CHAPTER 13

Fixed-Effect Versus Random-Effects Models

INTRODUCTION

In Chapter 11 and Chapter 12 we introduced the fixed-effect and random-effects models. Here, we highlight the conceptual and practical differences between them.

Consider the forest plots in Figures 13.1 and 13.2. They include the same six studies, but the first uses a fixed-effect analysis and the second a random-effects analysis. These plots provide a context for the discussion that follows.

DEFINITION OF A SUMMARY EFFECT

Both plots show a summary effect on the bottom line, but the meaning of this summary effect is different in the two models. In the fixed-effect analysis we assume that the true effect size is the same in all studies, and the summary effect is our estimate of this common effect size. In the random-effects analysis we assume that the true effect size varies from one study to the next, and that the studies in our analysis represent a random sample of effect sizes that could have been observed. The summary effect is our estimate of the mean of these effects.

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

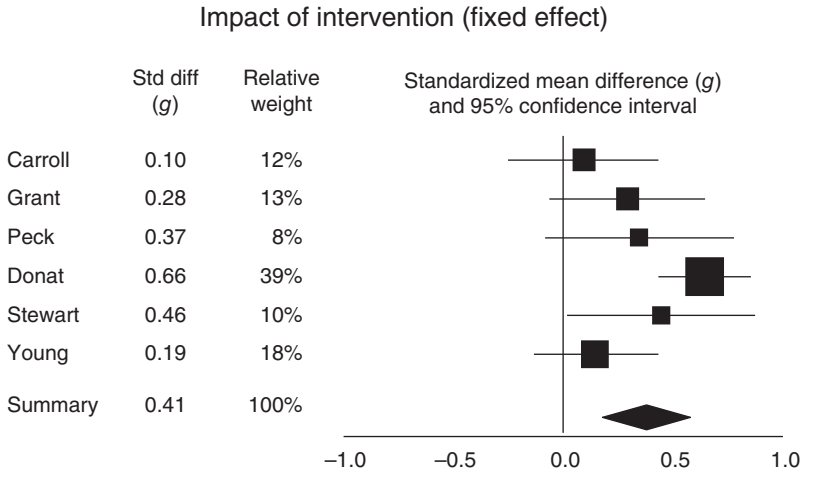

Figure 13.1 Fixed-effect model – forest plot showing relative weights.

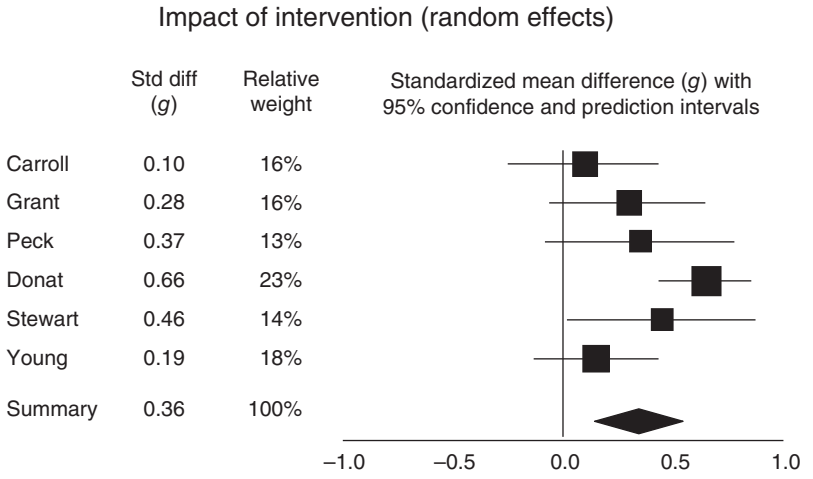

Figure 13.2 Random-effects model – forest plot showing relative weights.

ESTIMATING THE SUMMARY EFFECT

Under the fixed-effect model we assume that the true effect size for all studies is identical, and the only reason the effect size varies between studies is sampling error (error in estimating the effect size). Therefore, when assigning weights to the different studies we can largely ignore the information in the smaller studies since we have better information about the same effect size in the larger studies.

By contrast, under the random-effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about a different effect size, we want to be sure that all these effect sizes are represented in the summary estimate. This means that we cannot discount a small

study by giving it a very small weight (the way we would in a fixed-effect analysis). The estimate provided by that study may be imprecise, but it is information about an effect that no other study has estimated. By the same logic we cannot give too much weight to a very large study (the way we might in a fixed-effect analysis). Our goal is to estimate the mean effect in a range of studies, and we do not want that overall estimate to be overly influenced by any one of them.

In these graphs, the weight assigned to each study is reflected in the size of the box (specifically, the area) for that study. Under the fixed-effect model there is a wide range of weights (as reflected in the size of the boxes) whereas under the random-effects model the weights fall in a relatively narrow range. For example, compare the weight assigned to the largest study (Donat) with that assigned to the smallest study (Peck) under the two models. Under the fixed-effect model Donat is given about five times as much weight as Peck. Under the random-effects model Donat is given only 1.8 times as much weight as Peck.

EXTREME EFFECT SIZE IN A LARGE STUDY OR A SMALL STUDY

How will the selection of a model influence the overall effect size? In this example Donat is the largest study, and also happens to have the highest effect size. Under the fixed-effect model Donat was assigned a large share (39%) of the total weight and pulled the mean effect up to 0.41. By contrast, under the random-effects model Donat was assigned a relatively modest share of the weight (23%). It therefore had less pull on the mean, which was computed as 0.36.

Similarly, Carroll is one of the smaller studies and happens to have the smallest effect size. Under the fixed-effect model Carroll was assigned a relatively small proportion of the total weight (12%), and had little influence on the summary effect. By contrast, under the random-effects model Carroll carried a somewhat higher proportion of the total weight (16%) and was able to pull the weighted mean toward the left.

The operating premise, as illustrated in these examples, is that whenever τ^2 is nonzero, the relative weights assigned under random effects will be *more balanced* than those assigned under fixed effects. As we move from fixed effect to random effects, extreme studies will lose influence if they are large, and will gain influence if they are small.

CONFIDENCE INTERVAL

Under the fixed-effect model the only source of uncertainty is the within-study (sampling or estimation) error. Under the random-effects model there is this same source of uncertainty plus an additional source (between-studies variance). It follows that the variance, standard error, and confidence interval for the summary effect will always be larger (or wider) under the random-effects model than under the fixed-effect model (unless τ^2 is zero, in which case the two models are the same). In this example, the standard error is 0.064 for the fixed-effect model, and 0.105 for the random-effects model.

Consider what would happen if we had five studies, and each study had an infinitely large sample size. Under either model the confidence interval for the effect size in each

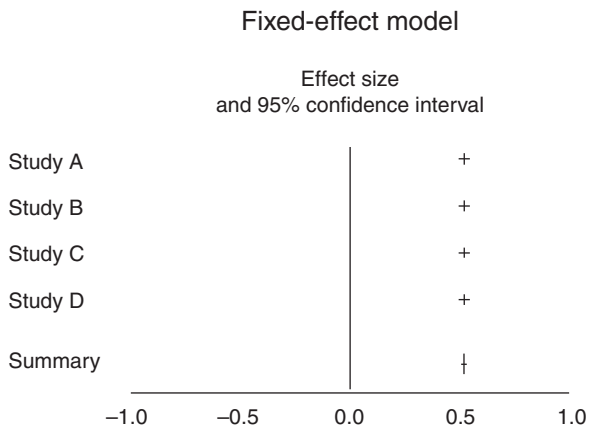

Figure 13.3 Very large studies under fixed-effect model.

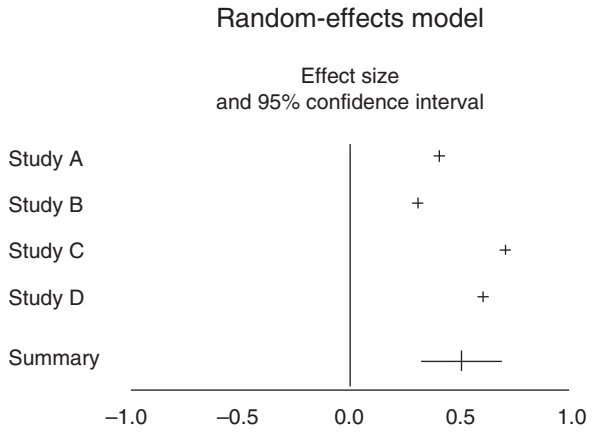

Figure 13.4 Very large studies under random-effects model.

study would have a width approaching zero, since we know the effect size in that study with perfect precision. Under the fixed-effect model the summary effect would also have a confidence interval with a width of zero, since we know the common effect precisely (Figure 13.3). By contrast, under the random-effects model the width of the confidence interval would not approach zero (Figure 13.4). While we know the effect in each study precisely, these effects have been sampled from a universe of possible effect sizes, and provide only an estimate of the mean effect. Just as the error within a study will approach zero only as the sample size approaches infinity, so too the error of these studies as an estimate of the mean effect will approach zero only as the number of studies approaches infinity.

More generally, it is instructive to consider what factors influence the standard error of the summary effect under the two models. The following formulas are based on a meta-analysis of means from k one-group studies, but the conceptual argument applies

to all meta-analyses. The within-study variance of each mean depends on the standard deviation (denoted σ) of participants' scores and the sample size of each study (n). For simplicity we assume that all of the studies have the same sample size and the same standard deviation (see Box 13.1 for details).

Under the fixed-effect model the standard error of the summary effect is given by

$$SE_M = \sqrt{\frac{\sigma^2}{k \times n}} \tag{13.1}$$

It follows that with a large enough sample size the standard error will approach zero, and this is true whether the sample size is concentrated on one or two studies, or dispersed across any number of studies.

Under the random-effects model the standard error of the summary effect is given by

$$SE_M = \sqrt{\frac{\sigma^2}{k \times n} + \frac{\tau^2}{k}} \tag{13.2}$$

The first term is identical to that for the fixed-effect model and, again, with a large enough sample size, this term will approach zero. By contrast, the second term (which reflects the between-studies variance) will only approach zero as the number of studies approaches infinity. These formulas do not apply exactly in practice, but the conceptual argument does. Namely, increasing the sample size within studies is not sufficient to reduce the standard error beyond a certain point (where that point is determined by τ^2 and k). If there is only a small number of studies, then the standard error could still be substantial even if the total n is in the tens of thousands or higher.

BOX 13.1 FACTORS THAT INFLUENCE THE STANDARD ERROR OF THE SUMMARY EFFECT

To illustrate the concepts with some simple formulas, let us consider a meta-analysis of studies with the very simplest design, such that each study comprises a single sample of n observations with standard deviation σ. We combine estimates of the mean in a meta-analysis. The variance of each estimate is

$$V_{Y_i} = \frac{\sigma^2}{n}$$

so the (inverse-variance) weight in a fixed-effect meta-analysis is

$$W_i = \frac{1}{\sigma^2/n} = \frac{n}{\sigma^2}$$

and the variance of the summary effect under the fixed-effect model the standard error is given by

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i} = \frac{1}{k \times n/\sigma^2} = \frac{\sigma^2}{k \times n}$$

BOX 13.1 CONTINUED

Therefore under the fixed-effect model the (true) standard error of the summary mean is given by

$$SE_M = \sqrt{\frac{\sigma^2}{k \times n}}$$

Under the random-effects model the weight awarded to each study is

$$W_i^* = \frac{1}{(\sigma^2/n) + \tau^2}$$

and the (true) standard error of the summary mean turns out to be

$$SE_M^* = \sqrt{\frac{\sigma^2}{k \times n} + \frac{\tau^2}{k}}.$$

THE NULL HYPOTHESIS

Often, after computing a summary effect, researchers perform a test of the null hypothesis. Under the fixed-effect model the null hypothesis being tested is that there is zero effect in *every study*. Under the random-effects model the null hypothesis being tested is that the *mean effect* is zero. Although some may treat these hypotheses as interchangeable, they are in fact different, and it is imperative to choose the test that is appropriate to the inference a researcher wishes to make.

WHICH MODEL SHOULD WE USE?

The selection of a computational model should be based on our expectation about whether or not the studies share a common effect size and on our goals in performing the analysis.

Fixed effect

It makes sense to use the fixed-effect model if two conditions are met. First, we believe that all the studies included in the analysis are functionally identical. Second, our goal is to compute the common effect size for the identified population, and not to generalize to other populations.

For example, suppose that a pharmaceutical company will use a thousand patients to compare a drug versus a placebo. Because the staff can work with only 100 patients at a time, the company will run a series of ten trials with 100 patients in each. The studies are identical in the sense that any variables which can have an impact on the outcome are the same across the ten studies. Specifically, the studies draw patients from a common pool, using the same researchers, dose, measure, and so on (we assume that there is no concern about practice effects for the researchers, nor for the different

starting times of the various cohorts). All the studies are expected to share a common effect and so the first condition is met. The goal of the analysis is to see if the drug works in the population from which the patients were drawn (and not to extrapolate to other populations), and so the second condition is met, as well.

In this example the fixed-effect model is a plausible fit for the data and meets the goal of the researchers. It should be clear, however, that this situation is relatively rare. The vast majority of cases will more closely resemble those discussed immediately below.

Random effects

By contrast, when the researcher is accumulating data from a series of studies that had been performed by researchers operating independently, it would be unlikely that all the studies were functionally equivalent. Typically, the subjects or interventions in these studies would have differed in ways that would have impacted on the results, and therefore we should not assume a common effect size. Therefore, in these cases the random-effects model is more easily justified than the fixed-effect model.

Additionally, the goal of this analysis is usually to generalize to a range of scenarios. Therefore, if one did make the argument that all the studies used an identical, narrowly defined population, then it would not be possible to extrapolate from this population to others, and the utility of the analysis would be severely limited.

A caveat

There is one caveat to the above. If the number of studies is very small, then the estimate of the between-studies variance (τ^2) will have poor precision. While the random-effects model is still the appropriate model, we lack the information needed to apply it correctly. In this case the reviewer may choose among several options, each of them problematic.

One option is to report the separate effects and not report a summary effect. The hope is that the reader will understand that we cannot draw conclusions about the effect size and its confidence interval. The problem is that some readers will revert to vote counting (see Chapter 33) and possibly reach an erroneous conclusion.

Another option is to perform a fixed-effect analysis. This approach would yield a descriptive analysis of the included studies, but would not allow us to make inferences about a wider population. The problem with this approach is that (a) we do want to make inferences about a wider population and (b) readers will make these inferences even if they are not warranted.

A third option is to take a Bayesian approach, where the estimate of τ^2 is based on data from outside of the current set of studies. This is probably the best option, but the problem is that relatively few researchers have expertise in Bayesian meta-analysis. Additionally, some researchers have a philosophical objection to this approach.

For a more general discussion of this issue see *When does it make sense to perform a meta-analysis* in Chapter 45.

MODEL SHOULD NOT BE BASED ON THE TEST FOR HETEROGENEITY

In the next chapter we will introduce a test of the null hypothesis that the between-studies variance is zero. This test is based on the amount of between-studies variance observed, relative to the amount we would expect if the studies actually shared a common effect size.

Some have adopted the practice of starting with a fixed-effect model and then switching to a random-effects model if the test of homogeneity is statistically significant. This practice should be strongly discouraged because the decision to use the random-effects model should be based on our understanding of whether or not all studies share a common effect size, and not on the outcome of a statistical test (especially since the test for heterogeneity often suffers from low power).

If the study effect sizes are seen as having been sampled from a *distribution* of effect sizes, then the random-effects model, which reflects this idea, is the logical one to use. If the between-studies variance is substantial (and statistically significant) then the fixed-effect model is inappropriate. However, even if the between-studies variance does not meet the criterion for statistical significance (which may be due simply to low power) we should still take account of this variance when assigning weights. If τ^2 turns out to be zero, then the random-effects analysis reduces to the fixed-effect analysis, and so there is no cost to using this model.

On the other hand, if one has elected to use the fixed-effect model *a priori* but the test of homogeneity is statistically significant, then it might be helpful to revisit the assumptions that led to the selection of a fixed-effect model.

Rice, Higgins, and Lumley (2018) have suggested that it would be helpful to distinguish between two versions of the fixed-effect analysis. The label 'fixed-effect' where 'effect' is in the singular applies to the case where all studies share a common true effect size and our goal is to make an inference about the one population represented in the analysis. The label 'fixed-effects' where 'effects' is in the plural applies to the case where the true effect size varies across studies and our goal is to make an inference to the set of populations included in the analysis. In both cases, our inference is limited to the population (or populations) included in the analysis, and the computational formula (and results) are identical under the two models. By contrast, under the random-effects model, we use the studies in the analysis to make an inference to a wider universe of comparable studies.

In this volume, we will focus on the fixed-effect (singular) model and the random-effects model. However, the reader should be aware that the fixed-effects (plural) model is also an option, and there are situations where we may prefer to use this model. This might be the case if we do not have a sufficient number of studies to employ the random-effects model reliably. This might also be the case if the analysis is being used to support the approval of a new drug, and we need to make an inference to the studies in the analysis, without generalizing to a larger universe. In these cases, researchers have sometimes tried to justify the use of the fixed-effect model by arguing that the studies share a common effect size, when in fact the studies

do not share a common effect size. A better option is to acknowledge that the effect size varies across studies and apply the fixed-effects (plural) model.

For a lengthy discussion of this issue, see Borenstein (2019); Rice et al. (2018).

CONCLUDING REMARKS

Our discussion of differences between the fixed-effect model and the random-effects model focused largely on the computation of a summary effect and the confidence intervals for the summary effect. We did not address the implications of the dispersion itself. Under the fixed-effect model we assume that all dispersion in observed effects is due to sampling error, but under the random-effects model we allow that some of that dispersion reflects real differences in effect size across studies. In the chapters that follow we discuss methods to quantify that dispersion and to consider its substantive implications.

SUMMARY POINTS

- A fixed-effect meta-analysis estimates a single effect that is assumed to be common to every study, while a random-effects meta-analysis estimates the mean of a distribution of effects.
- Study weights are more balanced under the random-effects model than under the fixed-effect model. Large studies are assigned less relative weight and small studies are assigned more relative weight as compared with the fixed-effect model.
- The standard error of the summary effect and (it follows) the confidence intervals for the summary effect are wider under the random-effects model than under the fixed-effect model.
- The selection of a model must be based solely on the question of which model fits the distribution of effect sizes, and takes account of the relevant source(s) of error. When studies are gathered from the published literature, the random-effects model is generally a more plausible match.
- The strategy of starting with a fixed-effect model and then moving to a random-effects model if the test for heterogeneity is significant is a mistake, and should be strongly discouraged.