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.