Preface

In his best-selling book *Baby and Child Care*, Dr. Benjamin Spock wrote 'I think it is preferable to accustom a baby to sleeping on his stomach from the beginning if he is willing'. This statement was included in most editions of the book, and in most of the 50 million copies sold from the 1950s into the 1990s. The advice was not unusual, in that many pediatricians made similar recommendations at the time.

During this same period, from the 1950s into the 1990s, more than 100,000 babies died of sudden infant death syndrome (SIDS), also called *crib death* in the United States and *cot death* in the United Kingdom, where a seemingly healthy baby goes to sleep and never wakes up.

In the early 1990s, researchers became aware that the risk of SIDS decreased by at least 50% when babies were put to sleep on their backs rather than face down. Governments in various countries launched educational initiatives such as the *Back to sleep* campaigns in the UK and the US, which led to an immediate and dramatic drop in the number of SIDS deaths.

While the loss of more than 100,000 children would be unspeakably sad in any event, the real tragedy lies in the fact that many of these deaths could have been prevented. Gilbert *et al*. (2005) write

> 'Advice to put infants to sleep on the front for nearly half a century was contrary to evidence available from 1970 that this was likely to be harmful. Systematic review of preventable risk factors for SIDS from 1970 would have led to earlier recognition of the risks of sleeping on the front and might have prevented over 10,000 infant deaths in the UK and at least 50,000 in the Europe, the USA and Australasia.'

AN ETHICAL IMPERATIVE

This example is one of several cited by Sir Iain Chalmers in a talk entitled *The scandalous failure of scientists to cumulate scientifically* (Chalmers, 2006). The theme of this talk was that we live in a world where the utility of almost any intervention will be tested repeatedly, and that rather than looking at any study in isolation, we need to look at the body of evidence. While not all systematic reviews carry the urgency of SIDS, the logic of looking at the body of evidence, rather than trying to understand studies in isolation, is always compelling.

Meta-analysis refers to the statistical synthesis of results from a series of studies. While the statistical procedures used in a meta-analysis can be applied to any set of data, the synthesis will be meaningful only if the studies have been collected

systematically. This could be in the context of a systematic review, the process of systematically locating, appraising, and then synthesizing data from a large number of sources. Or, it could be in the context of synthesizing data from a select group of studies, such as those conducted by a pharmaceutical company to assess the efficacy of a new drug.

If a treatment effect (or effect size) is consistent across the series of studies, these procedures enable us to report that the effect is robust across the kinds of populations sampled, and also to estimate the magnitude of the effect more precisely than we could with any of the studies alone. If the treatment effect varies across the series of studies, these procedures enable us to report on the range of effects, and may enable us to identify factors associated with the magnitude of the effect size.

FROM NARRATIVE REVIEWS TO SYSTEMATIC REVIEWS

Prior to the 1990s, the task of combining data from multiple studies had been primarily the purview of the narrative review. An expert in a given field would read the studies that addressed a question, summarize the findings, and then arrive at a conclusion – for example, that the treatment in question was, or was not, effective. However, this approach suffers from some important limitations.

One limitation is the subjectivity inherent in this approach, coupled with the lack of transparency. For example, different reviewers might use different criteria for deciding which studies to include in the review. Once a set of studies has been selected, one reviewer might give more credence to larger studies, while another gives more credence to 'quality' studies and yet another assigns a comparable weight to all studies. One reviewer may require a substantial body of evidence before concluding that a treatment is effective, while another uses a lower threshold. In fact, there are examples in the literature where two narrative reviews come to opposite conclusions, with one reporting that a treatment is effective while the other reports that it is not. As a rule, the narrative reviewer will not articulate (and may not even be fully aware of) the decision-making process used to synthesize the data and arrive at a conclusion.

A second limitation of narrative reviews is that they become *less useful as more information becomes available*. The thought process required for a synthesis requires the reviewer to capture the finding reported in each study, to assign an appropriate *weight* to that finding, and then to synthesize these findings across all studies in the synthesis. While a reviewer may be able to synthesize data from a few studies in their head, the process becomes difficult and eventually untenable as the number of studies increases. This is true even when the treatment effect (or effect size) is consistent from study to study. Often, however, the treatment effect will vary as a function of study-level covariates, such as the patient population, the dose of medication, the outcome variable, and other factors. In these cases, a proper synthesis requires that the researcher be able to understand how the treatment effect varies as a function of these variables, and the narrative review is poorly equipped to address these kinds of issues.

THE SYSTEMATIC REVIEW AND META-ANALYSIS

For these reasons, beginning in the mid 1980s and taking root in the 1990s, researchers in many fields have been moving away from the narrative review, and adopting systematic reviews and meta-analysis.

For systematic reviews, a clear set of rules is used to search for studies, and then to determine which studies will be included in or excluded from the analysis. Since there is an element of subjectivity in setting these criteria, as well as in the conclusions drawn from the meta-analysis, we cannot say that the systematic review is entirely objective. However, because all of the decisions are specified clearly, the mechanisms are transparent.

A key element in most systematic reviews is the statistical synthesis of the data, or the meta-analysis. Unlike the narrative review, where reviewers implicitly assign some level of importance to each study, in meta-analysis the weights assigned to each study are based on mathematical criteria that are specified in advance. While the reviewers and readers may still differ on the substantive meaning of the results (as they might for a primary study), the statistical analysis provides a transparent, objective, and replicable framework for this discussion.

The formulas used in meta-analysis are extensions of formulas used in primary studies, and are used to address similar kinds of questions to those addressed in primary studies. In primary studies we would typically report a mean and standard deviation for the subjects. If appropriate, we might also use analysis of variance or multiple regression to determine if (and how) subject scores were related to various factors. Similarly, in a meta-analysis, we might report a mean and standard deviation for the treatment effect. And, if appropriate, we would also use procedures analogous to analysis of variance or multiple regression to assess the relationship between the effect and study-level covariates.

Meta-analyses are conducted for a variety of reasons, not only to synthesize evidence on the effects of interventions or to support evidence-based policy or practice. The purpose of the meta-analysis, or more generally, the purpose of any research synthesis has implications for *when* it should be performed, what model should be used to analyze the data, what sensitivity analyses should be undertaken, and how the results should be interpreted. Losing sight of the fact that meta-analysis is a tool with multiple applications causes confusion and leads to pointless discussions about *what is the right way to perform a research synthesis*, when there is no single right way. It all depends on the purpose of the synthesis, and the data that are available. Much of this book will expand on this idea.

META-ANALYSIS IS USED IN MANY FIELDS OF RESEARCH

In medicine, systematic reviews and meta-analysis form the core of a movement to ensure that medical treatments are based on the best available empirical data. For example, The Cochrane Collaboration has published the results of over 3700 meta-analyses (as of January 2009) which synthesize data on treatments in all areas of

health care including headaches, cancer, allergies, cardiovascular disease, pain prevention, and depression. The reviews look at interventions relevant to neo-natal care, childbirth, infant and childhood diseases, as well as diseases common in adolescents, adults, and the elderly. The kinds of interventions assessed include surgery, drugs, acupuncture, and social interventions. BMJ publishes a series of journals on Evidence Based Medicine, built on the results from systematic reviews. Systematic reviews and meta-analyses are also used to examine the performance of diagnostic tests, and of epidemiological associations between exposure and disease prevalence, among other topics.

Pharmaceutical companies usually conduct a series of studies to assess the efficacy of a drug. They use meta-analysis to synthesize the data from these studies, yielding a more powerful test (and more precise estimate) of the drug's effect. Additionally, the meta-analysis provides a framework for evaluating the series of studies as a whole, rather than looking at each in isolation. These analyses play a role in internal research, in submissions to governmental agencies, and in marketing. Meta-analyses are also used to synthesize data on adverse events, since these events are typically rare and we need to accumulate information over a series of studies to properly assess the risk of these events.

In the field of education, meta-analysis has been applied to topics as diverse as the comparison of distance education with traditional classroom learning, assessment of the impact of schooling on developing economies, and the relationship between teacher credentials and student achievement. Results of these and similar meta-analyses have influenced practice and policy in various locations around the world.

In psychology, meta-analysis has been applied to basic science as well as in support of evidence-based practice. It has been used to assess personality change over the life span, to assess the influence of media violence on aggressive behavior, and to examine gender differences in mathematics ability, leadership, and nonverbal communication. Meta-analyses of psychological interventions have been use to compare and select treatments for psychological problems, including obsessive-compulsive disorder, impulsivity disorder, bulimia nervosa, depression, phobias, and panic disorder.

In the field of criminology, government agencies have funded meta-analyses to examine the relative effectiveness of various programs in reducing criminal behavior. These include initiatives to prevent delinquency, reduce recidivism, assess the effectiveness of different strategies for police patrols, and for the use of special courts to deal with drug-related crimes.

In business, meta-analyses of the predictive validity of tests that are used as part of the hiring process, have led to changes in the types of tests that are used to select employees in many organizations. Meta-analytic results have also been used to guide practices for the reduction of absenteeism, turnover, and counterproductive behavior, and to assess the effectiveness of programs used to train employees.

In the field of ecology, meta-analyses are being used to identify the environmental impact of wind farms, biotic resistance to exotic plant invasion, the effects of changes

in the marine food chain, plant reactions to global climate change, the effectiveness of conservation management interventions, and to guide conservation efforts.

META-ANALYSIS AS PART OF THE RESEARCH PROCESS

Systematic reviews and meta-analyses are used to synthesize the available evidence for a given question to inform policy, as in the examples cited above from medicine, social science, business, ecology, and other fields. While this is probably the most common use of the methodology, meta-analysis can also play an important role in other parts of the research process.

Systematic reviews and meta-analyses can play a role in designing new research. As a first step, they can help determine whether the planned study is necessary. It may be possible to find the required information by synthesizing data from prior studies, and in this case, the research should not be performed. Iain Chalmers (2007) made this point in an article entitled *The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews.*

In the event that the new study *is needed*, the meta-analysis may be useful in helping to design that study. For example, the meta-analysis may show that in the prior studies one outcome index had proven to be more sensitive than others, or that a specific mode of administration had proven to be more effective than others, and should be used in the planned study as well.

For these reasons, various government agencies, including institutes of health in various countries, have been encouraging (or requiring) researchers to conduct a meta-analysis of existing research prior to undertaking new funded studies.

The systematic review can also play a role in the publication of any new primary study. In the introductory section of the publication, a systematic review can help to place the new study in context by describing what we knew before, and what we hoped to learn from the new study. In the discussion section of the publication, a systematic review allows us to address not only the information provided by the new study, but the body of evidence as enhanced by the new study. Iain Chalmers and Michael Clarke (1998) see this approach as a way to avoid studies being reported without context, which they refer to as 'Islands in Search of Continents'. Systematic reviews would provide this context in a more rigorous and transparent manner than the narrative reviews that are typically used for this purpose.

THE INTENDED AUDIENCE FOR THIS BOOK

Since meta-analysis is a relatively new field, many people, including those who actually use meta-analysis in their work, have not had the opportunity to learn about it systematically. We hope that this volume will provide a framework that allows them to understand the logic of meta-analysis, as well as how to apply and interpret meta-analytic procedures properly.

This book is aimed at researchers, clinicians, and statisticians. Our approach is primarily conceptual. The reader will be able to skip the formulas and still understand, for example, the differences between fixed-effect and random-effects analysis, and the mechanisms used to assess the dispersion in effects from study to study. However, for those with a statistical orientation, we include all the relevant formulas, along with worked examples. Additionally, the spreadsheets and data files can be downloaded from the web at www.Meta-Analysis.com.

This book can be used as the basis for a course in meta-analysis. Supplementary materials and exercises are posted on the book's web site.

This volume is intended for readers from various substantive fields, including medicine, epidemiology, social science, business, ecology, and others. While we have included examples from many of these disciplines, the more important message is that meta-analytic methods that may have developed in any one of these fields have application to all of them.

Since our goal in using these examples is to explain the meta-analysis itself rather than to address the substantive issues, we provide only the information needed for this purpose. For example, we may present an analysis showing that a treatment reduces pain, while ignoring other analyses that show the same treatment increases the risk of adverse events. Therefore, any reader interested in the substantive issues addressed in an example should not rely on this book for that purpose.

AN OUTLINE OF THIS BOOK'S CONTENTS

Part 1 is an introduction to meta-analysis. We present a completed meta-analysis to serve as an example, and highlight the elements of this analysis – the effect size for each study, the summary effect, the dispersion of effects across studies, and so on. Our intent is to show where each element fits into the analysis, and thus provide the reader with a context as they move on to the subsequent parts of the book where each of the elements is explored in detail.

Part 2 introduces the effect sizes, such as the standardized mean difference or the risk ratio, that are computed for each study, and that serve as the unit of currency in the meta-analysis. We also discuss factors that determine the variance of an effect size and show how to compute the variance for each study, since this affects the weight assigned to that study in the meta-analysis.

Part 3 discusses the two computational models used in the vast majority of meta-analyses, the fixed-effect model and the random-effects model. We discuss the conceptual and practical differences between the two, and show how to compute a summary effect using either one.

Part 4 focuses on the issue of dispersion in effect sizes, the fact that the effect size varies from one study to the next. We discuss methods to quantify the heterogeneity, to test it, to incorporate it in the weighting scheme, and to understand it in a substantive as well as a statistical context. Then, we discuss methods to explain the heterogeneity. These include subgroup analyses to compare the effect in

different subgroups of studies (analogous to analysis of variance in primary studies), and meta-regression (analogous to multiple regression).

Part 5 shows how to work with complex data structures. These include studies that report an effect size for two or more independent subgroups, for two or more outcomes or time-points, and for two or more comparison groups (such as two treatments being compared with the same control).

Part 6 is used to address three separate issues. One chapter discusses the procedure called vote counting, common in narrative reviews, and explains the problems with this approach. One chapter discusses statistical power for a meta-analysis. We show how meta-analysis often (but not always) yields a more powerful test of the null than do any of the included studies. Another chapter addresses the question of publication bias. We explain what this is, and discuss methods that have been developed to assess its potential impact.

Part 7 focuses on the issue of why we work with effect sizes in a meta-analysis. In one chapter we explain why we work with effect sizes rather than *p*-values. In another we explain why we compute an effect size for each study, rather than summing data over all studies and then computing an effect size for the summed data. The final chapter in this part shows how the use of inverse-variance weights can be extended to other applications including Bayesian meta-analysis and analyses based on individual participant data.

Part 8 includes chapters on methods that are sometimes used in meta-analysis but that fall outside the central narrative of this volume. These include meta-analyses based on *p*-values, alternate approaches (such as the Mantel-Haenszel method) for assigning study weights, and options sometimes used in psychometric meta-analyses.

Part 9 is dedicated to a series of general issues related to meta-analysis. We address the question of when it makes sense to perform a meta-analysis. This Part is also the location for a series of chapters on separate issues such as reporting the results of a meta-analysis, and the proper use of cumulative meta-analysis. Finally, we discuss some of the criticisms of meta-analysis and try to put them in context.

Part 10 is a discussion of resources for meta-analysis and systematic reviews. This includes an overview of several computer programs for meta-analysis. It also includes a discussion of organizations that promote the use of systematic reviews and meta-analyses in specific fields, and a list of useful web sites.

WHAT THIS BOOK DOES NOT COVER

Other elements of a systematic review

This book deals only with meta-analysis, the statistical formulas and methods used to synthesize data from a set of studies. A meta-analysis can be applied to any data, but if the goal of the analysis is to provide a synthesis of a body of data from various sources, then it is usually imperative that the data be compiled as part of a systematic review.

A systematic review incorporates many components, such as specification of the question to be addressed, determination of methods to be used for searching the literature and for including or excluding studies, specification of mechanisms to appraise the validity of the included studies, specification of methods to be used for performing the statistical analysis, and a mechanism for disseminating the results.

If the entire review is performed properly, so that the search strategy matches the research question, and yields a reasonably complete and unbiased collection of the relevant studies, then (providing that the included studies are themselves valid) the meta-analysis will also be addressing the intended question. On the other hand, if the search strategy is flawed in concept or execution, or if the studies are providing biased results, then problems exist in the review that the meta-analysis cannot correct.

In Part 10 we include an annotated listing of suggested readings for the other components in the systematic review, but these components are not otherwise addressed in this volume.

Other meta-analytic methods

In this volume we focus primarily on meta-analyses of effect sizes. That is, analyses where each study yields an estimate of some statistic (a standardized mean difference, a risk ratio, a prevalence, and so on) and our goal is to assess the dispersion in these effects and (if appropriate) compute a summary effect. The vast majority of meta-analyses performed use this approach. We deal only briefly (see Part 8) with other approaches, such as meta-analyses that combine *p*-values rather than effect sizes. We do not address meta-analysis of diagnostic tests.

Further Reading

Chalmers, I. (2007). The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In P. Rothwell(ed.), *Treating Individuals*, ed. London: Lancet: 37–58.

Chalmers, I., Hedges, L.V. & Cooper, H. (2002). A brief history of research synthesis. *Evaluation in the Health Professions*. 25(1): 12–37.

Clarke, M, Hopewell, S. & Chalmers, I. (2007). Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. *Journal of the Royal Society of Medicine*. 100: 187–190.

Hunt, M. (1999). *How Science Takes Stock: The Story of Meta-analysis*. New York: Russell Sage Foundation.

Sutton, A.J. & Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Statistics in Medicine* 27: 625–650.