CHAPTER 3

Overview

Treatment effects and effect sizes
Parameters and estimates
Outline of effect size computations

TREATMENT EFFECTS AND EFFECT SIZES

The terms *treatment effects* and *effect sizes* are used in different ways by different people. Meta-analyses in medicine often refer to the effect size as a *treatment effect*, and this term is sometimes assumed to refer to odds ratios, risk ratios, or risk differences, which are common in meta-analyses that deal with medical interventions. Similarly, meta-analyses in the social sciences often refer to the effect size simply as an *effect size* and this term is sometimes assumed to refer to standardized mean differences or to correlations, which are common in social science meta-analyses.

In fact, though, both the terms *effect size* and *treatment effect* can refer to any of these indices, and the distinction between these terms lies not in the index itself but rather in the nature of the study. The term *effect size* is appropriate when the index is used to quantify the relationship between two variables or a difference between two groups. By contrast, the term *treatment effect* is appropriate only for an index used to quantify the impact of a deliberate intervention. Thus, the difference between males and females could be called an *effect size* only, while the difference between treated and control groups could be called either an *effect size* or a *treatment effect*.

While most meta-analyses focus on relationships between variables, some have the goal of estimating a mean or risk or rate in a single population. For example, a meta-analysis might be used to combine several estimates for the prevalence of Lyme disease in Wabash or the mean SAT score for students in Utah. In these cases the index is clearly not a treatment effect, and is also not an effect size, since *effect* implies a relationship. Rather, the parameter being estimated could be called simply a *single group summary*.

Introduction to Meta-Analysis. Michael Borenstein, L. V. Hedges, J. P. T. Higgins and H. R. Rothstein
 ISBN: 978-0-470-05724-7

Note, however, that the classification of an index as an *effect size* and/or a *treatment effect* (or simply a *single group summary*) has no bearing on the computations. In the meta-analysis itself we have simply a series of values and their variances, and the same mathematical formulas apply. In this volume we generally use the term *effect size*, but we use it in a generic sense, to include also treatment effects, single group summaries, or even a generic statistic.

How to choose an effect size

Three major considerations should drive the choice of an effect size index. The first is that the effect sizes from the different studies should be comparable to one another in the sense that they measure (at least approximately) the same thing. That is, the effect size should not depend on aspects of study design that may vary from study to study (such as sample size or whether covariates are used). The second is that estimates of the effect size should be computable from the information that is likely to be reported in published research reports. That is, it should not require the re-analysis of the raw data (unless these are known to be available). The third is that the effect size should have good technical properties. For example, its sampling distribution should be known so that variances and confidence intervals can be computed.

Additionally, the effect size should be substantively interpretable. This means that researchers in the substantive area of the work represented in the synthesis should find the effect size meaningful. If the effect size is not inherently meaningful, it is usually possible to transform the effect size to another metric for presentation. For example, the analyses may be performed using the log risk ratio but then transformed to a risk ratio (or even to illustrative risks) for presentation.

In practice, the kind of data used in the primary studies will usually lead to a pool of two or three effect sizes that meet the criteria outlined above, which makes the process of selecting an effect size relatively straightforward. If the summary data reported by the primary study are based on means and standard deviations in two groups, the appropriate effect size will usually be either the raw difference in means, the standardized difference in means, or the response ratio. If the summary data are based on a binary outcome such as events and non-events in two groups the appropriate effect size will usually be the risk ratio, the odds ratio, or the risk difference. If the primary study reports a correlation between two variables, then the correlation coefficient itself may serve as the effect size.

PARAMETERS AND ESTIMATES

Throughout this volume we make the distinction between an underlying effect size parameter (denoted by the Greek letter θ) and the sample estimate of that parameter (denoted by Y).

If a study had an infinitely large sample size then it would yield an effect size Y that was identical to the population parameter θ. In fact, though, sample sizes are finite and so the effect size estimate Y always differs from θ by some amount. The value of Y will vary from sample to sample, and the distribution of these values is the sampling distribution of Y. Statistical theory allows us to estimate the sampling distribution of effect size estimates, and hence their standard errors.

OUTLINE OF EFFECT SIZE COMPUTATIONS

Table 3.1 provides an outline of the computational formulas that follow.

These are some of the more common effect sizes and study designs. A more extensive array of formulas is offered in Borenstein *et al.* (2009).

Table 3.1 Roadmap of formulas in subsequent chapters.

Effect sizes based on means (Chapter 4)
Raw (unstandardized) mean difference (*D*)
Based on studies with independent groups
Based on studies with matched groups or pre-post designs
Standardized mean difference (*d* or *g*)
Based on studies with independent groups
Based on studies with matched groups or pre-post designs
Response ratios (*R*)
Based on studies with independent groups
Effect sizes based on binary data (Chapter 5)
Risk ratio (*RR*)
Based on studies with independent groups
Odds ratio (*OR*)
Based on studies with independent groups
Risk difference (*RD*)
Based on studies with independent groups
Effect sizes based on correlational data (Chapter 6)
Correlation (*r*)
Based on studies with one group