Meta-Analysis of Controlled Clinical Trials

Anne Whitehead

Medical and Pharmaceutical Statistics Research Unit,

The University of Reading, UK

JOHN WILEY & SONS, LTD

Meta-Analysis of Controlled Clinical Trials

Statistics in Practice

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above area. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Meta-Analysis of Controlled Clinical Trials

Anne Whitehead

Medical and Pharmaceutical Statistics Research Unit,

The University of Reading, UK

JOHN WILEY & SONS, LTD

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-98370-5

Typeset in 10/12pt Photina by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Biddles Ltd, Guildford, Surrey
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

To John

Contents

Preface xiii

1 Introduction 1

1.1 The role of meta-analysis 1
1.2 Retrospective and prospective meta-analyses 3
1.3 Fixed effects versus random effects 5
1.4 Individual patient data versus summary statistics 6
1.5 Multicentre trials and meta-analysis 7
1.6 The structure of this book 8

2 Protocol development 11

2.1 Introduction 11
2.2 Background 12
2.3 Objectives 12
2.4 Outcome measures and baseline information 13
2.5 Sources of data 13
2.6 Study selection 14
2.7 Data extraction 15
2.8 Statistical analysis 16
2.8.1 Analysis population 16
2.8.2 Missing data at the subject level 17
2.8.3 Analysis of individual trials 18
2.8.4 Meta-analysis model 19
2.8.5 Estimation and hypothesis testing 19
2.8.6 Testing for heterogeneity 19
2.8.7 Exploration of heterogeneity 20
2.9 Sensitivity analyses 20
2.10 Presentation of results 21

3 Estimating the treatment difference in an individual trial 23

3.1 Introduction 23
3.2 Binary data 25
3.2.1 Example: Stroke in hypertensive patients 25
3.2.2 Measurement of treatment difference 25

3.3 Survival data 32
3.3.1 Example: Mortality following myocardial infarction 32
3.3.2 Measurement of treatment difference 33
3.4 Interval-censored survival data 38
3.4.1 Example: Ulcer recurrence 38
3.4.2 Measurement of treatment difference 39
3.5 Ordinal data 42
3.5.1 Example: Global impression of change in Alzheimer's disease 42
3.5.2 Measurement of treatment difference 42
3.6 Normally distributed data 49
3.6.1 Example: Recovery time after anaesthesia 49
3.6.2 Measurement of treatment difference 50

4 Combining estimates of a treatment difference across trials 57

4.1 Introduction 57
4.2 A general fixed effects parametric approach 58
4.2.1 A fixed effects meta-analysis model 58
4.2.2 Estimation and hypothesis testing of the treatment difference 58
4.2.3 Testing for heterogeneity across studies 60
4.2.4 Obtaining the statistics via weighted least-squares regression 61
4.2.5 Example: Stroke in hypertensive patients 61
4.2.6 Example: Mortality following myocardial infarction 69
4.2.7 Example: Ulcer recurrence 73
4.2.8 Example: Global impression of change in Alzheimer's disease 78
4.2.9 Example: Recovery time after anaesthesia 82
4.3 A general random effects parametric approach 88
4.3.1 A random effects meta-analysis model 88
4.3.2 Estimation and hypothesis testing of the treatment difference 88
4.3.3 Estimation of τ^2 using the method of moments 90
4.3.4 Obtaining the statistics via weighted least-squares regression 91
4.3.5 Example: Mortality following myocardial infarction 91
4.3.6 Example: Global impression of change in Alzheimer's disease 93
4.3.7 Example: Recovery time after anaesthesia 93
4.3.8 A likelihood approach to the estimation of τ^2 94
4.3.9 Allowing for the estimation of τ^2 97

5 Meta-analysis using individual patient data 99

5.1 Introduction 99
5.2 Fixed effects models for normally distributed data 100
5.2.1 A fixed effects meta-analysis model 100
5.2.2 Estimation and hypothesis testing 101
5.2.3 Testing for heterogeneity in the absolute mean difference across studies 103
5.2.4 Example: Recovery time after anaesthesia 103
5.2.5 Modelling of individual patient data versus combining study estimates 105
5.2.6 Heterogeneity in the variance parameter across studies 105
5.3 Fixed effects models for binary data 107
5.3.1 A fixed effects meta-analysis model 107
5.3.2 Estimation and hypothesis testing 108
5.3.3 Testing for heterogeneity in the log-odds ratio across studies 109

5.3.4 Example: Stroke in hypertensive patients 110
5.3.5 Modelling of individual patient data versus combining study estimates 110
5.4 Fixed effects models for ordinal data 111
5.4.1 A fixed effects meta-analysis model 111
5.4.2 Estimation and hypothesis testing 113
5.4.3 Testing for heterogeneity in the log-odds ratio across studies 115
5.4.4 Example: Global impression of change in Alzheimer's disease 116
5.4.5 Modelling of individual patient data versus combining study estimates 117
5.4.6 Testing the assumption of proportional odds between treatments 117
5.4.7 A proportional odds model for studies and treatments 119
5.5 Fixed effects models for survival data 120
5.5.1 A fixed effects meta-analysis model 120
5.5.2 Estimation and hypothesis testing 121
5.5.3 Testing for heterogeneity in the log-hazard ratio across studies 122
5.5.4 Example: Mortality following myocardial infarction 122
5.5.5 Modelling of individual patient data versus combining study estimates 123
5.5.6 Testing the assumption of proportional hazards between treatments 124
5.5.7 A proportional hazards model for studies and treatments 124
5.6 Fixed effects models for interval-censored survival data 126
5.6.1 A fixed effects meta-analysis model 126
5.6.2 Estimation and hypothesis testing 127
5.6.3 Testing for heterogeneity in the log-hazard ratio across studies 127
5.6.4 Example: Ulcer recurrence 128
5.6.5 Modelling of individual patient data versus combining study estimates 128
5.6.6 Testing the assumption of proportional hazards between treatments across timepoints 129
5.6.7 A proportional hazards model for studies and treatments 130
5.7 The treatment difference as a random effect 131
5.8 Random effects models for normally distributed data 131
5.8.1 A random effects meta-analysis model 131
5.8.2 Estimation and hypothesis testing 132
5.8.3 Example: Recovery time after anaesthesia 133
5.8.4 The connection between the multilevel model and the traditional mixed effects linear model 134
5.9 Random effects models for binary data 136
5.9.1 A random effects meta-analysis model 136
5.9.2 Estimation and hypothesis testing 136
5.9.3 Example: Pre-eclampsia 139
5.10 Random effects models for other data types 142
5.10.1 A random effects meta-analysis model for ordinal data 142
5.10.2 Example: Global impression of change in Alzheimer's disease 143
5.11 Random study effects 144
5.11.1 Random study and study by treatment effects: normally distributed data 145
5.11.2 Example: Recovery time after anaesthesia 146
5.11.3 Random study and study by treatment effects: other data types 147
5.12 Comparisons between the various models 147

6 Dealing with heterogeneity 151

6.1 Introduction 151
6.2 The use of a formal test for heterogeneity 152
6.3 The choice between a fixed effects and a random effects model 153

6.4 When not to present an overall estimate of treatment difference 154
6.5 The choice of an appropriate measure of treatment difference 156
6.6 Meta-regression using study estimates of treatment difference 157
6.6.1 Example: Global impression of change in Alzheimer's disease 160
6.6.2 Example: Recovery time after anaesthesia 161
6.6.3 Extension to study estimates of treatment difference from subgroups 163
6.7 Patient-level covariates 165
6.7.1 Adjustment for imbalance in prognostic factors 165
6.7.2 Investigation of potential sources of heterogeneity 166
6.7.3 Example: Global impression of change in Alzheimer's disease 167
6.7.4 Meta-regression using individual patient data 168
6.7.5 Example: Recovery time after anaesthesia 168
6.8 An investigation of heterogeneity: Aspirin in coronary heart disease 170
6.9 A strategy for dealing with heterogeneity 174

7 Presentation and interpretation of results 175

7.1 Introduction 175
7.2 Structure of a report 176
7.2.1 Introduction 176
7.2.2 Methods 176
7.2.3 Results 178
7.2.4 Discussion 181
7.3 Graphical presentation 182
7.3.1 A confidence interval plot 183
7.3.2 A radial plot 186
7.4 Clinically useful measures of treatment difference 189
7.4.1 Simple transformations of the treatment difference parameter 190
7.4.2 Probability of doing better on treatment than on control 192
7.4.3 The number needed to treat 194

8 Selection bias 197

8.1 Introduction 197
8.2 An investigation of publication bias: Intravenous magnesium following acute myocardial infarction 199
8.3 A funnel plot 199
8.4 Statistical methods for the detection and correction of publication bias 205
8.4.1 A test of funnel plot asymmetry 205
8.4.2 Rosenthal's file-drawer method 208
8.4.3 Models for the probability of selection 210
8.5 Bias due to selective reporting within studies 213

9 Dealing with non-standard data sets 215

9.1 Introduction 215
9.2 No events in treatment arms of individual trials 216
9.3 Different rating scales or methods of assessment across trials 220
9.4 Different times of assessment across trials 225
9.5 Combining trials which report different summary statistics 228
9.5.1 Continuous outcomes 228
9.5.2 Ordinal data 231
9.5.3 Survival data 233

9.6 Imputation of the treatment difference and its variance 233
9.6.1 Absolute mean difference for continuous outcomes 233
9.6.2 The log-hazard ratio for survival data 235
9.7 Combining summary statistics and individual patient data 236
9.8 Combining *p*-values 237

10 Inclusion of trials with different study designs 241

10.1 Introduction 241
10.2 More than two treatment groups 242
10.2.1 A fixed effects meta-analysis model 242
10.2.2 A random effects meta-analysis model 243
10.2.3 Random study effects 244
10.2.4 Example: First bleeding in cirrhosis 245
10.3 Dose–response relationships 249
10.4 Multicentre trials 253
10.5 Cross-over trials 254
10.6 Sequential trials 255

11 A Bayesian approach to meta-analysis 259

11.1 Introduction 259
11.2 A Bayesian approach to the random effects model for study estimates 261
11.3 Choice of the prior distribution 263
11.4 Implementation using the BUGS software 265
11.4.1 Example: Recovery time after anaesthesia 267
11.5 Bayesian meta-regression 268
11.6 A Bayesian random effects model based on individual patient data 270
11.6.1 Normally distributed data 271
11.6.2 Binary data 273
11.6.3 Ordinal data 274
11.6.4 Study-level and patient-level covariates 276
11.6.5 Random study effects 276
11.7 Incorporating data from other treatment comparisons 279
11.8 An empirical prior distribution for the heterogeneity parameter 282

12 Sequential methods for meta-analysis 285

12.1 Introduction 285
12.2 A proactive cumulative meta-analysis 286
12.2.1 Choice of a sequential design 287
12.2.2 A fixed effects model 291
12.2.3 A random effects model 292
12.2.4 Example: The triangular test for a primary efficacy outcome 293
12.2.5 Estimation of the heterogeneity parameter 296
12.3 A reactive cumulative meta-analysis 296
12.3.1 Example: Endoscopic haemostasis for bleeding peptic ulcers 297
12.3.2 Alternative approaches to a formal stopping rule 303

Appendix: Methods of estimation and hypothesis testing 305

A.1 Introduction 305
A.2 The method of least squares 306

A.3 The method of weighted least squares 308
A.4 Iterative maximum likelihood estimation 308
A.5 Likelihood, efficient score and Fisher's information 311
A.6 Iteratively weighted least squares 312
A.7 Maximum likelihood methods for general linear mixed models 314
A.8 Iterative generalized least squares for normally distributed data 316
A.9 Marginal quasi-likelihood and penalized quasi-likelihood methods for discrete data 317

References **321**

Index **329**

Preface

Since the 1980s there has been an upsurge in the application of meta-analysis to medical research. Over the same period there have been great strides in the development and refinement of the associated statistical methodology. These developments have mainly been due to greater emphasis on evidence-based medicine and the need for reliable summaries of the vast and expanding volume of clinical research. Most meta-analyses within the field of clinical research have been conducted on randomized controlled trials, and the focus of this book is on the planning, conduct and reporting of a meta-analysis as applied to a series of randomized controlled trials.

There is wide variation in the amount and form of data which might be available for a meta-analysis. At one extreme lie individual patient data and at the other just a *p*-value associated with each test of the treatment difference. Consequently, a number of different approaches to the conduct of a meta-analysis have been developed, and this has given the impression that the methodology is a collection of distinct techniques. My objective has been to present the various approaches within a general framework, enabling the similarities and differences between the available techniques to be demonstrated more easily. In addition, I have attempted to place this general framework within mainstream statistical methodology, and to show how meta-analysis methods can be implemented using general statistical packages. Most of the analyses presented in this book were conducted using the standard statistical procedures in SAS. Other statistical packages, namely MLn, BUGS and PEST, were used for the implementation of some of the more advanced techniques.

In this book, the meta-analysis techniques are described in detail, from their theoretical development through to practical implementation. Emphasis is placed on the consequences of choosing a particular approach and the interpretation of the results. Each topic discussed is supported by detailed worked examples. The example data sets and the program code may be downloaded from either the Wiley website or my own (for details, see Section 1.6).

Meta-analyses have often been performed retrospectively using summary statistics from reports of individual clinical trials. However, the advantages of prospectively planning a meta-analysis are now being recognized. The advantages of using individual patient data are also well accepted. The techniques

covered in the book include those for conducting prospectively planned meta-analyses as well as retrospective meta-analyses. Methods based on individual patient data are included, as well as those based on study summary statistics. This book will be of relevance to those working in the public sector and in the pharmaceutical industry.

This book is based on a short course which has been presented numerous times to practicing medical statisticians over the last ten years and has also been influenced by my involvement in several large meta-analyses. I am grateful to colleagues with whom I have undertaken collaborative research, in particular, Andrea Bailey, Jacqueline Birks, Nicola Bright, Diana Elbourne, Julian Higgins, Rumana Omar, Rebecca Turner, Elly Savaluny, Simon Thompson and John Whitehead.

I am grateful to John Lewis, Stephen Senn, Sue Todd, John Whitehead and Paula Williamson for providing helpful comments and suggestions on earlier drafts of the book.

Anne Whitehead
Reading
2002

1

Introduction

1.1 THE ROLE OF META-ANALYSIS

Meta-analysis was defined by Glass (1976) to be 'the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings'. Although Glass was involved in social science research, the term 'meta-analysis' has been adopted within other disciplines and has proved particularly popular in clinical research. Some of the techniques of meta-analysis have been in use for far longer. Pearson (1904) applied a method for summarizing correlation coefficients from studies of typhoid vaccination, Tippet (1931) and Fisher (1932) presented methods for combining *p*-values, and Yates and Cochran (1938) considered the combination of estimates from different agricultural experiments. However, the introduction of a name for this collection of techniques appears to have led to an upsurge in development and application.

In the medical world, the upsurge began in the 1980s. Some of the key medical questions answered by meta-analyses at this time concerned the treatment of heart disease and cancer. For example, Yusuf *et al.* (1985) concluded that long-term beta blockade following discharge from the coronary care unit after a myocardial infarction reduced mortality, and the Early Breast Cancer Trialists' Collaborative Group (1988) showed that tamoxifen reduced mortality in women over 50 with early breast cancer. By the 1990s published meta-analyses were ubiquitous. Chalmers and Lau (1993) claimed: 'It is obvious that the new scientific discipline of meta-analysis is here to stay'. They reported a rise in the number of publications of meta-analyses of medical studies from 18 in the 1970s to 406 in the 1980s. Altman (2000) noted that Medline contained 589 such publications from 1997 alone.

The rapid increase in the number of meta-analyses being conducted during the last decade is mainly due to a greater emphasis on evidence-based medicine and the need for reliable summaries of the vast and expanding volume of clinical research. Evidence-based medicine has been defined as 'integrating individual clinical expertise with the best available external clinical evidence from systematic research' (Sackett *et al.*, 1997). A systematic review of the relevant external evidence provides a framework for the integration of the research, and meta-analysis offers a quantitative summary of the results. In many cases a systematic review will include a meta-analysis, although there are some situations when

this will be impossible due to lack of data or inadvisable due to unexplained inconsistencies between studies.

The Cochrane Collaboration, launched in 1993, has been influential in the promotion of evidence-based medicine. This international network of individuals is committed to preparing, maintaining and disseminating systematic reviews of research on the effects of health care. Their reviews are made available electronically in the Cochrane Database of Systematic Reviews, part of the Cochrane Library (http://www.update-software.com/cochrane).

Within the pharmaceutical industry, meta-analysis can be used to summarize the results of a drug development programme, and this is recognized in the International Conference on Harmonization (ICH) E9 guidelines (ICH, 1998). In accordance with ICH E9, meta-analysis is understood to be a formal evaluation of the quantitative evidence from two or more trials bearing on the same question. The guidelines indicate that meta-analysis techniques provide a useful means of summarizing overall efficacy results of a drug application and of analysing less frequent outcomes in the overall safety evaluation. However, there is a warning that confirmation of efficacy from a meta-analysis only will not usually be accepted as a substitute for confirmation of efficacy from individual trials. Certainly the magnitude of the treatment effect is likely to be an important factor in regulatory decision-making. If the treatment effect is smaller than anticipated, then statistical significance may not be reached in the individual trials. Even if statistical significance is reached in the meta-analysis, the magnitude of the treatment effect may not be *clinically* significant, and thus be considered insufficient for approval.

Fisher (1999) considered the two conditions under which one large trial might substitute for the two controlled trials usually required by the Food and Drug Administration (FDA) in the USA. The first relates to the strength of evidence for demonstrating efficacy. He showed that if the evidence required from the two controlled trials is that they should each be statistically significant at the two-sided 5% significance level, then the same strength of evidence is obtained from one large trial if it is statistically significant at the two-sided 0.125% level. The same type of argument could be applied to combining trials in a meta-analysis. It would seem reasonable to set a more stringent level of statistical significance corresponding to proof of efficacy in a meta-analysis than in the individual trials.

The second condition discussed by Fisher relates to evidence of replicability, and he proposes criteria which need to be met by the one large trial. A meta-analysis will always involve at least two trials, and it will be important to assess the consistency of the results from the individual trials. The extent of any inconsistencies amongst the trials will be influential in the choice of model for the meta-analysis and in the decision whether to present an overall estimate. These issues are discussed in detail in Chapter 6 of this book.

A recent 'Points to Consider' document (Committee for Proprietary Medicinal Products, 2001) has provided guidance on when meta-analyses might usefully be undertaken. Reasons include the following:

- To provide a more precise estimate of the overall treatment effects.
- To evaluate whether overall positive results are also seen in pre-specified subgroups of patients.
- To evaluate an additional efficacy outcome that requires more power than the individual trials can provide.
- To evaluate safety in a subgroup of patients, or a rare adverse event in all patients.
- To improve the estimation of the dose-response relationship.
- To evaluate apparently conflicting study results.

There is much to be gained by undertaking a meta-analysis of relevant studies before starting a new clinical trial. As Chalmers and Lau (1993) note, this allows investigators to ascertain what data are needed to answer the important questions, how many patients should be recruited, and even whether a new study is unnecessary because the questions may have already been answered. Meta-analysis also has a useful role to play in the generation of hypotheses for future studies.

The conduct of a meta-analysis requires a team, which should include both statisticians and knowledgeable medical experts. Whilst the statistician is equipped with the technical knowledge, the medical expert has an important role to play in such activities as identifying the trials, defining the eligibility criteria for trials to be included, defining potential sources of heterogeneity and interpreting the results.

Most meta-analyses within the field of medical research have been conducted on randomized controlled trials, and this is the focus of this book. Other application areas include epidemiological studies and diagnostic studies. The special problems associated with observational studies are outside the scope of this book, and the interested reader is referred to Chapter 16 of Sutton *et al.* (2000) and Chapters 12–14 of Egger *et al.* (2001).

Over the last twenty years there have been great strides in the development and refinement of statistical methods for the conduct of meta-analyses, as illustrated in the books by Sutton *et al.* (2000) and Stangl and Berry (2000). A number of different approaches have been taken, giving the impression that the methodology is a collection of distinct techniques. The present book is self-contained and describes the planning, conduct and reporting of a meta-analysis as applied to a series of randomized controlled trials. It attempts to present the various approaches within a general unified framework, and to place this framework within mainstream statistical methodology.

1.2 RETROSPECTIVE AND PROSPECTIVE META-ANALYSES

Meta-analyses are often performed retrospectively on studies which have not been planned with this in mind. In addition, many are based on summary statistics

which have been extracted from published papers. Consequently, there are a number of potential problems which can affect the validity of such meta-analyses.

A major limitation is that a meta-analysis can include only studies for which relevant data are retrievable. If only published studies are included, this raises concern about publication bias, whereby the probability of a study being published depends on the statistical significance of the results. Even if a study is published, there may be selective reporting of results, so that only the outcomes showing a statistically significant treatment difference are chosen from amongst the many analysed. If the outcomes of interest have not been defined or recorded in the same way in each trial, it may not be appropriate or possible to combine them. Even if identical outcomes have been recorded in each trial, the way in which the summary statistics have been calculated and reported may differ, particularly with regard to the choice of the subjects included and the mechanism of dealing with missing values. Matters can be improved if time and effort are devoted to obtaining data from all (or nearly all) of the randomized trials undertaken, irrespective of their publication status. Retrieving individual patient data from trial investigators is especially advantageous.

Typically, the objective of a meta-analysis is to estimate and make inferences about the difference between the effects of two treatments. This involves choosing an appropriate measure of the treatment difference, for example the log-odds ratio for binary data or the difference in means for normally distributed data, and calculating individual study estimates and an overall estimate of this difference. In a retrospective meta-analysis the available studies may vary in design, patient population, treatment regimen, primary outcome measure and quality. Therefore, it is reasonable to suppose that the true treatment difference will not be exactly the same in all trials: that is, there will be heterogeneity between trials. The effect of this heterogeneity on the overall results needs to be considered carefully, as discussed by Thompson (1994). Great care is needed in the selection of the trials to be included in the meta-analysis and in the interpretation of the results.

Prospectively planning a series of studies with a view to combining the results in a meta-analysis has distinct advantages, as many of the problems associated with retrospective meta-analyses then disappear. The individual trial protocols can be designed to be identical with regard to the collection of data to be included in the meta-analysis, and individual patient data can be made available.

In drug development, a co-ordinated approach to the trial programme, in which meta-analyses are preplanned, would seem to be a natural way to proceed. The results of a meta-analysis will be more convincing if it is specified prior to the results of any of the individual trials being known, is well conducted and demonstrates a clinically relevant effect.

Within the public sector, collaborative groups are beginning to form in order to conduct prospective meta-analyses. For example, the Cholesterol Treatment Trialists' Collaboration (1995) reported on their protocol for conducting an overview of all the current and planned randomized trials of cholesterol treatment regimens. In such cases it is unlikely that the meta-analysis can be planned before

the start of any of the trials, but certainly the preparation of a protocol prior to the analysis of any of them offers considerable advantages.

The conduct of both retrospective and prospective meta-analyses will be discussed in this book. Many of the analysis methods are common to both, although methodological difficulties tend to be fewer and more manageable for the prospective meta-analysis.

1.3 FIXED EFFECTS VERSUS RANDOM EFFECTS

One of the controversies relating to meta-analysis has concerned the choice between the fixed effects model and the random effects model for providing an overall estimate of the treatment difference. The topic has usually been discussed in the context of a meta-analysis in which the data consist of trial estimates of the treatment difference together with their standard errors. In the fixed effects model, the true treatment difference is considered to be the same for all trials. The standard error of each trial estimate is based on sampling variation within the trial. In the random effects model, the true treatment difference in each trial is itself assumed to be a realization of a random variable, which is usually assumed to be normally distributed. As a consequence, the standard error of each trial estimate is increased due to the addition of this between-trial variation.

The overall estimate of treatment difference and its confidence interval based on a fixed effects model provide a useful summary of the results. However, they are specific to the particular trials included in the meta-analysis. One problem is that they do not necessarily provide the best information for determining the difference in effect that can be expected for patients in general. The random effects model allows the between-trial variability to be accounted for in the overall estimate and, more particularly, its standard error. Therefore, it can be argued that it produces results which can be considered to be more generalizable. In principle, it would seem that the random effects model is a more appropriate choice for attempting to answer this question. However, there are some concerns regarding the use of the random effects model in practice. First, the random effects model assumes that the results from the trials included in the meta-analysis are representative of the results which would be obtained from the total population of treatment centres. In reality, centres which take part in clinical trials are not chosen at random. Second, when there are only a few trials for inclusion in the meta-analysis, it may be inappropriate to try to fit a random effects model as any calculated estimate of the between-study variance will be unreliable. When there is only one available trial, its analysis can only be based on a fixed effects model.

When there is no heterogeneity between trials both models lead to the same overall estimate and standard error. As the heterogeneity increases the standard error of the overall estimate from the random effects model increases relative to that from the fixed effects model. The difference between the overall estimates from the two approaches depends to a large extent on the magnitude of the

estimates from the large informative trials in relation to the others. For example, if a meta-analysis is based on one large study with a small positive estimate and several small studies with large positive estimates, the overall estimate from the random effects model will be larger than that from the fixed effects model, the difference increasing with increasing heterogeneity. The more conservative approach of the random effects model will in general lead to larger numbers of patients being required to demonstrate a significant treatment difference than the fixed effects approach.

It may be useful in many cases to consider the results from both a fixed effects model and a random effects model. If they lead to important differences in conclusion, then this highlights the need for further investigation. For example, this could be due to variability in study quality, differences in study protocols, or differences in the study populations.

When individual patient data are available the models can be extended to include the trial effect. As the trial effect may also be included as a fixed or random effect, this leads to an increased choice of models, as discussed by Senn (2000). These models are presented in detail in Chapter 5 of this book, and comparisons made between them.

1.4 INDIVIDUAL PATIENT DATA VERSUS SUMMARY STATISTICS

There is wide variation in the amount and form of data which might be available for a meta-analysis. At one extreme a common outcome measure may have been used in all studies, with individual data available for all patients. At the other extreme the only available data may be the p-value from each study associated with the test of a treatment difference, or, even worse, a statement in a published paper to the effect that the p-value was or was not smaller than 0.05. In between, we may be confronted with summary statistics from published papers, individual patient data based on similar but not identically defined outcome measures, or a mixture of individual patient data and summary statistics.

A meta-analysis using individual patient data is likely to prove more comprehensive and reliable than one based on summary statistics obtained from publications and reports. Such an analysis will benefit from a standardized approach being taken to the extraction of relevant data and to the handling of missing data. In addition, if data at a patient level, such as age, gender or disease severity, are available, the relationship between these and the treatment difference can be explored. To be successful, such a meta-analysis will usually involve a considerable amount of time devoted to the planning, data collection and analysis stages. The advantages of a prospectively planned meta-analysis now become apparent.

Pharmaceutical statisticians are often in a good position to perform a meta-analysis on individual patient data, as they will usually have access to all original data from trials on the company's own as yet unlicensed product. Even if the

meta-analysis is retrospective, data from the various trials will often have been stored electronically in similarly structured databases. Outside the pharmaceutical industry, the task is more daunting. Details of the practical issues involved in such an undertaking can be found in Stewart and Clarke (1995), a paper resulting from a workshop held by the Cochrane working group on meta-analysis using individual patient data.

Meta-analyses based on individual patient data have clear advantages over those based on extracted summary statistics. However, they are time-consuming and costly, and the situation may arise in which the additional resources needed to obtain individual patient data are not available or cannot be justified. Even if it is planned to obtain individual patient data, it may not be possible to obtain these from all relevant studies. Therefore, many meta-analyses are conducted using summary statistics collected from each trial.

If the purpose of the meta-analysis is to provide an overall estimate of treatment difference, an individual trial can only be included if there is sufficient information from that trial to calculate an estimate of the treatment difference and its standard error. In some cases the summary statistics which are available from a trial enable the same calculations to be performed as if individual patient data were available. For example, for a binary outcome knowledge of the number of successes and failures in each treatment group is sufficient.

Because of the variety of ways in which data are made available for meta-analyses, a number of different techniques for conducting meta-analyses have been developed. This book attempts to present the various approaches within a general framework, highlighting the similarities and differences.

1.5 MULTICENTRE TRIALS AND META-ANALYSIS

Multicentre trials are usually conducted to enable the required number of patients to be recruited within an acceptable period of time and to provide a wider representation of the patient population than would be found at a single centre. A multicentre trial will have been designed prospectively with a combined analysis of the data from all centres as its main objective. Individual centres are expected to follow a common protocol, at least with respect to collection of the main efficacy data. When a meta-analysis is to be undertaken on a series of clinical trials, in which a common outcome measure has been recorded and individual patient data are available, it could be analysed using the same linear modelling techniques as are applied to the analysis of a multicentre trial. Here 'trial' would play the role of 'centre'. On the other hand the analysis of a multicentre trial could be conducted using traditional meta-analysis methods, in which 'centre' plays the role of 'trial'.

There is a continuum from the true multicentre trial, in which all centres follow an identical protocol, to a collection of trials addressing the same general therapeutic question but with different protocols. The same statistical methods can be applied across the continuum, but the choice of the most appropriate

method and the validity of the results may vary. There are differences between the approaches *traditionally* applied to the analysis of multicentre trials and those applied in meta-analysis, as discussed by Senn (2000). This is perhaps because most of the meta-analyses which appear in the medical literature are retrospective and based on summary data from published papers. The differences relate to the way in which the trial estimates of treatment difference are combined and the choice between random and fixed effects models. These issues will be covered in Chapter 5.

1.6 THE STRUCTURE OF THIS BOOK

The focus of this book is on the planning, conduct and reporting of a meta-analysis as applied to a series of randomized controlled trials. It covers the approaches required for retrospective and prospective meta-analyses, as well as for those based on either summary statistics or individual patient data.

The meta-analysis techniques are described in detail, from their theoretical development through to practical implementation. The intention is to present the various statistical methods which are available within a general unified framework, so that the similarities and differences between them become apparent. This is done at a level that can be understood by medical statisticians and statistically minded clinicians and health research professionals. Emphasis is placed on the consequences of choosing a particular approach, the implementation of the chosen method and the interpretation of the results. For interested readers, the mathematical theory underlying the methods is summarized in the Appendix.

The methodology throughout this book is illustrated by examples. All of the methods presented can be implemented using mainstream statistical packages. Most of the analyses presented in the book were conducted using the standard statistical procedures in SAS (Version 8.0: website at http://www.sas.com). At appropriate places in the text, SAS code relating to the specification of the model is provided. For fitting random effects models when individual patient data are available and the response type is binary or ordinal, the program MLn (Version 1.0A) or its interactive Windows version MLwiN (Version 1.10: website at http://multilevel.ioe.ac.uk) was utilized. The interactive Windows version of BUGS, WinBUGS (Version 1.3: website at http://www.mrc-bsu.cam.ac.uk/bugs) was used for the Bayesian analyses and PEST 4 (website at http://www.rdg.ac.uk/mps/mps_home/software/software.htm) was used for the cumulative meta-analyses. For these other packages, the details of their implementation are discussed in the text. The example data sets and the program code for the analyses may be obtained electronically from the Wiley ftp site at ftp://ftp.wiley.co.uk/pub/books/whitehead and also from the author at http://www.rdg.ac.uk/mps/mps_home/misc/publications.htm.

There is now a wide range of software available specifically for performing a meta-analysis. These include both specialist packages and general statistical

packages with meta-analysis routines. They have not been used for the implementation of the methods presented in this book because they have a limited range of options and lack the flexibility to accommodate the more advanced statistical modelling techniques. A recent review of meta-analysis software has been undertaken by Sterne *et al.* (2001b) and the reader is referred to this for further details. This review updates a previous one by Egger *et al.* (1998).

The preparation of a protocol is an important first stage in the conduct of a meta-analysis, and the items which need to be considered for inclusion in the protocol are discussed in Chapter 2.

The main statistical methods used in performing a meta-analysis are described in Chapters 3–5. The methodology is presented in detail for the situation in which each trial has a parallel group design, and a comparison is to be made between two treatments each of which are studied in each trial. This is the most straightforward application and the most common in practice. Usually one treatment will be the newly developed treatment of interest and the other a placebo or standard treatment. The main emphasis is on estimating and making inferences about the difference between the effects of the two treatments.

Meta-analyses are being conducted for an increasing diversity of diseases and conditions, involving a variety of outcome measures. In this book five different types of outcome are discussed in detail, namely binary, survival, interval-censored survival, ordinal and normally distributed. Chapter 3 is divided into sections, each of which considers one particular type of data. For each data type, the choice of an appropriate measure of treatment difference is addressed, together with the methods of estimation which are traditionally used within the context of an individual clinical trial.

Chapter 4 presents a methodology for combining the trial estimates of a treatment difference, based on Whitehead and Whitehead (1991). This approach is of use primarily when data available for the meta-analysis consist of summary statistics from each trial. It may also be used when individual patient data are available, but in this case the more advanced statistical modelling techniques of Chapter 5 may be preferred. In Chapter 4, meta-analyses based on the fixed effects model are illustrated for the different data types. The extension to the random effects model is also presented.

Chapter 5 considers various models which can be fitted making full use of individual patient data. These models include terms for the trial effect, which can be assumed to be a fixed effect or a random effect. The pros and cons of each model are discussed, and comparisons made with models used for multicentre trials.

It is important to assess the consistency between the individual trial estimates of treatment difference. Chapter 6 discusses the issues involved in this assessment, and how the amount of heterogeneity might affect the choice of model for the meta-analysis or even whether to present an overall estimate at all. In some situations the treatment difference may be expected to vary from one level of a factor to another. Regression techniques can be used to explore this if additional data at the trial level or at the patient level are available. Such techniques are

described in this chapter. Finally, a strategy for dealing with heterogeneity is proposed.

The presentation and interpretation of results is addressed in Chapter 7. The QUOROM statement (Moher *et al.*, 1999) which provides guidance on the reporting of meta-analyses of clinical trials is used as a basis for the discussion of the structure of a report. Graphical displays, which have an important role to play, are described.

When judging the reliability of the results of a meta-analysis, attention should focus on factors which might systematically influence the overall estimate of the treatment difference. One important factor is the selection of studies for inclusion in the meta-analysis. Chapter 8 considers the possible reasons why some trials may be excluded from a meta-analysis and how the problems might be addressed, focusing particularly on publication bias.

Chapter 9 deals with some of the issues arising from non-standard data sets. These include the problems of having no events in one or more of the treatment arms of individual trials and the use of different rating scales or different times of assessment across trials. Ways of combining trials which report different summary statistics and of combining p-values when it is impossible to estimate the treatment difference are also discussed.

Although the main focus of the book is on parallel group studies comparing two treatments, it is often desirable to consider the inclusion of other types of study in the meta-analysis. Chapter 10 considers the incorporation of data from multicentre trials, cross-over trials and sequential trials. Also, the handling of multiple treatment comparisons and the investigation of dose – response relationships are discussed.

Most of the statistical methods presented in this book have been derived from a classical (frequentist) approach. Chapter 11 presents a Bayesian approach to meta-analysis. Comparisons are made with the results from the frequentist analyses.

A cumulative meta-analysis involves repeated meta-analyses following completion of a further one or more studies addressing the same question. Repeated meta-analyses are becoming more common, and are encouraged within the Cochrane Collaboration so that the information in the Cochrane Library can be kept up to date. An analogy can be made with the conduct of a sequential clinical trial, in which information about the treatment difference is updated by conducting interim analyses. Chapter 12 considers the role that sequential methods may play in the conduct of a cumulative meta-analysis. Application to prospectively planned meta-analyses is discussed.

2

Protocol Development

2.1 INTRODUCTION

Before starting a clinical trial it is standard practice to prepare a study protocol, specifying in detail the procedures to be followed. Likewise, it should be standard practice to prepare a protocol for conducting a meta-analysis, particularly as this is often a complex process. As is the case for an individual study, it may be necessary to make changes to the meta-analysis protocol due to unforeseen circumstances. Protocol amendments can be made for a meta-analysis, in the same way as they can for an individual trial. Such changes should be documented and their impact on the results discussed. In a meta-analysis protocol it will be necessary to state the key hypotheses of interest. This should not prevent the conduct of exploratory analyses, undertaken to explain the findings and to suggest hypotheses for future studies. However, when the results are reported it is important to make a clear distinction between the preplanned analyses and the exploratory analyses.

In the development of a new drug or medical intervention there is an obvious advantage in designing the clinical trial programme to take account of the need for a meta-analysis. Individual trial protocols can include common elements, such as identically defined outcome measures. Preparation of the protocol for a meta-analysis before the start of any of the trials is the ideal situation. Certainly the existence of a meta-analysis protocol is a reminder that the impact of changes to a study protocol needs to be considered on a global scale rather than on an individual trial basis. There will, of course, be times when the need for a meta-analysis will not be identified until after some or all of the trials have started. Provided that the meta-analysis protocol is prepared before results from any of the trials are available, this is unlikely to compromise the integrity of the meta-analysis in any important way.

The preparation of a protocol is perhaps even more crucial for a retrospective meta-analysis, or for one planned following the disclosure of the results from one or more trials. For such meta-analyses there is the possibility of bias being introduced due to study selection. In many cases it may only be possible to perform the meta-analysis on a subset of the studies because of inconsistency in the recording and/or reporting of outcome measures or incompatible trial

designs. Further, if the meta-analysis is restricted to data obtained from published papers, the overall treatment difference may be overestimated because studies with statistically significant results are more likely to be published than those without. If the meta-analysis is undertaken because of the announcement of some very positive results, this may lead to an overestimation of the treatment difference. As a consequence, more attention will need to be given in the protocol to addressing the implications of these potential biases for the meta-analysis.

This chapter is concerned with the content of a meta-analysis protocol. Many of the items discussed will be common to both prospective and retrospective meta-analyses, although for a retrospective analysis the investigation of selection bias will require specific attention. Comprehensive guidelines for undertaking systematic reviews have been produced (see, for example, Cook *et al.*, 1995; Deeks *et al.*, 1996; Clarke and Oxman, 2001). Their focus is on retrospective reviews and meta-analyses, usually undertaken on summary statistics extracted from published papers. In this chapter, the list of topics covered is similar to those which appear in these guidelines. However, the topics are discussed in the context of a prospective as well as a retrospective meta-analysis, and also for individual patient data as well as summary statistics.

2.2 BACKGROUND

Background information helps to set the scene for the meta-analysis. Topics which might be included are a definition of the disease or condition in question, its incidence, prognosis, public health importance and alternative available treatments. General information on the treatment being evaluated will relate to its mechanism of action, results from its use in other indications and the rationale for its use in the disease or condition in question. The results of earlier meta-analyses could be discussed. The reasons for undertaking the current meta-analysis should be provided.

2.3 OBJECTIVES

The main objectives of the meta-analysis should be stated. For example, in the case of a new treatment for Alzheimer's disease, the objective might be to evaluate the efficacy and safety of the new treatment, when administered for up to six months according to a particular dosing regimen to patients with mild to moderate Alzheimer's disease, where efficacy is assessed in terms of cognitive performance and clinical global impression, and safety is assessed in terms of the occurrence of adverse events. A brief description should be provided of the types of study which will be examined.

2.4 OUTCOME MEASURES AND BASELINE INFORMATION

A list of all of the outcome measures to be analysed, with definitions where appropriate, should be given. As in the case of an individual trial, it is advisable to specify which of the efficacy measures is the primary one, so that the problem associated with multiple testing – that is, too many false positives – can be minimized. Often assessments are repeated at various timepoints during the trial, and how these are to be dealt with should be mentioned. If the assessment at one particular timepoint is of primary interest this should be stated. For example, the primary efficacy measure in the Alzheimer's disease meta-analysis might be the change in the cognitive subscale of the Alzheimer's Disease Assessment Scale between baseline and the six-month assessment.

It will often be important to obtain data on baseline variables such as demographic characteristics, prognostic factors and baseline assessments of efficacy and safety measures. There are several ways in which such data may be useful. First, they can be used to check the comparability of patients allocated to each of the treatment arms in each study, enabling within-study and between-study comparisons to be made. Second, if individual patient data are available, an analysis of covariance may be performed in which adjustment is made for one or more baseline variables considered likely to have an important affect on the outcome measure. Such variables would be prespecified. Third, baseline variables may be used to investigate heterogeneity in the treatment difference across studies or subgroups.

2.5 SOURCES OF DATA

In order to minimize problems associated with selection bias, it is important to identify all trials which could potentially contribute to the meta-analysis. This part of the protocol should provide details of the search strategy to be employed. When the meta-analysis is preplanned no search strategy is required because the relevant trials are identified before they are undertaken. A pharmaceutical company undertaking a retrospective meta-analysis on one of its own unlicensed drugs is likely to know about all trials which have been undertaken with the drug. In this case the search strategy will be reasonably straightforward, and a list of the company data sources can be provided. However, in all other cases careful thought needs to be given to the search strategy. Possible information sources include online bibliographic databases of published and unpublished research, trial registries, expert informants and the pharmaceutical industry. The restrictions to be applied, such as, publication status, language of publication and the time-frame concerning the year of publication should be specified. For example, in a meta-analysis conducted to examine the benefits of adding salmeterol as opposed to increasing the dose of inhaled steroid in subjects with symptomatic asthma, the EMBASE, Medline and GlaxoWellcome databases were searched for

all relevant publications and abstracts from 1985 until 1998 in any language (Shrewsbury *et al.*, 2000). For further information about searching strategies, the reader is referred to Chapters 4–7 of Cooper and Hedges (1994) and Clarke and Oxman (2001).

2.6 STUDY SELECTION

The selection criteria for studies in the meta-analysis should be specified. If there is more than one hypothesis to be tested it may be necessary to define separate selection criteria for each one. In addition, for each hypothesis of interest, it may be desirable to create two groups of studies. The first group would consist of the primary studies on which the formal meta-analysis would be undertaken. The second group would consist of additional studies whose results may be included in a sensitivity analysis, or in a graphical presentation of individual study results. Such studies may involve different patient populations or treatment comparisons from the primary studies, or may have less appropriate designs. However, their results may still be informative.

Careful thought needs to be given to the selection criteria for the primary studies. If they are very strict, the results of the meta-analysis may only be applicable to a small subset of the patient population or to a very specific treatment regimen, whereas if they are too liberal, it may not be possible to combine the individual trial results in an informative way.

Typically, the selection criteria will define the treatment of interest and the relevant subject population. This should follow logically from the statement of the objectives of the meta-analysis. In addition, they may relate to the type of study design used. For example, the selection criteria used in the salmeterol meta-analysis mentioned in Section 2.5 were stated as follows: a randomized controlled trial; direct comparison between adding salmeterol to the current dose of inhaled steroid and increasing (at least doubling) the dose of the current inhaled steroid; study duration of 12 weeks or longer; subjects aged 12 years or older with symptomatic asthma on the current dose of inhaled steroids.

The assessment of the methodological quality of a trial may also be used to determine its eligibility for inclusion in the group of primary studies. The most important aspect of this assessment concerns the avoidance of bias in the estimation of the treatment difference of interest. Therefore, design issues, such as the method of randomizing subjects to treatment group, blinding, method of assessing patient outcome, follow-up of patients, and handling of protocol deviations and patient withdrawals from the trial, are likely to feature prominently. It may be appropriate to categorize studies according to how well they adhere to important methodological standards. For further discussion on the types of scoring systems which have been devised, the reader is referred to Moher *et al.* (1995).

In the report of a meta-analysis it will be necessary to include a list of studies which were excluded as well as a list of studies which were included. The reason

for exclusion should be provided for each excluded study. It may be advantageous to have more than one assessor decide independently which studies to include or exclude, together with a well-defined checklist and a procedure which will be followed when they disagree.

In some cases, new information may surface during the reading of the study reports which indicate a need to modify the study selection criteria.

2.7 DATA EXTRACTION

A specification of the data items to be extracted should be provided. It may be useful to produce an additional document which details the desired format for the data, the recommended coding and the data checking procedures.

A meta-analysis based on individual patient data is likely to provide the most reliable information, as it will not depend on the way in which individual trial results are reported. For such a meta-analysis the aim should be to obtain individual patient data from all randomized subjects in all relevant trials. This will enable a consistent approach to be taken towards the coding of data and the handling of missing data across all trials. If there is a common database structure for all trials, this will facilitate the integration of their data. However, for many retrospective meta-analyses the data are not centrally located, and considerable time and effort are required to collect all of the necessary items together. Stewart and Clarke (1995) discuss the practical aspects of data collection and data checking when data are being supplied by individual trialists.

In many cases meta-analyses are conducted using summary information from published papers or trial reports. Even if the plan is to collect individual patient data from all trials, there may be some trials for which this is not possible. Also, as part of a sensitivity analysis it may be desirable to include results from additional studies from which only summary information is available. In these situations, consideration needs to be given to the type of information which will be required. Take, for example, the case of a dichotomous outcome, in which the patient response is either 'success' or 'failure'. To use the meta-analysis methodology described in Chapter 4, a measure of treatment difference must be chosen. Suppose that the chosen measure is the log-odds ratio of success on the new treatment relative to placebo. A trial can only be included in the meta-analysis if the available data from the trial enable an estimate of the log-odds ratio and its variance to be calculated. Knowledge of the number of successes and failures in each treatment group in each trial is sufficient. However, if the only available data from a trial is the estimate of the difference in the success probabilities between the two treatment groups, the trial cannot be included. Further details about what constitutes sufficient information are provided in Chapter 3. In addition, Section 9.5 considers ways of combining trials which report different summary statistics and Section 9.6 ways of imputing estimates of the treatment difference and its variance.

If the data available for the meta-analysis are mainly summary statistics from trial reports and publications, then it may be possible to extract some useful additional information from the trialists. For example, the trialist may be able to clarify whether the reported analysis of a binary response was based on all randomized patients or on a selected subset. If the latter, the trialist may be able to provide the numbers of 'successes' and 'failures' amongst the excluded patients. A data collection form, detailing the information required, can be distributed to the trialists. The process of extracting additional information from trialists is facilitated by having as part of the meta-analysis team clinical experts who know the field and the trialists.

2.8 STATISTICAL ANALYSIS

The principal features of the statistical analysis should be included in the main protocol, although it may also be useful to produce separately a detailed statistical analysis plan. For each outcome variable to be analysed the following items should be considered.

2.8.1 Analysis population

The set of subjects who are to be included in the meta-analysis should be defined. This will usually be based on the intention-to-treat principle, which in respect of an individual trial specifies that all randomized patients should be included in the analysis as members of the treatment group to which they were randomized. This principle is important in preventing bias and providing an objective basis for statistical analysis.

In the ideal situation in which all randomized subjects satisfy all of the trial selection criteria, comply with all of the trial procedures and provide complete data, the intention-to-treat analysis is straightforward to implement. However, this ideal situation is unlikely to be achieved in practice. Provided that there is proper justification and that bias is unlikely to be introduced, it may be considered appropriate to exclude certain randomized subjects from the analysis set. In the ICH E9 (ICH, 1998) guidelines the term 'full analysis' set is used to describe the analysis set which is as complete as possible and as close as possible to the intention-to-treat ideal of including all randomized subjects.

Reports of clinical trials often include analyses undertaken on a second set of subjects, referred to as the 'per protocol' set. The 'per protocol' set is a subset of patients who are more compliant with the protocol. For example, they are not classified as major protocol violators, they complete a minimum period on study treatment and provide data for the primary efficacy analysis. Sometimes an analysis is undertaken on all subjects who complete the study period and provide data on the primary efficacy variable, referred to as a 'completers' analysis. This is

an example of a 'per protocol' analysis. Because adherence to the study protocol may be related to the treatment and to the outcome, analyses based on the 'per protocol' set may be biased. For example, in a comparison of a new treatment with placebo, if patients who cannot tolerate the new treatment withdraw early from the trial, the analysis based on the 'per protocol' set may produce a larger estimate of the treatment difference than that based on the 'full analysis' set. Therefore, whilst a meta-analysis based on a 'per protocol' set may be undertaken as part of a sensitivity analysis, the evidence from an analysis based on the 'full analysis' set will usually be more convincing.

Whilst it is envisaged that most meta-analyses will be undertaken to determine if one treatment is superior to another, some will be undertaken to determine if two treatments are equivalent. In the latter case, the conservative nature of the intention-to-treat approach may be inappropriate and the meta-analysis based on a 'per protocol' set should be looked at on a more equal footing with that based on the 'full analysis' set.

When the meta-analysis is to be conducted using individual patient data, it is desirable to obtain data from all randomized patients, so that the most appropriate analysis can be undertaken. Difficulties may arise when a meta-analysis is based on summary information from published papers or trial reports in which the various authors have chosen different criteria for their main analysis set. In particular, some papers may only provide results from a 'full analysis' set, whereas others may only provide results from a 'per protocol' set. In such situations it may be advisable to separate the studies using 'full analysis' sets from those using 'per protocol' sets, before ascertaining whether or not it would be appropriate to combine them.

The set of subjects to be included in the assessment of safety and tolerability is often defined as those subjects who received at least one dose of the study medication, and is sometimes referred to as the 'safety analysis' set. The 'safety analysis' set would seem to be an appropriate choice for a meta-analysis of safety and tolerability data.

2.8.2 Missing data at the subject level

Difficulties arise in the analysis of a clinical trial when data are missing from some subjects. The intention-to-treat principle defines the set of subjects to be included in the analysis, but does not specify how to deal with missing data. As for an individual trial, the effect of data missing at the subject level on the overall results from a meta-analysis will need to be addressed.

Some subjects who meet the criteria for the 'full analysis' set may not provide data on some of the outcomes of interest, including the primary efficacy variable. This could occur if a subject withdraws from treatment part-way through the study and provides no further data after this point or if the subject is lost to follow-up. One option is to perform the analysis of each outcome variable using

only those subjects who provide data on that particular variable. This means that the set of subjects contributing to each analysis may vary. More importantly, this approach relies on the assumption that data are missing at random, that is, the absence of a recorded value is not dependent on its actual value (see, for example, Little and Rubin, 1987). In particular, if the mechanisms for data being missing differs between the study treatments, then the exclusion of the subject from the analysis may introduce bias into the estimate of the treatment difference.

An alternative strategy is to substitute values for the missing data. If the outcome of interest is measured at various timepoints during the study, values from early timepoints can be used to impute data for the later missing values. Imputation techniques range from carrying forward the last observation to the use of complex mathematical models (see, for example, Rubin, 1987; Little, 1995). However, caution is required as imputation techniques may themselves lead to biased estimates of the treatment difference. In some trials data continue to be collected according to the intended schedule on patients who withdraw early from study treatment. Such data may be used in the analysis, although careful thought needs to be given to this as such patients may have received alternative medication.

If there is a substantial amount of missing data, the reliability of the analysis may be questioned. In this case it may be useful to undertake sensitivity analyses in which the effects of different imputation schemes are compared.

When the meta-analysis is to be performed using individual patient data, the planned method for dealing with missing data should be described. If no imputation is to be undertaken, then this should be stated.

When meta-analyses are based on summary information from published papers, the amount of missing data and the way in which they have been handled by the author may be factors for consideration in the assessment of the methodological quality of a trial.

2.8.3 Analysis of individual trials

It is important to present the results from the individual trials as well as the results from the meta-analysis. Individual trial summaries may not be the same as those presented in earlier trial reports and publications because it is desirable to take the same approach to the analysis of each of the trials and to make this consistent with the meta-analysis. When individual patient data are available a reanalysis using a common approach will often be possible. However, this is unlikely to be the case for meta-analyses based on summary information. In this situation one hopes that the summary information will permit the use of the same measure of treatment difference in all studies.

The chosen measure of treatment difference should be specified. For example, for binary data this might be the log-odds ratio or for continuous data it might

be the absolute difference in means. Details of the various measures of treatment difference which can be used for commonly occurring types of data are presented in Chapter 3.

2.8.4 Meta-analysis model

The proposed meta-analysis model should be specified, including which terms are to be treated as fixed effects and which random effects. Models which can be used for the combination of trial estimates of treatment difference are discussed in Chapter 4. A model which assumes that the parameter measuring treatment difference is the same across all trials is typically referred to as a 'fixed effects' model. A model which allows this parameter to act as a random variable taking different values from one trial to the next is typically referred to as a 'random effects' model. Issues relating to the choice of a fixed or random effects model are discussed in Chapter 6. When individual patient data are available the statistical modelling approach of Chapter 5 may be used. Within this framework it is straightforward to include additional covariates in the model, to enable adjustment for prognostic factors which are considered likely to affect the outcome data.

2.8.5 Estimation and hypothesis testing

The main hypotheses to be tested should be specified. For example, in the comparison of a new treatment against the standard treatment the null hypothesis of no treatment difference might be tested against the two-sided alternative of some difference between the two treatments. If the new treatment has been tested at more than one dose level, it may not be appropriate to combine the data from all doses together. There may be one dose level of prime interest. Alternatively, or additionally, it may be of interest to investigate the dose-response relationship.

2.8.6 Testing for heterogeneity

Meta-analyses are often performed retrospectively on studies which were not planned with this in mind. In many situations it might be expected that differences in the study protocols will produce heterogeneity. Also, even if the same protocols are used for all studies, variability in study quality, possibly due to mistakes in implementing the protocol, may give rise to heterogeneity. Therefore, it is common to include a test for heterogeneity in the treatment difference parameter across studies. A test for heterogeneity when trial estimates are being combined is presented in Chapter 4, and analogous tests based on individual patient data are presented in Chapter 5.