Introduction to Meta-Analysis

Michael Borenstein
Biostat, Inc, New Jersey, USA.

Larry V. Hedges
Northwestern University, Evanston, USA.

Julian P.T. Higgins
MRC, Cambridge, UK.

Hannah R. Rothstein
Baruch College, New York, USA.

WILEY
A John Wiley and Sons, Ltd., Publication

Contents

List of Tables

List of Figures

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 at 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

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

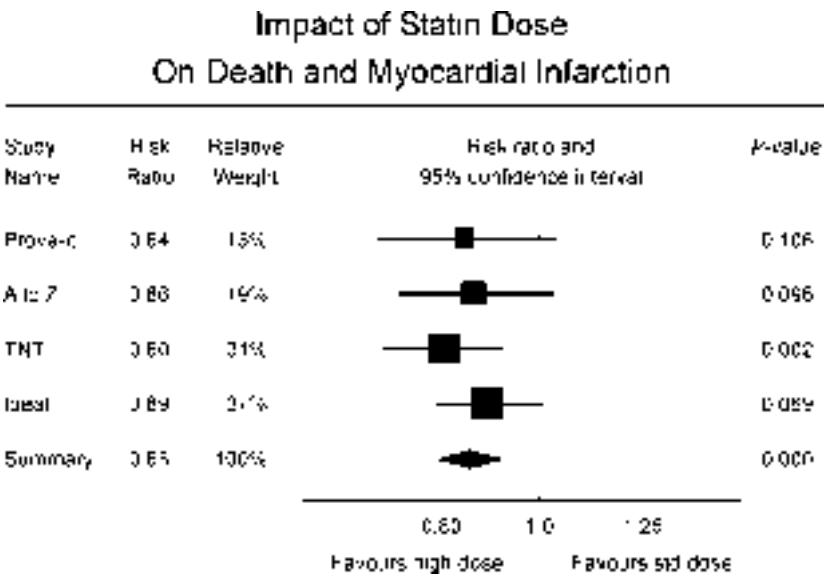

Figure 1.1 High-dose versus standard-dose of statins (adapted from Cannon *et al.*, 2006).

work with the effect sizes to assess the consistency of the effect across studies and to compute a summary effect.

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. 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 32.

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