CHAPTER 49

Comprehensive Meta-Analysis Software

INTRODUCTION

The screenshots in this volume are from the software Comprehensive Meta-Analysis (CMA). In this chapter, we provide an overview of this software and show how to use it to implement the ideas outlined in prior chapters. The same approach could be used with any other program as well. Our goal in this chapter is to provide a sense for the look-and-feel of the program. For the reader who would like to carry out the analyses, a step-by-step PDF is available on the book's website (www.Introduction-to-Meta-Analysis.com).

Full disclosure. The authors of this volume are also the developers of CMA, and some have a financial interest in the program. This chapter is a slightly modified

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

version of a chapter published in the text *Common Mistakes in Meta-Analysis – and How to Avoid Them* (Borenstein, 2019).

Comprehensive Meta-Analysis (CMA) is a computer program for meta-analysis that was developed with funding from the National Institutes of Health in the United States. The program was initially released in 2000 and has been updated on a regular basis since then.

CMA features a spreadsheet view and a menu-driven interface. As such, it allows a researcher to enter data and perform a simple analysis in a matter of minutes. At the same time, it offers a wide array of advanced features, including the ability to compare the effect size in subgroups of studies, to run meta-regression, to estimate the potential impact of publication bias, and to produce high-resolution plots. The program is designed to work with studies that compare an outcome in two groups or that estimate an outcome in one group. It is not intended for network meta-analyses nor for meta-analyses of diagnostic test accuracy.

FEATURES IN CMA

The first step in conducting a meta-analysis is to compute an effect size and variance for each study. Many programs will perform this computation automatically when the data are in the form of 2×2 tables or in the form of means and standard deviations for each group, but not for more complex data formats. By contrast, CMA will allow the user to enter data in more than one hundred formats. For example, the user can enter data as events and N in each group; or as an odds ratio and its confidence interval; or as a log risk ratio and its standard error. Or, the user can enter means and standard deviations for two independent groups; or the pre and post scores for a pre/post study; or the *p*-value from a *t*-test for two independent groups; and so on. Critically, the user may use a different format for each study. Thus, if one study reports means and standard deviations for two independent groups, a second reports a *p*-value based on two independent groups, and a third reports pre-scores and post-scores for a pre/post study, the user may enter data for each study in its own format. The program will apply the appropriate formula for each format to compute the effect size and its variance, and then include all the effects in the analysis.

The program allows the user to select from an array of effect size indices, including the odds ratio, risk ratio, risk difference, mean difference, standardized mean difference, correlation, hazard ratio, and prevalence, among others.

In the motivating example, each study provides one row of data. CMA also allows for the possibility that some (or all) studies will provide more than one row of data. There is an option for studies to report data for two or more outcomes, based on the same subjects. In the analysis, we could elect to look at either outcome alone. Or, we could tell the program to create a synthetic outcome which incorporates both measures, taking into account the fact that the two outcomes are not independent of each other.

Similarly, we can enter data for an outcome recorded at two or more time-points, which allows us to assess the impact at each time-point, and to see whether the effect

size changes over time. We can enter data for two or more independent subgroups within studies and then run the analysis using either subgroup or study as the unit of analysis. Finally, we can enter data for studies that employed one control group and multiple treatment groups.

TEACHING ELEMENTS

The program incorporates a number of features intended to make the computations as transparent as possible. On the data-entry screen, the user enters summary data and the program displays the effect size and its variance. Double-click on the computed values, and the program will show how those values were computed. On the analysis screen, there is also a tab labeled 'Calculations', which opens a window onto the calculations.

DOCUMENTATION

A manual is installed with the program. Each module in the program features an interactive guide that will walk the user through that module. Additionally, the website offers an array of PDFs and videos that show how to enter data, run the analysis, and then interpret the output. In each case, we also discuss how to report the data. The program's algorithms are discussed in this volume.

AVAILABILITY

The program's website is www.Meta-Analysis.com. The program may be downloaded and run for free as a trial, and the website lists rates for licenses. There are discounts available for nonprofit institutions and for students. The program is free for short-term workshops in meta-analysis and is available at a discount for semester-length classes in meta-analysis.

ACKNOWLEDGMENTS

Development of the program was funded by the National Institutes of Health in the United States under the following grants: MH052969 (Computer program for meta-analysis in mental health), AG021360 (Combining data types in meta-analysis), AG020052 (Publication bias in meta-analysis for mental health), AG024771 (Software for meta-regression), DA019280 (Forest plots for meta-analysis), AG029029 (Software for meta-analysis of diagnostic tests), and DA029351 (Software for meta-analysis with correlated outcomes). As a matter of policy, NIH does not endorse any product or software.

The program was developed by Michael Borenstein, Larry Hedges, Julian Higgins, and Hannah Rothstein. We gratefully acknowledge the contributions of Doug Altman, Betsy Becker, Jesse Berlin, Michael Brannick, Harris Cooper, Kay Dickersin,

Sue Duval, Roger Harbord, John Ioannidis, Jeff Valentine, Spyros Konstantopoulos, Mark Lipsey, Mike McDaniel, Fred Oswald, Terri Pigott, David Rindskopf, Stephen Senn, Will Shadish, Jonathan Sterne, Alex Sutton, Steven Tarlow, Thomas Trikalinos, Jack Vevea, Vish Viswesvaran, and David Wilson.

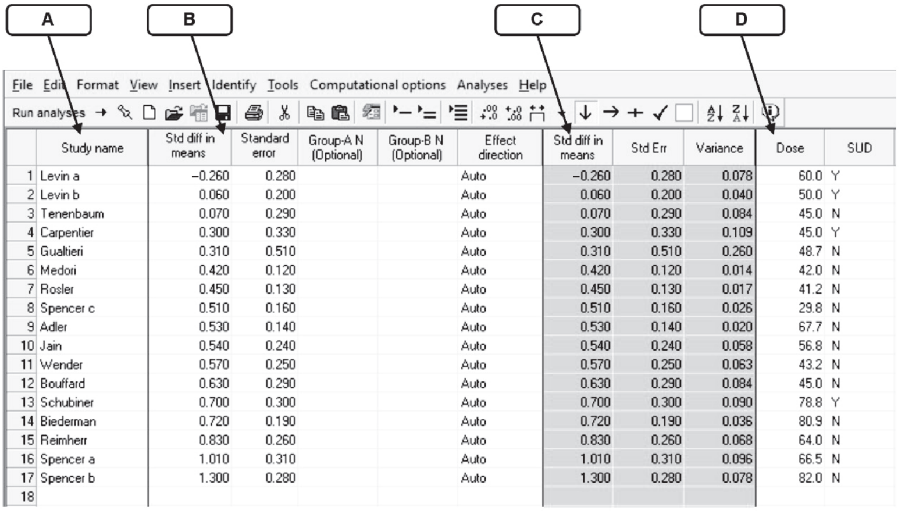
MOTIVATING EXAMPLE

To illustrate the program, we will use a meta-analysis of 17 studies that assessed the utility of methylphenidate for treating adults with attention deficit hyperactivity disorder (ADHD). In each study, patients who had been diagnosed with ADHD were randomly assigned to either methylphenidate or a placebo and then tested on a scale intended to assess cognitive function (Castells *et al*., 2011).

The effect size index is the standardized mean difference (*d*). In this context, a standardized mean difference of 0.20 would be considered trivial – this is a difference that shows up on the tests, but the patient might not be aware of any change. A standardized mean difference of 0.50 would be considered moderate – the patient would recognize that they were doing better than usual, and coworkers might be aware of a change. A standardized mean difference of 0.80 would be considered large – the patient would feel great, and the difference would be obvious enough that others might remark on it.

DATA ENTRY

Figure 49.1 shows the data-entry screen. For each study, enter the study name into column [A] and the summary data into the columns labeled [B]. The program displays



	Study name	Std diff in means	Standard error	Group-A N (Optional)	Group-B N (Optional)	Effect direction	Std diff in means	Std Err	Variance	Dose	SUD
1	Levin a	−0.260	0.280			Auto	−0.260	0.280	0.078	60.0	Y
2	Levin b	0.060	0.200			Auto	0.060	0.200	0.040	50.0	Y
3	Tenenbaum	0.070	0.290			Auto	0.070	0.290	0.084	45.0	N
4	Carpentier	0.300	0.330			Auto	0.300	0.330	0.109	45.0	Y
5	Gualtieri	0.310	0.510			Auto	0.310	0.510	0.260	48.7	N
6	Medori	0.420	0.120			Auto	0.420	0.120	0.014	42.0	N
7	Rosler	0.450	0.130			Auto	0.450	0.130	0.017	41.2	N
8	Spencer c	0.510	0.160			Auto	0.510	0.160	0.026	29.8	N
9	Adler	0.530	0.140			Auto	0.530	0.140	0.020	67.7	N
10	Jain	0.540	0.240			Auto	0.540	0.240	0.058	56.8	N
11	Wender	0.570	0.250			Auto	0.570	0.250	0.063	43.2	N
12	Bouffard	0.630	0.290			Auto	0.630	0.290	0.084	45.0	N
13	Schubiner	0.700	0.300			Auto	0.700	0.300	0.090	78.8	Y
14	Biederman	0.720	0.190			Auto	0.720	0.190	0.036	80.9	N
15	Reimherr	0.830	0.260			Auto	0.830	0.260	0.068	64.0	N
16	Spencer a	1.010	0.310			Auto	1.010	0.310	0.096	66.5	N
17	Spencer b	1.300	0.280			Auto	1.300	0.280	0.078	82.0	N
18											

Figure 49.1 Data-entry screen in CMA.

the standardized mean difference, standard error, and variance in the columns labeled [C]. We have also entered data for a series of moderator variables in the columns labeled [D], including the dose of methylphenidate (Dose), and whether the study enrolled patients who were abusing drugs (SUD).

In this example, the summary data entered for each study were the standardized mean difference and its standard error, because these are the data that had been reported. However, the user may elect to enter data in more than 100 formats. Similarly, we have elected to display the standardized mean difference, but the program will compute and display a wide array of effect size indices. The data may be entered directly into CMA or copied from another program such as Excel™.

To run the analysis, click [Run Analyses] on the toolbar.

BASIC ANALYSIS

Figure 49.2 shows the analysis screen. A tab at the bottom [E] may be used to switch between fixed-effect and random-effects meta-analyses. The fixed-effect model is appropriate when the intended inference is limited to the studies in the analysis. The random-effects model is appropriate when the intended inference is to the universe of comparable studies. In this example, we intend to generalize the results to the universe of comparable studies and have selected the random-effects model (see Part 3).

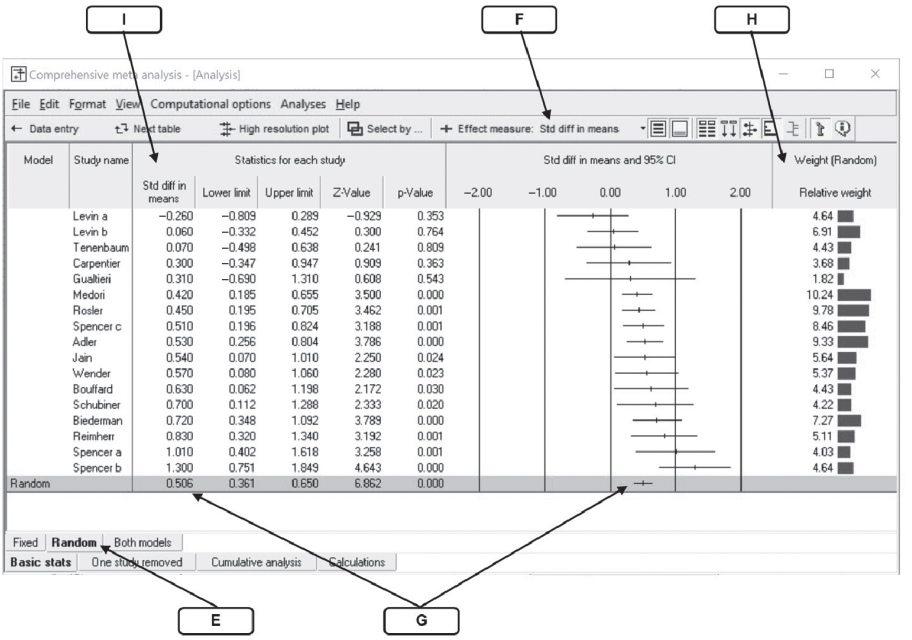


Figure 49.2 Basic analysis screen in CMA.

The toolbar [F] allows us to select the effect size index. Here, we have selected 'Standardized difference in means' (same as standardized mean difference).

WHAT IS THE *AVERAGE* EFFECT SIZE?

The program [G] displays the average effect size as 0.506 and the confidence interval as 0.361 to 0.650. Since this is a random-effects meta-analysis, this tells us that the average effect size in the universe of comparable studies is estimated as 0.506 and probably falls in the range of 0.361 to 0.650. The *Z*-value of 6.862 and the corresponding *p*-value of <0.001 test the null hypothesis that the average effect size in the universe of comparable studies is precisely zero. We can reject the null hypothesis and conclude that the average effect size is greater than zero – that the treatment is helpful. At the right [H], the program displays the relative weight assigned to each study when computing the combined effect size.

HOW MUCH DOES THE EFFECT SIZE VARY?

The *average* effect size represents a substantial clinical improvement. But to understand the potential utility of this intervention, we need to also know how much the effect size varies across populations. Is the intervention consistently effective or is the impact trivial in some populations and exceptional in others? Is the intervention always beneficial, or is it sometimes harmful?

To address these questions, we can click a tool on the menu bar [I] in Figure 49.2 and display the tables shown in Figure 49.3. The statistics at the top of Figure 49.3 [J] are the same as those in Figure 49.2 and address the *average* effect size. The statistics

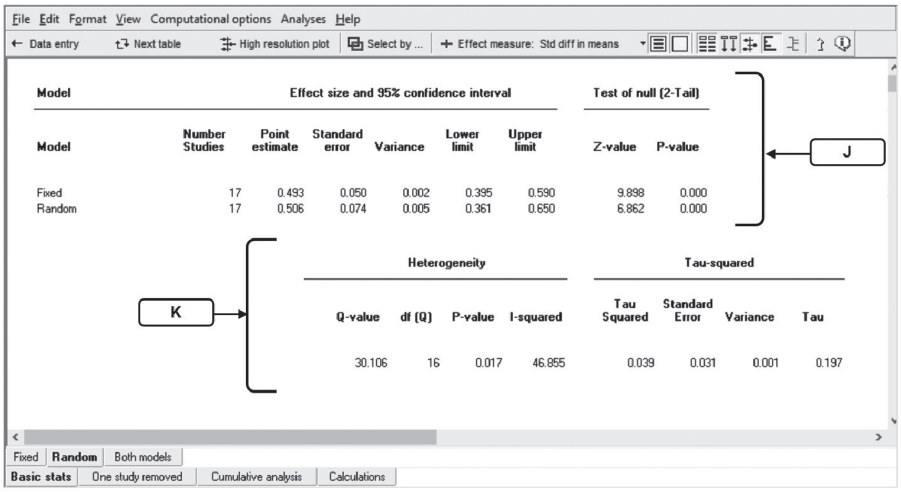


Figure 49.3 Average effect size (top), variation in effect size (bottom).

at the bottom of Figure 49.3 [K] address the *variation* in effect size across studies, as follows.

We can test the null hypothesis that all studies share a common effect size and that the variance in observed effects is due entirely to sampling error. The test statistic Q is 30.106 with 16 degrees of freedom and a corresponding p-value of 0.017. We conclude that the impact of methylphenidate is stronger in some populations than in others. However, the important question is not whether the effect size varies *at all*, but rather *how much* it varies. We turn to that now.

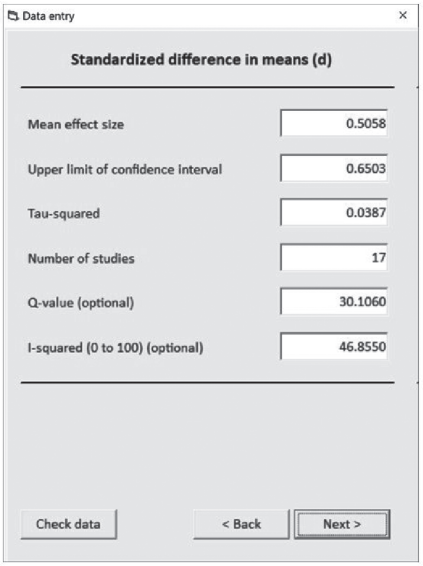
To get a general sense of the dispersion, we can start with the forest plot (Figure 49.2), where the observed effect sizes vary from -0.26 to +1.30. However, only some of this dispersion reflects variation in true effects (the variation that we care about), while the rest reflects variance due to sampling error. Returning to Figure 49.3, the I^2 statistic is the ratio of V_{TRUE} to $V_{OBSERVED}$, and as such, it provides some context for understanding the forest plot. When I^2 is low, the variance in the forest plot is mostly due to sampling error. When I^2 is high, the variance in the forest plot provides a reasonable estimate for the variance of true effects. Here, I^2 is around 47%, so the variation in true effects is somewhat less than the variation displayed in the forest plot. (There is a common belief that I^2 tells us how much the effect size varies, but this belief is incorrect. As explained in Chapters 19 and 20, I^2 is a proportion, not an absolute value.) The program displays T^2, the estimated variance of true effects (0.039) and T, the standard deviation of true effects (0.197).

PLOT SHOWING DISTRIBUTION OF EFFECTS

While most reports of meta-analyses tend to highlight the statistics outlined above, none of these statistics directly addresses the question 'What is the expected range of true effects for populations similar to those in the analysis?'

For that, we turn to the prediction interval. Version 4 of CMA (planned for release in 2021) will create the plot shown in Figure 49.4. The caption, *The true effect size in 95% of all populations falls in the interval 0.06 to 0.95,* is generated automatically. The plot displays the corresponding distribution of effects, which allows us to gauge the approximate proportion of effects in any given range. In this case, all effects fall to the right of zero (there are no populations where the treatment is harmful). Additionally, if we assume that 0.35 is the lower bound of a clinically important effect, we would conclude that the effect is clinically useful in more than 70% of all comparable populations.

A stand-alone program to create this plot may also be downloaded on the book's website. This program can be used with Review Manager™ and other programs, as well as CMA. We would enter the number of studies (17), the mean effect size (0.5058), the upper limit of the mean effect size (0.6503), and T^2 (0.0387). The program then generates the plot shown in Figure 49.4. Beginning with version 4 of CMA, this plot will be integrated into CMA. The stand-alone version will still be available for those using other software.



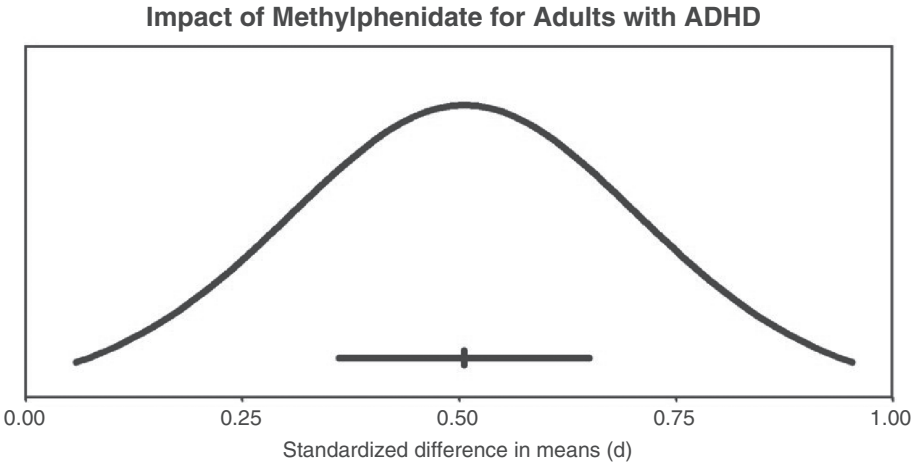


Figure 49.4 Plotting distribution of true effects. ADHD.

HIGH-RESOLUTION PLOT

Click the menu button labeled 'High-resolution plot' to create the plot displayed in Figure 49.5. Menus allow the user to extensively customize the plot and then export a copy directly to Microsoft™ Word™ or PowerPoint™.

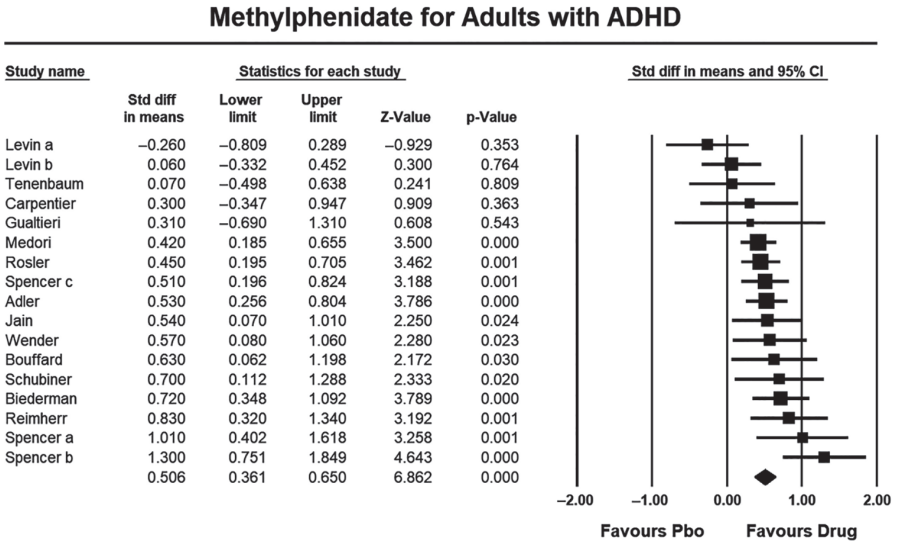


Figure 49.5 High-resolution plot in CMA.

SUBGROUP ANALYSIS

At this point, we have established that methylphenidate is more effective in some populations than in others, and we might want to identify factors associated with the magnitude of the effect. One possible factor is the nature of the population. Specifically, some studies excluded patients who were abusing drugs, while others enrolled these patients. We want to compute the effect size separately for each subgroup of studies and then to compare the two values (see Chapter 21).

We return to the analysis screen and use the 'Computational options' menu to 'Group by > SUD' [N]. The result is displayed in Figure 49.6.

First, we assess the mean effect of treatment for each subgroup of studies. For studies that excluded drug abusers [L], the combined effect size is 0.577 with a 95% confidence interval of 0.438 to 0.717 a *Z*-value of 8.090 and a *p*-value of < 0.001. For studies that included drug abusers [M], the combined effect size is 0.162 with a 95% confidence interval of −0.136 to +0.460, a *Z*-value of 1.064, and a *p*-value of 0.287.

Next, we want to compare the effect size in the two subgroups. That is, we want to ask whether the treatment's impact is different in studies that exclude drug abusers as compared with studies that include drug abusers. A button on the menu bar allows us to switch between the plot in Figure 49.6 and the details in Figure 49.7.

For the analyses comparing the impact of methylphenidate in studies that included drug abusers vs. studies that excluded drug abusers, we use a mixed-effects model, *at the bottom* of Figure 49.7. The subgroups are fixed, in the sense that we are comparing these two drugs specifically, and not generalizing to any other drugs. Within each

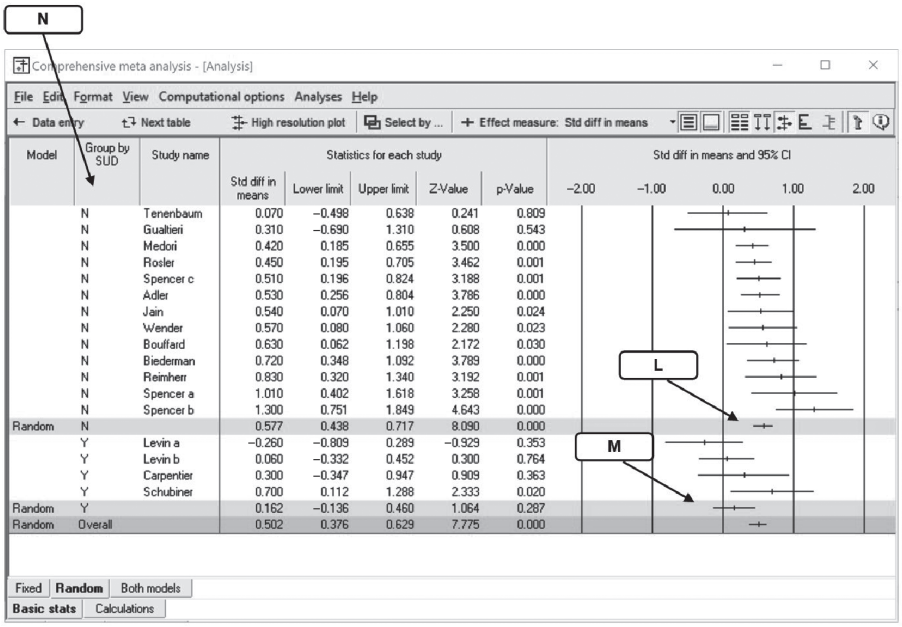


Figure 49.6 Impact of treatment as a function of subgroup: Forest plot.

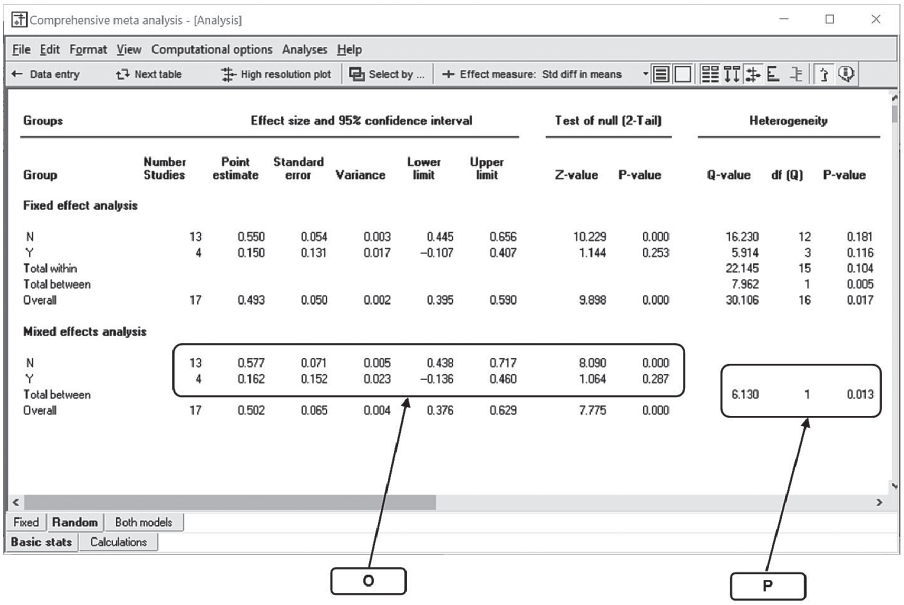


Figure 49.7 Impact of treatment as a function of subgroup: Statistics.

subgroup, the studies are a random sample of all relevant studies. Since the model is fixed at one level and random at the other level, it is called mixed-effects (Chapter 21).

The combined effect size for studies that exclude SUD patients is 0.577 with a 95% confidence interval of 0.438 to 0.717, while the combined effect size for studies that include SUD patients is 0.162 with a 95% confidence interval of −0.136 to + 0.460 [O]. To test the difference between the two effect sizes, we may use a *Q*-test. The *Q*-value for this difference is 6.130 with 1 degree of freedom and a *p*-value of 0.013 [P]. We conclude that the treatment is more effective in the subgroup of populations that exclude SUD patients and less effective in the subgroup of populations that includes these patients.

It is important to recognize that (with rare exceptions) subgroup comparisons in a meta-analysis are observational by nature and cannot prove a causal relationship (Chapter 23). In this example, it is *possible* that methylphenidate is more effective in populations that exclude SUD patients because it actually works better in these patients, which *would be* a causal relationship. But it is also possible that methylphenidate was more effective in the studies which excluded SUD patients for other reasons. For example, it is possible that these studies tended to employ a higher dose of methylphenidate, and it is the higher dose (rather than the fact that these patients abused drugs) that was responsible for the larger effect in these studies. We can use meta-regression to assess the relationship between SUD and effect, with Dose held constant. We turn to that now.

META-REGRESSION

In a primary study, we may use regression analysis to study the relationship between covariates and outcome. Similarly, in a meta-analysis we may use regression to study the relationship between covariates and effect size. In this case, the procedure is commonly called meta-regression. In a primary study, the unit of analysis is the *individual*, with covariates and outcome measured for each individual. In a meta-analysis, the unit of analysis is the study, with covariates and outcome measured for each study. However, with some modifications, the full arsenal of procedures that fall under the heading of 'regression' in primary studies is also available in meta-analysis (Chapter 22).

In the current example, we want to see whether the impact of methylphenidate is related to whether the study excluded patients who were abusing drugs (SUD) and/or the mean dose employed in the study (Dose). On the main analysis screen, we select 'Meta-regression 2' on the 'Analyses' menu. We define a regression with these two covariates, and the program displays the results in Figure 49.8. The results based on the random-effects model are shown here. The user may also choose to use a fixed-effect model though this is generally not recommended. The analyses displayed here are based on the Knapp–Hartung adjustment (Chapter 26).

The table at the top [Q] provides details for the relationship between SUD and effect size, with Dose held constant. The coefficient for SUD tells us that (with dose held constant) the mean effect size for studies that enrolled SUD patients is 0.4492 *d* lower than

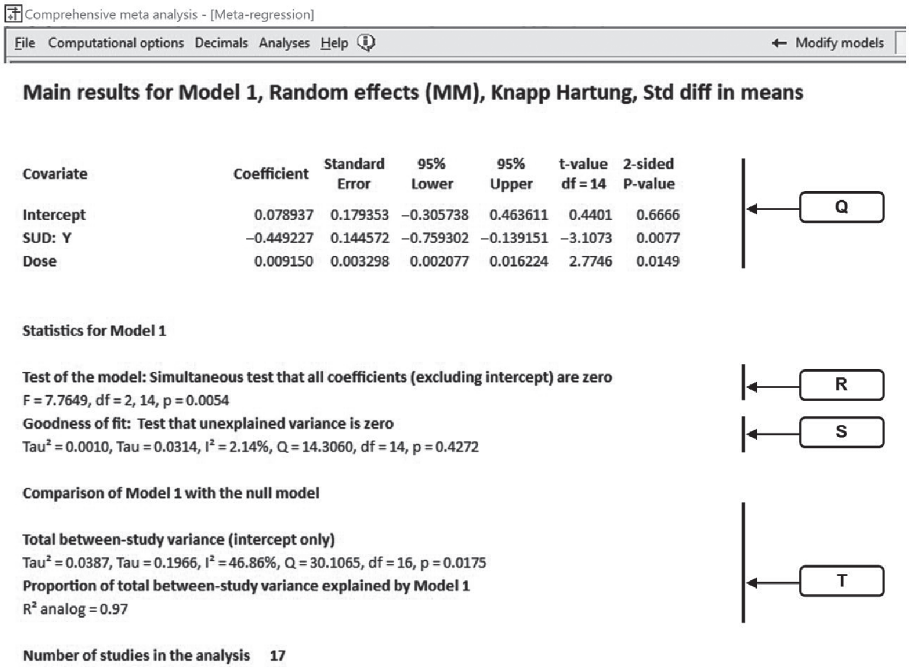


Figure 49.8 Results for regression, random effects.

for studies which excluded those patients. The confidence interval for the coefficient is −0.7593 to −0.1392, and a test of the null hypothesis that SUD is not related to effect size yields a *t*-value of −3.1073 with 14 *df*, and a corresponding *p*-value of 0.0077. This tells us that the relationship between SUD and effect size is not due to a confound with dose.

The coefficient for Dose is displayed as 0.0092. This tells us that for every one-unit increase in dose, the effect size will increase by approximately 0.01. The 95% confidence interval for the coefficient is 0.0021 to 0.0162, and a test of the null hypothesis that dose is not related to effect size yields a *t*-value of 2.7746 with 14 *df* and a *p*-value of 0.0149. This is plotted in Figure 49.9, where we see that the treatment is more effective in studies that employed a higher dose of the drug. Concretely, as the dose increases from 30 units to 82 units [points U to V on the regression line], the impact of treatment increases by 48 points.

As noted, in Figure 49.8, the table at the top [Q] displays statistics for the *unique* impact of *each* covariate. By contrast, the other sections display statistics for the *joint* impact of *all* covariates.

In the section labeled 'Test of Model' [R], we test the null hypothesis that *none* of the covariates explains any variation in effect size. The *F*-value for this test is 7.7649 with 2, 14 degrees of freedom and a corresponding *p*-value of 0.0054. We reject the null hypothesis and conclude that at least one of the covariates is related to effect size.

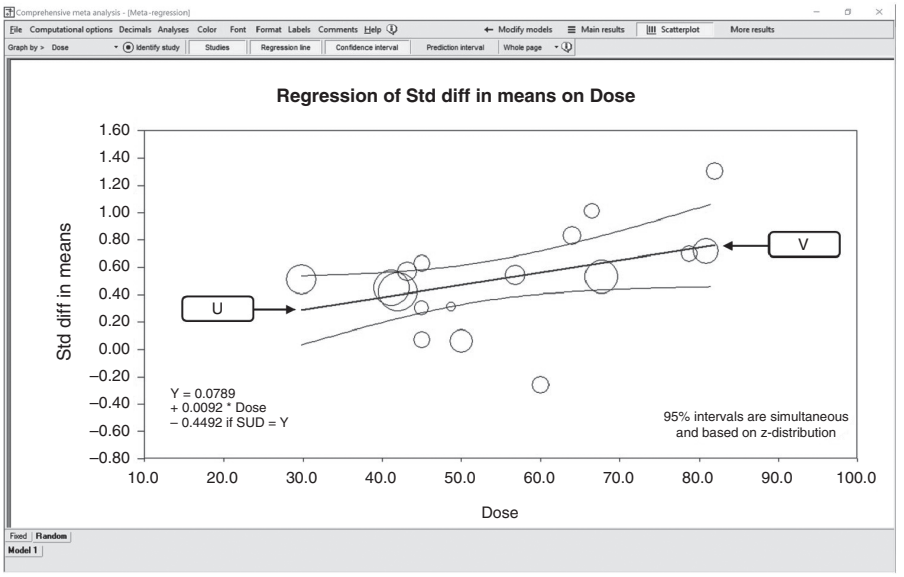


Figure 49.9 Regression of effect size on Dose, with SUD held constant.

The section labeled 'Goodness of fit' [S] addresses the residual variance. The estimated variance of true effects about the regression line (T^2) is 0.0010, and the standard deviation of true effects about the regression line (T) is 0.0314. The I^2 statistic is 2.14%, which tells us that some 2% of the observed variance about the regression lines based on these covariates reflects variation in true effects rather than sampling error. The test for heterogeneity yields a Q-value of 14.31 with 14 degrees of freedom and a corresponding p-value of 0.4272. We conclude that the variance of observed effects about the regression line could be due entirely to sampling error. Finally, the program displays R^2 as 0.97, which tells us that the covariates can explain some 97% of the initial variance in true effects.

The program offers a number of options for regression. It allows the user to include categorical covariates in the model (as it did here for SUD). In this case, the program will automatically create a set of dummy variables to represent the covariate. It allows the user to select either the Z-distribution or the Knapp–Hartung adjustment for computing confidence intervals and p-values (Knapp & Hartung, 2003) as discussed in chapter 26. It allows the user to estimate τ^2 using either the method of moments, maximum likelihood, or restricted maximum likelihood. It allows the user to define sets of covariates (for example, the linear and curvilinear impact of dose) and to assess the impact of the full set with other covariates held constant. It allows the user to define multiple prediction models and then compare them with each other.

As was true for analyses that compared subgroups, the relationships explored in meta-regression (with rare exceptions) are observational rather than causal. In this example, we attempted to identify the relationship between SUD and effect size while

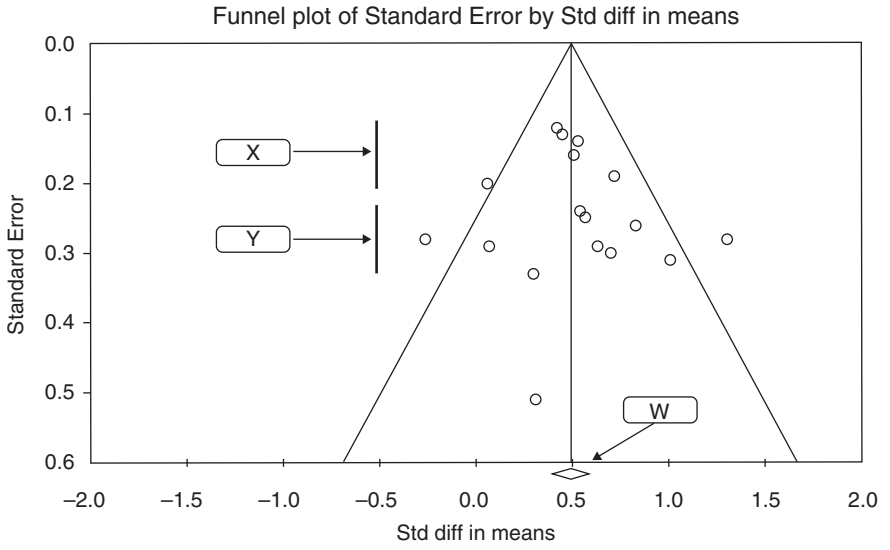


Figure 49.10 Funnel plot of observed effects.

controlling for dose and vice versa, but there may be other confounding variables that we have not considered.

PUBLICATION BIAS

To address the potential impact of publication bias, we can select 'Analyses > Publication bias' on the main analysis screen, to display Figure 49.10. This figure shows the effect size (on the *x*-axis) by the standard error (on the *y*-axis). The large studies appear at the top, and the smaller studies appear toward the bottom.

In this case, the sample size in most studies falls within a narrow range, and so it is not likely that the procedures normally employed to assess potential for publication bias would be effective. For purposes of this volume, we can nevertheless apply the Trim and Fill procedure (Chapter 35).

A vertical line denotes the average effect size [W]. If the effects are normally distributed, we would expect half the studies to fall on either side of the line. This is the case toward the top [X] but as we move toward the smaller studies [Y] there is a predominance of studies on the right and relatively few on the left. One possible reason for this could be that the studies toward the left were not statistically significant and therefore were not published and did not find their way into the analysis. In that case, the combined effect size is based on a biased subset of all actual studies and overestimates the true average effect size.

The Trim and Fill method (Duval & Tweedie, 2000b) employs an iterative procedure to identify the studies that may be missing. It then 'creates' these studies and inserts them into the analysis. These are displayed here as filled circles [Z] which are the

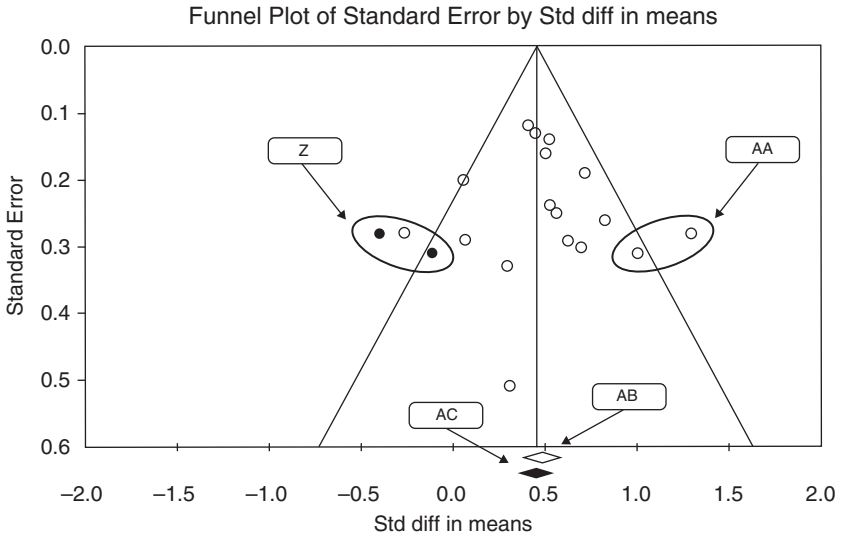


Figure 49.11 Funnel plot of observed and imputed effects.

mirror image of the actual studies [AA]. We can use all the studies (actual and imputed) to compute an adjusted estimate of the mean effect size. The initial estimate of the combined effect size was 0.506 [AB] but the adjusted value (included the imputed studies) is 0.442 [AC]. If the asymmetry was due to publication bias, then this adjustment yields an estimate of the unbiased effect size. The adjustment is minor, in the sense that the clinical meaning of 0.442 is essentially the same as the clinical meaning of 0.506. Note that there are reasons other than publication bias that may explain or contribute to asymmetry in funnel plots. As explained in Chapter 35, the Trim and Fill procedure should be seen as a sensitivity analysis. It is not intended to yield a "correct" effect size.

CMA also features other methods that are typically used with the aim of testing and/or adjusting for publication bias. These include the Egger test of the intercept, the Begg and Mazumdar rank correlation test, and Rosenthal's Fail-safe *N* (Rothstein *et al*., 2005; Sterne *et al*., 2011). The program can also generate a text report that explains how to interpret the results for each of the publication bias procedures.

EXPLAINING RESULTS

The following is how one might explain the results of this analysis.

Overview

This example is a re-analysis of a systematic review published by Castells *et al*. (2011). The analysis is based on seventeen studies that evaluated the effect of methylphenidate on cognitive function in adults with attention deficit hyperactivity disorder (ADHD).

In each study, patients were randomly assigned to either drug or a placebo and the researchers assessed the patients' cognitive function at the conclusion of treatment. The effect size index is the standardized mean difference (d). The results of this analysis will be generalized to comparable studies, and so the random-effects model was employed for the analysis.

In this context, a standardized mean difference of 0.20 would be considered trivial – this is a difference that shows up on the tests, but the patient might not be aware of any change. A standardized mean difference of 0.50 would be considered moderate – the patient would recognize that they were doing better than usual, and coworkers might be aware of a change. A standardized mean difference of 0.80 would be considered large – the patient would feel great, and the difference would be obvious enough that others might remark on it.

Does methylphenidate affect cognitive scores?

The standardized mean difference is 0.506. *On average*, methylphenidate increased cognitive functioning by 0.506 standard deviations as compared with placebo. The confidence interval for the standardized mean difference is 0.361 to 0.650, which tells us that the mean effect size in the universe of comparable studies could fall anywhere in this range. This range does not include an effect size of zero, which tells us that the mean effect size is probably not zero. Similarly, the Z-value for testing the null hypothesis (that d is 0.0) is 6.862, with a corresponding p-value of <0.001. Using the Knapp–Hartung adjustment, $t = 6.29$, $df = 16$, $p < 0.001$, and the 95% confidence interval is 0.335 to 0.676 (see Chapter 26). We can reject the null hypothesis and conclude that (on average) the drug does increase cognitive function in the universe of populations which are comparable to those in the analysis. Given the dispersion in effects (as discussed below), it is important to recognize that the mean effect size applies to this particular mix of studies and would be different for another mix of populations, dosages, and so on.

How much does the effect size vary across studies?

The Q-statistic provides a test of the null hypothesis that all studies in the analysis share a common effect size. If all studies shared the same effect size, the expected value of Q would be equal to the degrees of freedom (the number of studies minus 1). The Q-value is 30.106 with 16 degrees of freedom and p=0.017. We reject the null hypothesis that the true effect size is identical in all the studies. The I^2 statistic is 47%, which tells us that 47% of the variance in observed effects reflects variance in true effects rather than sampling error. The variance of true effects (T^2) is 0.039, and the standard deviation of true effects (T) is 0.197.

The prediction interval is 0.058 to 0.953. We would expect that in some 95% of all populations comparable to those in the analysis, the true effect size will fall in this range. Based on the context outlined above, there will be some populations where the impact of the treatment is trivial and some where it is large.

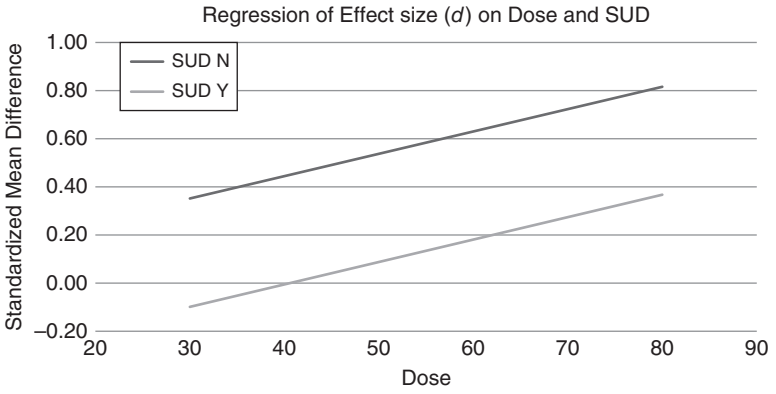


Figure 49.12 Regression of effect size (*d*) on Dose and SUD. Plot created in Excel (TM).

Is effect size related to Dose and/or SUD?

For every one mg. increase in dose there is an increase of roughly one point in the effect size, *d*. As dose increases by 50 mg. (from 0.30 to 0.80) the predicted effect size increases by 46 points. This is reflected by the slope of the prediction lines in Figure 49.12. For studies that enrolled SUD patients, the predicted effect size increases from −0.10 to +0.36. For studies that excluded SUD patients, the predicted effect size increases from 0.35 to 0.81.

The predicted effect size for studies that enrolled SUD patients is 45 points lower than for studies that excluded these patients. This is reflected in the difference between the two prediction lines in Figure. The bottom line (for studies that enrolled SUD patients) is 45 lower than the top line (for studies that included these patients).

The relationship between dose and effect size remains even after we partial SUD, and the relationship between SUD and effect size remains even after we partial dose. The coefficient for Dose as a predictor of effect size (with SUD held constant) is 0.0092 with a 95% confidence interval of 0.0021 to 0.1162. For a test of the null hypothesis that there is no relationship between dose and effect size, $t\ (14) = 2.7746$ and $p = 0.0149$. The coefficient for SUD (present) as a predictor of effect size (with dose held constant) is −0.4492 with a 95% confidence interval of −0.7593 to −0.1392. For a test of the null hypothesis that there is no relationship between dose and effect size, $t\ (14) = -3.1073$ and $p = 0.0077$.

Publication bias

While it is likely that some studies are missing from the analysis due to publication bias (since this is typically the case), the impact of missing studies in this analysis was probably minor. The Trim and Fill analysis suggests that there may be two missing studies, but if we impute these studies and include them in the analysis, the mean effect size shifts only slightly (from 0.506 to 0.442).