of the phase one is to check the consistency of summary estimates and prediction intervals derived from random effects MA and MR models. The utilization of such models producing identical summary estimates as underlying models in developing the framework of methodology ensures that each module of the framework yields outcomes those could be readily comparable with each other. Moreover, it highlights the differences in the results attributable to the differences in the methods used.

In this phase, an investigation is undertaken to verify the models producing identical pooled effect sizes and prediction intervals with several models based on MA and MR. Upcoming sections 8.7.1 - 8.7.2 describe each variant in detail.
8.7.1 Meta-analysis

This section explores the differences in the resulting summary estimates of the models based on random effects MA methods. These two models are different with respect to the number of input variables and the method used in estimating the heterogeneity. The existing evidence in table 8-7 is used to base the analysis on. Sections 8.7.1.1 – 8.7.1.2 discuss each model in turn.

8.7.1.1 Model 1 - Random Effects meta-analysis that estimates the heterogeneity using the Mantel-Haenszel method

This model is based on the random effects MA model that estimates the heterogeneity using the Mantel-Haenszel method.

Figure 8-1: The forest plot of the results of the meta-analysis of the existing evidence in which the heterogeneity is estimated using Mantel-Haenszel method.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
<th>%</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>cereghino-part1</td>
<td>8.52 (2.47, 29.41)</td>
<td>7.56</td>
<td></td>
</tr>
<tr>
<td>cereghino_part2</td>
<td>7.21 (2.40, 21.64)</td>
<td>9.35</td>
<td></td>
</tr>
<tr>
<td>shorvon_part1</td>
<td>4.54 (1.29, 15.92)</td>
<td>7.39</td>
<td></td>
</tr>
<tr>
<td>shorvon_part2</td>
<td>7.11 (2.38, 21.17)</td>
<td>9.45</td>
<td></td>
</tr>
<tr>
<td>Ben-menachem</td>
<td>3.87 (2.08, 7.22)</td>
<td>22.72</td>
<td></td>
</tr>
<tr>
<td>Tsai</td>
<td>6.46 (2.16, 19.32)</td>
<td>9.40</td>
<td></td>
</tr>
<tr>
<td>Xiao</td>
<td>1.34 (0.46, 3.87)</td>
<td>9.92</td>
<td></td>
</tr>
<tr>
<td>Wu</td>
<td>3.61 (1.99, 6.53)</td>
<td>24.21</td>
<td></td>
</tr>
<tr>
<td>Overall (I-squared = 17.9%, p = 0.288)</td>
<td>4.33 (3.02, 6.21)</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>with estimated predictive interval</td>
<td>. (2.15, 8.73)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis.
The existing evidence is analysed using this specific model and the pooled log OR and its 95% prediction interval are estimated. The forest plot in figure 8-1 shows the OR from and weights assigned to each individual trial along with pooled OR and its prediction interval. The summary results are given in summary table 8-8 in section 8.7.3.

### 8.7.1.2 Model 2- Random effects meta-analysis that estimates the heterogeneity using the inverse variance method

In this section, the pooled log OR and its 95% prediction intervals are estimated using the random effects MA model, which specifies the heterogeneity using inverse-variance fixed effects method.

**Figure 8-2**: The forest plot showing the results of the meta-analysis in which heterogeneity is estimated using Inverse variance method.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>cereghino-part1</td>
<td>8.52 (2.47, 29.41)</td>
<td>7.51</td>
</tr>
<tr>
<td>cereghino_part2</td>
<td>7.21 (2.40, 21.64)</td>
<td>9.30</td>
</tr>
<tr>
<td>shorvon_part1</td>
<td>4.54 (1.29, 15.92)</td>
<td>7.35</td>
</tr>
<tr>
<td>shorvon_part2</td>
<td>7.11 (2.38, 21.17)</td>
<td>9.40</td>
</tr>
<tr>
<td>Ben-menachem</td>
<td>3.87 (2.08, 7.22)</td>
<td>22.84</td>
</tr>
<tr>
<td>Tsai</td>
<td>6.46 (2.16, 19.32)</td>
<td>9.35</td>
</tr>
<tr>
<td>Xiao</td>
<td>1.34 (0.46, 3.87)</td>
<td>9.88</td>
</tr>
<tr>
<td>Wu</td>
<td>3.61 (1.99, 6.53)</td>
<td>24.37</td>
</tr>
<tr>
<td>Overall (I-squared = 17.1%, p = 0.295)</td>
<td>4.33 (3.02, 6.20)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

*NOTE: Weights are from random effects analysis.*
The forest plot in figure 8-2 shows the individual effect estimates, weights assigned to each study as well as the pooled effect estimate and its 95% prediction interval.

Following the computation of summary estimates from random effects MA based models, sections 8.7.2 below explores the same summary measures via MR methods.

8.7.2  **Meta-regression models without covariates**

Four distinct MR based models are developed here to estimate the pooled log OR and its prediction interval. These four models are different with respect to the methods used to estimate the heterogeneity ($\hat{\tau}^2$) and the use of Knapp and Hartung modification. Knapp and Hartung modification uses the t-distribution in place of the standard normal distribution in calculating the p-values and confidence intervals of the summary measures (Knapp and Hartung 2003).

To facilitate a direct comparison of the summary estimates from MA models, none of these MR models include a covariate. All models employed the same existing evidence specified in table 8-7.
8.7.2.1 **Model 3: Meta-regression based on method of moments estimation and with Knapp and Hartung modification**

This MR model estimates the between study variance component ($\hat{\tau}^2$) using the method of moments estimation technique. Further, it incorporates the Knapp and Hartung modification into the model as well.

8.7.2.2 **Model 4: Meta-regression based on method of moments estimation and without Knapp and Hartung modification**

This model performs the MR using the between study variance component ($\hat{\tau}^2$) estimated via the methods of moments estimation technique. This model does not specify the Knapp and Hartung modification and thus separating from the model specified in section 8.7.2.1.

8.7.2.3 **Model 5: Meta-regression based on Residual Maximum Likelihood and with Knapp and Hartung modification**

This MR model uses the Residual Maximum Likelihood (reml) estimation of the between study variation ($\hat{\tau}^2$) as well as Knapp and Hartung modification of the variance of the estimated coefficients. This model is different with the model described in section 8.7.2.1 only with respect of the method of estimation of the heterogeneity.
8.7.2.4 **Model 6: Meta-regression based on Residual Maximum Likelihood and without Knapp and Hartung modification**

This model performs the MR using the Residual Maximum Likelihood estimation of heterogeneity and without Knapp and Hartung modification. This model is similar to the model specified in section 8.7.2.2 in all respects except from the methods of estimation of the heterogeneity.

8.7.3 **Critical evaluation**

Table 8-8 below summarises the estimates of pooled effect sizes and prediction intervals obtained from all variants of MA and MR models.

**Table 8-8: Summary table reporting the pooled effect size and prediction intervals of all models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Analysis method</th>
<th>Estimation method</th>
<th>Modification</th>
<th>Pooled log odds ratio</th>
<th>Lower prediction limit</th>
<th>Pooled odds ratio</th>
<th>Upper prediction limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Meta Analysis</td>
<td>Mantel Haenszel</td>
<td></td>
<td>1.4662</td>
<td>2.15</td>
<td>4.3328</td>
<td>8.73</td>
</tr>
<tr>
<td>2</td>
<td>Meta Analysis</td>
<td>Inverse Variance</td>
<td></td>
<td>1.4653</td>
<td>2.18</td>
<td>4.3289</td>
<td>8.61</td>
</tr>
<tr>
<td>3</td>
<td>Meta Regression</td>
<td>Method of moments</td>
<td>Knapp &amp; Hartung</td>
<td>1.4653</td>
<td>2.16</td>
<td>4.3289</td>
<td>8.66</td>
</tr>
<tr>
<td>4</td>
<td>Meta Regression</td>
<td>Method of moments</td>
<td></td>
<td>1.4653</td>
<td>2.18</td>
<td>4.3289</td>
<td>8.61</td>
</tr>
<tr>
<td>5</td>
<td>Meta Regression</td>
<td>Restricted Maximum Maximum likelihood</td>
<td>Knapp &amp; Hartung</td>
<td>1.4451</td>
<td>2.75</td>
<td>4.2421</td>
<td>6.53</td>
</tr>
<tr>
<td>6</td>
<td>Meta Regression</td>
<td>Restricted Maximum Maximum likelihood</td>
<td></td>
<td>1.4451</td>
<td>2.86</td>
<td>4.2421</td>
<td>6.28</td>
</tr>
</tbody>
</table>
The prediction intervals derived from the MA model with Mantel-Haenszel estimation of heterogeneity are slightly wider than that with inverse-variance estimation. In MR models, prediction intervals are wider when method of moments estimation is used instead of restricted maximum likelihood estimations. The prediction intervals estimated from MR models with the Knapp and Hartung modification are slightly wider than without the modification.

All estimates are approximately equal, except those estimated from MR models, which use restricted maximum likelihood estimation methods (both with and without Knapp and Hartung modification). The MA model with heterogeneity estimated using inverse variance method produces an identical pooled effect estimate and prediction intervals to that of MR model with heterogeneity estimated using method of moments and without Knapp and Hartung modification. In the absence of covariates, this MR model reduces to a MA model with inverse variance heterogeneity estimates.

In conclusion, within this proposed framework of methodology to design new trials the random effects MA with heterogeneity estimated from inverse variance methods (model 2) and the MR models with methods of moments and without Knapp and Hartung modification (model 4) should be employed, to avoid inappropriate comparisons and to produce consistent results.

The next section conducts the synthesis of existing evidence, with special focus on ascertaining how trialists could have benefited had they conducted a synthesis prior embarking on to designing the trial in concern.
8.8 **Phase 2: The synthesis of existing evidence**

8.8.1 **Objectives**

The objective of phase 2 is to assess the evidence available to researchers (specifically the authors of trials available in the dataset) and to envisage the preliminary synthesis which could have been conducted in the planning stage before embarking on their respective trials.

Besides helping to identify if their particular research is required in the first place, this evaluation helps in determining the extent to which researches could have been benefited, had a synthesis carried out before designing each respective trials.

The trials in the existing evidence in table 8-7 have been carried out over a span of ten years, i.e. between 2000 and 2009. This analysis is carried out at three time points within this span of time, which is analogous to the synthesis that trialists could have carried out at the planning stage of their trial, with respective existing evidence at hand. Three different time points during this ten-year span, i.e. year 2000, 2006 and 2009 are selected to represent the beginning, the middle and the end of the time span.

Section 8.8.2 - 8.8.4 conducts the synthesis of trials conducted before three time points.
8.8.2 Trials published before the end of year 2000

8.8.2.1 Existing evidence

The existing evidence consists of three trials published before the end of year 2000. The list of trials along with number of events and non-events in both arms, and corresponding LEV dose are presented in the table 8-9 below.

Table 8-9: The list of trials published before the end of year 2000.

<table>
<thead>
<tr>
<th>Trial</th>
<th>PLC Events</th>
<th>PLC Non Events</th>
<th>LEV Events</th>
<th>LEV Non Events</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereghino1 et al</td>
<td>3</td>
<td>44</td>
<td>36</td>
<td>62</td>
<td>1000</td>
</tr>
<tr>
<td>Cereghino2 et al</td>
<td>4</td>
<td>44</td>
<td>40</td>
<td>61</td>
<td>3000</td>
</tr>
<tr>
<td>Shorvon1 et al</td>
<td>3</td>
<td>52</td>
<td>22</td>
<td>84</td>
<td>1000</td>
</tr>
<tr>
<td>Shorvon2 et al</td>
<td>4</td>
<td>53</td>
<td>37</td>
<td>69</td>
<td>2000</td>
</tr>
<tr>
<td>Ben-Menachem et al</td>
<td>15</td>
<td>90</td>
<td>71</td>
<td>110</td>
<td>3000</td>
</tr>
</tbody>
</table>

All three-arm trials have been split as though they consist of two two-arm trials, which brings the total number of trials up to five.
8.8.2.2 Results

The resulting forest plot of the initial random effects MA is presented in figure 8-3 below.

Figure 8-3: Forest plot of the random effects meta-analysis of trials carried out before the end of year 2000.

<table>
<thead>
<tr>
<th>Study</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>cereghino_part1</td>
<td>8.52 (2.47, 29.41)</td>
<td>11.77</td>
</tr>
<tr>
<td>cereghino_part2</td>
<td>7.21 (2.40, 21.64)</td>
<td>14.98</td>
</tr>
<tr>
<td>shorvon_part1</td>
<td>4.54 (1.29, 15.92)</td>
<td>11.48</td>
</tr>
<tr>
<td>shorvon_part2</td>
<td>7.11 (2.38, 21.17)</td>
<td>15.16</td>
</tr>
<tr>
<td>Ben-menachem</td>
<td>3.87 (2.08, 7.22)</td>
<td>46.61</td>
</tr>
<tr>
<td>Overall</td>
<td>5.21 (3.40, 7.97)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis.

The forest plot in above figure 8-3 suggests that, if LEV dose is ignored, evidence is found suggesting that efficacy of LEV and PLC are significantly different (p=2.81e^{-14} of the test of OR=1). A zero $I^2$ is found suggesting the no evidence of the variability between studies (p=0.701).

The association between the log OR and corresponding LEV dose is modelled via the MR approach using summary level data. Figure 8-4 below displays the relationship between the log OR and the LEV dose.
Figure 8-4: The association between the log odds ratio and LEV dose of trials published before the end of year 2000.

The results of this MR analysis suggest the absence of evidence to reject the null hypothesis of no relationship between the dose and the log OR (P-value of 0.552) at the 5% significant level. The dose coefficient of -0.0001873 with the standard error of 0.000259 is statistically non-significant although the regression plot 8-4 above suggests a negative relationship.

Further trials comparing specific LEV doses with PLC are warranted because there is not clear evidence of an association between the efficacy and the LEV dose. Having said that, ethical implications of administering PLC to patients should be considered before designing PLC controlled trials. Had Tsai et al, who designed the next trial with the similar scope (table 8-7) performed a similar analysis, the need of his trial would have been confirmed at the beginning of the trial.
The non-significant association of the efficacy and the LEV dose recommends conducting further relevant trials. A very similar analysis is carried out for the trials published before the end of year 2006 in section 8.8.3 below.

8.8.3 Trials published before the end of year 2006

8.8.3.1 Existing evidence

Only one directly relevant trial comparing a LEV dose and PLC has been published between the years 2000 and 2006, and included in the analysis (Tsai, Yen et al. 2006). Table 8-10 below presents the number of patients with events and non-events in both LEV and PLC arms of trials carried out before the end of year 2006.

Table 8-10: The list of trials comparing LEV and PLC published before the end of year 2006.

<table>
<thead>
<tr>
<th>Trial</th>
<th>LEV</th>
<th>PLC</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Non Events</td>
<td>Events</td>
</tr>
<tr>
<td>Cereghino1 et al</td>
<td>36</td>
<td>62</td>
<td>3</td>
</tr>
<tr>
<td>Cereghino2 et al</td>
<td>40</td>
<td>61</td>
<td>4</td>
</tr>
<tr>
<td>Shorvon1 et al</td>
<td>22</td>
<td>84</td>
<td>3</td>
</tr>
<tr>
<td>Shorvon2 et al</td>
<td>37</td>
<td>69</td>
<td>4</td>
</tr>
<tr>
<td>Ben-Menachem et al</td>
<td>71</td>
<td>110</td>
<td>15</td>
</tr>
<tr>
<td>Tsai et al</td>
<td>20</td>
<td>26</td>
<td>5</td>
</tr>
</tbody>
</table>

Section 8.8.3.2 presents the results.
8.8.3.2 Results

Figure 8-5 below shows the resulting forest plot of the random effects MA performed using the trials conducted before the end of year 2006. This meta-analysis provides the relevant standard errors of the log OR estimates of each individual trial, which are necessary to conduct the MR.

*Figure 8-5: Forest plot of the meta-analysis of trials published before the end of year 2006.*

<table>
<thead>
<tr>
<th>Study</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>cereghino_part1</td>
<td>8.52 (2.47, 29.41)</td>
<td>10.22</td>
</tr>
<tr>
<td>cereghino_part2</td>
<td>7.21 (2.40, 21.64)</td>
<td>13.02</td>
</tr>
<tr>
<td>shorvon_part1</td>
<td>4.54 (1.29, 15.92)</td>
<td>9.98</td>
</tr>
<tr>
<td>shorvon_part2</td>
<td>7.11 (2.38, 21.17)</td>
<td>13.18</td>
</tr>
<tr>
<td>Ben-menachem</td>
<td>3.87 (2.08, 7.22)</td>
<td>40.51</td>
</tr>
<tr>
<td>Tsai</td>
<td>6.46 (2.16, 19.32)</td>
<td>13.10</td>
</tr>
<tr>
<td>Overall (I-squared = 0.0%, p = 0.804) with estimated predictive interval</td>
<td>5.36 (3.60, 7.96)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

The forest plot above reveals that individual trials produce statistically significant effect estimates. In the absences of LEV dose, the pooled effect estimate (log OR) is statistically significant as well (p<0.000) suggesting the efficacy of LEV over PLC. Zero $I^2$ suggests absence of between study variability. It is further confirmed by the result of the heterogeneity chi-squared test (p= 0.804). All tests were carried out at 5% significance level.
The log OR is meta-regressed with LEV dose to examine the extent to which OR is related to the LEV dose. Figure 8-6 below depicts the result of the meta-regression model.

**Figure 8-6**: The plot showing the relationship between the log odds ratio and the dose administered for the trials published before the end of year 2006.

The dose coefficient i.e. -0.0001973 (standard error 0.0002555) is statistically non-significant at the 5% significance level (p=0.483), implying no evidence against the null hypothesis of no relation between the efficacy and LEV dose. Therefore, the observed variation in the effect sizes is attributable purely due to chance. These findings are not conclusive in their own right and hence designing further trials with adequate power are required to inform the conclusions.

Upcoming section 8.8.4 conducts the same analysis including all available trials.
8.8.4 **Trials published before the end of year 2009**

8.8.4.1 **Existing evidence**

Two more two-arm trials comparing LEV 3000 with PLC are found in the literature published between 2006 and 2009, i.e. trials by Xiao and Wu. Table 8-11 below shows the number of patients with events and the non-events in both LEV and PLC arms of each trial.

*Table 8-11: Trials published before the end of year 2009.*

<table>
<thead>
<tr>
<th>Trial</th>
<th>LEV</th>
<th>PLC</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Non events</td>
<td>Events</td>
</tr>
<tr>
<td>Cereghino1</td>
<td>36</td>
<td>62</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1000</td>
</tr>
<tr>
<td>Cereghino2</td>
<td>40</td>
<td>61</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>Shorvon1</td>
<td>22</td>
<td>84</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1000</td>
</tr>
<tr>
<td>Shorvon2</td>
<td>37</td>
<td>69</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000</td>
</tr>
<tr>
<td>Ben-Menachem</td>
<td>71</td>
<td>110</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>Tsai</td>
<td>20</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000</td>
</tr>
<tr>
<td>Xiao</td>
<td>13</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>Wu</td>
<td>57</td>
<td>45</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
</tr>
</tbody>
</table>
8.8.4.2 Results

Figure 8-7 below displays the resulting forest plot of the MA conducted using all trials conducted before the end of year 2009.

**Figure 8-7: The forest plot of the meta-analysis of trials published until the year 2009**

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>cereghino-part1</td>
<td>8.52 (2.47, 29.41)</td>
<td>7.56</td>
</tr>
<tr>
<td>cereghino_part2</td>
<td>7.21 (2.40, 21.64)</td>
<td>9.35</td>
</tr>
<tr>
<td>shorvon_part1</td>
<td>4.54 (1.29, 15.92)</td>
<td>7.39</td>
</tr>
<tr>
<td>shorvon_part2</td>
<td>7.11 (2.38, 21.17)</td>
<td>9.45</td>
</tr>
<tr>
<td>Ben-menachem</td>
<td>3.67 (2.08, 7.22)</td>
<td>22.72</td>
</tr>
<tr>
<td>Tsai</td>
<td>6.46 (2.16, 19.32)</td>
<td>9.40</td>
</tr>
<tr>
<td>Xiao</td>
<td>1.34 (0.46, 3.87)</td>
<td>9.92</td>
</tr>
<tr>
<td>Wu</td>
<td>3.61 (1.99, 6.53)</td>
<td>24.21</td>
</tr>
<tr>
<td>Overall (I-squared = 17.9%, p = 0.288)</td>
<td>4.33 (3.02, 6.21)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

The effect size of the trial reported by Xiao et al is not statistically significant when the dose is not accounted for. The heterogeneity accounts for 17.9% of the total variation as suggested by the \( I^2 \) statistic (p=0.288).

Figure 8-8 below illustrates the association of the log OR and the LEV dose. The log ORs of eight individual trials are modelled with LEV doses administered in a MR setting to examine the association between them.
Figure 8-8: The plot illustrating the association of the log odds ratio and the LEV dose administered of trials published before the end of year 2009.

The dose coefficient, i.e. -0.0003531 is not found to be statistically significant at the 5% level (p = 0.173), indicating the availability of no evidence to reject the null hypothesis of no relationship of the log OR with administered LEV dose.

Hence, this analysis is not conclusive in its own right and further trials addressing the same therapeutic question is required. The findings from sections 8.8.2-8.8.4 are critically evaluated in upcoming section 8.8.5.

8.8.5 Critical evaluation

This section aims to compare and critically evaluate the findings from three MR analyses conducted, including the trials published before the end of year 2000, 2006 and 2009. None of the analyses reports a statistically significant dose coefficient, at the 5% significant level. This finding implies that the trials conducted up until now (2009) have failed to detect a clinically and statistically meaningful relationship between the LEV dose and the efficacy. Hence, this
result highlights the need to conduct adequately powered future clinical trials addressing the same therapeutic question.

The remainder of this chapter develops methodologies to design new trials with specific dose level, including single trial with two-arms and three-arms as well as multiple trials with two-arms, by taking above recommendations on board.

Upcoming section 8.9 develops the methodology to design a single new trial with specified covariate (both two-arm and three-arm), based on the power of the subsequent MR.

### 8.9 Phase 3: Methodology for designing a single new trial

This phase develops the methodology of designing a single new trial with a specified dose. The methodology features designing both two-arm and three-arm trials. The single trial is characterised by having a PLC arm, and any of the three LEV doses as the second arm and/or third arm.

#### 8.9.1 Designing a new two-arm trial with specific covariates

##### 8.9.1.1 Objectives

The focus of this section is to design a single two-arm trial, using MR principles. Of two arms, one is always the control arm i.e. the placebo (PLC) and the other arm being the treatment arm which could be any particular LEV dose.
8.9.1.2 Existing evidence

All available trials comparing any LEV dose and PLC in treating partial epilepsy, published until the year 2009 are included in the existing evidence to form the basis of the analysis. Table 8-12 below lists the number of patients with events and non-events in both LEV and PLC groups along with the dose of all relevant trials included in the existing evidence.

**Table 8-12: The list of trials constituting the existing evidence.**

<table>
<thead>
<tr>
<th>Year</th>
<th>Trial</th>
<th>LEV Events</th>
<th>LEV Non events</th>
<th>PLC Events</th>
<th>PLC Non events</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>cereghino1</td>
<td>36</td>
<td>62</td>
<td>3</td>
<td>44</td>
<td>1000</td>
</tr>
<tr>
<td>2000</td>
<td>cereghino2</td>
<td>40</td>
<td>61</td>
<td>4</td>
<td>44</td>
<td>3000</td>
</tr>
<tr>
<td>2000</td>
<td>shorvon1</td>
<td>22</td>
<td>84</td>
<td>3</td>
<td>52</td>
<td>1000</td>
</tr>
<tr>
<td>2000</td>
<td>shorvon2</td>
<td>37</td>
<td>69</td>
<td>4</td>
<td>53</td>
<td>2000</td>
</tr>
<tr>
<td>2000</td>
<td>Ben-menachem</td>
<td>71</td>
<td>110</td>
<td>15</td>
<td>90</td>
<td>3000</td>
</tr>
<tr>
<td>2006</td>
<td>Tsai</td>
<td>20</td>
<td>26</td>
<td>5</td>
<td>42</td>
<td>2000</td>
</tr>
<tr>
<td>2009</td>
<td>Xiao</td>
<td>13</td>
<td>15</td>
<td>11</td>
<td>17</td>
<td>3000</td>
</tr>
<tr>
<td>2009</td>
<td>Wu</td>
<td>57</td>
<td>45</td>
<td>26</td>
<td>74</td>
<td>3000</td>
</tr>
</tbody>
</table>

8.9.1.3 Methods

The methodological approach adopted in designing a new single trial with two-arms is very similar to that used in designing trials using MA methods described in section 5.6 in Chapter 5. However, as set out in the objectives this chapter adopts a methodology based on random effects MR principles in the power computations. The step-by-step illustration to the methodology is as below.
**Step 1**

A random effects MA is performed of the data from existing trials (Existing evidence) excluding the dose information, to estimate the treatment effects (log OR) and corresponding standard errors (standard error of log OR) of individual trials.

**Step 2**

A transformation of the resulting ORs to the log scale is required as the sampling distribution of the log OR more closely follows the normal distribution. A dataset is constructed to include the data of each trial (The log OR, corresponding standard error and the dose) suitable to perform the MR analysis on.

**Step 3**

The treatment arm of the new trial (e.g. LEV 1000, 2000 or 3000) is then specified, which is the only information available about the new trial at the initial stage of the process. The information about the specific dose of the new trial is then incorporated into the dataset constructed in step 2, leaving other corresponding data fields of that record representing the new trial blank (i.e. unknown at present).
Step 4

Using the new dataset, the log OR is modelled with the dose in a MR setting with the log OR being the dependant/response variable and the dose being the independent variable/covariate.

Step 5

Here we specify the mean and the standard error of the predictive distribution from which an effect size of the new two-arm trial is drawn, based on the fitted values of the resulting regression coefficient of the MR. The mean of the predictive distribution is set to be the fitted log OR at the specified dose whilst the standard error is set to be the standard error attached to the fitted value.

Step 6

The decision is made as to the number of subjects in each arm, or in other words, the sample size of the new trial. An observation (log OR of the new trial) is then sampled from the predictive distribution. The event rate in the treatment group is derived from the mean of the event rate in the control group in each trial. The event rate in the treatment group is calculated using formulae 5.1 specified in the section 5.5 of Chapter 5.

Step 7

The subsequent binomial simulation, in which the number of observations is specified to be equal to the sample size, returns the number of events and the non-events in both treatment (LEV) and control arms (PLC) of the new trial.
Once the total number of subjects, events and non-events of both treatment and control groups are established, the set of data of the new trial has been obtained.

**Step 8**

The new trial, along with the existing trials, is subjected to a MR, to assess the statistical significance of the resulting dose coefficient, at 5% significance level. The process of computing power is associated with the null hypothesis that no relationship exists between the log OR and the dose.

**Step 9**

A new effect size is drawn and the procedure described in step 6-8 is repeated for 5000 times and the result of the significant test in each iteration is recorded. The power of the updated MR for the sample size of the new trial specified in step 6, is then computed as the proportion of the significant p-values.

**Step 10**

The procedure specified in step 6- step 9 is then repeated by varying the sample size of the new trial to explore the variability of power to the changes of sample size.

The methods specified here have been applied in designing new trials comparing different LEV doses with PLC and the results are presented in upcoming section 8.7.1.4.
8.9.1.4 Prediction models

Three different LEV doses are commonly prescribed for treating partial epilepsy as an adjunctive therapy, viz. LEV 1000mg, LEV 2000mg and LEV 3000mg. This suggests designing three combinations of two-arm trials as listed below, one arm of which is always being a PLC controlled arm.

1. LEV 1000 vs PLC
2. LEV 2000 vs PLC
3. LEV 3000 vs PLC

8.9.1.4.1 Prediction model 16: Designing a new trial between PLC and LEV 1000

This prediction model develops the methodology to design a new trial comparing PLC and LEV 1000 using the methods specified in section 8.9.1.3 above. The power of the updated MR at various sample sizes of the new trial is recorded and shown in table 8-13 in section 8.9.1.6 below.

8.9.1.4.2 Prediction model 17: Designing a new trial between PLC and LEV 2000

This prediction model designs a new trial comparing PLC and LEV 1000.

8.9.1.4.3 Prediction model 18: Designing a new trial between PLC and LEV 3000

The two-arm trial comparing PLC and LEV 3000 is designed in this section.
8.9.1.5 Results

The behaviour of power of the updated MR is observed with the changes in sample sizes of the new trial. The resulting power curves of all three two-arm combinations are presented in figure 8-9 below.

Figure 8-9: Power curves of the new two-arm trials between LEV 1000, 2000 and 3000 with PLC.

<table>
<thead>
<tr>
<th>Prediction model 16: Power of the updated MR after adding a new trial of LEV 1000 and PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A gradual increase in power is observed as the sample size of the new trial increases. Even at relatively low sample sizes (e.g. 100) of the new trial (one arm), the power of the updated MR is relatively high. The rate of power increase is falling after the sample size of the new trial reaches 1000.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 17: Power of the updated MR after adding a new trial of LEV 2000 and PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>The power remains zero for lower sample sizes, until the sample size in one arm reaches close to 500. Thereafter the power gradually increases with increased sample sizes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 18: Power of the updated MR after adding a new trial of LEV 3000 and PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>The power of the updated MR turns out to be zero for all sample sizes of the new trial. This is an interesting pattern triggering further investigations. An investigation is carried out in the section 8.9.1.7 to examine the underlying causes of this behaviour.</td>
</tr>
</tbody>
</table>
To facilitate an effective comparison of the results found thus far, a combined table and power curve are presented in section 8.9.1.6 below.

8.9.1.6 Extrapolation

The power of the updated MR at various sample sizes of the new trial comparing LEV 1000, 2000 and 3000 doses computed in previous section 8.9.1.4.1- 8.9.1.4.3 are combined here to aid the comparison. Although these three LEV doses (i.e. LEV 1000, 2000 and 3000) are commonly prescribed in practice, to learn the underlying causes of the observed behaviour of power, an extrapolation of the data is carried out to include LEV doses in between the commonly prescribed doses such as 2500, 3500, and extreme doses such as 4000 and 5000. The extrapolation aids in understanding the underlying behaviour of power.

Table 8-13: The power of the updated meta-regression including the new trial that consists of a PLC arm and one of the many LEV doses listed below as the second arm.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Power of the updated meta-regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One Arm</td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2 le</td>
</tr>
<tr>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>800</td>
<td>1600</td>
</tr>
<tr>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>2000</td>
<td>4000</td>
</tr>
<tr>
<td>3000</td>
<td>6000</td>
</tr>
<tr>
<td>4000</td>
<td>8000</td>
</tr>
<tr>
<td>5000</td>
<td>10000</td>
</tr>
</tbody>
</table>
The process of choosing LEV doses used in the analysis other than the commonly prescribed doses was done in a systematic way. Dose of LEV 2500 is first selected, which is the average of LEV 2000 and LEV 3000. Since the power remains zero of new trials with LEV 2500 arm, for all sample sizes concerned, LEV 2250 is then used, which is the midpoint of LEV 2000 and 2500, and still observes zero power at all sample sizes considered. The aim here is to detect the LEV dose at which the power becomes non-zero for any sample size in the range considered. The midpoint of LEV 2000 and LEV 2250, which is LEV 2125 is then used, and noticed that at larger sample sizes, non-zero power is observed.

Treatment arms with LEV 4000 and 5000 are then used to examine the behaviour of power and noticed power is relatively higher even at small sample sizes. Treatment arms with LEV 3500 and LEV 3250 are then assessed to identify at which point power becomes positive. Figure 8-10 below represents the corresponding combined power curves.
Section 8.9.1.7 below critically evaluates the findings of this section.

### 8.9.1.7 Critical evaluation

The results decisively confirm that the power of the updated MR varies with the LEV dose. The highest (LEV 5000) and the lowest (LEV 1000) doses report the highest power. LEV doses of 2250, 2500, 2750 and 3000 produce zero power at all sample sizes considered. Moreover, LEV doses 2125, 2000 and 3250 report zero power at smaller sample sizes of the new trial.

It becomes apparent that covariate values further from the weighted mean covariate value are likely to produce greater power than those lie closer to the weighted mean covariate.
In an attempt to investigate the reasoning for the observed power behaviour, upcoming sections 8.9.1.7.1 - 8.9.1.7.3 below explore the relationships between two of the three elements described in turn below.

**The leverage of the new trial**

This is the influence exerted by the new trial on the predicted value of the new regression line at the specific covariate.

**The predicted effect size**

This is the predicted effect size in the updated MR, i.e. the fitted log OR of the updated MR at the covariate of the new trial.

**The p-value of the significant test**

This is the p-value of the updated MR, based on which the power is computed.

For the convenience, the detailed analysis is carried out using 500 iterations and assuming 1000 subjects in each arm of the new trial.

**8.9.1.7.1 Exploring the association between the predicted effect size and the leverage**

The relationship between the predicted effect size in the updated MR and the leverage in each observation is assessed and shown in figure 8-11 below.
Figure 8-11: The association between the leverage of the new trial and the predicted effect size (log OR) in the updated meta-regression following the inclusion of the new trial.

Above analysis assists in examining how the leverage fluctuates with covariates. The maximum leverage exerted by the new trial is always higher whenever the new trial involves LEV doses away from the weighted mean covariate. The leverage is lowest when the new trial includes a covariate near the weighted mean. The maximum leverage is exerted when the new trial includes a covariate of LEV 5000.

A larger proportion of observations have the highest leverage in the LEV 2500 scenario, compared other scenarios where the proportion of observations with the highest leverage is diminishing as the covariate deviate from the weighted mean.
The range within which the predicted effect size fluctuates is narrowest in LEV 2500 scenario. This range begins to widen as the covariate moves away from the weighted mean. The widest range within which the leverage fluctuates is seen when LEV 2500 is involved. As the covariate moves away from the weighted mean this range begins to narrow down.

The figure below combines the leverage of the new effect size for various LEV doses.

*Figure 8-12: Combined curve between the leverage of the new trial and LEV dose.*

The above plot assists in examining the spread of the leverage of a new trial on its predicted value. The spread is widest for covariates near the weighted mean and begin to narrow down as the covariate moves away from the weighted mean.

Moreover, the minimum and the maximum leverage for each covariate increase as the covariate moves away from the weighted mean. The next section explores the association between the leverage and the p-value.
8.9.1.7.2 Exploring the association between the leverage and the p-value

The aim here is to assess the impact of the leverage of the new trial on the significance of the subsequent MR. To this end, this section examines the association between the leverage of the new observation and the p-value of the subsequent MR following the addition of a new trial. A line is drawn across the y-axis when the p-value is at 0.05, to differentiate significant p-values.

*Figure 8-13: The association of the p-value and the leverage.*

More than one p-value is often associated with a particular leverage point. The number of p-values associated at the highest leverage point is larger. Three noteworthy characteristics of the behaviour of the p-value and the leverage are noticed as the LEV dose moves away from the weighted mean covariate.
Firstly, the range within which the p-value varies is widening. Secondly, the highest leverage in each LEV dose is increasing. Thirdly, the range within which the leverage fluctuates is narrowing down.

If the LEV dose closer to the weighted mean covariate is selected, none of the p-values falls into the significant region. As the LEV dose moves away from the weighted mean covariate, the number of p-values falling into the significant region increases.

The association between the predicted effect size and the p-value is explored in upcoming section.

### 8.9.1.7.3 Exploring the association between the predicted effect size and the p-value

The aim of this section is to assess the relation between the predicted effect size in the updated MR and corresponding p-value. To this end, changes to the p-value and the predicted effect size of the subsequent MR are observed and plotted in figure 8-14 below.
**Figure 8.14:** The association between the p-value and the predicted effect size in the updated meta-regression for all LEV doses considered.

In summary, for lower LEV doses, the higher the predicted log OR the lower the p-value. For higher LEV doses, the lower the predicted log OR the lower the p-value. As expected, the leverage associated with covariates further from the weighted mean is also higher.

As above figure indicated, predicted log ORs below a certain limit, i.e. 1.04 in the case of LEV 5000 and 1.07 in the case of LEV 4000, are always found statistically significant. Conversely, predicted log ORs above a certain limit, i.e. 1.82 in the case of LEV 1000 and 1.78 in the case of LEV 2000, are always associated with significant p-values. The power of the updated MR is correlated with the gradient of the regression line yield in the updated MR. The predicted log ORs beyond these thresholds are associated with steeper gradients of the
regression line, which in turn contributes towards significant p-values and positive power.

8.9.1.7.4 Exploring the association between the predicted effect size in the updated MR and the dose

It is vital to understand the association of the predicted effect size of the updated MR and the covariate in investigating the power behaviour observed in table 8-13. Figure 8-15 below shows the relationship between the predicted effect size and the dose. The regression line of the existing MR is also included in this figure to support the comparison.

Figure 8-15: The predicted effect size in the updated MR for various LEV doses and the regression line in the existing meta-regression.
The range within which the predicted log OR fluctuates becomes wider as the covariate moves away from the weighted mean. When the new trial includes LEV dose of 2500, this range reaches its narrowest level.

8.9.1.7.5 Summary

Power of the existing meta-regression

It has already been observed (section 8.8.4) that the regression line in the existing MR produces a negative gradient (-0.0003531). Moreover, the regression slope is not statistically significant (p-value of 0.173), implying that the confidence interval (-0.000912 to 0.000206) of the regression coefficient includes zero. Figure 8-16 below shows the regression line of the existing MR including the confidence and prediction intervals.

Figure 8-16: Existing meta-regression including the confidence and prediction intervals.
As expected, due to approximately zero heterogeneity estimate, the resulting prediction interval coincides with the confidence interval at all doses and prediction effects including the random effects lie on the regression line. The confidence/prediction interval narrows down as the dose increases because of the decreased standard error in the estimation of dose coefficient.

**Power of the updated meta-regression**

To be able to produce power in the subsequent MR, the inclusion of the new trial should influence to an extent that the resulting regression line becomes steeper than that in the existing MR. The updated MR is different to the existing MR in few aspects. Firstly, the predicted effects in the updated MR are not as the same as that fitted in the existing MR. Moreover, the proportion of weights assigned to each trial in the weighted regression is changing as well. With the inclusion of a new trial, the proportion of weights assigned to each existing trial declines as well.

**Two arm trial of LEV 1000 and PLC**

When a new trial with LEV 1000 is added into the existing MR, the predicted values at LEV 2000 and LEV 3000 do not change as much. The predicted effect sizes at these covariates change slightly because of the change in the proportion of weights assigned to individual trials at those covariates. However, the inclusion of the new trial cause to a substantial change in the fitted log OR at LEV 1000. As seen in figure 8-15, the predicted log OR of the updated MR in LEV 1000 scenario varies in a wider range. Moreover, the new trial exerts a greater leverage on its predicted value as well (figure 8-11). The wider
fluctuation of the predicted log OR allows the regression line of the updated MR to take a wider range of gradient values. This causes the regression line to be steeper more often than not, in a given simulation. Hence, the power of the updated MR following the inclusion of a two arm trial of LEV 1000 and PLC is higher for all sample sizes considered.

**Two arm trial of LEV 2000/ PLC or LEV 3000/PLC**

When a new trial of LEV 2000 or LEV 3000 is added into the existing MR, the range within which the predicted effect size of the updated MR fluctuates is narrower compared to that in the case of LEV 1000 (figure 8-15). Moreover, the predicted effect sizes of the updated MR are not far off from those of the existing MR. This is due to the lighter leverage exerted by the new trial on its fitted value. In this case, the range of values that the gradient of the regression line of the subsequent MR takes is smaller in comparison to that after adding a new trial with a covariate further from the mean. This could limit the potential values that the gradient takes and consequently the possibility of including the null value in the confidence interval of the regression coefficient is increasing. These factors collectively affect the phenomenon of observing zero or nearly zero power, whenever a new trial with a covariate closer to the mean is added in to the existing MR.

Thus far, the power implications of a new single two-arm trial have been explored. The upcoming section 8.9.2 develops the methodology to design a new single trial of three arms. Resulting power of the new trial when it is added to the existing MR is reported against the sample size of the new trial.
8.9.2  **Phase 4 : Designing new three-arm trials using meta-regression methods**

8.9.2.1  **Objectives**

It is of interest to explore the impact of the inclusion of a three-arm trial into the MR on the power of the updated MR. In the clinical perspective, it is vital to allocate the limited numbers of patients available into a trial that is most likely to detect the treatment effect, if it exists.

The aim here in this section is to develop a methodology to design three-arm trials using existing evidence via MR model. The three-arm trial is designed in such a way that it features a common placebo arm and two remaining arms being any two of three available LEV doses. This is an attempt to evaluate whether it is beneficial to design three-arm trial instead of two-arm trials, based on the power of the updated MR including the new trial.

8.9.2.2  **Existing evidence**

Throughout this chapter, the same set of existing trials is used to base the development of the methodology on which is tabulated in table 8-12 above.

8.9.2.3  **Methods**

A three-arm trial is designed as if it constitutes two separate two-arm trials. This can be seen as the exact reversal of trial splitting mechanism carried out in section 8.5.4 to reduce three-arm trials into two two-arm trials.
The majority of the mechanism used in designing three-arm trials is similar to that of designing two-arm trials which is detailed in section 8.9.1.3, although the details exclusive to design three-arm trials are described in this section.

**Step 1 and 2**

Steps 1 and 2 in this procedure are as same as those described in section 8.9.1.3 of this chapter.

**Step 3**

This procedure requires specifying two covariates representing two covariates compared in the new trial along with PLC. Thus, two new rows representing two new trials are incorporated into the existing dataset. At this stage, only the covariates are known regarding the new trials but not the log ORs and their standard errors. Therefore, two fields in the dataset pertaining to the unknown variables are left black of the new two trials.

**Step 4**

Using the dataset, a MR is carried out to model the effect size (Dependant variable) and the dose (Predictor variable) to predict the fitted values and the standard errors of the regression line at specified new covariates.
**Step 5**

The fitted effect sizes and their standard errors are subsequently treated as the mean and the standard error of the predictive distribution from which the effect size of the new trials are drawn.

**Step 6**

At this design stage, two new trials look unbalanced in terms of the sample size, as sample sizes in the treatment arms in each of two new trials are as twice as much as the sample size in the control arm. However, if combined, these two new two-arm trials form a three-arm trial with equal number of subjects in each arm.

**Step 7**

Effect sizes representing new two-arm trials are drawn from each predictive distribution. Event rates in treatment arms of each trial are then computed using the equation 5.1 specified in section 5.5 of Chapter 5. The number of patients with events and non-events in both trials are simulated using binomial distribution as specified in equation 5.2 and 5.3 of Chapter 5. Once the number of patients with events and non-events are known this is deemed as the data set of new trials have been obtained.

**Step 8**

The power of the MR is based on the null hypothesis that states the dose coefficient is zero or the confidence interval of the dose coefficient includes
zero. The dataset including two new trials is then meta-regressed and the statistical significance of the test is evaluated in each simulation.

**Step 9**

This procedure (step 7-8) is iterated for a large number of times (5000 in this instance) and the power is then computed as the proportion of significant number of p-values. The whole procedure (step 3-9) is then repeated, by varying the sample size of the new trial and the power at each sample size is then determined.

The next section presents the power of the updated MR including the new three-arm trial at various sample sizes.

**8.9.2.4 Prediction models**

Three combinations of PLC controlled three-arm trials could be designed using available three doses of LEV as listed below.

1. PLC with LEV 1000 and LEV 2000
2. PLC with LEV 1000 and LEV 3000
3. PLC with LEV 2000 and LEV 3000

Upcoming sections 8.9.2.4.1 - 8.9.2.4.3 describe the process of design of each three-arm trial in turn.
8.9.2.4.1 Prediction model 19: Designing a new trial comparing PLC and LEV 1000 with 2000

The three arm trial comparing PLC with LEV 1000 and 2000 is designed using the methods specified in section 8.9.2.3 above and the power of the updated MR including the new three-arm trial is computed for various sample sizes of the new three-arm trial, and presented in table 8-14 below.

8.9.2.4.2 Prediction model 20: Designing a new trial comparing PLC and LEV 1000 with 3000

This prediction model designs a new three-arm trial comparing LEV 1000 and 3000 with PLC.

8.9.2.4.3 Prediction model 21: Designing a new trial comparing PLC and LEV 2000 with 3000

A new three-arm trial comparing LEV 2000 and 3000 with PLC is designed in this prediction model.

8.9.2.5 Results

The power results of the updated MR after adding the new three-arm trial for all combinations of three-arm trials are summarised in table 8-14 to assist in the explanation and to aid the comparison.
Table 8.14: The power of the updated meta-regression following the addition of three-arm PLC controlled trials.

<table>
<thead>
<tr>
<th>Sample size (each arm)</th>
<th>Prediction model 19 PLC with LEV1000 and LEV2000</th>
<th>Prediction model 20 PLC with LEV1000 and LEV3000</th>
<th>Prediction model 21 PLC with LEV2000 and LEV3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>26.30</td>
<td>30.14</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>43.00</td>
<td>50.06</td>
<td>5.48</td>
</tr>
<tr>
<td>300</td>
<td>49.64</td>
<td>56.50</td>
<td>18.00</td>
</tr>
<tr>
<td>500</td>
<td>55.34</td>
<td>65.86</td>
<td>30.08</td>
</tr>
<tr>
<td>800</td>
<td>58.16</td>
<td>71.58</td>
<td>42.52</td>
</tr>
<tr>
<td>1000</td>
<td>59.82</td>
<td>73.58</td>
<td>47.40</td>
</tr>
<tr>
<td>2000</td>
<td>61.54</td>
<td>79.08</td>
<td>61.02</td>
</tr>
<tr>
<td>3000</td>
<td>62.12</td>
<td>82.02</td>
<td>65.08</td>
</tr>
<tr>
<td>4000</td>
<td>64.02</td>
<td>82.96</td>
<td>68.34</td>
</tr>
<tr>
<td>5000</td>
<td>65.16</td>
<td>83.96</td>
<td>69.82</td>
</tr>
</tbody>
</table>

Figure 8.17 below shows the power curve in each individual prediction model.
**Figure 8-17**: The power curve of the three-arm trial designed using prediction model 19, 20 and 21.

<table>
<thead>
<tr>
<th>Prediction model 19: Power after adding a three arm trial of LEV 1000 and 2000 with PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>The power is comparatively high even at relatively small sample sizes of the new trial. A gradual increase in power is noticed as the sample size increases.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 20: Power after adding a three arm trial of LEV 1000 and 3000 with PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>The power of the updated MR increases as the sample size of the new trial increase. A relatively high power is reported even at lower sample sizes of the new trial.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction model 21: Power after adding a three arm trial of LEV 2000 and 3000 with PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>The power is zero when the sample size of the new trial is as small as 100 per arm, and gradually increases as the sample size increases. The rate at which power increases is falling as the sample size increases. Within this region of interest, the power does not reach a plateau.</td>
</tr>
</tbody>
</table>

Figure 8-18 below combines power curves of updated MR of all three-arm trials.
Figure 8-18: Comparative power curves of three-arm trials.

8.9.2.6 Critical evaluation

This section explored the power implications of adding a new three-arm trial into the existing MR. As mentioned previously a three-arm trial is designed as if it consists of two two-arm trials. This fact is fundamental in the interpretation and evaluation of results obtained. The PLC arms in two trials are assumed to be independent to avoid the need to account for the correlation between the PLC arms in two trials although they are derived from the same three-arm trial.

8.9.2.6.1 Key findings

The subsequent MR reports the highest power when a new trial with LEV 1000 and 3000 is added into the existing MR. The power is always greater when a new trial with LEV 1000 is involved than when a new trial with either LEV 2000 or 3000 is involved, for sample sizes below 2000.
In the updated MR the power is zero for sample sizes below 100.

8.9.2.6.2 **Factors influencing the power of the updated meta-regression after including a three-arm trial**

The scenario of three-arm trials is different to that of two-arm trials in that the uncertainty in the estimation of residual heterogeneity is reduced because of the increased number of trials in the analysis. Moreover, depending on the size of the new trials (hence the weights assigned to them) the predicted effect sizes of the updated MR at both covariate values are changed as well.

**Adding a new three arm trial of PLC with LEV 1000 and 3000**

In this proposed methodology, adding a new three arm trial of PLC with LEV 1000 and 3000 means indirectly adding two new two arm trials of PLC vs. LEV 1000 and PLC vs. LEV 2000 into the existing MR.

In this specific scenario, the confidence interval of the regression coefficient or the gradient of the regression line connecting the predicted mean log ORs at LEV 1000 and LEV 3000 is steeper in more simulations than that connecting the predicted mean log ORs at LEV 2000 and 3000 and that connecting LEV 2000 and 1000. Steeper gradients contribute to a gain in power.

**Adding a three-arm trial of PLC with LEV 1000 and 2000**

Due to the less dispersion in predicted log ORs involving LEV 2000, the inclusion of a new three-arm trial involving LEV 2000 makes the gradient of the regression line connecting LEV 2000 with either LEV 1000 or 3000 less steep in
a larger proportion of simulations. Moreover, the close proximity of LEV 2000 with both LEV 1000 and 3000 has equally caused the gradient of the regression line to become less steep whenever LEV 2000 is involved.

Given approximately equal distance from the weighted mean of the covariates (2439) LEV 2000 and LEV 3000 exert approximately equal leverages in the regression analysis. That is the influence of the two new trials on the predicted log OR at LEV 2000 and at LEV 3000 is significantly less than that in the case of LEV 1000. Due to the distance from the weighted mean the leverage exerted when involving LEV 1000 is greater than both LEV 2000 and 3000. Hence trials involving LEV 1000 tend to produce more power in the updated MR at least when the sample size is below 2000.

The close proximity of LEV 1000 and 2000 does not let the gradient of the regression line to take a wide range of values as in the case of LEV 1000 and 3000. Hence the power after including a three arm trial for LEV 1000 and 3000 is always the highest amongst them.

**Adding a three-arm trial of PLC with LEV 2000 and LEV 3000**

The subsequent MR following the inclusion of a trial comparing LEV 2000 and 3000 reports the lowest power. Given the close proximity between LEV 2000 and 3000, a small change in the fitted log ORs does not cause a considerable difference in the gradient connecting the new fitted log ORs. Moreover, the leverage exerted by adding LEV 2000 and 3000 is not greater than that exerted by LEV 1000, due to their proximity to the weighted mean of the dose
covariates (2439). These two factors collectively cause the subsequent MR to produce the lowest power following the inclusion of LEV 2000 and 3000 trial.

8.9.2.6.3 **Recommended trial designs**

Whenever LEV 3000 is involved, designing a three-arm trial is preferred over two-arm trial, because the updated MR including the new two-arm trial has shown no power in any sample size considered (figure 8-10). However, three-arm trials involving LEV 3000, when added into the existing MR, show increased power. Due to the increased power even at smaller sample sizes, designing a three-arm trial involving LEV 1000 and 3000 is preferred over a trial involving LEV 2000 and 3000.

When designing a trial involving LEV 2000 a three-arm trial with LEV 1000 is preferred over a placebo controlled two-arm trial, if the available sample size is smaller. For larger sample sizes available, whenever possible LEV 1000 should be incorporated in the three-arm trial with LEV 2000 to obtain substantial power, although the difference between these two types of trials is not substantial.

Designing a two-arm trial is preferred over a three-arm trial, provided the available sample size is below 1500, when LEV 1000 is involved. If available sample is approximately over 1500, designing a three-arm trial with LEV 3000 has shown to produce increased power.

8.9.2.6.4 **Future work**

Moreover, further investigations are required to explain the reasons for yielding slightly higher power in the subsequent MR after adding a trial comparing LEV
2000 and 3000 than when the trial comparing LEV 1000 and 3000, when the sample size goes beyond 2000.

This process is simplified in that adding two new two-arm trials is assumed similar to that of adding a new three-arm trial into the existing MR. However, the impact of the correlation of treatment effect in three-arm trial on the power of the updated MR is not considered.

The impact of adding a new single trial (both two-arm and three-arm trials) on power of the subsequent MR has been investigated in preceding sections 8.9.1 and 8.9.2. It is also of interest to examine power implications when multiple trials are added to the existing MR. Hence, methods are developed to design two and five new two-arm trials and to examine their power implications. Sections 8.10.1 and 8.10.2 develop the methodology to design two trials and five trials respectively, by way of illustration.

**8.10 Phase 5: Methodology to design multiple new trials**

The number of available patients for a trial is often deemed a scarce resource and perhaps limited in certain occasions such that they should be allocated with utmost care. The objective in designing multiple trials is to discover better approaches of allocating available patients across trials. It is of interest to discover whether the optimum allocation of patients is to design a large trial or several smaller trials.
8.10.1 Designing two trials of two arms using meta-regression methods

8.10.1.1 Objective
This section intends to design two two-arm trials, and explore the impact of power following the addition of two trials into the existing MR. The focus is to establish whether designing two trials instead of a single trial from available patients, would bring about additional benefits, such as identifying a small but clinically meaningful effect, if one exists.

8.10.1.2 Existing evidence
Throughout this chapter the same set of trials listed in table 8-7 is used to form the basis of the analysis.

8.10.1.3 Methods
The methods employed in designing two trials of two-arms are very similar to that of designing three-arm trials, which are in fact modelled as two trials of two arms. The only difference is in specifying the covariate (dose). Designing two trials of two arms requires specifying only a single covariate because two trials are assumed to be of the same type.
8.10.1.4 Prediction models

Three prediction models are developed in this section of designing two new trials of two-arms between PLC and LEV as listed below.

1. PLC and LEV 1000
2. PLC and LEV 2000
3. PLC and LEV 3000

8.10.1.4.1 Prediction model 22: Designing two new trials comparing PLC and LEV 1000

Two new two-arm trials comparing PLC and LEV 1000 are designed by employing the methods specified in section 8.9.2.3 and added to the existing MR, and the power of the updated MR is listed in table 8-15 below in section 8.10.1.5.

8.10.1.4.2 Prediction model 23: Designing two new trials comparing PLC and LEV 2000

This prediction model develops the methodology to design two new two-arm trials between PLC and LEV 2000.

8.10.1.4.3 Prediction model 24: Designing two new trials comparing PLC and LEV 3000

Two new two-arm trials between PLC and LEV 3000 is designed in this prediction model.
8.10.1.5 Results

The power of the subsequent MR after adding two new two-arm trials, for all three combinations of new trials is presented in table 8-15.

Table 8-15: A comparison of power of the subsequent meta-regression following the insertion of two new trials.

<table>
<thead>
<tr>
<th>Sample size (All available)</th>
<th>PLC and LEV 1000</th>
<th>PLC and LEV 2000</th>
<th>PLC and LEV 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>21.30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>240</td>
<td>40.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>360</td>
<td>46.86</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>480</td>
<td>50.92</td>
<td>0.82</td>
<td>0</td>
</tr>
<tr>
<td>720</td>
<td>54.66</td>
<td>4.70</td>
<td>0</td>
</tr>
<tr>
<td>960</td>
<td>55.32</td>
<td>7.72</td>
<td>0</td>
</tr>
<tr>
<td>1920</td>
<td>54.74</td>
<td>12.16</td>
<td>3.06</td>
</tr>
<tr>
<td>2880</td>
<td>53.10</td>
<td>13.90</td>
<td>3.40</td>
</tr>
<tr>
<td>3840</td>
<td>52.78</td>
<td>14.16</td>
<td>4.10</td>
</tr>
<tr>
<td>4800</td>
<td>51.82</td>
<td>14.94</td>
<td>4.82</td>
</tr>
</tbody>
</table>

Figure 8-19 below presents the individual power curves of prediction models 22-24.
Figure 8-19: The power curves after two new two-arm trials between PLC and LEV 1000, 2000 and 3000 are added to the existing meta-regression.

Prediction model 22: Power of updated MR after adding two new trials of PLC & LEV 1000

The power of the updated MR after adding two new trials comparing PLC and LEV 1000 declines as the total available sample size exceeds 960. Further investigations are required before arriving at any conclusions. However, the power of the updated MR after adding two new trials is lower than that after adding one new trial with the same dose (table 8-13), for all sample sizes considered.

Prediction model 23: Power of updated MR after adding two new trials of PLC & LEV 2000

The power of the updated MR remains zero until the available total sample size reaches above 400. A gradual increase in power is observed as available total sample size increases above 480. The power reaches a maximum of 14.9 when the available sample size reaches 4800.

Prediction model 24: Power of updated MR after adding two new trials of PLC & LEV 3000

The power of the updated MR is zero until available sample size reaches 1900. Relatively low power is reported even for large sample sizes.

The figure 8-20 below, presents the combined power curve.
Figure 8-20: Combined power curves of the subsequent meta-regression following inclusion of two new two-arm trials.

This section explores the power of the subsequent MR following the addition of two same sized two-arm new trials. The x-axis of above scatter plot in figure 8-20 represents the total available sample size.

In respect of the subsequent MR including new trials comparing LEV 3000 and PLC, the power becomes non-zero after sample size reaches 1000 as opposed to power remaining zero for all sample sizes in the context of adding a single new trial (figure 8-10).

The subsequent MR following the insertion of two new LEV 1000 trials produced the highest power followed by LEV 2000 and LEV 3000, at all sample sizes considered. This very same trend is observed in the power of the subsequent MR after addition a single new trial as seen in figure 8-10 of section 8.9.1.6.
The power results of the subsequent MR having added two new two-arm trials to the existing MR are proven to be encouraging, particularly in 3000 LEV dose.

It is of interest to explore power implications of subsequent MRs after adding more than two trials. To this end, section 8.10.2 below explores designing five new two-arm trials.

### 8.10.2 Phase 6: Designing five trials using meta-regression methods

#### 8.10.2.1 Objectives

This section focuses on designing five new two-arm trials and updating the existing MR including five new trials, to examine the extent to which the power of the updated MR fluctuates with available sample size.

The secondary objective is to discover whether the practice of designing multiple trials (two trials and five trials) from available sample size, instead of designing a single trial is associated with increased power.

#### 8.10.2.2 Existing evidence

The same set of trials listed in table 8-7 is employed as the existing evidence throughout this chapter.

#### 8.10.2.3 Methods

The methods employed are exactly similar to that used in designing two two-arm trials, apart from designing five new two-arm trials.
8.10.2.4 Prediction models

Three prediction models are developed in this section to design five new two-arm trials based on the power of the updated MR using existing evidence. Each prediction model belongs to one of the three combinations listed below.

1. PLC and LEV 1000
2. PLC and LEV 2000
3. PLC and LEV 3000

Each prediction model is described in turn in upcoming sections 8.10.2.4.1-8.10.2.4.3.

8.10.2.4.1 Prediction model 25: Designing five new trials comparing PLC and LEV 1000

Five placebo controlled two-arm trials with the second arm being LEV 1000 are designed based on the methods specified in section 8.9.2.3 and using existing evidence listed in table 8-7. The power of the updated MR is evaluated and presented in table 8-16 below.

8.10.2.4.2 Prediction model 26: Designing five new trials comparing PLC and LEV 2000

This section develops the methodology of designing five new two-arm trials comparing PLC with LEV 2000.
8.10.2.4.3 Prediction model 27: Designing five new trials comparing PLC and LEV 3000

Prediction model 27 develops the methodology to design five new two-arm trials between PLC and LEV 3000.

8.10.2.5 Results

The power of the updated MR after adding five new trial comparing PLC with any of three LEV doses, are presented in table 8-16.

Table 8-16: The power of the updated meta-regression after adding five new two-arm trials between PLC and any of LEV 1000, 2000 or 3000.

<table>
<thead>
<tr>
<th>Sample size (Total)</th>
<th>Power of the updated meta-regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLC and LEV 1000</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>21.16</td>
</tr>
<tr>
<td>240</td>
<td>37.42</td>
</tr>
<tr>
<td>360</td>
<td>47.76</td>
</tr>
<tr>
<td>480</td>
<td>53.64</td>
</tr>
<tr>
<td>720</td>
<td>58.78</td>
</tr>
<tr>
<td>960</td>
<td>61.32</td>
</tr>
<tr>
<td>1920</td>
<td>61.90</td>
</tr>
<tr>
<td>2880</td>
<td>60.84</td>
</tr>
<tr>
<td>3840</td>
<td>58.38</td>
</tr>
<tr>
<td>4800</td>
<td>57.76</td>
</tr>
</tbody>
</table>

Figure 8-21 below describes each power curve in detail.
**Figure 8-21:** The power curve after adding five new two-arm trials between PLC and LEV 1000, 2000 and 3000 are added to existing meta-regression.

**Prediction model 25:** Power of updated MR after adding five new trials of PLC & LEV 1000

The power of the updated MR increases drastically as the sample size of the new trial increases. However, for available total sample size beyond 960, the power is beginning to decline with increased sample size.

**Prediction model 26:** Power of updated MR after adding five new trials of PLC & LEV 2000

The power of the updated MR at any stage is not as large as when a new trial of LEV 1000 and PLC is added into. In fact, the power remains zero when all available sample size is 120. Thenceforth, an increase in power is observed as the sample size increases, although the rate at which the power increases is dropping as the size of the new trial increases. Markedly the power is dropping after the sample size increases beyond 2880.

**Prediction model 27:** Power of updated MR after adding five new trials of PLC & LEV 3000

The power remains very low even for very larger total sample sizes, i.e. below 4%. When the available total sample size is 120, the power remains zero. A gradual increase of power is seen thenceforth. However, the pattern of the power when the sample size goes beyond 1920 requires further investigations.

Figure 8-22 below combines all three individual power curves representing prediction models 25-27.
The power of the subsequent MR after adding five new trials is highest for new trials comparing PLC with LEV 1000 followed by LEV 2000 and LEV 3000. This trend is in accord with that has been observed after adding a single (section 8.9.1) and two new two-arm trials (section 8.10.1) into the existing MR. As was observed in the subsequent MR after adding two new trials, positive power is seen following the inclusion of five new LEV 3000 trials as well.

Thus far in this chapter the power of the updated MR after including two-arm trials of three different doses in relation to a particular design, i.e. a single trial of two-arms, two trials of two-arms and five trials of two-arms, have been considered. It is of interest to determine what design produces the most power, in relation to a particular LEV dose. Hence, the following section 8.11 conducts a dose-wise comparison exploring the power of different designs, in relation to a particular LEV dose.
8.11 A dose-specific comparison considering power of varying number of trials

The intention of this comparison is to support researchers in making decisions about the required number of similar trials to achieve the required power in the updated MR. Moreover, this comparison helps the optimum allocation of available limited sample size by comparing the power of the updated MR after adding one, two and five trials comparing PLC with a particular dose, designed from available sample sizes. Further, it exhibits the best choice of trial design, in relation to a particular dose, given the limited sample size.

8.11.1 LEV 1000

Table 8-17 below displays the power of the updated MR, after adding one, two and five PLC controlled two-arm trials of LEV 1000, designed from specific sample sizes provided.
Table 8-17: Power of the updated meta-regression after adding one, two and five two-arm trials between PLC and LEV 1000.

<table>
<thead>
<tr>
<th>Sample Size (Total)</th>
<th>Power of the updated meta-regression following the inclusion of trial/s of PLC and LEV 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Trial</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>26.76</td>
</tr>
<tr>
<td>240</td>
<td>45.10</td>
</tr>
<tr>
<td>360</td>
<td>52.90</td>
</tr>
<tr>
<td>480</td>
<td>55.64</td>
</tr>
<tr>
<td>720</td>
<td>61.62</td>
</tr>
<tr>
<td>960</td>
<td>64.02</td>
</tr>
<tr>
<td>1920</td>
<td>67.92</td>
</tr>
<tr>
<td>2880</td>
<td>70.32</td>
</tr>
<tr>
<td>3840</td>
<td>71.50</td>
</tr>
<tr>
<td>4800</td>
<td>70.50</td>
</tr>
</tbody>
</table>

Three lines in plot 8-23 below represent the power after adding one, two and five two-arm trials between PLC and LEV 1000, to the existing MR.

Figure 8-23: The power curve of updated meta-regression after adding one, two and five new two-arm trials between PLC and LEV 1000.
As far as placebo controlled two-arm trial with LEV 1000 is concerned, designing a single trial is the best option irrespective of the available sample size. Designing five trials produces slightly more power than designing two trials, when the size of the sample exceeds 700.

### 8.11.2 LEV 2000

One, two and five two-arm trials between PLC and LEV 2000 designed from available limited sample size are added to the existing MR. The power of the updated MR is tabulated below in table 8-18 against the available sample size.

**Table 8-18**: Power of the updated meta-regression after adding one, two and five two-arm trials between PLC and LEV 2000.

<table>
<thead>
<tr>
<th>Sample Size (Total)</th>
<th>Power of the updated meta-regression following the inclusion of new trial/s comparing PLC and LEV 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Trial</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>0</td>
</tr>
<tr>
<td>240</td>
<td>0</td>
</tr>
<tr>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>480</td>
<td>0</td>
</tr>
<tr>
<td>720</td>
<td>0</td>
</tr>
<tr>
<td>960</td>
<td>0.66</td>
</tr>
<tr>
<td>1920</td>
<td>12.06</td>
</tr>
<tr>
<td>2880</td>
<td>16.48</td>
</tr>
<tr>
<td>3840</td>
<td>19.88</td>
</tr>
<tr>
<td>4800</td>
<td>20.24</td>
</tr>
</tbody>
</table>

Tabulated data in table 8-18 above is presented graphically in figure 8-24.
**Figure 8-24**: The power curve of updated meta-regression after adding one, two and five new two-arm trials between PLC and LEV 2000.

For sample sizes below 1920, adding 5 new trials produce the highest power followed by adding 2 trials and 1 trial. For sample sizes above 1920, designing 1 trial produces the most power followed by 2 trials and 5 trials.

### 8.11.3 LEV 3000

Table 8-19 below displays the power of the updated MR having added one, two and five two-arm PLC controlled LEV 3000 trials, against the sample size available to design new trials.
Table 8-19: Power of updated meta-regression after adding one, two and five two-arm trials between PLC and LEV 3000.

<table>
<thead>
<tr>
<th>Sample Size (Total)</th>
<th>Power of the updated meta-regression after adding New trial/s comparing PLC and LEV 3000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Trial</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>0</td>
</tr>
<tr>
<td>240</td>
<td>0</td>
</tr>
<tr>
<td>360</td>
<td>0</td>
</tr>
<tr>
<td>480</td>
<td>0</td>
</tr>
<tr>
<td>720</td>
<td>0</td>
</tr>
<tr>
<td>960</td>
<td>0</td>
</tr>
<tr>
<td>1920</td>
<td>0</td>
</tr>
<tr>
<td>2880</td>
<td>0</td>
</tr>
<tr>
<td>3840</td>
<td>0</td>
</tr>
<tr>
<td>4800</td>
<td>0</td>
</tr>
</tbody>
</table>

The corresponding power curve is given below in figure 8-25. Three curves in different colours represent the power of the updated MR following the inclusion of one, two and five trials.

Figure 8-25: Power curve of the updated meta-regression after adding one, two and five new two-arm trials between PLC and LEV 3000.
When a smaller sample size is available, designing five new trials is beneficial in yielding greater power than designing a single or two trials. Conversely, designing two trials is beneficial at larger sample sizes. However, designing a single PLC controlled two-arm trial of LEV 3000 is not beneficial, at any sample size.

8.12 Critical evaluation

As far as trials with arms of either LEV 2000 or 3000 are concerned, adding multiple trials into the existing MA leads to a gain in power in the subsequent MR. For sample sizes below a certain value, 1900 in LEV 2000 and 2700 in LEV 3000, designing multiple trials from available sample size is associated with higher power.

This pattern is changing as the available sample size increases above said limits. Reasons for which require further investigations and it is a direction of possible future work. However, this may be attributable to the extent of the variation in the heterogeneity in the observed effect sizes following the inclusion of five new trials.

These results are encouraging and suggest designing multiple trials if involving either LEV 2000 or LEV 3000, to optimise the power. Obviously, the cost involved is greater in designing five trials instead of two trials from available patients. Another future area of related research is to evaluate whether the benefits outweigh the cost of designing multiple trials.
Although the within study error remains the same across all similar sized multiple trials, the between study variance is proportionately reduced, when allocating available subjects into multiple trials instead to a single trial. The reduced between study variation in multiple trials causes to increase the power of the subsequent MR by reducing the standard error of the regression coefficient.

In contrast, when involving LEV 1000, the power of the subsequent MR diminishes when adding multiple new trials into the existing MR. The power is highest after adding one trial and lowest after adding two trials. Surprisingly, power after adding five trials is lower than that after adding one trial but higher than that after adding two trials. The real cause of this behaviour requires further investigations.

We appreciate that these power curves are far from smooth. The error bars around the individual simulation results are relatively larger and allow the graph to depart quite some way from smoothness. Running 10,000 simulations instead of 5000 simulations carried out here, would have produced more smoother curves as illustrated in figure 5-6 of Chapter 5.
8.13 Discussion

This chapter embeds a methodology of designing new trials based on MR principles, into the methodological framework being developed throughout the thesis. This chapter concentrated on the power of the subsequent MR following the insertion of a single, two and five new two-arm trials into the existing MR. In addition to that, the power of the subsequent MR having included a new three-arm trial is investigated.

8.13.1 Key findings of this chapter

Descriptive analysis carried out with trials performed until the end of year 2000, 2006 and 2009 in section 8.8 suggests that with increased LEV doses the efficacy of LEV as an adjunctive therapy in treating partial epilepsy decreases. However, this relationship is not statistically significant at 5%.

8.13.1.1 A single two arm trial

As far as the power of the subsequent MR following the insertion of a single two-armed trial is concerned, the new trial with LEV doses further from the weighted mean (LEV 1000 and 5000) produces higher power. The power is negligible or in fact, zero, of trials with LEV doses closer to the weighted mean (LEV 2000 and 3000).

8.13.1.2 Three arm trials

In the case of three-arm trials, the power of the subsequent MR is highest after adding a new trial with LEV 1000 and 3000 followed by adding a new trial with
LEV 1000 and 3000 and a new trial with LEV 2000 and LEV 3000, in to the existing MR.

### 8.13.1.3 Multiple two-arm trials

Interesting results have been found after including multiple trials into the existing MR. When involving LEV 1000, the power is highest after adding a single trial, followed by adding five new trials and adding two new trials. In contrast, adding more trials involving LEV 2000 and 3000 is associated with a gain in power.

### 8.13.2 Related similar work

Rotondi et al (Rotondi, Donner and Koval 2012) seek to adapt Sutton et al's (Sutton, A. J., Cooper et al. 2007) described approach towards MR contexts. This is the closest relevant work to that presented in this chapter albeit they are not directly comparable. Two methods adopt different methods in the estimation of the heterogeneity. They adopted a Restricted Maximum Likelihood method whereas we used methods of moments estimation. They demonstrated of the methodology using log relative risk while we used log OR. They adopted a variance minimisation approach in the hypothesis test whilst the approach here is based on statistical significance.

Further, the two approaches use different distributional assumptions in the sampling distribution of the new effect size. They assumed a t-distribution in contrast to the normal distribution adopted in here. The use of t-distribution in place of the normal distribution is found to be theoretically superior as the t-
distribution accounts for the uncertainty in both the mean and the variance whilst the normal distribution account for the uncertainty in the mean only. As observed in section 5.8.8 of the Chapter 5 the use of t-distribution approximation yielded approximately the same power results to that of the normal distribution in our contexts. Therefore, for simplicity only the normal approximations are adopted in the later part of the thesis. Although the differences are subtle in the context considered here, we acknowledge that differences could be larger in other contexts.

Rotondi et al further moved on to develop cluster-randomised trials based on their variance minimisation approach (Rotondi, Donner and Koval 2012). We took a different direction and explored the power implications of three-arm trials and multiple two-arm trials after adding into the existing MR.

However, two similarities between the methodologies were found in that both assumed that the variance of the effect size parameters is known without error despite their being estimated from data. Further, both methods used data to estimate the control group event rate.

### 8.13.3 Strengths and limitations of this methodology

The application of this methodology may prove invaluable when information about study level covariates is available. It also enables to explore power implications of three-arm trials when information of only two-armed trials is available. As a rule of thumb, the number of trials in a MR analysis should be large enough to make up a ratio of at least ten trials per one covariate (Vittinghoff and McCulloch 2007). In this illustrative example, the ratio of trials to
covariates is relatively small and it might have affected the results, although the focus is on the methodology itself rather than the results of the illustrative example.

In the example used in this chapter in illustrating the methodology every patient receives the same dose in a specific arm; dose is not subject to within study variation. However, in a real life situation in which aggregate data is used as covariates this methodology poses a few limitations as the covariate in the new planned study is subject to within study variation. Therefore, extra care should be placed when choosing an appropriate distribution for the covariates. This whole process makes the interpretation of the resulting conclusions even harder. Therefore, this approach provides limited benefits when covariates are highly variable across individuals (Rotondi, Donner and Koval 2012).

Another limitation in the methodology of designing three-arm trials is that it ignores the correlation in the treatment effects for simplicity. A direct extension to this work could be to improve the methodology to be able to account for the said correlation.

When combining historic trials with contemporary trials extra caution should be exercised, because the illness severity of patients in two groups of trials may be different. If the treatment efficacy depends on the baseline risk, any analysis not considering this fact will be biased (Thompson, S. G., Smith and Sharp 1997). Therefore, a step forward would be to include baseline risk as a covariate into the methodology to partially account for the heterogeneity between trials.
8.13.4 Potential future work

It is peculiar that in some sections (figure 8-21 and figure 8-23) of this chapter the power declines with increased sample sizes. Although not directly comparable, Rotondi et al found a similar phenomenon in their work (Rotondi, Donner and Koval 2012) as well, where the percentage reduction of the variance (which is analogous to power in this work) is less in some larger sample sizes than a smaller sample sizes. The direction of immediate future work is to investigate the reasons for the said drop of power as sample size increases.

The methods developed here are specific to continuous covariates only. In future work, the methods could be generalised to allow both discrete and continuous covariates. Another promising area of future research would be to extend this methodology to include multiple covariates. Moreover, it assists in exploring the implications of both main effects and interaction effects of covariates. (Raudenbush and Bryk 1985)

It is not trivial to detect a specific pattern explaining the variation in the heterogeneity parameter when multiple trials are added into existing MR. We suggest further investigations focusing on examining the variation in the heterogeneity to interpret unusual findings in the updated MR particularly when multiple trials are involved.

The process of choosing the most appropriate existing evidence to base the original analysis is not trivial. Even if previous trials are meant to produce the
same outcome measures of interest, perhaps the scale or the duration of measurements may vary, making trial results incomparable unless adjusted.

As described in section 8.5.3.1 of this chapter the outcome of interest i.e. 50% responders rate has not been produced in consistent basis across existing trials. In the efficacy analysis, three different types of populations have been used in five different trials, namely intention to treat (ITT), inferential intention to treat, and Per Protocol population (PPP). Moreover, in some trials, the data obtained during both the titration and evaluation period was used in the analysis, whereas in some trials only the data obtained during the evaluation period was analysed. Therefore, to produce valid results it is advisable to identify subtle discrepancies between outcome measures to ascertain their comparability before implementing this methodology in practice.
8.14 Conclusions

Research aiming to examine plausible sources of heterogeneity between observed treatment effects is often less attractive to funders compared to those aiming to determine the overall effect estimate. This approach establishes a less costly but reliable methodology in prioritising areas of future research for example in prioritising specific treatment doses, and populations.

In summary, investigators exploring the role of study level covariates as a source of heterogeneity should consider designing small multiple trials rather than a single larger trial if LEV 2000 or 3000 is involved, to optimise the power of the test. However, when the size of available sample is small, even adding multiple trials do not gain enough power in the subsequent MR.

A single trial should be designed from available sample size for new trials involving LEV 1000, to yield maximum power. When planning to design three arm trials covariates further from the weighted mean should be selected, if the power of the subsequent MR is the prime interest.

Despite these theoretical recommendations, consideration should be given to the cost involved in designing multiple trials. Moreover, the benefits offered from multiple trials should be evaluated over the costs involved in designing them using an economical model such as expected value of information theory (Welton, Ades et al. 2008).

Despite increased power as far as designing a single trial is concerned in trials involving higher doses such as LEV 5000, high dose trials would not be
recommended on ethical grounds. Doses stronger than the standard maximum
dose, i.e. LEV 3000 may be linked with more side effects when given to
patients. Therefore, the potential benefits attached to increased power in the
designing of LEV 5000 trials should be compromised with the likely side effects
linked with higher doses.

It should however be noted that because of the observational nature of MR the
strength of the association found are limited and not as strong as that derived
from randomised trials.

So far, this thesis has examined the power implications of trials designed based
on frequentist approaches of meta-analysis, indirect comparisons, mixed
treatment comparisons and meta-regression principles. The next chapter
explores Bayesian approaches to meta-analysis to inform designing of new
trials.
Designing of future trials based on existing evidence using Bayesian methods

9.1 Introduction

This chapter adopts Bayesian evidence synthesis models to develop the methodology of designing future trials based on existing evidence. The motivation to the work in this chapter is explained in section 9.2. The objectives of this chapter are outlined in section 9.3. A brief introduction to Bayesian statistics is given in section 9.4. Section 9.5 discusses about Markov Chain Monte Carlo methods and section 9.6 briefly describes of the WinBUGS software. Section 9.7 discusses the approaches to Bayesian evidence synthesis methods. Section 9.8 describes the methods employed. Section 9.9 develops the methodology using Bayesian MA methods. Within that, sections 9.9.4 and 9.9.5 use fixed and random effects MA methods respectively. Section 9.10 develops the methodology of designing new trials using Bayesian MR methods. The critical evaluation of the methods is given in section 9.11. A general discussion of the findings of this chapter is given in the section 9.12. Section 9.13 concludes the chapter.
9.2 Motivation

The Bayesian paradigm is increasingly becoming popular and well established in providing methods to synthesise evidence from several clinical trials. Carlin implemented and developed a Bayesian MA approach to a 2x2 table, which is regarded as the earliest attempt at MA using Bayesian framework (Carlin 1992). Since then, various methodologies and extensions based on the Bayesian methods have emerged particularly in modelling and understanding the sources of heterogeneity (Thompson, S. G., Smith and Sharp 1997), MA of the results of diagnostic tests (Rutter and Gatsonis 2001), and application to cluster randomised trials with continuous response (Spiegelhalter, D. J. 2001).

The methodological development in predicting and designing new trials based on Bayesian methods has received less attention (DerSimonian, Rebecca 1996). Joseph implemented a non power based Bayesian approach of sample size determination (Joseph, Du Berger and Belisle 1997). Sutton et al discussed a criterion for sample size determination based on power. Although Bayesian methods become increasingly adopted in designing new trials (Chen and Pei 2009), widely accepted Bayesian approach guiding the design of a new trial has not been established yet.

The lack of the availability of widely accepted Bayesian methodology guiding the design of new trials has motivated the work undertaken in this chapter. The objectives of this chapter are outlined in section 9.3 below.
9.3 Objectives

The central objective of this chapter is to develop methodologies to design future trials based on evidence synthesis models using the Bayesian framework. These methodologies aim to estimate the sample size of a new trial to achieve a certain power of the subsequent MA including the new trial. An explicit list of aims is outlined below.

- Initial aim is to develop the methodologies based on FE models of Bayesian MA.
- The work conducted by Sutton et al with respect to the Bayesian RE meta-analysis is replicated with the same set of data, to examine if the same zero power results are found.
- To examine the sensitivity of power to the changes in prior distributions, several prior distributions are placed on $\tau^2$ via a sensitivity analysis in the random effects MA approach.
- Subsequently, the Bayesian aspect of MR is used to develop the methodology of designing new trials.

Having set out the objectives, the next section discusses the relevant methods to achieve the above-mentioned objectives.

9.4 Introduction to Bayesian statistics

The Bayesian paradigm emphasises learning from data. The data have a specific role in updating the belief of what is known about the parameters and the hypothesis. The Bayesian framework treats probability as a statement of
uncertainty and all unknown quantities as random variables (O’Hagan and Luce 2003). It allows the incorporation of prior belief about an unknown quantity into the analysis by means of a prior distribution. A full probability model is made for all observable and unobservable quantities by specifying a joint probability distribution to the data. To construct a posterior distribution, which is the joint probability distribution of all unknown parameters conditional on the data, Bayes theorem, which is the foundation of the Bayesian Statistics is used as in the equation 9.1 below.

\[
P(\theta|\text{data}) = \frac{P(\text{data}|\theta)P(\theta)}{P(\text{data})}
\]  

(9.1)

where, \(\theta\) represents a vector of unknown parameters required to be estimated and \(\text{data}\) denotes the observed data.

\(P(\text{data}|\theta)\) represents the statistical model part of the analysis and quantitatively describes the assumption that the observed data is generated based on the parameters.

\(P(\theta)\) denotes the prior belief of \(\theta\) in the form of a distribution and represents the prior assumption about \(\theta\) as well. In practice, the process of specifying appropriate prior distributions may not be trivial.

\(P(\theta|\text{data})\) corresponds to the posterior distribution of \(\theta\) which is a combination of the prior knowledge and the information provided by the data about \(\theta\). An examination of different characteristics of samples drawn from the posterior distribution is necessary to make inferences about unknown parameters. The greater the number of samples drawn from the posterior distribution, the more
accurate the resemblance of the sampling distribution to the posterior distribution.

\( P(data) \) represents the marginal distribution which does not depend on the parameter \( \theta \). If the model contains numerous parameters, it is difficult to evaluate this component.

An alternative form of the equation 9.1 is given below in equation 9.2.

\[
P(\theta|data) \propto P(\theta) L(\theta|data)
\]

(9.2)

\( L(\theta|data) \) - This corresponds to the likelihood function which summarises all information about \( \theta \) available in the sample (Jackman 2000).

The equation 9.2 specifies that the posterior distribution is proportionate to the product of the prior distribution and the likelihood. For non-informative prior distributions, the likelihood function predominantly determines the posterior distribution. Hence, parameter estimates of the Bayesian analysis using non-informative priors do not significantly differ from the frequentist estimates (Jackman 2000).

The advent of the Markov Chain Monte Carlo methods (MCMC) has led to an increasing popularity in the use of Bayesian statistics primarily because of the flexibility and the computational feasibility offered to the user. Section 9.5 below briefly reviews the MCMC methods.
9.5 **Markov Chain Monte Carlo Methods (MCMC)**

The Monte Carlo (MC) methods, often used in simulating systems, are a class of computational algorithms that use repeated random sampling to compute the results. MCMC is used to solve numerical problems for which analytical solutions are hard to find.

Markov Chain is a series of random variables \((X_1, X_2, \ldots, X_n)\) generated from a Markov process. A random variable is in a Markov process if the transitional probability between the different values in the state space depends only on the random variable’s current state, i.e. the only information required to predict the future of the random variable is its current status. The knowledge on the past state of the random variable does not change the transitional probability.

Bayesian inference entails evaluating summaries, such as moments and densities of the posterior distribution. Joint posterior distributions are usually high dimensional and closed form solutions are not often available. These high dimensional functions are required to be integrated with respect to \(\theta\). The MCMC methods avoid these integrations by drawing repeated samples from a set of full conditional distributions, which under Ergodic theory, converge to marginal posterior distribution. (Lunn, Thomas et al. 2000).

A Markov chain needs to be started at some point by assigning initial values to the chain. Although in theory (Spiegelhalter, D. J., Abrams and Myles 2004) any initial value could be assigned to these unknown parameters, in practice reasonable initial values need to be assigned to improve the convergence and in particularly to avoid numerical problems.
Assessing the convergence of the Markov chain, especially with many dimensions, is not always straightforward, although none or lack of convergence can be diagnosed by the behaviour of the sampled values. A chain moving in a steady trajectory does not necessarily mean samples are drawn from the correct posterior as chain might be stuck in one particular area due to wrong choice of initial values. Use of a multiple chains with diverse set of initial values has been advised (Spiegelhalter, D. J., Abrams and Myles 2004) to ascertain all chains are coming from the same equilibrium distribution that will then assumed to be the same posterior distribution of interest.

MCMC methods are implemented via WinBUGS software, and section 9.6 below reviews WinBUGS software in relation to the Bayesian model fitting.

9.6 **WinBUGS – Windows Version of Bayesian Inference Using Gibbs Sampling software**

Advances in software particularly with the advent of WinBUGS which uses MCMC methods, have led to a rapid growth in the use of MA model fitting under the Bayesian framework (Spiegelhalter, D.J. Thomas, A. Best, N.G. 2000). These models are greatly benefitted by the flexibility of WinBUGS particularly with respect to specifying the parametric models allowing full uncertainty of all quantities and allowing natural means for interpreting the results.

WinBUGS models consist of joint distribution of all unknown quantities such as parameters and observed quantities such as data. A posterior distribution is obtained by conditioning the model on data, through Bayes theorem. Unknown
quantities of the model are then estimated by marginalising the posterior
distribution using MCMC simulation methods. This requires integration of
various functions involving \( \theta | data \), with respect to \( \theta \). These integrals are
evaluated using MC simulations.

A series of values for each unknown quantity are then generated via MC
simulation from the posterior distribution and inferences about the unknown
quantity are then made. With sufficiently many observations, a true picture of
the distribution can be made.

The sampling method used in WinBUGS depends on the type of the model
being estimated. In the simplest case, direct sampling from the posterior
distribution is allowed. Methods such as Metropolis-Hastings or Gibbs sampling
seem appropriate in situations that are more complex where direct sampling
from the posterior distribution is not feasible.

This chapter develops and implements methodologies to design new trials
based on Bayesian evidence synthesis models; these models are reviewed in
section 9.7 below.

9.7 Bayesian evidence synthesis

A single study is often not conclusive in its own right and therefore does not
inform policy decisions. A collection of evidence from different sources is always
essential before arriving at policy decisions. The Bayesian framework allows a
flexible model based approach in combining evidences across different sources.
The form in which the conclusions are made in Bayesian perspective
contributes naturally to the decision making process (Spiegelhalter, David J., Myles et al. 1999).

Under the label of ‘confidence profile method’ Eddy (Eddy, D. M., Hasselblad and Shachter 1990) presented a framework for combining different study designs using Bayesian evidence synthesis. The practical take up of these techniques seems low due to the complexity and limitations in software. Importance of adjusting for potential confounders in Bayesian hierarchical models when synthesis evidence from randomised and non randomised studies has been outlined (McCarron, Pullenayegum et al. 2010).

Many authors have considered using Bayesian methods in designing randomised trials. Chen and Pei who investigated the association between TNF-α and a certain polymorphism, revealed that existing trials in their analysis are not adequately powered to make valid statistical statements. Hence they simulated a new trial with adequate power to base the inferences and conclusions on (Chen and Pei 2009). Sutton et al (Sutton, A. J. and Abrams 2001) and Abrams et al (Abrams and Jones 1995) extensively reviewed the Bayesian evidence synthesis.

The use of Bayesian methods has received much attention in number of areas in medical research including the MA and MR. This is mainly due to the flexibility allows by the Bayesian methods in both incorporating prior beliefs into the analysis and the advent of the MCMC methods in dealing with more complex models (Abrams and Jones 1995). The Bayesian approach of inference enables to incorporate existing information into the analysis, by
specifying a prior distribution. In Bayesian RE meta-analysis the between study variation is explicitly modelled.

Combining objective data with subjective opinion (in the prior) has always been a criticism made of Bayesian analysis (Gurrin, Kurinczuk and Burton 2001). Hence, in Bayesian MA, it has now become a common practice to use non-informative priors particularly in the analysis of the main effects. Non-informative priors make all the values in the range equally likely to be the unknown quantity.

9.8 **Methods**

Smith et al (Smith, T.C., Spiegelhalter and Thomas 1995) introduced a MA model specific to binary outcomes modelled on the OR scale and is fitted using Bayesian MCMC formulation. The methodology developed here extends the model specified by Smith et al in 1995.

The procedure used to estimate power is quite numerically intensive because a nested simulation approach is employed to evaluate every synthesis within the simulation framework. In considering this nested approach, the evaluation is designed to incorporate two separate WinBUGS programs (Sutton, A. J., Cooper et al. 2007).

The first program simulates the number of events and non-events of both treatment and control arms of the simulated new trial. The results from this will subsequently be fed in to the second program designed to synthesise all trials including the new trial. The step-by-step guidance to the process of developing
the methodology of designing new trial using the Bayesian approach is described below.

9.8.1 First program

This program conducts the initial synthesis, specifies the predictive distribution of the new trial and designs the new trial.

Step 1

An initial synthesis is carried out of the existing evidence.

Step 2

The number of subjects in both treatment ($n_{\text{treatment,control}}$) and control arm ($n_{\text{treatment,new}}$) of the new trial is then specified.

Step 3

The predictive distribution of the new trial is specified. The MCMC methods are employed to simulate $n$ number of samples from the predictive distribution representing an effect size of the new trial ($\theta_{\text{new}}$) in an iterative process.

Step 4

The probability of an event occurring in the treatment group of the new trial ($P_{\text{treatment,new}}$) can then be computed using equation 5.1 specified in section 5.6 of Chapter 5 (Sutton, A. J., Cooper et al. 2007).
**Step 5**

The number of events in each arm of the trial is modelled using the binomial distribution as specified in equation 5.2 and 5.3 of section 5.6 in Chapter 5.

The number of events and non-events in all \( n \) number of simulations is determined, which represent the 2x2 table of the outcome of a new trial.

**Step 6**

In each simulation, a new trial is designed and the number of events and non-events in both the treatment and control group is then fed into the program 2.

### 9.8.2 Second program

This program simultaneously conducts and evaluates \( n \) number of updated syntheses including the new trial, where \( n \) is equivalent to the number of simulations.

**Step 7**

A large number of samples are drawn from the posterior distribution of the parameter of interest and the mean of the posterior distribution is recorded.

**Step 8**

The middle 95% of the credible interval of the mean of the parameter of interest of each simulation is assessed to examine if the null value is included. The
proportion of rejections of the null hypothesis is counted and treated as the power of the test.

**Step 9**

The whole process (step 2-8) is then repeated with a different sample size of the new trial and the power of the updated synthesis is determined for that specific sample size of the new trial.

### 9.8.3 Power of the test

The assessment of the middle 95% of the credible interval (between 2.5% and 97.5%) of the posterior distribution of the parameter of interest is required in determining the power of the test. If the null value is contained within the interval then the two treatments are not regarded as different. Thus, the null hypothesis is not rejected and the test is found to be statistically non-significant. Alternatively, if the null value is not contained within the interval, consequently the null hypothesis is rejected in favour of the alternative hypothesis. Thus, the test is found to be statistically significant and hence that simulation results contribute towards the estimate of statistical power.

This approach, however, assumes equal probabilities in either tail area of the posterior distribution, and does not work well when the posterior distribution is skewed. Alternatively, the Highest Posterior Density (HPD) intervals, which adjusts so that the probability ordinates at each end of the interval are identical could be used.
Section 9.9 below implements the methodology of designing new trials using Bayesian MA methods.

9.9 Designing a new trial based on existing evidence using Bayesian meta-analysis model

9.9.1 Existing evidence

To enable the comparability and to explore the causes of the discrepancies between the results yielded using the Bayesian and the frequentist analysis, the very same existing evidence used in the frequentist FE (section 5.7) and frequentist RE (section 5.8) is used in this section as well.

9.9.2 Null hypothesis

The power of the test of the mean effect is determined based on the null hypothesis that two treatments are not different or are equally effective.

9.9.3 Event rate in the control group

The probability of an event occurring in the control group of the new trial $P_{control\_new}$ is assumed to be equal to the probability of an event occurring in the control group of the existing trials, which is calculated to be 0.352 as mentioned in the frequentist approach in section 5.5.3.
9.9.4 Prediction model 28 - Designing a new trial based on the fixed effects Bayesian meta-analysis methods

9.9.4.1 Rationale

The FE model assumes no heterogeneity between primary trials included in the MA (Sutton, A., Abrams et al. 2000), implying that all trials are estimating the same underlying treatment effect. The methodology developed here is based on binary outcomes measured on OR scale although it could be extendable to other outcome measures of interest.

9.9.4.1.1 Bayesian fixed effects model

The event rate in the control group ($\mu_i$) is a function of the probability of an event in the control group ($P_{control,i}$) of the $i^{th}$ trial as specified in equation 9.3 below.

$$\logit (P_{control,i}) = \mu_i$$

(9.3)

Under the FE model, the underlying effect parameter ($\theta$) is assumed a constant across all trials. $\theta$ denotes the natural logarithm of the OR.

$$\logit (P_{treatment,i}) = \mu_i + \theta$$

(9.4)

Both $\mu_i$ and $\theta$ are unknown and require prior distributions.
9.9.4.1.2 **Specifying the predictive distribution of the new trial**

In the FE model, the predictive distribution of an effect size of the new trial \( (\theta_{new}) \) is specified as below.

\[
\theta_{new} \sim \text{normal} \left[ \theta, SE(\theta) \right]
\]  

(9.5)

9.9.4.1.3 **Specifying prior distributions**

The FE Bayesian approach requires specifying prior densities for \( \theta \) and \( \mu_i \).

Non-informative normal prior densities are assigned with zero means and very large variances on both \( \theta \) and \( \mu_i \) as specified in equation 9.6 and 9.7.

\[
\theta \sim \text{normal} \left[ 0, 10^6 \right] \]  

(9.6)

\[
\mu_i \sim \text{normal} \left[ 0, 10^6 \right] \]  

(9.7)

9.9.4.2 **Methods**

A new trial is designed using the methods specified in section 9.8 above and the power of the updated FE meta-analysis is reported for various sample sizes of the new trial.

9.9.4.3 **Results**

Table 9.1 below shows the power of the updated MA computed using the Bayesian FE model for several sample sizes of the new trial. To enable a convenient comparison, corresponding power estimates of the frequentist FE method (section 5.7) are also incorporated into the table. This analysis is based on the dataset presented in table 5-1.
Chapter 9  Designing new trials using Bayesian evidence synthesis methods

**Table 9-1**: The power results comparison of the Bayesian and frequentist FE meta-analysis for corresponding sample sizes of the new trial.

<table>
<thead>
<tr>
<th>Sample size of the new trial (one arm)</th>
<th>Fixed Effects MA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Power of the updated MA from the Bayesian FE model</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>25.28</td>
</tr>
<tr>
<td>300</td>
<td>49.88</td>
</tr>
<tr>
<td>500</td>
<td>60.08</td>
</tr>
<tr>
<td>800</td>
<td>65.92</td>
</tr>
<tr>
<td>1000</td>
<td>67.94</td>
</tr>
<tr>
<td>2000</td>
<td>75.76</td>
</tr>
<tr>
<td>3000</td>
<td>81.94</td>
</tr>
<tr>
<td>4000</td>
<td>83.46</td>
</tr>
<tr>
<td>5000</td>
<td>84.34</td>
</tr>
</tbody>
</table>

Figure 9-1 below represents the corresponding power curve.

**Figure 9-1**: Two power curves of the Bayesian and frequentist FE meta-analysis of corresponding sample sizes of the new trial.

The assumption that every trial is estimating the same underlying effect parameter is practically implausible. In practice, most trials have different design...
and clinical characteristics leading to have different underlying effect size. RE model of meta-analysis address these situations by accounting for the variation between trials (DerSimonian, R. and Kacker 2007). The section 9.9.5 below implements the methodology of designing new trials based on existing evidence, using Bayesian RE meta-analysis model.

9.9.5 **Designing a new trial based on existing evidence using Bayesian random effects meta-analysis model**

9.9.5.1 **Rationale**

9.9.5.1.1 **Bayesian random effects model**

Bayesian approach to RE meta-analysis has implemented a hierarchical model, which models both within and between study variations. The first level of the hierarchy indicates that the observed outcome measures ($Y_i$) are normally distributed about the underlying effect size ($\theta_i$) of primary trials (Sutton, A., Abrams et al. 2000). This association is illustrated in equation 9.8 below.

$$Y_i \sim Normal(\theta_i, \frac{\sigma^2_i}{n_i}) \quad (9.8)$$

where $i$ indexes the $k$ studies and $\frac{\sigma^2_i}{n_i}$ denotes the standard error of the estimation of the effect size and $n_i$ represents the sample size of the $i^{th}$ trial.

The second level of the hierarchy relates the underlying treatment effects in the $i^{th}$ ($\theta_i$) to the overall mean ($\theta$) of the distribution from which the individual effect
sizes are assumed to be drawn. (Sutton, A., Abrams et al. 2000). The second level of the hierarchy is specified as below in equation 9.9.

\[ \theta_i \sim Normal(\theta, \tau^2) \]  

(9.9)

Based on RE principles, trial specific treatment effects are a random sample from a normal distribution with mean \( \theta \) and variance \( \tau^2 \), where \( \tau^2 \) represents the variation in the underlying effect size parameters between trials.

The same association as specified in the fixed effect settings is found between the \( P_{control,i} \) and \( \mu_i \) in the random effects settings as well.

\[ \text{logit} \ (P_{control,i}) = \mu_i \]  

(9.10)

Similar to the FE settings, \( \mu_i \) represents the event rate in the control arm and \( P_{control,i} \) denotes the probability of an event in the control group of the \( i^{th} \) trial. However, to represent the fact that each trial is estimating a trial specific effect size \( (\theta_i) \), the relevant association is modified as in equation 9.11 below.

\[ \text{logit} \ (P_{treatment,i}) = \mu_i + \theta_i \]  

(9.11)

9.9.5.1.2 Specifying the predictive distribution

This model assumes the effect size of the new trial is normally distributed with the mean equivalent to the pooled effect size and the variance equivalent to the heterogeneity of existing MA.

\[ \theta_{new} \sim Normal(\theta, \tau^2) \]  

(9.12)
Equation 9.12 above specifies the predictive distribution of a new trial in the Bayesian analysis.

**9.9.5.1.3 Specifying the prior distributions**

The Bayesian model presented here requires estimating $\theta$, $\tau$ and $\mu_i$ using the MCMC formulation. Therefore, these unknown quantities must be assigned prior distributions.

As given in equation 9.13, a non-informative normal prior is specified on $\mu_i$ with a zero mean and a very large variance.

$$\mu_i \sim \text{Normal}(0, 10^6) \quad (9.13)$$

A common practice in placing a prior on $\theta$ is to assign a vague prior or a uniform prior over the whole real line. However, subjectivity may be involved in incorporating pertinent information external to MA, e.g. from observational information. Alternatively, prior distributions could be modelled explicitly from generalised synthesis methods (Sutton, A., Abrams et al. 2000).

$$\theta \sim \text{Normal}(0, 10^6) \quad (9.14)$$

In this analysis, two distinct priors, the uniform and inverse gamma distributions are specified for $\tau^2$ to explore the sensitivity of the results to different priors.

$$\tau \sim \text{Uniform}(0, 2) \quad (9.15)$$
Section 9.9.5.2 below conducts a sensitivity analysis examining the impact of the different prior distributions of $\tau$ on power estimates of the updated MA, using Bayesian RE.

### 9.9.5.2 The sensitivity analysis on different prior distributions of the between study standard deviation

The standard deviation of the random effects ($\tau$) plays a key role in the accurate estimation of the uncertainty concerning the mean effect size ($\theta$) and in the calculation of effect size of a future trial ($\theta_{new}$). The empirical Bayes approach that ignores the uncertainty in the estimation of $\tau$ leads to produce narrower intervals than the usual (Spiegelhalter, D. J., Abrams and Myles 2004). The specification of a precise prior distribution for $\tau$ is particularly important when the analysis includes a smaller number of trials.

In this context, we propose to specify a non-informative prior with uniform distribution and an empirically based inverse gamma distribution on $\tau$. This process of placing different priors resembles a sensitivity analysis which allows exploring the sensitivity of the power of the updated MA to the changes of the prior distributions placed on $\tau$.

Sections 9.9.5.2.1 and 9.9.5.2.2 upcoming below develop the methodology of designing new trials based on existing evidence using Bayesian methods by assessing the sensitivity of the results to different prior distributions of the between study standard deviation.
9.9.5.2.1 Prediction model 29 - Designing a new trial based on random effects Bayesian meta-analysis methods with an inverse gamma prior on $\tau^2$

A gamma prior distribution is specified on the precision or inverse variance of the random effects. An inverse gamma prior distribution is specified on $\tau^2$ as illustrated in equation 9.17 below.

\[
\frac{1}{\tau^2} \sim \text{Gamma}(0.001, 0.001) \tag{9.17}
\]

This prediction model is characterised with an inverse gamma prior distribution with a low and equal scale and shape parameters (0.001) on the variance of the random effects.

The power of the updated MA is then computed and the results are given in the table 9-2 below. The corresponding power estimates of the updated MA using frequentist RE model (section 5.8.10) are also incorporated to assist in the comparison.
Table 9-2: Power of the updated MA of the prediction model based on Bayesian RE meta-analysis model that specifies an inverse Gamma prior on $\tau^2$, along with that of frequentist counterpart, at different sample sizes of the new trial.

<table>
<thead>
<tr>
<th>Sample size of the new trial (one arm)</th>
<th>Random Effects MA</th>
<th>Power of the updated MA from the Bayesian Random Effects model</th>
<th>Power of the updated MA from the Frequentist Random Effects model (Both Within and between study Variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
<td>0</td>
<td>16.84</td>
</tr>
<tr>
<td>800</td>
<td>0</td>
<td>0</td>
<td>28.40</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>29.76</td>
</tr>
<tr>
<td>2000</td>
<td>0</td>
<td>0</td>
<td>35.64</td>
</tr>
<tr>
<td>3000</td>
<td>0</td>
<td>0</td>
<td>37.04</td>
</tr>
<tr>
<td>4000</td>
<td>0</td>
<td>0</td>
<td>37.40</td>
</tr>
<tr>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>37.80</td>
</tr>
</tbody>
</table>

The power of the updated MA remains zero for all sample sizes considered.
Figure 9-2: Power curve of the updated MA of the prediction model based on Bayesian RE meta-analysis model that specifies an inverse Gamma prior on \( \tau^2 \), along with that of frequentist counterpart, at different sample sizes of the new trial.

The power of the updated MA using the Bayesian analysis is estimated to be zero at all sample sizes considered of the new trial. Clearly, further investigations into the discrepancies between results are needed.

Section 9.9.5.2.2 below, place an uniform prior distribution on \( \tau \) to identify whether the zero power estimates are attributable to the use of inverse gamma prior of \( \tau^2 \).

9.9.5.2.2 Prediction model 30 - Designing a new trial based on random effects Bayesian meta-analysis methods with an uniform prior on \( \tau \)

The implication of specifying a uniform prior distribution on \( \tau \) is such that \( \tau \) is equally likely to take any value in the range. To be meaningful, a range of the uniform distribution has to be specified as illustrated in equation 9.18 below.

\[
\tau \sim Uniform (0,2) \quad (9.18)
\]
This prediction model develops a new trial using the Bayesian RE principles with a uniform prior distribution of the standard deviation of the random effects restricted to limits of 0 and 2.

Similar to the prediction model with an inverse gamma prior (section 9.5.2.2.1), the statistical power of the updated MA is found to be zero for all sample sizes considered of the new trial from which we can conclude that zero power is not an artefact of using the inverse gamma prior.

9.9.5.3 Critical evaluation of the sensitivity analysis

Three methods are illustrated in literature for choosing an appropriate prior distribution for the between study standard deviation, viz. by elicitation of opinion, by summary of evidence and the use of reference priors. The elicitation of opinion on prior distributions in healthcare settings requires evidential and consensus support as opposed to merely expressing the personal opinion. Values of $\tau$ between 0.1 and 0.5 may deem to be reasonable. The specification of a subjective prior on $\tau$ often requires validating if $\tau=0$ is a plausibility i.e. no heterogeneity between trials.

Pauler and Wakefield implemented a method of placing a subjective prior (a half normal prior) on $\tau$. Nevertheless, we do not pursue use of a prior based on elicitation of opinion here because of the unavailability of a formal technique of validation. Thus far, no formal techniques have been developed in assessing the validity of a subjective prior.
The summary of evidence method entails determination of a prior distribution of \( \tau \) based on the values used in previous hierarchical models with similar contexts. Higgins and Whitehead (Higgins, J. P. T. and Whitehead 1996) used empirical distributions of past \( \tau \)s and revealed that \( \tau^{-2} \) has a Gamma \([1,0.35]\) distribution. Transforming this into \( \tau \) scale, they propose a root inverse gamma distribution for \( \tau \).

In general, standard reference priors are placed on the sampling variance, i.e. a uniform prior on \( \log(\sigma^2) \). However, they are not suitable for random effects standard deviation as it will result in improper posterior distributions (Spiegelhalter, D. J., Abrams and Myles 2004). Even though universally accepted reference or non-informative priors for \( \tau \) are not available, methods have been developed to place non-informative priors on \( \tau \) or \( \tau^2 \). Inverse gamma and uniform priors are the commonly used in practice (Spiegelhalter, D.J. Thomas, A. Best, N.G. 2000) which is used in here by way of a sensitivity analysis.
9.10 Designing a new trial based on existing
evidence using Bayesian random effects
meta-regression model

9.10.1 Existing evidence

For consistency, the same existing evidence used in developing the
methodology using the frequentist MR approach is used in here. The data is
listed in table 8-8 of Chapter 8.

9.10.2 Null hypothesis

The power of the test is determined based on the null hypothesis that the
regression coefficient of the updated MR is zero, i.e. the covariate is not
influential in determining the effect size.

9.10.3 Event rate in the control group

The event rate in the control group in the Bayesian approach is assumed to be
the same as that used in the frequentist analysis, which is 0.146.
9.10.4 Prediction model 31: Designing a new trial based on random effects Bayesian meta-regression methods

9.10.4.1 Rationale

9.10.4.1.1 Centring the covariates

The process of centring covariates is common in Bayesian MR to reduce the auto correlation between successive values of the MCMC chain. The centred covariate \(cov_{cen}\) is achieved by subtracting the mean covariate \(cov_{mean}\) value from the covariate value \(cov_i\) of each individual trial. High auto correlation between the successive values of the MCMC chain leads to sample more values than required when autocorrelation is negligible.

\[
cov_{cen} = cov_i - cov_{mean} \tag{9.19}
\]

However, a transformation of parameters is required to remove the centring in interpreting the results. If \(d_{cen}\) represents the mean of the distribution of the effect size with centred covariate and \(d_{uncen}\) denotes the true effect size at zero covariate, the removal of centring could be achieved by

\[
d_{uncen} = d_{cen} - \beta * cov_{mean} \tag{9.20}
\]

where \(\beta\) is the regression coefficient, the value of which determines the power of the test. The value of \(\beta\) is unknown and a prior distribution needs to be specified for this unknown quantity.
9.10.4.1.2 Bayesian meta-regression model

The event rate in the control group ($\mu_i$) is a function of the probability of an event in the control group ($P_{\text{control},i}$) of the $i^{th}$ trial as specified in equation 9.21 below.

$$\logit (P_{\text{control},i}) = \mu_i$$

(9.21)

$\mu_i$ is an unknown quantity which requires specifying a prior distribution. Under the Bayesian RE model, each trial is assumed estimating a different true underlying effect parameter ($\theta_i$). The product of the regression coefficient $\beta$ and the centred covariate is included to equation 9.21 specified above in RE meta-analysis model to represent the Bayesian RE model as below.

$$\logit (P_{\text{treatment},i}) = \mu_i + \theta_i + \beta * \text{cov}_{\text{cen}}$$

(9.22)

$\theta_i$ is normally distributed with a mean equivalent to $d_{\text{cen}}$ and with a variance equivalent to the residual heterogeneity $\tau^2$ as in equation 9.23 below.

$$\theta_i \sim \text{Normal} (d_{\text{cen}} , \tau^2)$$

(9.23)

Both $d_{\text{cen}}$ and $\tau^2$ are unknown and hence prior distributions are required.

9.10.4.1.3 Prediction distribution of the new trial

An effect size of the new trial ($\theta_{\text{new}}$) is assumed normally distributed with mean equivalent to $d_{\text{new}}$ and the variance equivalent to $\tau^2$. 


where $\tau^2$ represents the residual heterogeneity. $d_{\text{new}}$ is the mean of the distribution of effect sizes of trials with the new covariate value ($\text{cov}_{\text{new}}$) and can be defined as below in equation 9.25 below.

$$d_{\text{new}} = d_{\text{uncen}} + \beta \cdot \text{cov}_{\text{new}}$$

### 9.10.4.1.4 Specification of prior distributions

The RE Bayesian meta-regression approach requires specifying prior densities for $\mu_i$, $d_{\text{cen}}$, $\tau$, and $\beta$. Non-informative normal prior densities are assigned with zero means and very large variances on both $\mu_i$, $d_{\text{cen}}$ and $\beta$ as specified in equations 9.26 - 9.28.

\[
\mu_i \sim \text{normal}[0, 10^6]
\]

\[
d_{\text{cen}} \sim \text{normal}[0, 10^6]
\]

\[
\beta \sim \text{normal}[0, 10^6]
\]

A uniform prior with limits of 0 and 2 is assigned on $\tau$ as below.

\[
\tau \sim \text{uniform}[0, 2]
\]
**9.10.4.2 Methods**

The methods specified in section 9.8 are used to design a new trial. However, in MR approach the specific covariate (LEV dose) of the new trial in addition to the sample size needs to be specified in step 1.

**9.10.4.3 Results**

The power of the updated Bayesian MR including the new trial with different covariates (LEV doses) for various sample sizes of the new trial is listed in Table 9-3 below. Corresponding frequentist results are also included in the table to highlight the differences in the results between the two approaches.

*Table 9-3: A comparison of the power of the updated Bayesian and frequentist meta-regression.*

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>LEV 1000</th>
<th>LEV 2000</th>
<th>LEV 3000</th>
<th>LEV 3500</th>
<th>LEV 4000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BAYESIAN</td>
<td>FREQUENTIST</td>
<td>BAYESIAN</td>
<td>FREQUENTIST</td>
<td>BAYESIAN</td>
</tr>
<tr>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>100</td>
<td>2.70</td>
<td>41.24</td>
<td>0.00</td>
<td>0.06</td>
<td>20.16</td>
</tr>
<tr>
<td>200</td>
<td>2.26</td>
<td>54.74</td>
<td>0.00</td>
<td>0.12</td>
<td>23.86</td>
</tr>
<tr>
<td>300</td>
<td>2.32</td>
<td>59.50</td>
<td>0.00</td>
<td>0.20</td>
<td>24.80</td>
</tr>
<tr>
<td>500</td>
<td>2.22</td>
<td>65.32</td>
<td>0.00</td>
<td>1.04</td>
<td>26.48</td>
</tr>
<tr>
<td>800</td>
<td>4.42</td>
<td>67.74</td>
<td>0.00</td>
<td>9.16</td>
<td>27.52</td>
</tr>
<tr>
<td>1000</td>
<td>3.30</td>
<td>68.08</td>
<td>0.02</td>
<td>12.74</td>
<td>29.58</td>
</tr>
<tr>
<td>2000</td>
<td>2.42</td>
<td>70.74</td>
<td>0.10</td>
<td>18.68</td>
<td>34.16</td>
</tr>
<tr>
<td>3000</td>
<td>5.22</td>
<td>71.76</td>
<td>0.10</td>
<td>21.80</td>
<td>35.36</td>
</tr>
<tr>
<td>4000</td>
<td>4.88</td>
<td>72.30</td>
<td>0.12</td>
<td>22.32</td>
<td>39.08</td>
</tr>
<tr>
<td>5000</td>
<td>5.48</td>
<td>73.08</td>
<td>0.10</td>
<td>23.34</td>
<td>39.76</td>
</tr>
</tbody>
</table>

Figure 9-3 below compares the power of the updated MR at different covariate values.
Figure 9-3: Combined power curve of the Bayesian meta-regression analysis.

When the new trial includes covariates near the weighted mean, i.e. LEV 2000 and 3000, the lowest power of the updated MR is reported. As the covariate moves away from the weighted mean covariate, the power of the updated MR is increasing. This pattern is very similar to that observed in the frequentist MR context as well. Despite being very similar in the overall pattern, the one-to-one comparison of the frequentist and Bayesian power is quite different at each covariate, i.e. the scale on which the increase operates is quite different in the two cases. Figures 9-4 and 9-5 below compare the power of the updated frequentist and Bayesian MR at each covariate level.
Figure 9-4: Comparison of the power of the updated frequentist and Bayesian meta-regression at LEV doses 1000, 2000 and 3000.

**LEV 1000**

The power of both frequentist and Bayesian approaches is non-zero at all sample sizes.

The power of the frequentist approach is substantially larger than the Bayesian counterpart at all sample sizes considered.

**LEV 2000**

The power of both frequentist and Bayesian approaches is zero for sample sizes until 300. Beyond that, the power in the frequentist approach begins to be non-zero and increases with the sample size. However, the power of the Bayesian approach begins non-zero when the sample size reaches about 1000 and stays constant as sample size increases.

**LEV 3000**

When new trial includes a covariate of LEV 3000, the frequentist approach produces zero in all sample sizes.

In contrast, despite being smaller, the comparative Bayesian approach produces non-zero power, which increases as the sample size increases.

Figure 9-5 below compares the power of the updated frequentist and Bayesian MR at LEV 3500 and 4000.
**Figure 9-5**: Comparison of the power of the updated frequentist and Bayesian meta-regression at LEV doses 3500 and 4000.

| LEV 3500 | The power of both frequentist and Bayesian approaches increases along with the sample sizes, when the new trial includes a covariate of LEV 3500. However, the rate at which power increases and the range within which the power fluctuates are higher in the frequentist approach compared to those of the Bayesian approach. |
| LEV 4000 | In LEV 4000 scenario, the power of the Bayesian approach stays rather constant over the range of sample size. In the frequentist approach the power is higher at even smaller sample sizes and is increasing as the sample size increase. |

The section below critically evaluates the results of this chapter.
9.11 Critical evaluation

The Bayesian approach facilitates the process of making predictions, particularly in designing future trials using results from the current synthesis. The MCMC formulations allow a natural way of incorporating uncertainty in the estimation of parameters into the analysis. This is reflected by the wider interval estimations compared to those from the frequentist counterpart.

The accurate estimation of \( \tau^2 \) is paramount as \( \tau^2 \) is used in the process of the prediction of the effect size in a new study. The estimation of parameters in the frequentist model involves the use of the method of moments formulation, which is a frequentist based approach. This approach however assumes that \( \tau^2 \) is known without errors, even though \( \hat{\tau}^2 \) which is estimated from the data acts as a surrogate for \( \tau^2 \) in the analysis.

With the hierarchical nature of the model presented here, the normality assumptions together with the considerable number of trials in any specific MA make MCMC methods particularly appropriate in addressing these scenarios (Sutton, A., Abrams et al. 2000).

The use of the Bayesian MA model benefits from not having to compute the summary measures of the individual studies. Moreover, in the case of zero events, complete events, or perhaps missing events in any of these trials, application of the continuity correction is not required since the computation of individual summary statistics is not required. (Sutton, A. J. and Abrams 2001)
The methods employed here are characterised as hybrid frequentist-Bayesian in nature (Spiegelhalter, D. J., Abrams and Myles 2004), such that the method involves specifying prior distributions for unknown quantities (Bayesian nature) and decision about rejecting the null hypothesis is based on significance tests (Frequentist nature).

The method based on a simulation approach is preferred over the closed form solutions in estimating power because different closed form solutions are required for different outcome/MA combinations. The simulation framework adopted here offers the luxury of using either the frequentist method (considered in Chapter 5) or the Bayesian method (considered here in Chapter 9) of specification and estimation.

The general discussion of the findings of this chapter is in the next section.
9.12 Discussion

9.12.1 Key findings of this chapter
This chapter adopts a Bayesian approach to develop methodologies of designing future trials based on existing evidence. As expected, the power estimates of the updated MA obtained via both the frequentist (section 5.7) and the Bayesian FE (section 9.9.4) methods are approximately equal.

The updated MA using Bayesian RE method produces zero power estimates for all sample sizes considered (sections 9.9.5.2.1 and 9.9.5.2.2) regardless of the prior distribution placed on the heterogeneity.

The power results of the Bayesian and frequentist MR results are broadly similar in the sense that the phenomenon of increasing power as the new trial includes a covariate further from the weighted mean is common to both approaches. However, at individual covariate level, the power of the two approaches is disparate.

9.12.2 Related similar work
The power results of the Bayesian RE meta-analysis found here are inconsistent with the findings of Sutton et al (Sutton, A. J., Cooper et al. 2007), the work of which this chapter extends. Sutton et al reported that the power of the updated RE meta-analysis, when an inverse gamma prior is specified on $\tau^2$, is around 0% and 1.3% when the sample size is 100 and 800 respectively. As
the sample size increases, the power reach a plateau around 5%. Contrastingly, corresponding power reported in this chapter (section 9.9.5.2) is found to be zero at every sample size considered. The differences in the results could be due to the differences in the assumptions of control group events rate in two studies. We took the average control group event rate of existing trials, which is 0.354 and Sutton et al assumed it to be 0.20. To reconcile the differences, a simulation study may be recommended.

Further, these results are different from corresponding frequentist power estimates reported in Chapter 5 as well. Chapter 10 further investigates as to why these results are disparate.

9.12.3 **Critical evaluation of the results**

The Bayesian FE approach specifying non-informative priors on both $\theta$ and $\mu_i$, resembles a frequentist FE model as the data or the likelihood predominantly dictates the shape of the posterior distribution (Smith, T.C., Spiegelhalter and Thomas 1995). Therefore, the resulting Bayesian estimates and intervals will match the frequentist counterparts leading to produce almost identical results provided the sampling variation is negligible.

Nevertheless, resulting power estimates of the updated MA via the frequentist and Bayesian RE models produce disparate results. The Bayesian and frequentist RE methods are primarily different in the way they handle the uncertainty in the estimation of $\tau^2$. The frequentist model assumes that $\tau^2$ is known without errors although it is estimated from the data. The Bayesian MA
explicitly models $\tau^2$ in a unified framework and automatically propagates the estimation of the uncertainty in $\tau^2$ into the analysis. Hence, the phenomenon of yielding zero power estimates is at least partly attributable to the handling of uncertainty in the estimation of $\tau^2$. Section 10.4.1 in Chapter 10 examines the impact of uncertainty in the estimation of $\tau^2$ by manually restricting the uncertainty to specific values in the Bayesian analysis.

The incorporation of the uncertainty in the estimation of $\tau^2$ into this analysis is recommended for two reasons; firstly, because $\tau^2$ is used in the prediction of the new trial and secondly, the uncertainty is higher as the number of studies in the analysis is smaller, which is six trials in the case of MA. Although, the frequentist methods that incorporates the full uncertainty in the estimation of $\tau^2$ into the analysis (Hardy and Thompson 1996) (Biggerstaff and Tweedie 1997) have been available, the Bayesian MCMC methods are preferred over them as the MCMC methods automatically account for the uncertainty in the estimation.

### 9.12.4 Future work

The specification of non-informative priors with no explicit information causes the data to dominate in the determination of the posterior distribution. The posterior distributions of $\theta$ and $\tau$ for which non-informative priors were specified are therefore predominantly data dominated. The existing evidence used here comprises only six trials in the case of MA and eight trials in the case of MR and thus not much information is propagated into the analysis. The uncertainty in the estimation of unknown quantities is substantial despite the fact that it is based on 5000 samples from the posterior using MCMC methods. This raises
the need for informative prior distributions instead of non-informative priors used in the illustrative example.

The specification of informative priors by elicitation of opinion in a multi parameter context is not straightforward. Therefore, we refrain from eliciting any prior distributions in here. This is mainly because of the absence of a mechanism to validate our elicitations. Instead, non-informative priors have been used. As future work, we suggest developing informative priors to be placed on $\theta$ and $\tau$.

The one to one comparison of the Bayesian and frequentist MA that involves a degree of heterogeneity tends to produce discrepant results. Chapter 10 attempts to uncover the reasons for the discrepancies of the findings.
Chapter 9  
Designing new trials using Bayesian evidence synthesis methods

9.13 Conclusions

This chapter develops and implements a methodology to design new trials based on updated MA and MR using Bayesian methods. A simulation-based approach was undertaken in calculating required sample size by achieving certain power in the updated synthesis, including the new trial.

Someone with a degree of scepticism about the use of RE meta-analysis model may use the methodology based on Bayesian or frequentist FE model in designing future trials. The methodology based on the FE Bayesian model is practically advantageous in scenarios where the existing evidence comprises of a set of very similar trials with no hint of heterogeneity. Since both methodologies based on the frequentist and Bayesian FE models reported very similar power estimates researchers in favour of both frequentist, and Bayesian may use this model in practice, provided the FE assumption be justified. Both frequentist and Bayesian MR methods could be used in designing future trials with a specific covariate.

The methodology based on the Bayesian RE model is unlikely to be pragmatically embraced to design new trials unless refinements are made to it, especially eliciting prior distributions and addressing issues of handling the uncertainty of the estimates. Although non-informative priors are used for convenience, eliciting a proper prior may prove advantageous. The posterior distributions of unknown quantities are not impacted enough by the limited information available in the set of six trials. It is argued that, in situations where not enough prior evidence is available, posterior needs to be impacted more by the prior distributions derived using elicitation.
Handling the uncertainty in the estimation of $\tau^2$ is not trivial and should be performed with care. Both the location and the uncertainty of $\tau^2$ for the existing trials may be influential in the design of the next study.

In summary, research focusing on methodological developments to address the uncertainty in the estimation of $\tau^2$ is required before embarking on to design new trials basing the Bayesian RE model. As indicated above, more work is required before Bayesian MA methods can be employed as a basis of designing future studies. The importance and appropriateness of using quantitative methods for designing future studies in the light of evidence accumulated to date should not be underplayed.

As emphasised before, the power estimates from the Bayesian RE results are dramatically differ from that of corresponding frequentist results. The following chapter attempts a reconciliation of the power estimates of Bayesian and frequentist MA approaches.
10.1 Introduction

The power estimates of the random effects models using frequentist (Chapter 5) and Bayesian (Chapter 9) methods turn out to be substantially different. This chapter attempts to reconcile the differences and to investigate the reasons for the discrepancies between the results. Following this introduction, section 10.2 discusses the motivation for the work in this chapter. The objectives of this chapter are set out in section 10.3. Section 10.4 describes the methods employed. Three Bayesian models restricting the uncertainty are designed in section 10.4.1 to reconcile the power estimates of the two approaches. Section 10.4.4 looks at the credible/confidence limits focusing on finding reasons as to why these two approaches produce drastically different estimates. An investigation is conducted in section 10.4.5 into the causes of the bimodality shape of the distribution on the mean of the pooled OR of the updated MA. The results are critically evaluated in section 10.5. The discussion and the conclusion of this chapter are given in section 10.6 and 10.7 respectively.
10.2 Motivation

The power of the updated MA including the new trial reported in frequentist (table 5-3) and Bayesian RE analysis (table 9-3) is considerably different. The power estimates of the updated Bayesian RE meta-analysis at every sample size of the new trial turned out to be zero (figure 9-2), although the magnitude of the power yielded in the frequentist RE meta-analysis model accounted for both within and between study variation is larger and positively correlated with the sample size of the new trial (figure 5-4). This chapter is attempting to reconcile the power estimates obtained via the comparative frequentist and Bayesian methods.

10.3 Objectives

The main objective of this chapter is to identify the causes of and to rationalise the discrepancies in power estimates between the frequentist and the Bayesian methods.

During this process of reconciliation of the discrepancies, a bimodality nature is observed in the distribution of the mean of the pooled OR of the frequentist analysis. This has led the later part of this chapter to perform an investigation into the causes of the said bimodality.

10.4 Methods

To achieve the main objectives set out in section 10.3 above, the following two methods are employed.
• Suggesting several Bayesian models to design new trials with restricted uncertainty in the estimation of the between study variation.

• Comparing confidence limits of the pooled OR from the frequentist approach (The random effects MA approach accounted for both within and between study variations) with corresponding credible limits of the Bayesian approach.

Section 10.4.1 below, describes the process of designing Bayesian models with restricted uncertainty aiming to reconcile the frequentist and the Bayesian results.

### 10.4.1 Designing Bayesian models restricting the uncertainty in the estimation of $\tau$

Before embarking on the exploration into the reasons for the said disagreements, it is worth identifying the aspects that separate the two approaches. They primarily differ with respect to how they handle the uncertainty in the estimation of the between study variation parameter ($\tau^2$).

The Bayesian approach treats $\tau^2$ as unknown quantity with a distribution and accounts for the uncertainty in the estimation of $\tau^2$ by incorporating the prior probability distribution into the analysis. In contrary to that, the frequentist approach treats as if the parameter ($\tau^2$) is exactly known (a point estimate value) even though the estimator ($\hat{\tau}^2$) is used in place, which is estimated from the data without considering the uncertainty in the estimation process.
It is hypothesised that the discrepancies in the power estimates of the updated MA can be attributed to the differences in the approaches in handling the uncertainty in the estimation of $\tau^2$. If the uncertainty is removed in the Bayesian model to a degree that resembles a frequentist model, both approaches should yield approximately equal power estimates. Since the uncertainty in the estimation of $\tau^2$ is given a distributional form, removing the full uncertainty in the Bayesian model is not trivial. However, it is more feasible to restrict the uncertainty as opposed to remove it altogether. In doing so, the aim is to ascertain that the restriction of the uncertainty causes the model to produce Bayesian estimates similar to those from the frequentist approach.

The uncertainty is restricted by centring the prior distribution of $\tau$ to a pre-specified value by setting known values to the upper and lower limits of the uniform distribution. Alternatively, a profile likelihood method which considers the uncertainty in the heterogeneity estimate in a frequentist perspective (Hardy and Thompson 1996) could be used in the comparison.

Three Bayesian models have been designed to include different restriction criterion and power estimates of the updated MA are examined to explore the relation between the uncertainty and the power estimates. The principles and the rational of each model are outlined below.
10.4.1.1 **Model 1 – Prior distribution of $\tau$ is centred on the mean of $\tau$ of the frequentist random effects analysis**

The aim here is to design a Bayesian model analogous to the frequentist model in respect of how the uncertainty in the estimation of between study standard error ($\tau$) is accounted for.

The prior distribution of $\tau$ is centred on the value of $\hat{\tau}$ obtained in the frequentist approach i.e. 0.3270. The lower (0.3200) and the upper (0.3350) limits of the uniform distribution are specified such that the midpoint of the interval represents the $\hat{\tau}$ obtained in the frequentist approach. Equation 10.1 below specifies the distribution of $\tau$.

$$\tau \sim Uniform[0.3200, 0.3350] \quad (10.1)$$

To explore the extent of variations of power to changes to the interval of the uniform distribution, the boundaries of the uniform distribution are set close values i.e. 0.3245 and 0.3295 to narrow the interval. Thus the prior distribution of $\tau$ is specified as below in equation 10.2.

$$\tau \sim Uniform[0.3245, 0.3295] \quad (10.2)$$

10.4.1.2 **Model 2 – Prior distribution of $\tau$ is centred on the mean of $\tau$ of the Bayesian random effects analysis**

This model focuses restricting the uncertainty in the estimation of $\tau$ around the mean value of the posterior distribution of $\tau$ obtained (0.582) in the Bayesian RE analysis. The lower and the upper limits of the uniform distribution, i.e. the
prior distribution of $\tau$ are set values 0.5800 and 0.5850 as specified in equation 10.3.

$$\tau \sim Uniform[0.5800, 0.5850]$$  \hspace{1cm} (10.3)

The limits of the uniform distribution were set further apart (0.5750 and 0.5900) to facilitate an investigation into the behaviour of power to changes in the limits of the uniform distribution as specified in equation 10.4.

$$\tau \sim Uniform[0.5750, 0.5900]$$  \hspace{1cm} (10.4)

10.4.1.3 Model 3 – Prior distribution of $\tau$ is centred on the median of $\tau$ of the Bayesian random effects analysis

This section attempts to design a Bayesian model that limits the uncertainty in the estimation of the $\tau$, to values closer to the median (0.511) of the posterior distribution of $\tau$ obtained in the Bayesian analysis accounted for the full uncertainty.

The midpoint of the interval of the uniform distribution is specified to be 0.511 by setting the lower (0.5094) and the upper (0.5143) limits to locate at equal distance from the centre value. The revised prior distribution of $\tau$ is as specified below in equation 10.5.

$$\tau \sim Uniform[0.5094, 0.5143]$$  \hspace{1cm} (10.5)

To discover the impact on power of varying the interval from which values of $\tau$ is sampled, the interval is broadened as below in equation 10.6 by setting relevant lower and upper limits.

$$\tau \sim Uniform[0.5050, 0.5200]$$  \hspace{1cm} (10.6)
10.4.2 Results

The power results of the Bayesian models designed with different prior distributions of the between study standard deviation for a new trial with 5000 subjects in each arm are shown in table 10-1 below. For completeness and comparability, power estimates obtained via the frequentist framework and the power when accounting the full uncertainty in Bayesian framework are also included in the table.

Table 10-1: Power of the updated MA based on the Bayesian RE model in which the uniform prior of $\tau$ is centred on the

Model 1 – $\hat{\tau}$ obtained in the frequentist MA

Model 2 – Mean of the posterior distribution of the $\tau$ obtained in the Bayesian RE MA

Model 3 – Median of the posterior distribution of the $\tau$ obtained in the Bayesian RE MA

<table>
<thead>
<tr>
<th>Model number</th>
<th>Framework</th>
<th>Prior distribution Of $\tau$ is centred on</th>
<th>Centre value of the prior distribution of $\tau$</th>
<th>Type of Prior distribution of $\tau$</th>
<th>Lower limit of the Uniform Distribution</th>
<th>Upper limit of the uniform distribution</th>
<th>Power of the updated meta-analysis</th>
<th>Gap between the lower and upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequentist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37.8</td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td>Uniform</td>
<td>0.0000</td>
<td>2.0000</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>Bayesian</td>
<td>$\hat{\tau}$ of frequentist</td>
<td>0.3270</td>
<td>Uniform</td>
<td>0.3200 – 0.3350</td>
<td>18.14</td>
<td>0.0150</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Bayesian</td>
<td>$\hat{\tau}$ of frequentist</td>
<td>0.3270</td>
<td>Uniform</td>
<td>0.3245 – 0.3295</td>
<td>17.72</td>
<td>0.0050</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Bayesian</td>
<td>Mean of the Posterior distribution of $\tau$</td>
<td>0.5824</td>
<td>Uniform</td>
<td>0.5800 – 0.5850</td>
<td>3.48</td>
<td>0.0050</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Bayesian</td>
<td>Mean of the Posterior distribution of $\tau$</td>
<td>0.5824</td>
<td>Uniform</td>
<td>0.5750 – 0.5900</td>
<td>3.54</td>
<td>0.0150</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Bayesian</td>
<td>Median of Posterior distribution of $\tau$</td>
<td>0.5110</td>
<td>Uniform</td>
<td>0.5094 – 0.5143</td>
<td>5.42</td>
<td>0.0049</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Bayesian</td>
<td>Median of Posterior distribution of $\tau$</td>
<td>0.5110</td>
<td>Uniform</td>
<td>0.5050 – 0.5200</td>
<td>5.58</td>
<td>0.0150</td>
<td></td>
</tr>
</tbody>
</table>
The frequentist MA approach reported the highest power estimates. The Bayesian approach reported the most power when the point estimate of $\tau$ obtained via frequentist analysis is substituted as the centre value of the uniform prior. The least power is recorded when the mean of the posterior distribution of $\tau$ is used as the centre of the uniform distribution.

### 10.4.3 Critical evaluation

The highest power of 37.8% reported in the frequentist analysis is a sizable difference from the zero power reported in the Bayesian model accounted for the full uncertainty. As the midpoint of the prior distribution of $\tau$ increases, the power of the updated MA is falling. In each model, reducing the information in the prior by broadening the interval between the lower and upper limits has resulted in a slight increase in power.

The forthcoming section 10.4.4 compares the lower and upper confidence limits of the pooled effect size of the updated MA of the frequentist approach with corresponding 2.5% and 97.5% credible limits of the Bayesian approach. The aim is to identify any systematic pattern revealing the cause of the disagreement of power estimates computed via the Bayesian and frequentist RE model.
10.4.4 The comparison of the frequentist confidence limits and Bayesian credible limits of the pooled OR of the updated MA

The power estimates obtained via the frequentist and Bayesian RE models are disparate. This triggers the need for further investigations exploring the causes of discrepancies.

The statistical power increases based on the ability of the hypothesis test to correctly reject the null hypothesis. The association between the confidence interval and the rejection of the null hypothesis (OR=1) is such that the null hypothesis is not rejected when the confidence interval contains the null value of the outcome measure, i.e. OR=1 in this case. The behaviour of the confidence/credible interval, to a certain extent, dictates the shape of the power curves. Hence, the confidence/credible limits of the pooled OR of the updated frequentist/Bayesian MA are compared to discover the reasons for the discrepancies.

The specification of vague prior distributions leads credible and confidence intervals to coincide with each other. In theory, the credible interval within the 2.5% and 97.5% credible limits construct a 95% credible interval coincides with 95% confidence interval in the frequentist analysis, provided vague priors are specified (Spiegelhalter, D. J., Abrams and Myles 2004).

In theory, the mean effect estimates, i.e. the pooled OR (θ), the upper confidence/credible limits and the lower confidence/credible limits obtained from the frequentist and Bayesian methods with vague priors should coincide,
although the two approaches treat $\tau^2$ differently. Overlaid histograms drawn in upcoming sections are based on 5000 iterations to discover the empirical behaviour of these confidence/credible limits, at specified sample sizes of the new trial.

10.4.4.1 The lower confidence / 2.5 % credible limit of the pooled Odds ratio

The section compares the lower confidence limits of the pooled OR ($\theta$) from the frequentist method and the 2.5% credible limits of the pooled OR ($\theta$) from the Bayesian method by way of overlaid histograms. This visual inspection is carried out for three specific sample sizes of the new trial i.e. 1000, 2000 and 5000 in each arm.
Figure 10-1: The overlaid histograms of lower confidence limits and 2.5% credible limits of the pooled OR ($\theta$), when the new trial contains 1000, 2000 and 5000 subjects in each arm.

Sample size 1000

It is observed that the 2.5% credible limits are relatively low than the corresponding lower confidence limits, leading to produce wider credible intervals, given the upper limits remain constant.

Sample size 2000

Similar to in the case of sample size 1000, the mean of the distribution of credible limits are smaller than that of confidence limits, suggesting a wider credible limit, if the upper limit remains constant.

Sample size 5000

The mean of the 2.5% credible limits is relatively lower than their counterpart mean of the lower confidence limits, implying comparatively wider credible intervals.
10.4.4.1.1 Overall comparisons

Table 10-2 below summarises the magnitudes of mean of the distributions of lower confidence limit and 2.5% credible limits of $\theta$ at specific sample sizes of the new trial.

Table 10-2: The summary table of the means of the frequentist lower confidence limits and the Bayesian 2.5% credible limits for different sample sizes of the new trial.

<table>
<thead>
<tr>
<th>Sample size of the new trial</th>
<th>The mean of the distribution</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequentist Lower confidence limit</td>
<td>Bayesian 2.5% credible limit</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.5555304</td>
<td>0.4264554</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.5572581</td>
<td>0.4235775</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>0.5594645</td>
<td>0.4222330</td>
<td></td>
</tr>
</tbody>
</table>

The mean of the 2.5% credible limits are lower than the mean of lower confidence limits of the pooled OR, in all sample sizes considered. As the sample size of the new trial increases, the mean of lower confidence limits are beginning to increase whereas the mean of 2.5% credible intervals start to decrease. This leads confidence intervals to get narrower and credible intervals to get wider, provided the upper limits remain unchanged (The behaviour of the upper confidence and 97.5% credible limits is explored in section 10.4.4.2). The wider intervals are more likely to include the null value and consequently more likely to produce low statistical power.

Section 10.4.4.2 below, compares the distribution of the upper confidence limit of the frequentist approach and the distribution of upper 97.5% credible limits of the Bayesian approach, at three different sample sizes of the new trial.
10.4.4.2 **The upper confidence / 97.5 % credible limit**

The upper confidence limits and the 97.5% credible limits of the pooled OR of the update MA after adding a new trial with specified number of subjects (1000, 2000 and 5000) in each arm are compared and represented graphically in figure 10-2 below.
Figure 10-2: The distributions of the frequentist upper confidence limits and the Bayesian 97.5% credible limits of the $\theta$ of the updated MA, when a new trial of 1000, 2000 and 5000 subjects in each arm is added.

**Sample size 1000**

The mean of upper confidence limits of the frequentist analysis is lower than the mean of the counterpart 97.5% credible limits of the Bayesian approach.

**Sample size 2000**

Similar to when the new trial contains 1000 subjects as illustrated above, the mean of the 97.5% credible limits are greater than that of upper confidence limits reported in the frequentist approach.

**Sample size 5000**

The mean of the upper confidence limits of the pooled effect estimates of the frequentist approach is marginally higher than that of the frequentist counterpart.
10.4.4.2.1 Overall comparisons

The means of distributions of upper confidence/credible limits of $\theta$ at specific sample sizes are summarised in the table 10-3 below.

**Table 10-3**: Summary table of means of the frequentist upper confidence limit and the Bayesian 97.5% credible limit of the $\theta$ of the updated MA, when a new trial of certain size is added to the existing MA.

<table>
<thead>
<tr>
<th>Sample size of the new trial</th>
<th>The mean of the distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequentist Upper confidence limit</td>
</tr>
<tr>
<td>1000</td>
<td>1.108611</td>
</tr>
<tr>
<td>2000</td>
<td>1.104455</td>
</tr>
<tr>
<td>5000</td>
<td>1.104417</td>
</tr>
</tbody>
</table>

The means of the 97.5% credible limits of the pooled effect estimates of the updated MA is larger in Bayesian analysis, compared to corresponding limits in the frequentist analysis, in all sample sizes concerned. As the sample size increases, the mean of the upper confidence limits of the frequentist approach is dropping. However, a systematic pattern is not observed in the Bayesian mean. The results suggest that the Bayesian analysis is associated with wider intervals and consequently lower power.

Section 10.4.4.3 compares the distributions of means of the pooled OR of the frequentist and Bayesian approaches, focusing on rationalising the power discrepancies.
10.4.4.3 The mean of the pooled odds ratio

The aim in this section is to compare the distributions of the mean effect size i.e. the pooled OR ($\theta$) of the updated MA including the new simulated trial obtained via the frequentist and the mean of the posterior distribution of the pooled OR via the Bayesian analysis methods, with 1000, 2000 and 5000 sizes of samples in each arm. This comparison ascertains whether the both analysis methods report approximately the same means or disparate means.
Chapter 10

The comparison of the frequentist and Bayesian results

Figure 10-3: The distribution of the pooled OR after adding a new trial of size 1000, 2000 and 5000 in each arm of the Bayesian and frequentist approaches.

Sample size 1000

The Bayesian and frequentist distributions report approximately equal mean magnitudes of the pooled OR. Both distributions appear to be bimodal, particularly the frequentist distribution. Moreover, the Bayesian distribution shows a larger variance than the frequentist distribution.

Sample size 2000

As in the case of sample size 1000 above, the distribution of pooled OR reported via the two approaches do not substantially differ. The bimodality nature observed when the sample size of the new trial is 1000, persists in here as well. However, the Bayesian distribution in here shows more bimodality than when the size of the new trial is 1000. Furthermore, the spread of the Bayesian distribution is wider than in the frequentist approach.

Sample size 5000

The visual inspection does not suggest a much of a difference between two means obtained via two analysis methods. Both the frequentist and the Bayesian distributions appear to be bimodal as well.
10.4.4.3.1 Overall comparisons

Table 10-4 below summarises the magnitudes of the mean of the pooled OR distributions obtained via the frequentist and the Bayesian approaches for specified sample sizes of the new trial.

Table 10-4: Table comparing the mean of pooled OR distribution reported using frequentist and Bayesian analysis, for specific sample sizes of the new trial.

<table>
<thead>
<tr>
<th>Sample size of the new trial</th>
<th>The mean of the distribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequentist mean of the pooled OR</td>
<td>Bayesian mean of the pooled OR</td>
</tr>
<tr>
<td>1000</td>
<td>0.7817154</td>
<td>0.7922453</td>
</tr>
<tr>
<td>2000</td>
<td>0.7810225</td>
<td>0.7877902</td>
</tr>
<tr>
<td>5000</td>
<td>0.7821078</td>
<td>0.7862028</td>
</tr>
</tbody>
</table>

The mean of the pooled OR in the Bayesian context is slightly larger than that in the frequentist context, in all sample sizes considered. As the sample size increases, the mean of the pooled OR in the Bayesian approach tends to decrease. In the frequentist approach, the mean of the pooled effect estimate does not show an apparent behaviour. The locations of the means of the pooled OR in the updated MA via the frequentist and the Bayesian approaches are approximately equal; suggesting the location of the mean of the pooled effect size is not influential in determining the power.

The bimodality nature of both the frequentist and the Bayesian distributions is highlighted in the results and further investigations are carried out in section 10.4.5 to discover the causes and reasoning for this behaviour.

Following the comparison of the frequentist pooled effect estimates, lower and upper confidence limits with corresponding Bayesian counterparts in this section, forthcoming section 10.4.4.4 critically evaluates the findings.
10.4.4.4 Critical evaluations

The statistical power of the updated MA reported in the Bayesian RE analysis is comparatively lower (effectively zero) than the frequentist RE analysis. When the full uncertainty in the estimation of $\tau^2$ is accounted for the resulting statistical power turns out to be zero. When the uncertainty in $\tau^2$ is not fully accounted for, power increases. This disagreement is therefore partially explained by the uncertainty in the estimation of $\tau^2$.

Since the power and the confidence/credible intervals are associated, exploring the confidence/credible intervals might assist in identifying any potential causes of the disagreement of the power estimates. It is important to discover if the confidence/credible intervals are overlapped to rationalise the disagreement in the power estimates.

The results revealed that the Bayesian credible and the frequentist confidence intervals are centred on the approximately the same location, although the Bayesian intervals are relatively wider than that of the frequentist intervals. This is to reflect the additional uncertainty accounted for in the Bayesian approach.

In all sample sizes concerned, the Bayesian credible intervals of the pooled OR are wider than the frequentist confidence intervals. The wider confidence intervals are likely to include the null value and often result in retaining the null hypothesis. Thus, the statistical power in the Bayesian analysis is more likely to be lesser than that in the frequentist analysis.

As an alternative to the 2.5% and 97.5% credible limits of the posterior distribution, the 2.5% and 97.5% limits of the HPD interval is preferable.
because of it ordinates identical probabilities at either side of the posterior distribution. However, HPD has not been considered here because a HPD interval may be made up of disjoint intervals due to the bimodal nature of the posterior distribution.

As observed in figure 10-3, the distribution of pooled OR of the updated MA is identified to have a bimodality nature in both the frequentist and the Bayesian analysis. The next section is dedicated to explore the causes of this unexpected nature of these distributions.

10.4.5 **Rationalising the bi-modality of the Bayesian and frequentist distributions of the pooled OR of the updated MA**

The shape of the distribution of the pooled OR of the updated MA of sample sizes looked at, i.e. 1000, 2000 and 5000 in each arm, is shown to be bi-modal, for this specific dataset examined (Figure 10-3), particularly the distribution of the frequentist analysis.

In the random effects MA settings, the underlying true effects in each trial are assumed a random sample of a normal distribution. The predictive distribution, from which an effect size of a new trial is drawn, is assumed normally distributed too. However, the resulting pooled OR of the subsequent MA combining existing trials with the new trial is found to be bi-modally distributed.

Upcoming sections investigate the causes for the said behaviour while focusing on determining the circumstances where the said behaviour exists. As the first
step into the investigation, section 10.4.5.1 below explores the distribution of the pooled OR and the pooled log OR of the updated MA, to discover the point at which the bimodality is beginning to appear.

10.4.5.1 Exploring the distributions of the pooled odds ratio and the pooled log odds ratio of the updated MA

This section aims to assess the maximum size of the new trial needs to be added into the existing MA before the bimodality begins to appear in the OR distribution of the updated MA. A series of new trials of size 50, 100, 300 and 500 are added in to the existing MA.

The distributions of the OR and the log OR of the updated MA are monitored using histograms and presented in figure 10-4 below. The OR (0.771) and the log OR (-0.259) of the existing MA are also incorporated in respective plots as reference points.
**Figure 10-4:** The histograms of the probability distribution of the odds ratio and corresponding log odds ratio of the updated MA when a new trial of size 50 and 100 is added to the existing MA.

Figure 10-4 above depicts that both the pooled OR and log OR of the updated MA are normally distributed, when a new trial of size 50 or 100 is added into the existing MA.
Figure 10-5: The histograms of the distributions of the pooled odds ratio and the pooled log odds ratio of the updated MA, following an inclusion of a new trial of size 300 and 500 in each arm.

Figure 10-5 above indicates the departure from normality of both distributions (The pooled OR and the log OR) following the inclusion of a new trial of size 300 in each arm, although the bimodality is not beginning to appear. However, a clear-cut bimodality appears after including a new trial of 500 subjects in each arm. This suggests that the bimodality begins to appear when the new sample size lies between 300 and 500. This range happens to be the same range within which the power of the updated random effects MA in the frequentist approach (section 5.8.11 of Chapter 5) begins to turn positive, suggesting this is a range, which needs further investigations. Figure 10-6 below shows the distributions of the pooled OR of the updated MA for sample sizes 340, 380 and 420 of the new trial.
The bimodality begins to appear when the sample size of the new trial reaches 340. However, bimodality is more apparent at sample size 380 and above. The power of the updated MA is reliant on the fact that the confidence interval of the pooled OR includes the null value, i.e. OR=1. The further the pooled OR from the null value the less likely that the confidence interval includes the null value which in turn increases the power.

At 380, the bimodality is clearly visible and the density around OR of 0.7 increases as well. The confidence intervals of the ORs chosen around this area are less likely to include the null value. Because of the higher density of ORs chosen around this area the power is beginning to become positive. As the sample sizes increases, the bimodality is even more clearly visible leading to the generation of an even larger density of ORs around this area and increases the power gradually.

The next section 10.4.5.2 is dedicated to examine the distributions of treatment group event rates of the new trial, focusing on to identify reasons for the bimodality.
10.4.5.2 Exploring the distributions of the treatment group event rate in the new trial

This section examines the distribution of the event rate in the treatment group of a new trial of size 50, 100, 300, 500, 1000, 2000 and 5000 in one arm. The results are presented in separate histograms in figure 10-7 below.
Chapter 10

The comparison of the frequentist and Bayesian results

Figure 10-7: The histograms of distributions of the event rate in the treatment group of the new trial including a sample of 50, 100, 300, 500, 1000, 2000 and 5000.

Sample size 50

Sample size 100

Sample size 300

Sample size 500

Sample size 1000

Sample size 2000

Sample size 5000
The distribution of the treatment group event rate in a new trial is positively skewed regardless of the number of subjects in the new trial.

This section explores the distributions of the treatment group events rate of a new trial. Section 10.4.5.3 below inspects the distributions of the heterogeneity of the updated MA.

10.4.5.3 Exploring the distributions of the heterogeneity of the updated MA

The aim of this section is to examine the distributions of the heterogeneity of the updated MA, after adding a new trial of size 50, 100, 300, 500, 1000, 2000 and 5000 in each arm, to find out the reasons driving the said bimodality. The heterogeneity of the existing MA is also included in histograms as a reference point.
Figure 10-8: The distribution of the heterogeneity of the updated MA, following the inclusion of a new trial of size 50, 100, 300, 500, 1000, 2000 and 5000 into the existing MA.
Figure 10-8 above indicates that the shape of the distribution of the heterogeneity of the subsequent MA, does not vary with the sample size of the new trial.

**10.4.5.4 Exploring the distributions of the odds ratio and the log odds ratio of the new trial**

This section aims to identify the extent of the variation in the OR and the log OR with the sample size of the new trial, by observing the histograms presented in figure 10-9 below. As a reference point the OR (0.771) and the log OR (-0.259) of the existing MA are included in the plots as well.
Figure 10-9: The histograms of the odds ratio and the log odds ratio of the new trial of sample sizes 50, 100, 300 and 500.
Chapter 10  The comparison of the frequentist and Bayesian results

*Figure 10-10* - *The histograms of the odds ratio and the log odds ratio of the new trial of sample sizes 1000, 2000 and 5000*

The distribution of the OR of the new trial is positively skewed at all sample sizes of the new trial considered. As expected, the log OR of the new trial is normally distributed regardless of the size of the new trial.

Thus far, the distributions of several statistics have been looked at focusing on identifying reasons driving the said bimodality. Sections 10.4.5.5 - 10.4.5.7
upcoming explore the relationship between two variables aiming to indentify any cause of the bimodality.

10.4.5.5 Exploring the scatter plots between the pooled OR of the updated MA and OR of the new trial

The aim here is to comprehend the bimodality nature of the distribution of mean of the pooled OR, from the information derived from exploring the relationship between the pooled OR of the updated MA and the OR of the new trial. The corresponding scatter plots are presented in figure 10-11 below. The OR of the existing MA, i.e. 0.771 is also included in scatter plots as a reference point.
Figure 10-11: The scatter plots of the OR of the new trial against the pooled OR of the updated MA.
As the OR of the new trial increases, the pooled OR of the updated MA increases as well, indicating a positive correlation between the two ORs considered. However, the rate of this increase is lower for new ORs above one, which is the region indicating that antibiotics are harmful. This relationship becomes more apparent for increased sample sizes of the new trial.

In relatively small sample sizes of the new trial, for a given new OR, the magnitude of the pooled OR varies in a broader range in the sense that the number of different values that the pooled OR could take is larger. Conversely, as the sample size of the new trial increases, the said range narrows down and the number of individual different values assigned to the pooled OR is falling.

This section looked at the association between the OR of the new trial and the pooled OR of the updated MA using scatter plots.

### 10.4.5.6 Exploring the scatter plots between the heterogeneity of the updated MA and the OR of the new trial

It is difficult to comprehend the whys and wherefores of the bimodality behaviour in the distribution of the pooled OR of the updated MA, without identifying the association between the heterogeneity of the updated MA and the OR of the new trial. Corresponding scatter plots are presented in figure 10-12 and subsequently will be scrutinised to discover possible reasons explaining the bimodality nature.

The OR and the heterogeneity of the existing MA are also incorporated into each plot as reference points.
Chapter 10  
The comparison of the frequentist and Bayesian results

**Figure 10-12**: The scatter plots of the OR of the new trial against the heterogeneity of the updated MA.

<table>
<thead>
<tr>
<th>Sample Size 50</th>
<th>Sample Size 100</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Scatter Plot" /></td>
<td><img src="image2.png" alt="Scatter Plot" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Size 300</th>
<th>Sample Size 500</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Scatter Plot" /></td>
<td><img src="image4.png" alt="Scatter Plot" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Size 1000</th>
<th>Sample Size 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.png" alt="Scatter Plot" /></td>
<td><img src="image6.png" alt="Scatter Plot" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Size 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7.png" alt="Scatter Plot" /></td>
</tr>
</tbody>
</table>
The figure 10-12 above presents scatter plots between the OR of the new trial and the heterogeneity of the updated MA for several sample sizes of the new trial. The smaller sample sizes of the new trial are associated with lower heterogeneity of the updated MA. This is clearly evident from above figure 10-12 in which the heterogeneity of the updated MA takes values up to 0.6 when the sample size is 50 and up to 0.8 when the sample size is 100 and up to 1 for the rest of the sample sizes.

When new ORs are selected from around the pooled OR of the existing MA, i.e. 0.771, adding larger new trials seems to reduce the heterogeneity of the updated MA. At larger sample sizes of the new trial, as the OR of the new trial deviates from the 0.771, the heterogeneity tends to increase in the updated MA. However, the inclusion of smaller sized new trials with ORs further from mean does not decreases the heterogeneity of the updated MA.

10.4.5.7 Exploring the scatter plots between the pooled OR and the heterogeneity of the updated MA

The association between another combination of variables i.e. the pooled OR and the heterogeneity of the updated MA is explored to discover the causes of the bimodality of the distribution of the pooled OR of the updated MA. As reference points, the OR and the heterogeneity of the existing MA are added in to the plots.
**Figure 10-13:** The scatter plots between the heterogeneity and the pooled OR of the updated MA for various sample sizes of the new trial.
For smaller sample sizes of the new trial, the resulting heterogeneity of the updated MA is lower if the pooled OR of the updated MA is nearer the pooled OR of the existing MA (0.78). However, for small sample sizes of the new trial, even a slighter deviation of the OR of the updated MA from the pooled OR of the existing MA causes the updated MA to become highly heterogeneous. In other words, smaller heterogeneity in smaller sample sizes are associated with pooled ORs closer to the mean of the predictive distribution.

As the sample size of the new trial increases, the low heterogeneity of the updated MA is associated with wider range of updated ORs indicating that the heterogeneity of the updated MA is not necessarily declining after adding new trials with ORs further away from the mean.

The figure 10-14 below, combines all scatter plots shown in above figure 10-13 focusing on to compare and to contrast between each individual plot.
Chapter 10

The comparison of the frequentist and Bayesian results

Figure 10-14: The combined scatter plot between the heterogeneity of the updated MA and the pooled OR of the updated MA for different sample sizes of the new trial.

The heterogeneity of the updated MA remains relatively low for a larger proportion of observations in each sample size. However, as the size of the new trial increases, the range within which the heterogeneity varies is widening.

For smaller sample sizes, the heterogeneity of the updated MA remains lower only when the pooled OR of the new trial is chosen closer to the mean of the predictive distribution of the new trial. In contrast to that, for larger sample sizes of the new trial, the heterogeneity remains low for a wider range of values of the pooled OR of the updated MA. This range is broadening as the sample size of the new trial increases.
10.5 Critical evaluation

As manifested in figures 10-11 and 10-12, the diffused scatter plots of the OR of the new trial against the OR and the heterogeneity of the updated MA turn into more concentrated scatter plots as the sample size of the new trial increases. In other words, the number of values that the OR of the updated MA (figure 10-11) and the heterogeneity of the updated MA (figure 10-12) could take for a given value of the OR of the new trial is reducing as the sample size of the new trial increases. When a larger number of scatter points concentrated into one point, the density at that point increases and subsequently reflected with higher bars in the histograms. The concentration of these scatter points could partly be responsible for the observed bimodality in the distribution of the OR of the updated MA.

As figure 10-11 illustrates, a positive correlation exists in the relationship between the OR of the new trial and the OR of the updated MA, although it is more apparent as the sample size of the new trial increases. However, the correlation is less strong when ORs of the new trial are smaller, i.e. the smaller OR of new trial do not necessarily associate with smaller OR of the updated MA. Nevertheless, as the sample size increases, a smaller OR of the new trial is necessarily associated with a smaller OR of the updated MA (figure 10-11). Further, as the sample size increases, the OR of the new trial clearly dictates the OR of the updated MA indicating the dominance of a larger new trial in the updated MA.

Moreover figure 10-11 shows that the rate at which the OR of the updated MA increases with the OR of new trial is declining as the OR of the new trial
exceeds one, i.e. OR >1, which is the region signifying that the antibiotic is harmful. The updated MA after adding larger new trials does not suggest that the antibiotic is harmful as the OR of the updated MA does not exceed one, except in a very few instance when the sample size of the new trial is 50 and 100 as indicated in figure 10-11.

The figure 10-12 illustrates that adding a smaller new trial does not influence the range within which the heterogeneity of the updated MA fluctuates as much as it does after adding a larger new trial. In other words, the heterogeneity of the updated MA after adding a smaller new trial could take a wider range of values regardless of the value of the OR of the new trial (sample size 50 and 100 of figure 10-12). Figure 10-12 further illustrates that when the OR of the new trial is sampled close to the OR of the existing MA, the heterogeneity of the updated MA begins to concentrate. When added a new trial larger than 300 chosen around the OR of the existing MA, the heterogeneity of the updated MA does not exceed the heterogeneity of the existing MA (sample size 500 and above in figure 10-12). In larger new trial, as the OR of the new trial moves further from the OR of the existing MA, the resulting heterogeneity of the updated MA increases. Figure 10-12 further shows that the variability in the heterogeneity of the updated MA after adding a new trial with an OR around the OR of the existing MA declines as the sample size of the new trial increases.

The combined figure 10-14 indicates that the range within which the heterogeneity of the updated MA fluctuates is increasing as the sample size of the new trial increases. It further suggests that the range within which the OR of the updated MA fluctuates is decreasing as the sample size of the new trial
increases. Moreover, figure 10-14 illustrates that the range within which the minimum heterogeneity of the updated MA is reported broadens as the sample size increases.

The minimum heterogeneity of the updated MA is reported for a variety of ORs of the new trial (figure 10-12) as well as when the OR of the updated MA is exactly equal to the OR of the existing MA (figure 10-13). Because of the dominance of the larger new trial in the updated MA, the heterogeneity in the updated MA remains relatively low for a wide range of ORs of the updated MA selected around the OR of the existing MA, as indicated in sample sizes 500 and above of figure 10-13. This rise in the proportion of observations with low heterogeneity in the updated MA contributes to increasing the power of the updated MA as the sample size of the new trial increases.
10.6 Discussion

This chapter attempts a reconciliation of power estimates obtained from comparative prediction models based on frequentist and Bayesian RE methods. The majority of this chapter commits to discover the reasoning or to rationalising zero power estimates of the Bayesian RE model which Sutton et al did not perform. The difference of two modelling approaches lies in how they treat the uncertainty in the estimation of the heterogeneity in their individual analysis. This forms the basis for this reconciliation. The reconciliation is attempted via developing a series of Bayesian models analogues to corresponding frequentist models.

Further, a comparison between the confidence intervals of the frequentist analysis and corresponding credible intervals of the Bayesian analysis was done in an attempt to explain the aforementioned discrepancies. The later part of this chapter attempts to explore the causes of the bi-modality nature of the distribution of the pooled ORs of the updated MA, obtained in the frequentist analysis.

In the original Bayesian analysis described in Chapter 9, the uniform distribution of $\tau$ (which is the prior distribution placed on $\tau$) has the potential to vary between 0 and 2 leading to account for the full uncertainty around the estimation. Three Bayesian models developed here in this chapter attempt to restrict the uncertainty by centring the uniform distribution around some known specific parameter estimates relevant to this analysis i.e. the mean heterogeneity (0.327) obtained in the frequentist analysis (section 5.8.4.2 of Chapter 5) and the mean (0.5824) and the median (0.511) heterogeneity
obtained in the original Bayesian analysis (section 9.9.5.2 of Chapter 9). Within each of these models, the sensitivity of the limits of the uniform distribution on power is assessed by changing the distance between the upper and lower limits.

The analogous Bayesian model with the frequentist component of heterogeneity (centred on 0.327) substantially increases the power i.e. up to 18.14 (limits are 0.015 apart) and 17.72 (limits are 0.005 apart) compared to zero when the full uncertainty is accounted for. Nonetheless, this is considerably lower when compared to that obtained in the frequentist approach (37.8).

These two models are distinct only with respect to the uncertainty accounted in the estimation of $\tau^2$, in that the Bayesian model includes a very little amount of uncertainty whilst the frequentist model does not account for the uncertainty. Importantly, both models are characterised with the same amount of heterogeneity. Thereby proving the disagreement is more likely due to the additional slightest amount of uncertainty accounted for in the Bayesian model. This concludes that even a slighter amount of uncertainty in the estimation of $\tau^2$ causes to have reduced power.

However, two analogues Bayesian models (centred on 0.5824 and 0.5110) do not increase the power significantly. Yet, the model that centred the uniform distribution around the mean of the heterogeneity of the Bayesian model managed to lift the power from zero, up to 3.48 when the limits are 0.005 apart and up to 3.54 when the limits are 0.015 apart. The model that centred the uniform distribution around the median of the heterogeneity managed to
increase the power from zero up to 5.42 when the limits are 0.005 apart and up
to 5.58 when the limits are 0.015 apart respectively.

The above findings clearly suggest the impact of both the magnitude of and the
uncertainty in the estimation of the heterogeneity on the power of the updated
MA. As expected, the power declines as the magnitude of the heterogeneity
increases. The key finding here is the decline in power as the limits of the
uniform distribution get closer. As a direction of future work, it would be
desirable to estimate the power of a few more different values of \( \tau \) before
concluding that the pattern is systematic. This is not pursued here due to time
constraints. It can be concluded that the location and the uncertainty of \( \tau \),
certainly influence the power.

This study does not consider restricting the uncertainty in the estimation of \( \tau^2 \),
when placing the inverse gamma prior on \( \tau^2 \). Another relevant path of future
work would be to develop methods facilitating the restriction the inverse gamma
prior enabling to take values only in a specified range. This would enable an
insight into the sensitivity of power estimates between different methods.

The 95% confidence interval of the frequentist analysis and 95% credible limit of
the Bayesian counterpart (with vague priors) of the pooled ORs should coincide.
Nonetheless, the 2.5% credible interval is always found to be lower than the
lower confidence limit whilst the 97.5% credible interval is always found to be
larger than the upper confidence limit. This causes the credible intervals to be
wider than the corresponding confidence intervals. This further confirms the
power results found in Chapter 9, as wider credible limits more likely to include
the null value, which gives rise to the decision of not rejecting the null hypothesis and directly contribute towards low power.

The location of the mean of the pooled ORs of the updated MA of the frequentist model does not far off from their Bayesian counterparts. However, the distribution of the pooled ORs of the updated MA tends to turn into a bi-modal distribution as the new trial begins to include over 300 samples. This gives rise to another direction of further investigation aiming to examine the association between interrelated parameter estimates involving the pooled ORs and log ORs of the new trial, the pooled ORs and log ORs of updated MA, treatment group event rate and the heterogeneity of the updated MA.

In the examination of the shape of the distributions, the distributions of the treatment group event rate are found positively skewed for all sample sizes considered. Moreover, the log OR of the new trial is found normally distributed for all sample sizes considered. However, the distributions of the heterogeneity of the updated MA look negatively skewed for all sample sizes considered.

When the sample size of the new trial is smaller, a larger frequency of the ORs of the new trial is more likely to be chosen from the close proximity to the pooled OR of existing MA. Even for new trials with ORs close to the mean of the predictive distribution, the between study variability of the updated MA, varies widely. The distribution of the pooled OR of the updated MA looks normally distributed as well.

However, as the sample size of the new trial increases, the ORs of the new trial and the ORs of the updated MA begins to correlate positively. When a new trial
with a smaller OR is added, the resulting OR of the existing MA is more likely to be smaller and vice versa. This correlation begins to appear when the sample size of new trial is between 300 and more defined in sample sizes above 300. Moreover, when a new trial with OR closer to the mean of the predictive distribution is added the between study variability of the OR of the updated MA remains very low. However, the between study variability of the updated MA begin to increase when a new trial with an OR further from the mean is added.

The investigations carried out in this chapter further aid in rationalising the reasons for obtaining zero power until the sample size of the new trial reaches around 380 in the RE model that accounted for both within and between study variability in the frequentist analysis in Chapter 5.
10.7 Conclusions

The power of the mean test of the updated MA is very reliant on the magnitude and uncertainty in the estimation of the heterogeneity. The illustrative example has shown that even a slightest amount of uncertainty contributes to a larger reduction in power. In addition to that, the magnitude of the heterogeneity is also found influential in determining the power.

The between study variance of the updated MA is heavily depend on the sample size of the new trial. Increased heterogeneity in the updated MA negatively contributes towards power. Adding smaller new trials does not increase the power of the updated MA as the heterogeneity of the updated MA remains high.

Larger trials, when added into the existing MA can dominate the updated MA. The heterogeneity of the updated MA remains low for a wider range of ORs in the updated MA. New larger trials with ORs near the pooled OR of the existing MA could cause to increase the power of the updated MA.

This chapter compares and contrast the disparate power results obtained using the frequentist and Bayesian approaches. The upcoming Chapter 11 presents the final discussion of the thesis as a whole and draws conclusions.
CHAPTER 11

General discussion and conclusions

This thesis has taken the form of a review of published literature, and a survey assessing current practices of using previous evidence in designing new trials, followed by developing the methodological framework guiding the design of a new trial based on evidence synthesis models. In this discussion, the key findings of the thesis are discussed by relating them to relevant published work, the strengths and limitations of the methodology are discussed by critically evaluating the methodology and the directions of potential future work and extensions are set out. The final section of this chapter draws together overall conclusions from the thesis.

11.1 Key findings

The main purpose of medical and health related research should be to put the gained knowledge into future applications. The present work attempts to synthesise all accessible previous evidence addressing a particular research question, and put the knowledge into informing the design of a new trial addressing the same therapeutic question. The methodological framework developed here provides guidance in determining the sample size of a new trial needed to achieve a certain power in the new trial or in the subsequent synthesis.

The survey conducted in Chapter 4 concludes that systematic reviews and meta-analyses are underutilized in the planning and designing stages of new
trials. These findings trigger the need for a methodological framework guiding use previous evidence in designing new trials, particularly the sample size calculations. This framework of methodologies for designing new trials consists of a variety of evidence synthesis methods including MA, indirect comparisons, mixed treatment comparisons and meta-regression. Regardless of the evidence synthesis method adopted to develop the methodology, the power of both the new trial in isolation and the subsequent synthesis depends on the variance of the predictive distribution of the effect size of the new trial.

Throughout this thesis, attention is focused on the power of the updated evidence synthesis model, except in indirect comparison settings, which consider the power of the new head-to-head trial in isolation. The behaviour of the power of a new head-to-head trial with the changes to the variance of the predictive distribution in indirect comparison settings (section 6.7.2 of Chapter 6) is very similar to that of the frequentist MA context that considers the power of the new trial alone (section 5.9 of Chapter 5). In both contexts, the power increases as the variability in the predictive distribution increases.

The behaviour of the power of the mixed treatment comparison meta-analysis with the changes to the variability in the predictive distribution (section 7.7 of Chapter 7) is similar to the context that considers the power of the updated meta-analysis in the frequentist meta-analysis approach (section 5.8 of Chapter 5). In both these instances, the increased variability of the predictive distribution has led to a decrease in power of the updated meta-analysis. Importantly, when the new trial is designed using the FE method, the maximum power is reported in all sample sizes considered.
Chapter 8 of this thesis uses meta-regression methods in designing new trials, when the observed heterogeneity is thought of explained by the study level covariates. The leverage of the new trial is influential in determining the power of the subsequent meta-regression. As shown using the running example, the prediction distribution of the new trial widens as the covariate of the new trial moves away from the weighted mean covariate. The power results of the updated meta-regression indicate that it would be prudent in certain situations to design multiple smaller trials instead of a single large trial from available patients. Despite theoretical recommendations, economic considerations, such as if the benefits of multiple trials should outweigh the cost of designing them need to be accounted for in practice.

Both the frequentist and Bayesian approaches produce approximately equal power in the updated meta-analysis, when the new trial is designed using the FE method. However, they produce markedly different power results when the new trial is designed using the RE method. Following the inclusion of a new trial designed based on RE methods, the Bayesian approach reports zero power at all sample sizes of the new trial considered. This phenomenon does not change with different non-informative prior distributions specified on the standard deviation of random effects.
11.2 Related research

The power of the test of the mean effect in a meta-analysis has been investigated before (Hedges, L. V. and Pigott 2001). In addition to that the power of statistical tests for moderators in meta-analysis has also been explored (Hedges, L.V. and Pigott 2004). Neither of them discussed the specific context of informing the trial design based on the power of the updated meta-analysis, which is the concern throughout this thesis.

As discussed in detail in section 5.12.2 in Chapter 5, the most recent and possibly the most relevant related work to the meta-analysis aspect of present work is that by Roloff and Higgins (Roloff and Higgins 2011). Thorlund and colleagues also proposed an approach to determine the sample size of a future trial in the settings of prospective meta-analysis (Thorlund, Anema and Mills 2010). Hinchliffe et al (Hinchliffe, Crowther et al. 2012) also explored the use of meta-analysis methods to inform the design of studies of diagnostic test accuracy.

Rotondi et al proposed a methodology to inform the design of a new trial based on the updated meta-regression (Rotondi, Donner and Koval 2012). Their work is based on a variance minimisation approach as opposed to the significant test approach we used in Chapter 8. A detailed comparison of their approach with the methodology proposed here is given in section 8.13.2 of Chapter 8.
11.3 **Critical appraisal of the proposed methodology**

Before this methodological framework was developed, the need for this kind of methodology was justified by reviewing the extent of current practices of using previous evidence via a survey. The survey concluded that the previous evidence is underutilized in practice, particularly in informing the design of new trials.

Although systematic reviews often suggest the need for further trials upon obtaining non significant findings, the recommendations are rarely explicit (Roloff, Higgins and Sutton 2012). The use of our approach helps determining when to update the systematic review. For example, in situations where our approach suggests the need of a larger trial to achieve a certain power in the updated synthesis, updating the systematic review upon publishing a small sized clinical trial is not necessarily helpful.

The adherence to this proposed methodology may assist investigators to justify the need for their new trial by referring to what is already known about a particular subject, as required in funding proposals as well as by major journals. In this era of evidence-based medicine, the importance of the evidence from an individual trial alone is increasingly devalued compared with the totality of evidence. A change in the guidelines in which the funders would want to see the extent of the impact of an individual trial on the power of the updated synthesis including the new trial is likely in the future.
Further, this methodology features a unique property in that it considers the power of the subsequent evidence synthesis model including the new trial designed based on the existing evidence. The only exception is in the use of indirect comparison methods to design new head-to-head trial described in Chapter 6, due to the practical limitation in finding a set of existing head-to-head trials into which the new head-to-head trial is added. Considering the advantages of exploring the association of the power of the new trial and the variance of the predictive distribution of the new trial in a methodological perspective, section 5.9 of Chapter 5 examines the power of the new trial in isolation in addition to the power of the updated MA.

MA is often perceived as a retrospective activity, although its importance as a prospective analysis is yet to be recognized. This framework of methodology reinforces the importance of MA as a prospective tool, in planning and designing a new trial in the light of the totality of evidence.

Although this methodological framework is specifically developed here on the OR scale, this could be readily manipulated to accommodate other binary outcome scales, such as risk ratio, by primarily altering the formulae specified in equation 5.1 as required. Moreover, this methodology offers the users the luxury of choosing between the frequentist and Bayesian methods. Depending on the circumstances and the preferences, the choice could be made.

The use of random effects MA has been criticised in the presence of covariates to explain the heterogeneity. Taking this on board, Chapter 8 considers designing new trials using meta-regression models.
Section 11.3.1 **Limitations of the proposed methodology**

When the number of available patients to be included in a trial is known or fixed beforehand, the potential benefits of this proposed framework are limited. The methodology could be used only in determining the power of the MA including the new trial with fixed sample size as opposed to calculating a sample size required to achieve a certain power in the updated synthesis.

Although the Bayesian aspect of trial design explored here includes a sensitivity analysis, it is however limited to the sensitivity of prior distribution of the heterogeneity parameter. Ideally, the sensitivity of other choices of parameters could be evaluated alongside the vulnerability of the results to the changes to assumptions.

Whilst this methodology is explicitly developed to inform two/multi-arm parallel designs, the framework could be extended to be compatible with other complex designs. A careful consideration of parameterization is required when multiple parameters are involved in complex designs.

Whenever a three-arm trial is involved, i.e. both in designing a three-arm trial (section 8.5.4 of Chapter 8) and in splitting a three-arm trial (section 8.9.2 of Chapter 8), this approach treats the three-arm trial as two independent two-arm trials. This is a simplified approach in which the PLC arm is split into multiple components or multiple components of PLC arm are combined into a single component. This approach has inherent shortcomings when none exactly divisible number of events or/and non events exist in the PLC arm, e.g. odds number of events/non events. A more advanced and statistically sound
approach, which accounts for the correlation between two treatments in the three-arm trial when splitting, may enhance the reliability of the results.

11.4 Recommendations and Future work

11.4.1 Future applications

The trials included in the common cold example used in Chapters 5 and 9 spans over almost 60 years. Due to the rapid advances in the therapies and clinical practices over time, the treatment regimes used in earlier trials are not as effective as those in latest trials. Use of such trials in the analysis have been criticised on the inconsistency of the outcome measured (Sutton, A. J., Cooper et al. 2007). While acknowledging the criticism, Sutton et al counteracted by arguing that they consider only the face value of the MA to base the development of new methodology. However, it is recommended that anyone who is trying to use this methodology in practice should assess the validity of the previous evidence before embarking on designing trials.

A general recommendation with respect to this methodology is that in the synthesis of outcome measures from different trials, the comparability of outcome measures of each trial should carefully be ascertained. In each trial the outcome measures are tailored to their own requirements. Due to this diverse and sparse nature of the outcome measures, it is paramount to ensure the comparability of the outcome measures before the synthesis is taken place.
This methodology could retrospectively be applied to a few MA contexts available in the literature, focusing on identifying the point at which no more new trials are required addressing the same therapeutic question. Moreover, this process facilitates computing the power of each new trial in the context of subsequent MA, and advising the trialists on the number of additional patients required to make the subsequent MA significance.

In the direction of future work, it is aimed to make this methodology accessible to non-statisticians by developing a sophisticated user-friendly purpose-built software package that embeds all computer programs developed during the course of this thesis.

### 11.4.2 Future methodological aspect

The methodology developed in this thesis uses the observed heterogeneity estimate derived from data in place of the population parameter. When the number of trials included in the synthesis is limited, the validity of this approach should be evaluated via a sensitivity analysis. The robustness of the models developed here to violations of assumptions is an important pathway of future research in the field.

Particularly in the Bayesian framework, when the number of trials in the synthesis is small, the likelihood is less influential in the estimation of the heterogeneity parameter and results yielded might be very sensitive to the specified priors. Future work could be to conduct a sensitivity analysis to observe how the inferences are influenced by changes in prior distributions (Sutton, A. J. and Abrams 2001).
In addition to that, this framework could be expandable to be able to deal with trials reporting multiple outcomes. The methodology would assist in designing new trials by accounting the correlation between the multiple outcome parameters using multivariate MA models.
11.5 Conclusions

This thesis has illustrated and critically explored the potential of various forms of MA and regression models in informing the design of future trials, particularly the sample size calculations. The results produced and the recommendations outlined in this thesis are primarily based on the example datasets used. Particularly, the recommendations to design new placebo controlled trial should be interpreted with care. Placebo-controlled trials are used as existing evidence and hence naturally led to recommend designing new placebo-controlled trials.

It is appreciated that patients' welfare is the utmost priority and recommendation of new placebo-controlled trials in the presence of active treatment is unethical. Therefore, application of this methodological framework in a variety of examples is required before more generalised conclusions are made.

In conclusion, on the basis of results obtained here based on power of updated meta-analyses, frequentist methods are preferable over Bayesian methods in designing future trials using previous evidence. Whenever the assumption of common effect holds, the FE methods should be used in favour of the RE model in designing new trials using previous evidence. The approach based on meta-regression methods should be used instead of MA methods, if the information about covariates is available. In situation where direct evidence is not available new trials could be designed using indirect comparison methods. In addition to that, if both direct and indirect evidence is available, new trials could be designed based on mixed treatment comparison methods.

We encourage investigators to consider future meta-analyses when planning and designing a new trial. However, it is immensely challenging to persuade
trialists into the habit of considering the future MA before planning and
designing of future trials. It is hoped that this framework of methodologies
together with that by Roloff et al (Roloff and Higgins 2011) and Rotondi et al
(Rotondi, Donner and Koval 2012) lay the foundation towards changing trialist’s
perception on designing new trials in the totality of evidence. Simulation-based
power computation within MA is an area of active research. We look forward to
see further developments of related methodologies in the future.
Appendixes

Additional information in relation to the survey reported in chapter 4

Appendix A1: The list of trial reports included in the first sample of the survey


A2 Darboe M K, Thurnham D I et al, Effectiveness of an early supplementation scheme of high-dose vitamin A versus standard WHO protocol in Gambian mothers and infants: a randomised controlled trial. The Lancet, 2007; 9579 2088-2096


A6 Coombes R C, Kilburn L S et al, Survival and safety of exemestane versus tamoxifen after 2–3 years' tamoxifen treatment (Intergroup Exemestane Study): a randomised controlled trial, The Lancet, 9561, 559-570

A7 Smith I, Procter M et al, 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial, The Lancet, 9555, 29-36
**Appendix A2: The list of trials included in the second sample of the survey**

A8 Kinmonth A, Wareham NJ et al, Efficacy of a theory-based behavioural intervention to increase physical activity in an at-risk group in primary care (ProActive UK): a randomised trial, The Lancet, 9606, 41-48


Bibliography


epilepsy, cognitive disorders, or an anxiety disorder during clinical trials."


Tonner, F. (2008). "A double-blind, randomized, placebo-controlled 5 parallel groups, confirmatory trial on the efficacy and safety of L059 (levetiracetam) used as add-on therapy in patients from 16 to 65 years with epilepsy with partial onset seizures under treatment with anti-epileptic drugs."


