CHAPTER 1

How a Meta-Analysis Works

INTRODUCTION

Figure 1.1 illustrates a meta-analysis that shows the impact of high dose versus standard dose of statins in preventing death and myocardial infarction (MI). This analysis is adapted from one reported by Cannon *et al*. and published in the *Journal of the American College of Cardiology* (2006).

Our goal in presenting this here is to introduce the various elements in a meta-analysis (the effect size for each study, the weight assigned to each effect size, the estimate of the summary effect, and so on) and show where each fits into the larger scheme. In the chapters that follow, each of these elements will be explored in detail.

INDIVIDUAL STUDIES

The first four rows on this plot represent the four studies. For each, the study name is shown on the left, followed by the effect size, the relative weight assigned to the study for computing the summary effect, and the *p*-value. The effect size and weight are also shown schematically.

Effect size

The effect size, a value which reflects the magnitude of the treatment effect or (more generally) the strength of a relationship between two variables, is the unit of currency in a meta-analysis. We compute the effect size for each study, and then work with the effect sizes to assess the consistency of the effect across studies and to compute a summary effect.

Introduction to Meta-Analysis, Second Edition.
Michael Borenstein, Larry V. Hedges, Julian P.T. Higgins, and Hannah R. Rothstein.

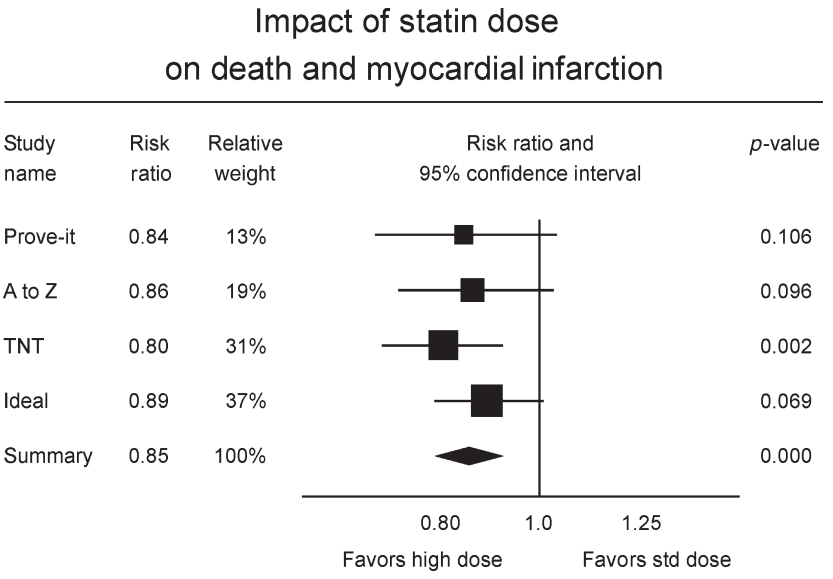

Figure 1.1 High dose versus standard dose of statins (adapted from Cannon *et al.*, 2006).

The effect size could represent the impact of an intervention, such as the impact of medical treatment on risk of infection, the impact of a teaching method on test scores, or the impact of a new protocol on the number of salmon successfully returning upstream. The effect size is not limited to the impact of interventions, but could represent *any relationship* between two variables, such as the difference in test scores for males versus females, the difference in cancer rates for persons exposed or not exposed to second-hand smoke, or the difference in cardiac events for persons with two distinct personality types. In fact, what we generally call an *effect size* could refer simply to the estimate of a single value, such as the prevalence of Lyme disease.

In this example the effect size is the risk ratio. A risk ratio of 1.0 would mean that the risk of death or MI was the same in both groups, while a risk ratio less than 1.0 would mean that the risk was lower in the high-dose group, and a risk ratio greater than 1.0 would mean that the risk was lower in the standard-dose group.

The effect size for each study is represented by a square, with the location of the square representing both the direction and magnitude of the effect. Here, the effect size for each study falls to the left of center (indicating a benefit for the high-dose group). The effect is strongest (most distant from the center) in the *TNT* study and weakest in the *Ideal* study.

Note. For measures of effect size based on ratios (as in this example) a ratio of 1.0 represents no difference between groups. For measures of effect based on differences (such as mean difference), a difference of 0.0 represents no difference between groups.

Precision

In the schematic, the effect size for each study is bounded by a confidence interval, reflecting the precision with which the effect size has been estimated in that study. The confidence interval for the last study (*Ideal*) is noticeably narrower than that for the first study (*Prove-it*), reflecting the fact that the *Ideal* study has greater precision. The meaning of precision and the factors that affect precision are discussed in Chapter 8.

Study weights

The solid squares that are used to depict each of the studies vary in size, with the size of each square reflecting the weight that is assigned to the corresponding study when we compute the summary effect. The *TNT* and *Ideal* studies are assigned relatively high weights, while somewhat less weight is assigned to the *A to Z* study and still less to the *Prove-it* study.

As one would expect, there is a relationship between a study's precision and that study's weight in the analysis. Studies with relatively good precision (*TNT* and *Ideal*) are assigned more weight while studies with relatively poor precision (*Prove-it*) are assigned less weight. Since precision is driven primarily by sample size, we can think of the studies as being weighted by sample size.

However, while precision is one of the elements used to assign weights, there are often other elements as well. In Part 3 we discuss different assumptions that one can make about the distribution of effect sizes across studies, and how these affect the weight assigned to each study.

p-values

For each study we show the *p*-value for a test of the null hypothesis. There is a necessary correspondence between the *p*-value and the confidence interval, such that the *p*-value will fall under 0.05 if and only if the 95% confidence interval does not include the null value. Therefore, by scanning the confidence intervals we can easily identify the statistically significant studies. The role of *p*-values in the analysis, as well as the relationship between *p*-values and effect size, is discussed in Chapter 37.

In this example, for three of the four studies the confidence interval crosses the null hypothesis, and the *p*-value is greater than 0.05. In one (the *TNT* study) the confidence interval does not cross the null hypothesis, and the *p*-value falls under 0.05.

THE SUMMARY EFFECT

One goal of the synthesis is usually to compute a summary effect. Typically we report the effect size itself, as well as a measure of precision and a *p*-value.

Effect size

On the plot the summary effect is shown on the bottom line. In this example the summary risk ratio is 0.85, indicating that the risk of death (or MI) was 15% lower for patients assigned to the high dose than for patients assigned to standard dose.

The summary effect is nothing more than the weighted mean of the individual effects. However, the mechanism used to assign the weights (and therefore the meaning of the summary effect) depends on our assumptions about the distribution of effect sizes from which the studies were sampled. Under the fixed-effect model, we assume that all studies in the analysis share the same true effect size, and the summary effect is our estimate of this common effect size. Under the random-effects model, we assume that the true effect size varies from study to study, and the summary effect is our estimate of the mean of the distribution of effect sizes. This is discussed in Part 3.

Precision

The summary effect is represented by a diamond. The location of the diamond represents the effect size while its width reflects the precision of the estimate. In this example the diamond is centered at 0.85, and extends from 0.79 to 0.92, meaning that the actual impact of the high dose (as compared to the standard) likely falls somewhere in that range.

The precision addresses the accuracy of the summary effect as an estimate of the true effect. However, as discussed in Part 3, the exact meaning of the precision depends on the statistical model.

p -value

The *p*-value for the summary effect is 0.00003. This *p*-value reflects both the magnitude of the summary effect size and also the volume of information on which the estimate is based. Note that the *p*-value for the summary effect is substantially more compelling than that of any single study. Indeed, only one of the four studies had a *p*-value under 0.05. The relationship between *p*-values and effect sizes is discussed in Chapter 37.

HETEROGENEITY OF EFFECT SIZES

In this example the treatment effect is consistent across all studies (by a criterion explained in Chapter 16), but such is not always the case. A key theme in this volume is the importance of assessing the dispersion of effect sizes from study to study, and then taking this into account when interpreting the data. If the effect size is consistent, then we will usually focus on the summary effect, and note that this effect is robust across the domain of studies included in the analysis. If the effect size varies modestly, then we might still report the summary effect but note that the true effect in any given study could be somewhat lower or higher than this value. If the effect varies

substantially from one study to the next, our attention will shift from the summary effect to the dispersion itself.

Because the dispersion in observed effects is partly spurious (it includes both real difference in effects and also random error), before trying to interpret the variation in effects we need to determine what part (if any) of the observed variation is real. In Part 4 we show how to partition the observed variance into the part due to error and the part that represents variation in true effect sizes, and then how to use this information in various ways.

In this example our goal was to estimate the summary effect in one set of populations. In some cases, however, we will want to compare the effect size for one subgroup of studies versus another (say, for studies that used an elderly population versus those that used a relatively young population). In other cases we may want to assess the impact of putative moderators (or covariates) on the effect size (say, comparing the effect size in studies that used doses of 10, 20, 40, 80, 160 mg.). These kinds of analyses are also discussed in Part 4.

SUMMARY POINTS

- To perform a meta-analysis we compute an effect size and variance for each study, and then compute a weighted mean of these effect sizes.
- To compute the weighted mean we generally assign more weight to the more precise studies, but the rules for assigning weights depend on our assumptions about the distribution of true effects.