PART 5

Complex Data Structures

Overview

Thus far we have assumed that each study contributes one (and only one) effect size to a meta-analysis. In this section we consider cases where studies contribute more than one effect size to the meta-analysis. These usually fall into one of the following types.

Multiple independent subgroups within a study

Sometimes a single study will report data for several cohorts of participants. For example, if researchers anticipate that the treatment effect (e.g., drug versus placebo) could vary by age, they might report the treatment effect separately for children and for adults. Similarly, if researchers anticipate that the treatment effect could vary by disease stage they might report the effect separately for patients enrolled with early-stage disease and for those enrolled with late-stage disease.

The defining feature here is that the subgroups are independent of each other, so that each provides unique information. For this reason, we can treat each subgroup as though it were a separate study, which is sometimes the preferred method. However, there are sometimes other options to consider, and we will be discussing these as well.

Multiple outcomes or time-points within a study

In some cases researchers will report data on several related, but distinct outcomes. A study that looked at the impact of tutoring might report data on math scores and also on reading scores. A study that looked at the association between diet and cardiovascular disease might report data on stroke and also on myocardial infarction. Similarly, a study that followed subjects over a period of time may report data using the same scale but at a series of distinct time-points. For example, studies that looked at the impact of an intervention to address a phobia might collect data at one month, six months, and twelve months.

The defining feature here is that *the same* participants provide data for the different outcomes (or time-points). We cannot treat the different outcomes as though they were independent as this would lead to incorrect estimates of the variance for the summary effect. We will show how to correct the variance to take account of the relationship among the outcomes.

More than one comparison group within a study

Sometimes, a study will include several treatment groups and a single control group. For example, one effect size may be defined as the difference between the placebo group and drug *A*, while another is defined as the difference between the same placebo group and drug *B*.

The defining feature here is similar to multiple outcomes, in that some participants (those in the control group) contribute information to more than one effect size. The methods proposed for dealing with this problem are similar to those proposed for multiple outcomes. They also include some options that are unique to the case of multiple comparisons.

How this Part is organized

The next three chapters address each of these cases in sequence. Within each chapter we first show how to combine data to yield a *summary* effect, and then show how to look at *differences* in effects.

The worked examples in these chapters use the fixed-effect model. We adopt this approach because it involves fewer steps and thus allows us to focus on the issue at hand, which is how to compute an effect size and variance. Once we have these effect sizes we can use them for a fixed-effect or a random-effects analysis, and the latter is generally more appropriate.

In the worked examples we deliberately use a generic effect size rather than specifying a particular effect size such as a standardized mean difference or a log odds ratio. The methods discussed here can be applied to any effect size, including those based on continuous, binary, correlational, or other kinds of data. As always, computations for risk ratios or odds ratios would be performed using log values, and computations for correlations would be performed using the Fisher's *z* transformed values.

Independent Subgroups within a Study

Introduction
Combining across subgroups
Comparing subgroups

INTRODUCTION

The first case of a complex data structure is the case where studies report data from two or more independent subgroups.

Suppose we have five studies that assessed the impact of a treatment on a specific type of cancer. All studies followed the same design, with patients randomly assigned to either standard or aggressive treatment for two months. In each study, the results were reported separately for patients enrolled with stage-1 cancer and for those enrolled with stage-2 cancer. The stage-1 and stage-2 patients represent two independent subgroups since each patient is included in one group or the other, but not both.

If our goal was to compute the summary treatment effect for all stage-1 patients and, separately, for all stage-2 patients, then we would perform two separate analyses. In this case we would treat each subgroup as a separate study, and include the stage-1 studies in one analysis and the stage-2 studies in the other.

This chapter addresses the case where we want to use data from two or more subgroups in the *same* analysis. Specifically,

- We want to compute a summary effect for the impact of the intervention for stage-1 and stage-2 patients combined.
- Or, we want to compare the effect size for stage-1 patients versus stage-2 patients.

Table 23.1 Independent subgroups – five fictional studies.

Study	Subgroup	ES	Variance
Study 1	Stage 1	0.300	0.050
	Stage 2	0.100	0.050
Study 2	Stage 1	0.200	0.020
	Stage 2	0.100	0.020
Study 3	Stage 1	0.400	0.050
	Stage 2	0.200	0.050
Study 4	Stage 1	0.200	0.010
	Stage 2	0.100	0.010
Study 5	Stage 1	0.400	0.060
	Stage 2	0.300	0.060

COMBINING ACROSS SUBGROUPS

The defining feature of independent subgroups is that each subgroup contributes independent information to the analysis. If the sample size within each subgroup is 100, then the effective sample size across two subgroups is 200, and this will be reflected in the precision of the summary effect. However, within this framework we have several options for computing the summary effect.

We shall pursue the example of five studies that report data separately for patients enrolled with stage-1 or stage-2 cancer. The effect size and variance for each subgroup are shown in Table 23.1 and are labeled simply *ES* and *Variance* to emphasize the point that these procedures can be used with any effect size. If the outcome was continuous (means and standard deviations), the effect size might be a standardized mean difference. If the outcome was binary (for example, whether or not the cancer had metastasized), the effect size might be the log risk ratio.

Using subgroup as unit of analysis (option 1a)

One option is simply to treat each subgroup as a separate study. This is shown in Table 23.2, where each subgroup appears on its own row and values are summed across the ten rows.

Then, using formulas (11.3) to (11.5),

$$M = \frac{76.666}{413.333} = 0.1855,$$

with variance

$$V_M = \frac{1}{413.333} = 0.0024$$

Table 23.2 Independent subgroups – summary effect.

Study	Subgroup	ES	Variance	WT	ES*WT
Study 1	Stage 1	0.30	0.05	20.000	6.000
	Stage 2	0.10	0.05	20.000	2.000
Study 2	Stage 1	0.20	0.02	50.000	10.000
	Stage 2	0.10	0.02	50.000	5.000
Study 3	Stage 1	0.40	0.05	20.000	8.000
	Stage 2	0.20	0.05	20.000	4.000
Study 4	Stage 1	0.20	0.01	100.000	20.000
	Stage 2	0.10	0.01	100.000	10.000
Study 5	Stage 1	0.40	0.06	16.667	6.667
	Stage 2	0.30	0.06	16.667	5.000
Sum				413.333	76.667

and standard error

$$SE_M = \sqrt{0.0024} = 0.0492.$$

Using study as unit of analysis (option 1b)

A second option is to compute a composite score for each study and use this in the analysis, as in Figure 23.1. The unit of analysis is then the study rather than the subgroup.

Computing a combined effect across subgroups within a study
The mean and variance of the composite within a study are computed by performing a fixed-effect meta-analysis on the subgroups for that study. For study 1, this is shown in Figure 23.1 and in Table 23.3.

We apply formulas (11.3) and (11.4) to yield a mean effect

$$M = \frac{8.0000}{40.0000} = 0.2000$$

Study	Subgroup	Effect size	Variance	Mean	Variance
Study 1	Stage-1	0.30	0.05	0.200	0.025
	Stage-2	0.10	0.05		

Figure 23.1 Creating a synthetic variable from independent subgroups.

Table 23.3 Independent subgroups – synthetic effect for study 1.

Subgroup	Effect Y	Variance V_Y	Weight W	Computed WY
Stage 1	0.30	0.05	20.000	6.000
Stage 2	0.10	0.05	20.000	2.000
Sum			40.000	8.000

with variance

$$V_M = \frac{1}{40.0000} = 0.0250.$$

Note that the variance for the study (0.025) is one-half as large as the variance for either subgroup (0.050) since it is based on twice as much information.

This procedure is used to form a composite effect size and variance for each study, as shown in Table 23.4. Then, we perform a meta-analysis working solely with these study-level effect sizes and variances.

At this point we can proceed to the meta-analysis using these five (synthetic) scores. To compute a summary effect and other statistics using the fixed-effect model, we apply the formulas starting with (11.3). Using values from the line labeled *Sum* in Table 23.4,

$$M = \frac{76.667}{413.333} = 0.1855$$

with variance

$$V_M = \frac{1}{413.333} = 0.0024$$

Table 23.4 Independent subgroups – summary effect across studies.

Study	Subgroup	ES	Variance	ES	Variance	WT	ES*WT
Study 1	Stage 1	0.300	0.050	0.200	0.025	40.000	8.000
	Stage 2	0.100	0.050				
Study 2	Stage 1	0.200	0.020	0.150	0.010	100.000	15.000
	Stage 2	0.100	0.020				
Study 3	Stage 1	0.400	0.050	0.300	0.025	40.000	12.000
	Stage 2	0.200	0.050				
Study 4	Stage 1	0.200	0.010	0.150	0.005	200.000	30.000
	Stage 2	0.100	0.010				
Study 5	Stage 1	0.400	0.060	0.350	0.030	33.333	11.667
	Stage 2	0.300	0.060				
Sum						413.333	76.667

and standard error

$$SE_M = \sqrt{0.0024} = 0.0492.$$

Note that the summary effect and variance computed using study as the unit of analysis are identical to those computed using subgroup as the unit of analysis. This will always be the case when we use a fixed effect analysis to combine effects at both steps in the analysis (within studies and across studies).

However, the two methods will yield different results if we use a random effects analysis to combine effects across studies. This follows from the fact that T^2 (which is used to compute weights) may be different if based on variation in effects from study to study than if based on variation in effects from subgroup to subgroup. Therefore, the decision to use subgroup or study as the unit of analysis should be based on the context for computing T^2. Consider the following two cases.

- Case 1: Five researchers studied the impact of an intervention. Each researcher selected five schools at random, and each school is reported as an independent subgroup within the study. In this situation, between-school variation applies just as much within studies as across studies. We would therefore use subgroup as the unit of analysis (option 1a).
- Case 2: Five researchers have published papers on the impact of an intervention. Each researcher worked in a single school, and the subgroups are grade 1, grade 2, and so on. We expect the effects to be relatively consistent within a school, but to vary substantially from one school to the next. To allow for this between-school variation we should use a random-effects model only *across studies*. To properly estimate this component of uncertainty we would use study as the unit of analysis (option 1b).

Recreating the summary data for the full study (option 2)

Options 1a and 1b differ in the unit of analysis, but they have in common that the effect size is computed *within* subgroups. Another option (option 2) is to use the summary data from the subgroups to recreate the data for the study as a whole, and then use this summary data to compute the effect size and variance.

When the subgroup data are reported as 2×2 tables we can simply collapse cells to recreate the data for the full sample. That is, we sum cell A over all subgroups to yield an overall cell A, and repeat the process for cells B, C, and D.

When the subgroup data are reported as means, standard deviations, and sample size for each treatment group the combined sample size is summed across subgroup. For example, for treatment group 1,

$$n_1 = n_{11} + n_{12}, \tag{23.1}$$

the combined mean is computed as the weighted mean (by sample size) across groups,

$$\overline{X}_1 = \frac{n_{11}\,\overline{X}_{11} + n_{12}\,\overline{X}_{12}}{n_{11} + n_{12}}\,, \tag{23.2}$$

and the combined standard deviation is computed as

$$S_1 = \sqrt{\frac{(n_{11}-1)\,S_{11}^2 + (n_{12}-1)\,S_{12}^2 + \dfrac{n_{11}n_{12}}{n_{11}+n_{12}}\,(\overline{X}_{11} - \overline{X}_{12})^2}{n_{11} + n_{12} - 1}}\,, \tag{23.3}$$

where $\overline{X}_{11}$, $\overline{X}_{12}$ are the means in subgroups 1 and 2 of treatment group 1; S_{11}, S_{12} the standard deviations, and n_{11}, n_{12} the sample sizes; of subgroups 1 and 2.

When the subgroup data are reported as correlations, analogous formulas exist to recreate the correlation for the full study, but these are beyond the scope of this book.

Option 2 is sometimes used when some studies report summary data for all subjects combined, while others break down the data by subgroups. If the researcher believes that the subgroup classifications are unimportant, and wants to have a uniform approach for all studies (to compute an effect size from a single set of summary data) then this option will prove useful.

However, it is important to understand that this is a fundamentally different approach than other options. To return to the example introduced at the start of this chapter, under options 1a and 1b the effect size was computed *within* subgroups, which means that the effect size is *the impact of intervention controlling for cancer stage* (even if we then merge the effect sizes to yield an overall effect). By contrast, under option 2 we merge the summary data and *then* compute an effect size. Therefore the effect is 'the impact of intervention *ignoring* cancer stage'.

When the studies are randomized trials, the proportion of participants assigned to each treatment is typically constant from one subgroup to the next. In this case there is not likely to be a confounder between treatment and subgroup, and so either approach would be valid. By contrast, in observational studies the proportion of exposed subjects may vary from one subgroup to the next, which would yield confounding between exposure and subgroup. In this case option 2 should *not* be used. (This is the same issue discussed in Chapter 33, under the heading of Simpson's paradox.)

COMPARING SUBGROUPS

When our goal is to *compare* the effect size in different subgroups (rather than *combine* them) we have two options, as follows.

Using subgroup as unit of analysis

One option is simply to treat each subgroup as a separate study, where each study is classified (in this example) as stage 1 or stage 2. We then compute a summary

effect for all the stage 1 effects, another for all the stage 2 effects, and then compare the two using a Z-test or analysis of variance as discussed in Chapter 19.

A second option is to compute the effect size within subgroups for each study, and then to compute the difference in effects within each study. In this case each study will contribute one effect to the analysis, where the effect is the difference between subgroups.

The first option is a more general approach. It allows us to work with studies that report data for any subgroup or combination of subgroups (one study has subgroups *A* and *B*, another *B* and *C*, and so on), and then to use all relevant subgroups to compute the summary effect.

The second option can only be used if all studies report data on the same two subgroups, which is relatively rare. When this option *can* be used, however, it will usually yield a more precise estimate of the difference in effects in random effects analyses, and is also desirable because differences in effects are not confounded by possible differences between studies.

SUMMARY POINTS

- When we have independent subgroups within a study, each subgroup contributes independent information. Therefore, if the sample size within each subgroup is 100, then the effective sample size across five subgroups is 500. In this sense, independent subgroups are no different than independent studies.
- To compute a summary effect we typically compute the effect within subgroups and then either use these effects as the unit of analysis, or merge effects within each study and use study as the unit of analysis. A second option is to combine the summary data from all subgroups to recreate the original study-level data, and then compute an effect size from this data. The second approach should be used only in limited circumstances.
- To compare effects across subgroups we typically use subgroup as the unit of analysis. In some cases we may also be able to compute the difference between subgroups in each study, and use study as the unit of analysis.

Multiple Outcomes or Time-Points within a Study

Introduction
Combining across outcomes or time-points
Comparing outcomes or time-points within a study

INTRODUCTION

The second case of a complex data structure is the case where a study reports data on more than one outcome, or more than one time-point, where the different outcomes (or time-points) are based on the same participants.

For example, suppose that five studies assessed the impact of tutoring on student performance. All studies followed the same design, with students randomly assigned to either of two groups (tutoring or control) for a semester, after which they were tested for proficiency in reading and math. The effect was reported separately for the reading and the math scores, but within each study *both outcomes were based on the same students*.

Or, consider the same situation with the following difference. This time, assume that each study tests only for reading but does so at two time-points (immediately after the intervention and again six months later). The effect was reported separately for each time-point but *both measures were based on the same students*.

For our purposes the two situations (multiple outcomes for the same subjects or multiple time-points for the same subjects) are identical, and we shall treat them as such in this discussion. We shall use the term *outcomes* throughout this chapter, but the reader can substitute *time-points* in every instance.

If our goal was to compute a summary effect for the impact of the intervention on reading, and *separately* for the impact of the intervention on math scores, we would simply perform two separate meta-analyses, one using the data for reading and the other using the data for math. The issues we address in this chapter are how to

proceed when we want to incorporate both outcomes in the same analysis. Specifically,

- We want to compute a summary effect for the intervention on *Basic skills*, which combines the data from reading and math.
- Or, we want to investigate the *difference* in effect size for reading versus math.

In either case, the issue we need to address is that the data for reading and math are not independent of each other and therefore the errors are correlated.

COMBINING ACROSS OUTCOMES OR TIME-POINTS

The data for the five fictional studies are shown in Table 24.1. In study 1, for example, the effect size for reading was 0.30 with a variance of 0.05, and the effect size for math was 0.10 with a variance of 0.05.

While it might seem that we could treat each line of data as a separate study and perform a meta-analysis with ten *studies*, this is problematic for two reasons. One problem is that in computing the summary effect across studies this approach will assign more weight to studies with two outcomes than to studies with one outcome. (While this problem does not exist in our set of studies, it would be a problem if the number of outcomes varied from study to study.)

The second, and more fundamental problem, is that this approach leads to an improper estimate of the precision of the summary effect. This is because it treats the separate outcomes as providing independent information, when in fact the math and reading scores come from the same set of students and therefore are not independent of each other. If the outcomes are positively correlated (which is almost always the case with effects that we would want to combine), this approach underestimates the error (and overestimates the precision) of the summary effect.

Table 24.1 Multiple outcomes – five fictional studies.

Study	Outcome	ES	Variance
Study 1	Reading	0.300	0.050
	Math	0.100	0.050
Study 2	Reading	0.200	0.020
	Math	0.100	0.020
Study 3	Reading	0.400	0.050
	Math	0.200	0.050
Study 4	Reading	0.200	0.010
	Math	0.100	0.010
Study 5	Reading	0.400	0.060
	Math	0.300	0.060

Table 24.2 Creating a synthetic variable as the mean of two outcomes.

Study	Outcome	Effect size	Variance	Mean	Variance
Study 1	Math	0.30	0.05	0.20	?
	Reading	0.10	0.05		

Note. If the correlation between outcomes is negative, this approach will over-estimate the error (and underestimate the precision) of the summary effect. The solutions presented below will work for this case as well, but in the discussion we assume that we are dealing with a positive correlation.

To address these problems, rather than treating each outcome as a separate unit in the analysis, we'll compute the mean of the outcomes for each study, and use this synthetic score as the unit of analysis. In Table 24.2 we show this schematically for study 1.

We start with summary data for two outcomes (math and reading), and compute an effect size and variance for each. If the data are continuous (means and standard deviations on the exam) the effect size might be Hedges' g. If the data are binary (number of students passing the course) the effect size might be a log risk ratio. And so on. Then, we compute a synthetic effect size for *Basic skills* which incorporates both the math and reading effects. The method used to compute this effect size and its variance is explained below.

Since every study will be represented by one score in the meta-analysis regardless of the number of outcomes included in the mean, this approach solves the problem of more weight being assigned to studies with more outcomes. This approach also allows us to address the problem of non-independent information, since the formula for the variance of the synthetic variable will take into account the correlation among the outcomes.

Computing a combined effect across outcomes

Our notation will be to use Y_1, Y_2 etc. for effect sizes from different outcomes or time points within a study, and Y_j to refer to the jth of these. Strictly, we should use Y_{ij}, for the jth outcome (or time-point) in the ith study. However, we drop the i subscript for convenience. The effect size for *Basic skills* is computed as the mean of the reading and math scores,

$$\overline{Y} = \frac{1}{2}(Y_1 + Y_2). \tag{24.1}$$

This is what we would use as the effect estimate from this study in a meta-analysis. Using formulas described in Box 24.1, the variance of this mean is

$$V_{\overline{Y}} = \frac{1}{4}\left(V_{Y_1} + V_{Y_2} + 2r\sqrt{V_{Y_1}}\sqrt{V_{Y_2}}\right) \tag{24.2}$$

BOX 24.1 COMPUTING THE VARIANCE OF A COMPOSITE OR A DIFFERENCE

1. The variance of the sum of two correlated variables

If we know that the variance of Y_1 is V_1 and the variance of Y_2 is V_2, then

$$\text{var}(Y_1 + Y_2) = V_1 + V_2 + 2r\sqrt{V_1}\sqrt{V_2},$$

where r is the correlation coefficient that describes the extent to which Y_1 and Y_2 co-vary. If Y_1 and Y_2 are inextricably linked (so that a change in one determines completely the change in the other), then $r = 1$, and the variance of the sum is roughly twice the sum of the variances. At the other extreme, if Y_1 and Y_2 are unrelated, then $r = 0$ and the variance is just the sum of the individual variances. This is because when the variables are unrelated, knowing both gives us twice as much information, and so the variance is halved compared with the earlier case.

2. The impact of a scaling factor on the variance

If we know the variance of X, then the variance of a scalar (say c) multiplied by X is given by

$$\text{var}(cX) = c^2 \times \text{var}(X).$$

3. The variance of the mean of two correlated variables

Combining 1 with 2, we can see that the variance of the mean of Y_1 and Y_2 is

$$\text{var}\left(\frac{1}{2}(Y_1 + Y_2)\right) = \left(\frac{1}{2}\right)^2 \text{var}(Y_1 + Y_2) = \frac{1}{4}\left(V_1 + V_2 + 2r\sqrt{V_1}\sqrt{V_2}\right).$$

4. The variance of the sum of several correlated variables

If we know Y_i has variance V_i for several variables $i = 1, \ldots, m$, then the formula in 1 extends as follows:

$$\text{var}\left(\sum_{i=1}^{m} Y_i\right) = \sum_{i=1}^{m} V_i + \sum_{i \neq j}\left(r_{ij}\sqrt{V_i}\sqrt{V_j}\right)$$

where r_{ij} is the correlation between Y_i and Y_j.

5. The variance of the mean of several correlated variables

Combining 4 with 2, we can see that the variance of the mean of several variables is

$$\text{var}\left(\frac{1}{m}\sum_{i=1}^{m} Y_i\right) = \left(\frac{1}{m}\right)^2 \text{var}\left(\sum_{i=1}^{m} Y_i\right) = \left(\frac{1}{m}\right)^2 \left(\sum_{i=1}^{m} V_i + \sum_{i \neq j}\left(r_{ij}\sqrt{V_i}\sqrt{V_j}\right)\right).$$

6. The variance of the difference between two correlated variables

If we know that the variance of Y_1 is V_1 and the variance of Y_2 is V_2, then

$$\text{var}(Y_1 - Y_2) = V_1 + V_2 - 2r\sqrt{V_1}\sqrt{V_2},$$

BOX 24.1 CONTINUED

where r is the correlation coefficient that describes the extent to which Y_1 and Y_2 co-vary. If Y_1 and Y_2 are inextricably linked (so that a change in one determines completely the change in the other), then $r = 1$, and the variance of the difference is close to zero. At the other extreme, if Y_1 and Y_2 are unrelated, then $r = 0$ and the variance is the sum of the individual variances. If the correlation is $r = 0.5$ then the variance is approximately the average of the two variances.

---

where r is the correlation between the two outcomes. If both variances V_{Y_1} and V_{Y_2} are equal (say to V), then (24.2) simplifies to

$$V_{\overline{Y}} = \frac{1}{2}V(1 + r). \qquad (24.3)$$

In the running example, in study 1 the effect sizes for math and reading are 0.30 and 0.10, the variance for each is 0.02. Suppose we know that the correlation between them is 0.50. The composite score for *Basic skills* ($\overline{Y}$) is computed as

$$\overline{Y} = \frac{1}{2}(0.30 + 0.10) = 0.2000,$$

with variance (based on (24.2))

$$V_{\overline{Y}} = \frac{1}{4}(0.05 + 0.05 + 2 \times 0.50 \times \sqrt{0.05} \times \sqrt{0.05}) = 0.0375,$$

or, equivalently (using (24.3)),

$$V_{\overline{Y}} = \frac{1}{2} \times 0.05 \times (1 + 0.50) = 0.0375.$$

Using this formula we can see that if the correlation between outcomes was zero, the variance of the composite would be 0.025 (which is half as large as either outcome alone) because the second outcome provides entirely independent information. If the correlation was 1.0 the variance of the composite would be 0.050 (the same as either outcome alone) because all information provided by the second outcome is redundant. In our example, where the correlation is 0.50 (*some* of the information is redundant) the variance of the composite falls between these extremes. When we were working with independent subgroups (earlier in this chapter) the correlation was zero, and therefore the variance of the composite was 0.025.

These formulas are used to create Table 24.2, where the variance for each composite is based on formula (24.2) and the weight is simply the reciprocal of the variance.

At this point we can proceed to the meta-analysis using these five (synthetic) scores. To compute a summary effect and other statistics using the fixed-effect model, we apply the formulas starting with (11.3). Using values from the line labeled *Sum* in Table 24.3,

Table 24.3 Multiple outcomes – summary effect.

Study	Outcome	ES	Variance	ES	Correlation	Variance	Weight	ES*WT
Study 1	Reading	0.300	0.050	0.200	0.500	0.038	26.667	5.333
	Math	0.100	0.050					
Study 2	Reading	0.200	0.020	0.150	0.600	0.016	62.500	9.375
	Math	0.100	0.020					
Study 3	Reading	0.400	0.050	0.300	0.600	0.040	25.000	7.500
	Math	0.200	0.050					
Study 4	Reading	0.200	0.010	0.150	0.400	0.007	142.857	21.429
	Math	0.100	0.010					
Study 5	Reading	0.400	0.060	0.350	0.800	0.054	18.519	6.481
	Math	0.300	0.060					
Sum							275.542	50.118

$$M = \frac{50.118}{275.542} = 0.1819,$$

with variance

$$V_M = \frac{1}{275.542} = 0.0036.$$

The average difference between the tutored and control groups on *Basic skills* is 0.1819 with variance 0.0036 and standard error 0.060. The 95% confidence interval for the average effect is 0.064 to 0.300. The *Z*-value for a test of the null is 3.019 with a two-sided *p*-value of 0.003.

Working with more than two outcomes per study

These formulas can be extended to accommodate any number of outcomes. If *m* represents the number of outcomes within a study, then the composite effect size for that study would be computed as

$$\overline{Y} = \frac{1}{m}\left(\sum_j^m Y_j\right), \tag{24.4}$$

and the variance of the composite is given by

$$V_{\overline{Y}} = \left(\frac{1}{m}\right)^2 \mathrm{var}\left(\sum_{j=1}^m Y_i\right) = \left(\frac{1}{m}\right)^2 \left(\sum_{j=1}^m V_i + \sum_{j\neq k}\left(r_{jk}\sqrt{V_j}\sqrt{V_k}\right)\right) \tag{24.5}$$

as derived in Box 24.1. If the variances are all equal to *V* and the correlations are all equal to *r*, then (24.5) simplifies to

$$V_{\overline{Y}} = \frac{1}{m}V(1 + (m-1)r). \tag{24.6}$$

Impact of the correlations on the combined effect

One issue to consider is what happens as the correlation moves toward 1.0. Continuing with the simplified situation where all observations within a study have the same variance (*V*) and all pairs of observations within the study have the same correlation (*r*), if the *m* observations are independent of each other ($r = 0$), the variance of the composite is *V/m*. If the *m* observations are not independent of each other, then the variance of the composite is *V/m* times a correction factor. We will refer to this correction factor as the *variance inflation factor* (*VIF*), which is

$$VIF = 1 + (m - 1)r, \tag{24.7}$$

where *m* is the number of observations and *r* is the correlation between each pair. An increase in either *m* or *r* (or both) will result in a higher inflation of the variance compared with treating the different outcomes as independent of each other.

In Table 24.4 we explore how the variance inflation factor depends on the value of *r*, the correlation coefficient. For the purposes of this illustration we assume the simplistic situation of a study having just two outcomes ($m = 2$) with the same variance ($V = 0.2$) for each outcome. Each column in the table (A-E) corresponds to a different correlation coefficient between these outcomes.

The variance of the composite for the study is

$$V_{\bar{Y}} = \frac{1}{m} V \times VIF. \tag{24.8}$$

Taking as an example column C, the correlation is $r = 0.50$, and the variance inflation factor is

$$VIF = 1 + (2 - 1) \times 0.50 = 1.5000.$$

Table 24.4 Multiple outcomes – Impact of correlation on variance of summary effect.

	A	B	C	D	E
Effect size (here assumed identical for all outcomes)	0.4	0.4	0.4	0.4	0.4
Number of outcomes (*m*)	2	2	2	2	2
Variance of each outcome (*V*)	0.2	0.2	0.2	0.2	0.2
Correlation among outcomes (*r*)	0.000	0.250	0.500	0.750	1.000
Variance inflation factor (*VIF*)	1.000	1.250	1.500	1.750	2.000
Variance of composite	0.100	0.125	0.150	0.175	0.200
Standard error of composite	0.316	0.354	0.387	0.418	0.447
Standard error inflation factor	1.000	1.118	1.225	1.323	1.414
Lower limit	−0.220	−0.293	−0.359	−0.420	−0.477
Upper limit	1.020	1.093	1.159	1.220	1.277
p-value (2-tailed)	0.206	0.258	0.302	0.339	0.371

Thus, the variance of the composite score for the study is

$$V_{\bar{Y}} = \frac{1}{2} \times 0.2 \times 1.5000 = 0.150.$$

As we move from left to right in the table (from a correlation of 0.00 to 1.00) the variance inflation factor (*VIF*) and (by definition) the variance double. If the inflation factor for the variance moves from 1.00 to 2.00, it follows that the inflation factor for the standard error (which is the square root of the variance) will move from 1.00 to 1.44. Therefore, the width of the confidence interval will increase by a factor of 1.44 (and, correspondingly, the *Z*-value for the test of the null hypothesis for this study would decrease by a factor of 1.44).

When the correlation is unknown

This table also provides a mechanism for working with synthetic variables when we don't know the correlation among outcomes. Earlier, we assumed that the correlation between math and reading was known to be 0.50, and used that value to compute the standard error of the combined effect and the related statistics. In those cases where we don't know the correlation for the study in question, we should still be able to use other studies in the same field to identify a plausible range for the correlation. We could then perform a sensitivity analysis and might assume, for example, that if the correlation falls in the range of 0.50 to 0.75 then the standard error probably falls in the range of 0.39 to 0.42 (columns C to D in the table).

Researchers who do not know the correlation between outcomes sometimes fall back on either of two 'default' positions. Some will include both math and verbal scores in the analysis and treat them as independent. Others would use the average of the reading variance and the math variance. It is instructive, therefore, to consider the practical impact of these choices.

Treating the two outcomes as independent of each other yields the same precision as setting the correlation at 0.00 (column A). By contrast, using the average of the two variances yields the same precision as setting the correlation at 1.00 (column E). In effect, then, researchers who adopt either of these positions as a way of bypassing the need to specify a correlation, are actually adopting a correlation, albeit implicitly. And, the correlation that they adopt falls at either extreme of the possible range (either zero or 1.0). The first approach is almost certain to underestimate the variance and overestimate the precision. The second approach is almost certain to overestimate the variance and underestimate the precision. In this context, the idea of working with a *plausible* range of correlations rather than the *possible* range offers some clear advantages.

As we noted at the outset, exactly the same approach applies to studies with multiple outcomes and to studies with multiple time-points. However, there could be a distinction between the two when it comes to deciding what is a plausible range of correlations. When we are working with different outcomes at a single point in time, the

plausible range of correlations will depend on the similarity of the outcomes. When we are working with the same outcome at multiple time-points, the plausible range of correlations will depend on such factors as the time elapsed between assessments and the stability of the relative scores over this time period.

One issue to consider is what happens if the correlations between multiple outcomes are higher in some studies than in others. This variation will affect the *relative weights* assigned to different studies, with *more weight* going to the study with a *lower correlation*. In the running example the variances for reading and math were the same in studies 1 and 3, but the correlation between reading and math was higher in study 3. Therefore, study 3 had a higher variance and was assigned less weight in the meta-analysis.

COMPARING OUTCOMES OR TIME-POINTS WITHIN A STUDY

We now turn to the problem of investigating *differences* between outcomes or between time-points. To extend the current example, suppose that each study reports the impact of the intervention for math and for reading and we want to know if the impact is stronger for one of these outcomes than for the other. Or, each study reports the effect at 6 months and 12 months, and we want to know if the effect changes over time.

When our goal was to compute a combined effect based on both outcomes our approach was to create a synthetic variable for each study (defined as the mean of the effect sizes) and to use this as the effect size in the analysis. We will follow the same approach here, except that the synthetic variable will be defined as the difference in effect sizes rather than their mean.

The approach is shown in Table 24.5. As before, we start with summary data for two outcomes (math and reading), and compute an effect size and variance for each. Then, we compute a synthetic effect size, which is now the *difference* between the two effects and its variance, as explained below.

This approach allows us to address the problem of correlated error, since the formula for the variance of the synthetic variable will take into account the correlation between the outcomes.

Computing a variance for correlated outcomes

Whenever we use sample data to estimate a difference, the variance reflects the error of our estimate. If we compute the difference of two *unrelated* outcomes, each with

Table 24.5 Creating a synthetic variable as the difference between two outcomes.

Study	Outcome	Effect size	Variance	Difference	Variance
Study 1	Math Reading	0.30 0.10	0.05 0.05	0.20	?

variance V, then the variance of the difference is $2V$, which incorporates the two sources of error. By contrast, if we compute the difference of two (positively) related outcomes, then some of the error is redundant, and so the total error is less than $2V$. If the correlation between outcomes is 0.50, the variance of the difference would be equal to V, and as the correlation approaches 1.00, the variance of the difference would approach zero. The operating principle is that the higher the correlation between the outcomes, the lower the variance (the higher the precision) of the difference. The formula for the variance of a difference from correlated outcomes (see Box 24.1) is

$$V_{Y_{diff}} = V_{Y_1} + V_{Y_2} - 2r\sqrt{V_{Y_1}}\sqrt{V_{Y_2}}. \tag{24.9}$$

In words, we sum the two variances and then *subtract* a value that reflects the correlated error.

Note the difference from the formula for variance of a mean, where we *added* the correlated error and included a scaling factor. When we *combine* positively correlated outcomes, a higher correlation between outcomes results in *a higher variance*. By contrast, when we compute the *difference* between positively correlated outcomes, a higher correlation between outcomes results in *a lower variance*.

To understand why, suppose that we assess patients using two different measures of depression. If a patient is having a particularly good day when the measures are taken, both scores will tend to be higher than the patient's average. If the patient is having a bad day, both will tend to be lower than the patient's average. If we compute a *combined* effect, the error will *build* as we increase the number of outcomes. If this is a *good* day, both measures will over-estimate the patient's level of functioning. By contrast, if we compute a *difference*, we *subtract* one effect from the other, and the day-to-day variation is removed.

Computing a difference between outcomes

With this as background, we can return to the running example, and discuss the computation of the synthetic effect size and its variance.

The difference between reading and math is computed as

$$Y_{diff} = Y_1 - Y_2, \tag{24.10}$$

with variance

$$V_{Y_{diff}} = V_{Y_1} + V_{Y_2} - 2r\sqrt{V_{Y_1}}\sqrt{V_{Y_2}}. \tag{24.11}$$

If both variances are equal to the same variance, V, then (24.11) simplifies to

$$V = 2V(1 - r). \tag{24.12}$$

In the running example, the difference between reading and math in study 1 is computed as

$$V_{diff} = 0.30 - 0.10 = 0.2000,$$

with variance

$$V_{Y_{diff}} = 0.05 + 0.05 - 2 \times 0.50 \times \sqrt{0.05}\sqrt{0.05} = 0.05,$$

or equivalently,

$$V_{Y_{diff}} = 2(0.05)(1 - 0.50) = 0.05.$$

These formulas are used to create Table 24.6, where the variance for each composite is based on formula (24.11) and the weight is simply the reciprocal of the variance.

At this point we can proceed to the meta-analysis using these five (synthetic) scores. The scores happen to represent difference scores, but the same formulas apply. Under the fixed-effect model the formulas starting with (11.3) yield a summary effect

$$M = \frac{27.750}{232.500} = 0.1194,$$

with variance

$$V_M = \frac{1}{232.500} = 0.0043.$$

The average difference between the effect size for reading and the effect size for math is 0.1194 with variance 0.0043 and standard error 0.066. The 95% confidence interval for the average difference is -0.009 to 0.248. The Z-value for a test of the null is 1.820 with a two-sided p-value of 0.069.

Working with more than two outcomes per study

The formulas presented for a difference based on two outcomes can be extended to accommodate any number of outcomes using contrasts. For example, we could look

Table 24.6 Multiple outcomes – difference between outcomes.

		ES	Variance	ES	Correlation	Variance	Weight	ES*WT
Study 1	Reading	0.300	0.050	0.200	0.500	0.050	20.000	4.000
	Math	0.100	0.050					
Study 2	Reading	0.200	0.020	0.100	0.600	0.016	62.500	6.250
	Math	0.100	0.020					
Study 3	Reading	0.400	0.050	0.200	0.600	0.040	25.000	5.000
	Math	0.200	0.050					
Study 4	Reading	0.200	0.010	0.100	0.400	0.012	83.333	8.333
	Math	0.100	0.010					
Study 5	Reading	0.400	0.060	0.100	0.800	0.024	41.667	4.167
	Math	0.300	0.060					
Sum							232.500	27.750

at the difference between (a) math scores and (b) the mean of reading and verbal scores. However, this is beyond the scope of this volume.

Impact of the correlations on the combined effect

One issue to consider is what happens if the correlation between two outcomes is higher in some studies than in others. This variation will affect the *relative weights* assigned to different studies, with more weight going to the study with a higher correlation. In the running example the variances for reading and math were the same in studies 1 and 3, but the correlation between reading and math was higher in study 3. Therefore, study 3 had a lower variance and was assigned *more* weight in the meta-analysis. This is the opposite of what happens for a composite.

A second issue to consider is what happens as the set of correlations as a whole moves toward 1.0. Continuing with the simplified situation where both observations within a study have the same variance (V), if the two observations are independent of each other, the variance of the composite is $2V$. If the observations are not independent of each other, then the variance of the composite is $2V$ times a correction factor. We will refer to this correction factor as the *variance inflation factor* (*VIF*),

$$VIF = 1 - r, \tag{24.13}$$

where r is the correlation between the two components. An increase in r will result in a deflation of the variance compared with treating the different outcomes as independent of each other.

In Table 24.7 we explore how the variance inflation factor depends on the value of r, the correlation coefficient. For the purposes of this illustration we assume the simplistic situation of a study having the same variance ($V_Y = 0.2$) for each outcome. Each column in the table (A-E) corresponds to a different correlation coefficient between these outcomes.

The variance of the difference between the outcomes is

$$V_{Y_{diff}} = 2 \times V_Y \times VIF. \tag{24.14}$$

Taking as an example column C, the correlation is $r = 0.50$ and the variance inflation factor is

$$VIF = 1 - 0.50 = 0.5000.$$

Thus the variance of the composite score for one study is

$$V_{Y_{diff}} = 2 \times 0.20 \times 0.50 = 0.20.$$

As we move from left to right in the table (from a correlation of 0.00 to 0.75) the variance inflation factor (*VIF*) moves from 1.00 to 0.25. If the inflation factor for the variance moves from 1.00 to 0.25, it follows that the inflation factor for the standard error (which is the square root of the variance) will move from 1.00 to 0.50.

Table 24.7 Multiple outcomes – Impact of correlation on the variance of difference.

	A	B	C	D	E
Difference (Y)	0.4	0.4	0.4	0.4	0.4
Variance of each outcome (V)	0.2	0.2	0.2	0.2	0.2
Correlation between outcomes	0.000	0.250	0.500	0.750	1.000
Variance inflation factor	1.000	0.750	0.500	0.250	0.000
Variance of difference ($V_{Y_{diff}}$)	0.400	0.300	0.200	0.100	0.000
Standard error of difference	0.283	0.245	0.200	0.141	0.003
Standard error inflation factor	1.000	0.866	0.707	0.500	0.010
Lower limit	−0.155	−0.0802	0.008	0.124	0.394
Upper limit	0.955	0.8802	0.792	0.676	0.406
p-value (2-tailed)	0.158	0.103	0.046	0.005	0.000

Therefore, the confidence interval will narrow by 50 %, and the Z-value for the test of the null will double.

Note. In this example we focused on correlations in the range of 0.0 to 0.75, columns A-D in the table. As the correlation approaches 1.0 (column E) the variance will approach zero. This means that the width of the confidence interval will approach zero, the Z-value will approach infinity, and the p-value will approach zero. These apparent anomalies reflect what would happen if all error were removed from the equation. In Table 24.7, the correlation displayed as 1.000 is actually entered as 0.9999.

When the correlation is unknown
Table 24.7 also provides a mechanism for working with synthetic variables when we don't know the correlation among outcomes. Earlier, we assumed that the correlation between math and reading was known to be 0.50, and used that value to compute the standard error of the difference and related statistics. In those cases where we don't know the correlation for the study in question, we should still be able to use other studies in the same field to identify a plausible range for the correlation. We could then perform a sensitivity analysis and say, for example, that if the correlation falls in the range of 0.50 to 0.75 then the two-tailed p-value probably falls in the range of 0.046 to 0.005 (columns C to D in the table).

Researchers who do not know the correlation between outcomes sometimes treat the outcomes as coming from independent subgroups. It is instructive, therefore, to consider the practical impact of this choice. Treating the two outcomes as independent of each other yields the same precision as setting the correlation at 0.00 (column A). In effect, then, researchers who take this approach as a way of bypassing the need to specify a correlation, are actually adopting a correlation, albeit implicitly. And, the correlation that they adopt is zero. As such, it is almost certain to overestimate the variance and underestimate the precision of the

difference. In this context, (as in the case of a combined effect) the idea of working with a *plausible range of correlations* offers some clear advantages.

SUMMARY POINTS

- When we have effect sizes for more than one outcome (or time-point) within a study, based on the same participants, the information for the different effects is not independent and we need to take account of this in the analysis.
- To compute a summary effect using multiple outcomes we create a synthetic effect size for each study, defined as the mean effect size in that study, with a variance that takes account of the correlation among the different outcomes. We then use this effect size and variance to compute a summary effect across studies. Higher correlations yield *less* precise estimates of the summary effect.
- To compute the difference in effects we create a synthetic effect size for each study, defined as the *difference* between effect sizes in that study, with a variance that takes account of the correlation among the different outcomes. We then use this effect size and variance to assess the difference in effect sizes. Higher correlations yield *more* precise estimates of the difference in effects.

Further Reading

Cooper, H. (1982) Scientific Guidelines for conducting integrative research reviews. *Review of Educational Research*, 52(2), 291–302.

Glaser, L. & Olkin, I. (1994) Stochastically dependent effect sizes. In Cooper, H. & Hedges, L. V., *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Multiple Comparisons within a Study

Introduction
Combining across multiple comparisons within a study
Differences between treatments

INTRODUCTION

The final case of a complex data structure is the case where studies use a single control group and several treatment groups. For example, suppose we are working with five studies that assessed the impact of tutoring on student performance. Each study included three groups – a control group (a free study period), intervention *A* (tutoring focused on that day's school lesson) and intervention *B* (tutoring based on a separate agenda).

If our goal was to compute a summary effect for *A* versus control and separately for *B* versus control, we would simply perform two separate meta-analyses, one using the *A* versus control comparison from each study, and one using the *B* versus control comparison from each study.

The issues we address in this chapter are how to proceed when we want to incorporate both treatment groups in the same analysis. Specifically,

- We want to compute a summary effect for the active intervention (combining *A* and *B*) versus control
- Or, we want to investigate the difference in effect size for intervention *A* versus intervention *B*

COMBINING ACROSS MULTIPLE COMPARISONS WITHIN A STUDY

The issue we need to address is that the effect for *A* versus control and the effect for *B* versus control are not independent of each other. If each group (*A*, *B* and control) has 200 participants and we treated the two effects as independent, our effective sample size would appear to be 800 (since we count the control group twice) when in fact the true sample size is 600.

The problem, and the solution, are very similar to the ones we discussed for multiple outcomes (or time-points) within a study. If our goal is to compare *any treatment* versus control, we can create a composite variable which is simply the mean of A versus control and B versus control. The variance of this composite would be computed based on the variance of each effect size as well as the correlation between the two effects. At that point, all the formulas for combining data from multiple outcomes would apply here as well.

The difference between multiple outcomes and multiple comparison groups is the following. In the case of multiple outcomes, the correlation between outcomes could fall anywhere in the range of zero (or even a negative correlation) to 1.0. We suggested that the researcher work with a range of plausible correlations, but even this approach would typically yield a nontrivial range of possible correlations (say, 0.25 to 0.75) and variances. By contrast, in the case of multiple comparison groups, the correlation can be estimated accurately based on the number of cases in each group. For example, if group A, group B and the control group each have 200 participants, the correlation between A versus control and B versus control is 0.50. (This follows from the fact that the correlation between group A and group B is 0, while the correlation between control and control is 1, yielding a combined correlation midway between the two, or 0.50.) Therefore, we can work with a correlation of 0.50 without the need to conduct a sensitivity analysis based on a range of possible correlations.

This approach can be extended for the case where the sample size differs from one group to the next. For example, we can use a weighted mean of the effects rather than a simple mean, to give more weight to the treatments with more subjects. In this case, we would also need to adjust the variance to take account of the weighting. If the sample size differs from group to group the correlation will no longer be 0.50, but can be estimated precisely based on the data, without the need to resort to external correlations.

An alternate approach for working with multiple comparison groups is to collapse data from the treatment groups and use this data to compute an effect size and variance. If we are working with binary data from one control group and two treatment groups in a 2×3 table we would collapse the two treatment groups to create a 2×2 table, and then compute the effect size from that. Or, if we are working with means and standard deviations for three groups (A, B and control) we would collapse the data from A and B to yield a combined mean and standard deviation, and then compute the effect size for the control group versus this merged group (see option 2 for independent subgroups, (23.1), (23.2), and (23.3)). This approach will yield essentially the same results as the method proposed above.

DIFFERENCES BETWEEN TREATMENTS

We now turn to the problem of investigating *differences* between treatments when the different treatments use the same comparison group. To extend the current example, suppose that we have two treatment groups (A and B) and a control group. We want to know if one of the treatments is superior to the other.

While the approach used for computing a *combined* effect with multiple comparisons was similar to that for multiple outcomes, this is not the case when we turn to *differences* among the treatments. In the case of multiple outcomes, we had an effect size for reading (defined as the difference between treated and control) and an effect size for math (the difference between treated and control). Our approach was to work with the difference between the two effect sizes. In the case of multiple comparisons the analogous approach would be to compute the effect size for *A* versus control, and the effect size for *B* versus control, and then work with the difference between the two effect sizes.

While this approach would work, a better (potentially more powerful) approach is to ignore the control group entirely, and simply define the effect size as the difference between *A* and *B*. If we are working with binary data we would create a 2 x 2 table for *A* versus *B* and use this to compute an effect size. If we are working with means and standard deviations we would compute an effect size from the summary data in these two groups, ignoring the control group entirely.

This approach will only work if all studies have the same groups (here, *A* and *B*), which allows us to create the same effect size (*A* versus *B*) for each study. In practice, we are likely to encounter problems since some studies might compare *A* versus *B*, while others compare *A* versus control and still others compare *B* versus control. Or, some studies might include more than two comparison groups. Methods developed to address these kinds of issues are beyond the scope of this book, but are covered in the further readings.

SUMMARY POINTS

- When a study uses one control group and more than one treatment group, the data from the control group is used to compute more than one effect size. Therefore, the information for these effect sizes is not independent and we need to take this into account when computing the variance.
- To compute a combined effect, ignoring differences among the treatments, we can create a synthetic effect size for each study, defined as the mean effect size in that study (say, the mean of treatment *A* versus control and of treatment *B* versus control), with a variance that takes account of the correlation among the different treatments. We can then use this synthetic effect size and variance to compute a summary effect across studies. This is the same approach used with multiple outcomes.
- To look at differences among treatments the preferred option is to perform a direct comparison of treatment *A* versus treatment *B*, removing the control group from the analysis entirely. In some cases this will not be possible for practical reasons. In this case we can revert to the synthetic effect size, or can apply advanced methods.

Further Reading

Caldwell, D.M., Ades A.E. & Higgins, J.P.T. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 331: 897–900.

Glass, G.V, McGaw, B., & Smith, M.L., (1981). *Meta-analysis in Social Research*. Beverly Hills: Sage Publications.

Higgins J.P.T. & Green, S. (eds) (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons, Ltd.

Salanti G, Higgins J, Ades A.E. Ioannidis J.P.A. (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 17: 279–301.

Notes on Complex Data Structures

Introduction
Summary effect
Differences in effect

INTRODUCTION

In this Part we discussed three cases where studies provide more than one unit of data for the analysis. These are the case of multiple independent subgroups within a study, multiple outcomes or time-points based on the same subjects, and two or more treatment groups that use the same comparison group.

SUMMARY EFFECT

One issue we addressed was how to compute a summary effect using all of the data. For independent subgroups this meant looking at the impact of treatment versus no treatment, and ignoring any differences between stage-1 patients and stage-2 patients. For multiple outcomes this meant looking at the impact of the intervention on basic skills, and ignoring any differences between the impact on math versus reading. For multiple treatment groups it meant looking at the impact of treatment and control, and ignoring any differences among variants of the treatment.

In all cases, the key issue was the need to address any possible redundancy of information, since the precision of the combined effect is strongly dependent on the amount of information. To highlight the difference among the three cases, we used the same numbers in the worked examples as we moved from one chapter to the next. For example, for study 1, we assumed a variance of 0.05 between the two units of information, whether for two independent groups or for two outcomes.

The formula for the variance of a composite is the same in all cases, namely (in its simple form)

$$V_{\bar{Y}} = \frac{1}{2} V_Y (1 + r).$$ (26.1)

Consider, then, how this formula plays out for independent subgroups, for multiple outcomes, and for multiple comparisons.

- For multiple independent subgroups r is always zero, and so the variance of the composite is 0.025.
- For multiple outcomes r can take on any value. We assume that it falls somewhere in the range of 0.00 to 1.00, and assume further that the researcher can provide a range of plausible values for r. For example, we might have evidence that r is probably close to 0.50, with a plausible range of 0.25 to 0.75. In that case the variance would probably be close to 0.038, with a plausible range of 0.031 to 0.044.
- For multiple comparisons r can be determined based on the number of treatment groups and the number of subjects in each groups. If there is one control group and two treatment groups, and the subjects are divided evenly across groups, then r would be 0.50 and the variance would be 0.038. In other cases r could move either up or down, but can always be computed, and therefore we don't need to work with a range of values.

DIFFERENCES IN EFFECT

The second issue we addressed was how to look at the difference between effects.

For independent subgroups this meant looking at whether the treatment was more effective for one of the subgroups (stage-1 or stage-2 patients) than the other. For multiple outcomes this meant looking at whether the intervention had more of an impact on one outcome (reading or m ath) than the other. For multiple treatment groups it meant looking at whether one of the treatments was more effective than the other.

Again, the key issue was the need to address any possible redundancy of information, since the precision of the difference is strongly dependent on the amount of information. To highlight the difference among the three cases we used the same numbers in the worked examples as we moved from one chapter to the next. For example, we assumed a variance of 0.05 for each subgroup, or for each outcome, or for each comparison.

The formula for the variance of a difference is the same in all cases, namely (in its simple form)

$$V_{Y_{diff}} = 2V_Y (1 - r).$$ (26.2)

This formula incorporates the term $(1-r)$, which means that a higher correlation will yield a more precise estimate of the difference. This is the reverse of the case for

a composite effect, where the formula incorporates the term $(1 + r)$, and a higher correlation will yield a less precise estimate.

- For multiple independent subgroups r is always zero, and so the variance of the difference is 0.100.
- For multiple outcomes r can take on any value. We assume that it falls somewhere in the range of 0.00 to 1.00, and assume further that the researcher can provide a range of plausible values for r. For example, we might have evidence that r is probably close to 0.50, with a plausible range of 0.25 to 0.75. In that case the variance would probably be close to 0.050, with a plausible range of 0.075 to 0.025.
- For multiple comparisons we have essentially the same situation as for multiple outcomes, except that we can actually compute the correlation needed for the formula. However, there are other approaches available that allow for head-to-head comparisons of the treatment groups.