Example name	ADHD Analysis
Effect size	Standardized mean difference
Analysis type	Basic, subgroups, regression
Level	Intermediate

Synopsis

This analysis assesses the impact of methylphenidate on cognitive function in adults with Attention Deficit Hyperactivity Disorder (ADHD). The analysis includes seventeen studies where patients were randomized to the intervention or to a control condition. The outcome was some measure of cognitive function, and a positive difference reflects better function for the treated group.

If all studies had measured cognitive function using the same diagnostic test, we would have performed the analysis using scores on this scale. However, each study used a unique test to assess cognitive function. Therefore, the reviewers transformed the mean difference for each study into a standardized mean difference, d.

A standardized mean difference of 0.20 would be considered a small effect. Patients who improved by this amount would probably not be aware of any change in their functioning. A standardized mean difference of 0.50 would be considered a medium effect. Patients who improved by this amount would be aware that they were doing better than usual, and their colleagues might notice a change. A standardized mean difference of 0.58 would be considered a large effect. Patients who improved by this amount would be very cognizant of their improvement, and their colleagues might be likely to remark on the change.

I show

- How to enter data for the standardized means difference
- How to run a random-effects analysis
- How to quantify the dispersion in effects
- How to see the weight assigned to each study
- How to understand the statistics for the summary effect and the dispersion
- How to create a high-resolution plot
- How to compare the effect size in two subgroups of studies
- How to assess the relationship between a continuous variable and outcome

The example that follows is based on the systematic review cited below. This example is not intended to be representative of all analyses in the review. For details of the review, please see the original.

Note that the original analysis includes eighteen studies, but for purposes of this exercise I excluded a study that was missing data on the moderators.

ORIGINAL RESEARCH ARTICLE

CNS Drugs 2011; 25 (2): 157-169
1172-7047/11/0002-0157/$49.95/0

Efficacy of Methylphenidate for Adults with Attention-Deficit Hyperactivity Disorder

A Meta-Regression Analysis

Xavier Castells,[1,2] *Josep Antoni Ramos-Quiroga*,[1,3] *David Rigau*,[4] *Rosa Bosch*,[1] *Mariana Nogueira*,[1] *Xavier Vidal*[5] and *Miguel Casas*[1,3]

1 Psychiatry Department, Hospital Universitari Vall d'Hebron, Barcelona, Spain
2 Unit of Clinical Pharmacology, Department of Medical Sciences, School of Medicine, Universitat de Girona, Girona, Spain
3 Psychiatry Department, Universitat Autònoma de Barcelona, Barcelona, Spain
4 Iberoamerican Cochrane Center, Barcelona, Spain
5 Department of Pharmacology, Therapeutics and Toxicology, Universitat Autònoma de Barcelona, Barcelona, Spain

CNS Drugs. 2011 Feb;25(2):157-69. doi: 10.2165/11539440-000000000-00000

The data for this exercise are located at

www.Common-Mistakes-in-Meta-Analysis.com

Navigate to the folder for ADHD analysis and download the following

This PDF
The data for this analysis in an Excel format
The data for this analysis in CMA format
The Excel file Prediction intervals for this data

To download a trial of CMA go to
www.Meta-Analysis.com

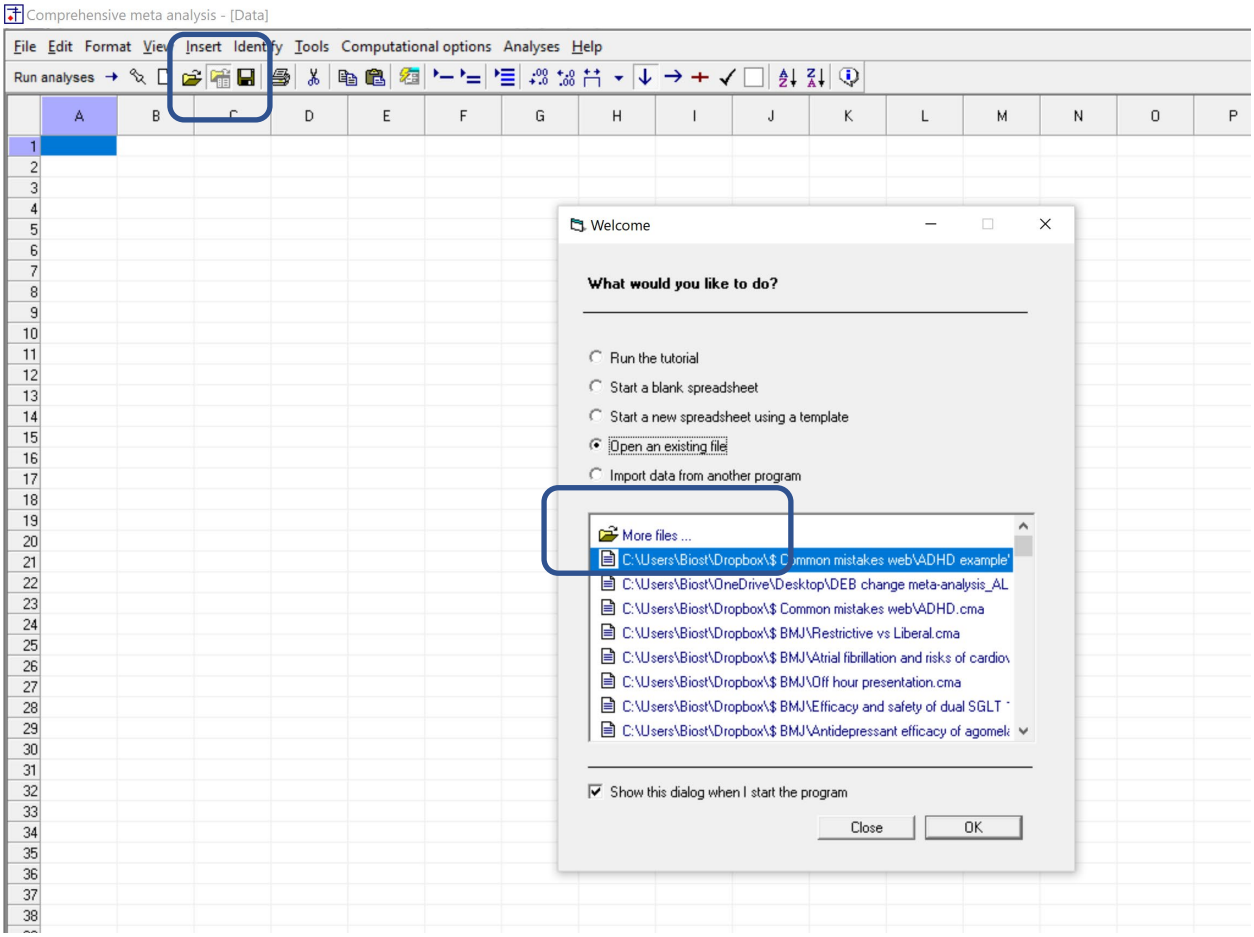
To open a CMA file > Open file from within CMA

You cannot open the file by clicking on it

If you would like to start with a blank file and enter the data, proceed with the next page

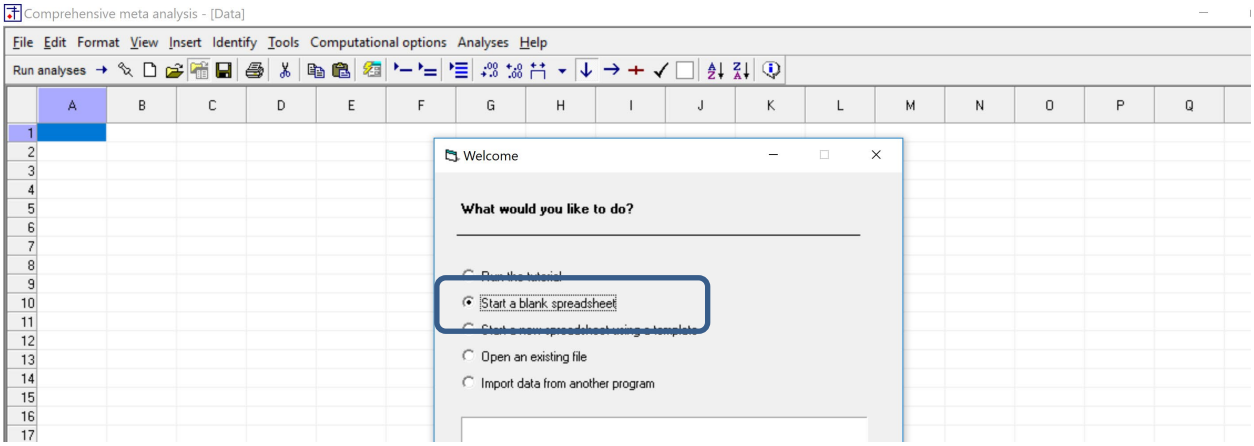
If you would like to simply open the file with the ADHD data, proceed with this page

- Start the CMA program
- Click the icon for the Opening Wizard
- Click More Files
- Locate the file ADHD.CMA
- Proceed with the analysis on Page 21

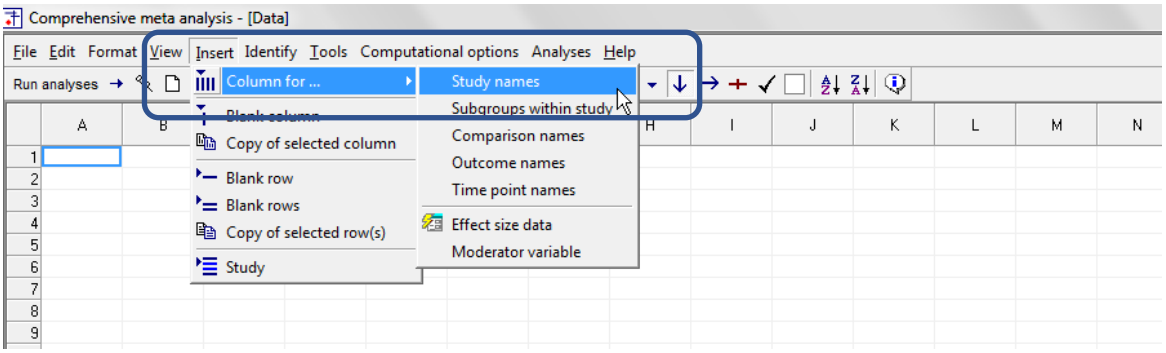


To create a new file from scratch, start here.

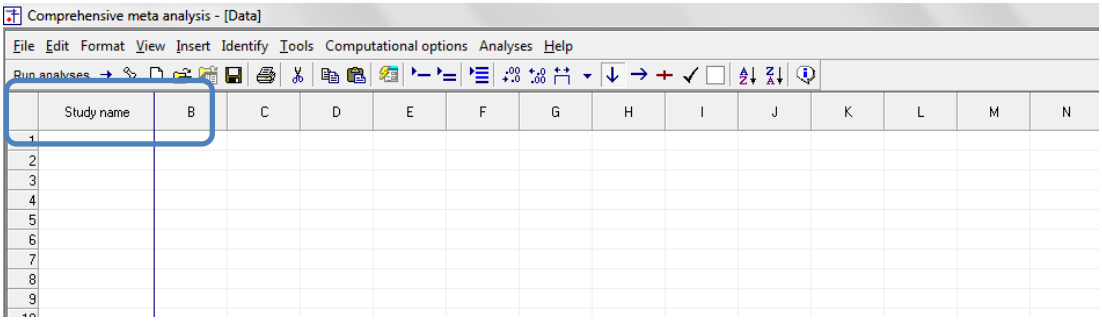
- Select the option [Start a blank spreadsheet]
- Click [Ok]



Click Insert > Column for > Study names



The screen should look like this

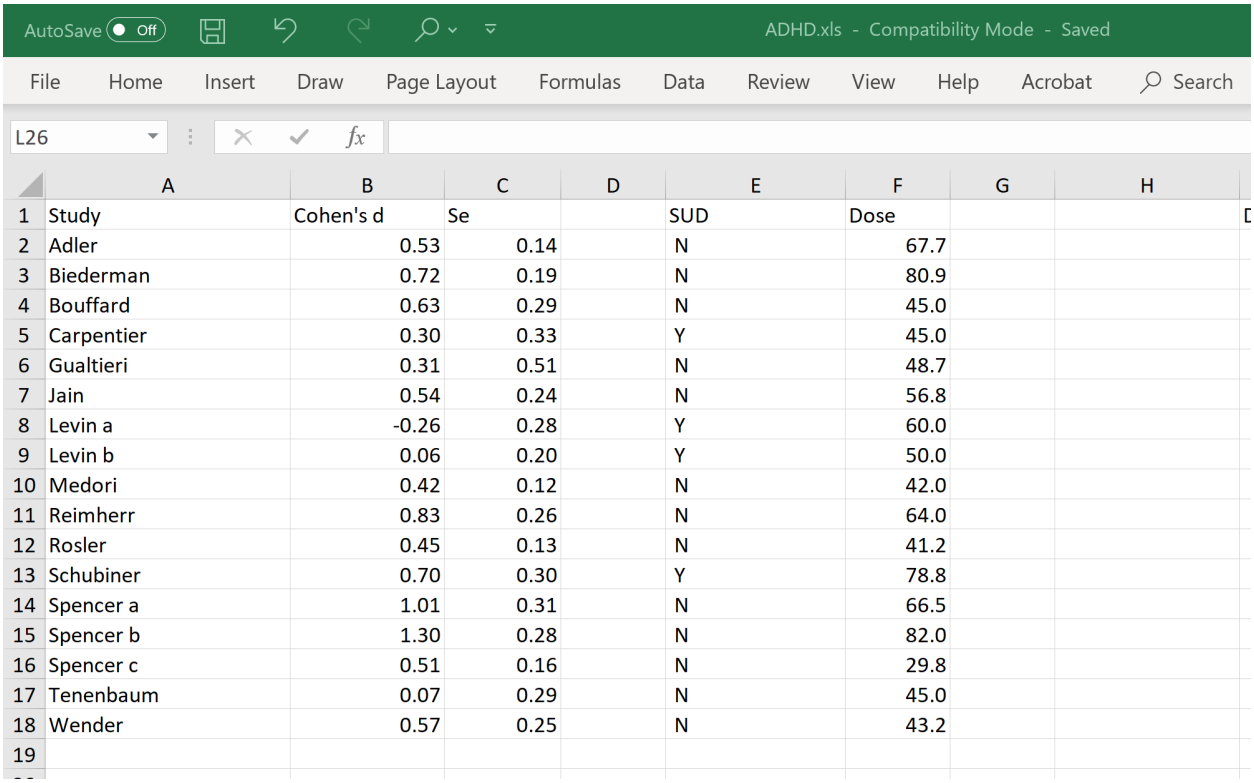


Next, we need to create columns that will hold the summary data for each study.

To do that, we need to know what format the data are in – for example, do we have the mean and standard deviation for each group, or the number of events in each group, or something else.

We can open an Excel file with the data, to see what format was provided.

The Excel file is called ADHD.xls.



	A	B	C	D	E	F	G	H
1	Study	Cohen's d	Se		SUD	Dose		
2	Adler	0.53	0.14		N	67.7		
3	Biederman	0.72	0.19		N	80.9		
4	Bouffard	0.63	0.29		N	45.0		
5	Carpentier	0.30	0.33		Y	45.0		
6	Gualtieri	0.31	0.51		N	48.7		
7	Jain	0.54	0.24		N	56.8		
8	Levin a	-0.26	0.28		Y	60.0		
9	Levin b	0.06	0.20		Y	50.0		
10	Medori	0.42	0.12		N	42.0		
11	Reimherr	0.83	0.26		N	64.0		
12	Rosler	0.45	0.13		N	41.2		
13	Schubiner	0.70	0.30		Y	78.8		
14	Spencer a	1.01	0.31		N	66.5		
15	Spencer b	1.30	0.28		N	82.0		
16	Spencer c	0.51	0.16		N	29.8		
17	Tenenbaum	0.07	0.29		N	45.0		
18	Wender	0.57	0.25		N	43.2		
19								

Typically, we would start with the raw mean and standard deviation for each study. We would enter these values into CMA, and CMA would compute the standardized mean difference and variance.

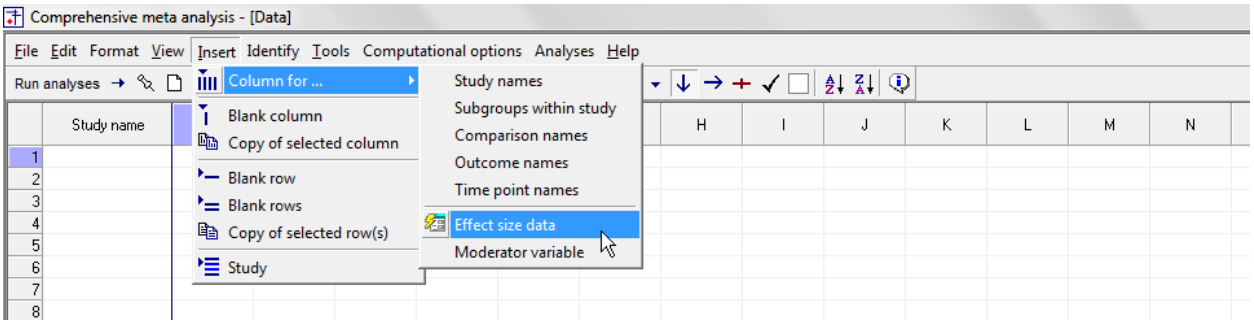
In this example, however, we are working from a paper that reported the standardized mean difference (Cohen's *d*) and its standard error (SE)

There are also two columns (SUD and Dose) that hold data for moderator variables.

Leave the Excel file open, since we will have the option of copying the data from Excel into CMA, rather than typing it into CMA.

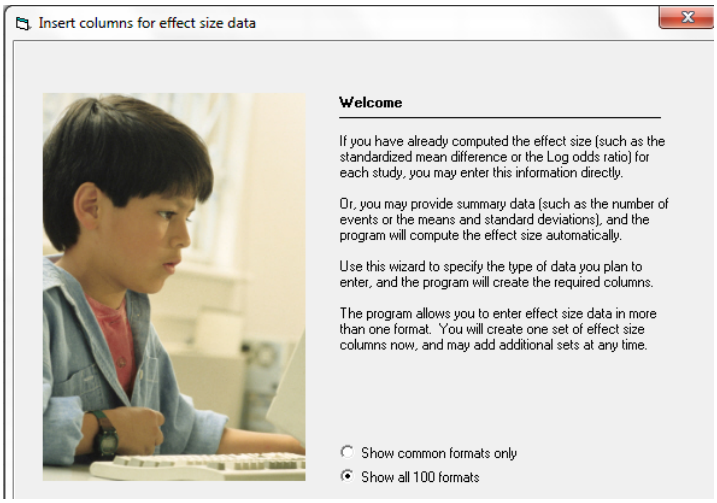
Return to CMA and create the columns that correspond to this Excel file.

Click Insert > Column for > Effect size data

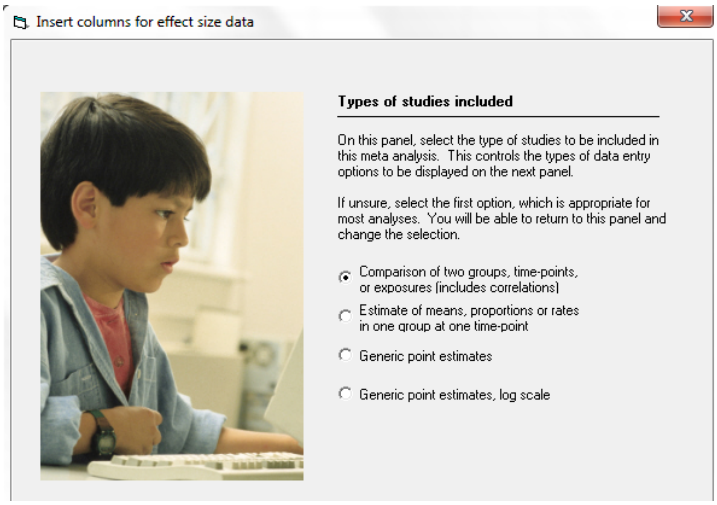


The program displays this wizard

Select [Show all 100 formats]
Click [Next]

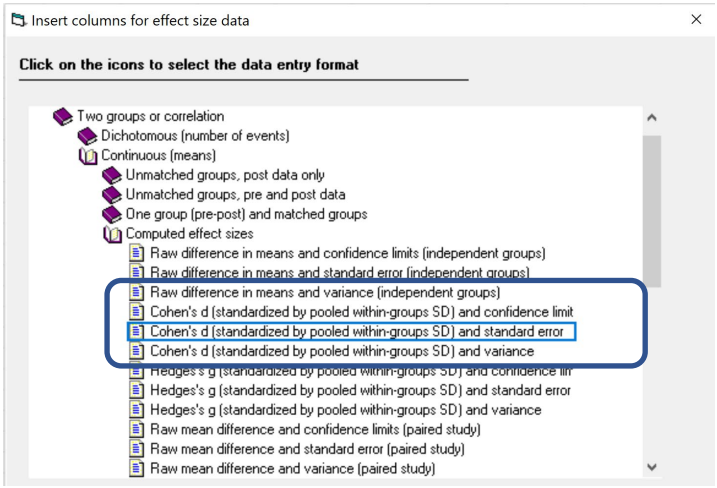


Select [Comparison of two groups…]
Click [Next]



Drill down to

Continuous (means)
Computed effect sizes
Hedges's *g* and standard error

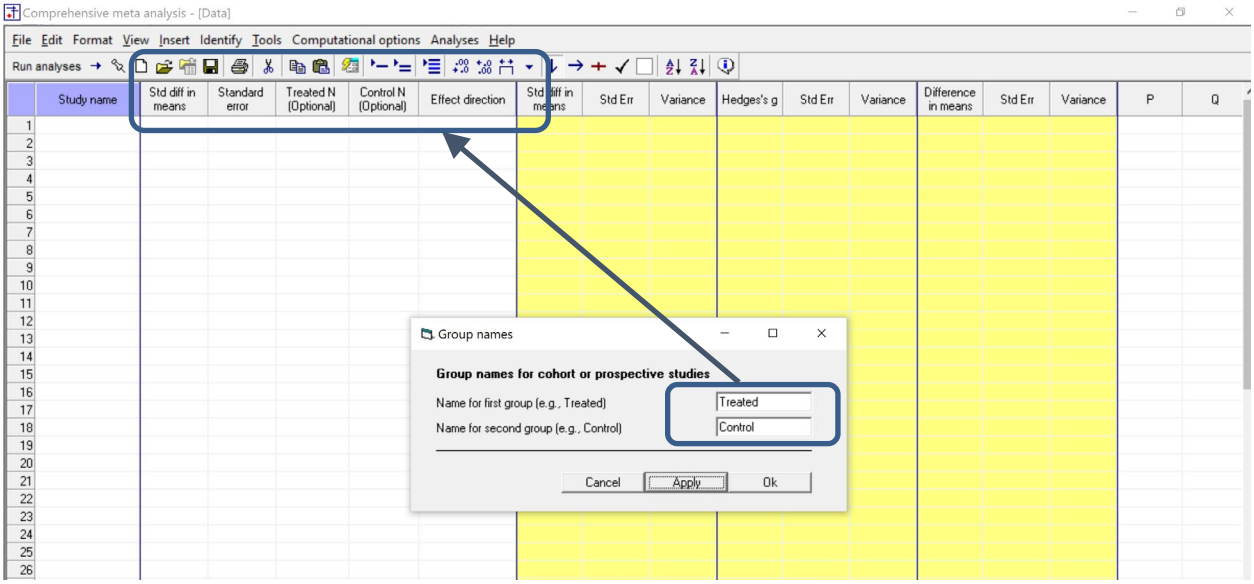


The program creates the columns that correspond to the selected format

Then, it displays this wizard. Enter the following labels into the wizard

- First group > Treated
- Second group > Control

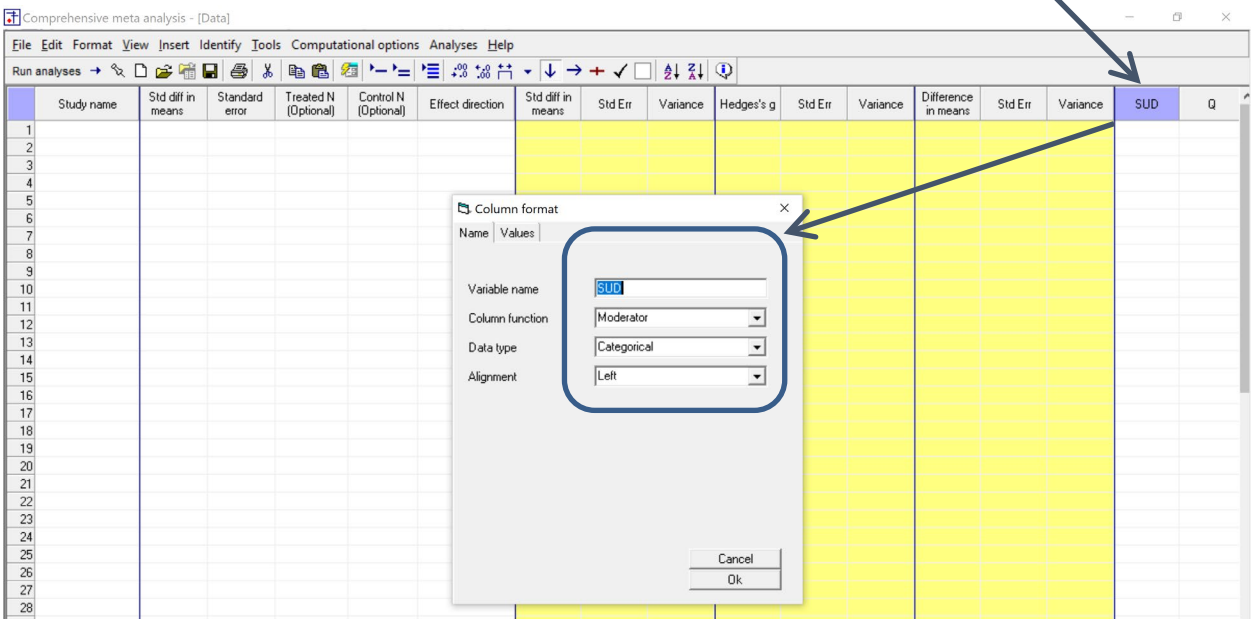
Click [Ok] and the program will copy the names into the grid



We want to create a column to hold the moderator “SUD”, which stands for substance abuse disorder. Each study will be coded Yes (Y) if that study enrolled patients with SUD, and No (No) if it excluded such patients.

Right-click on the first open column (P) and select Column Properties

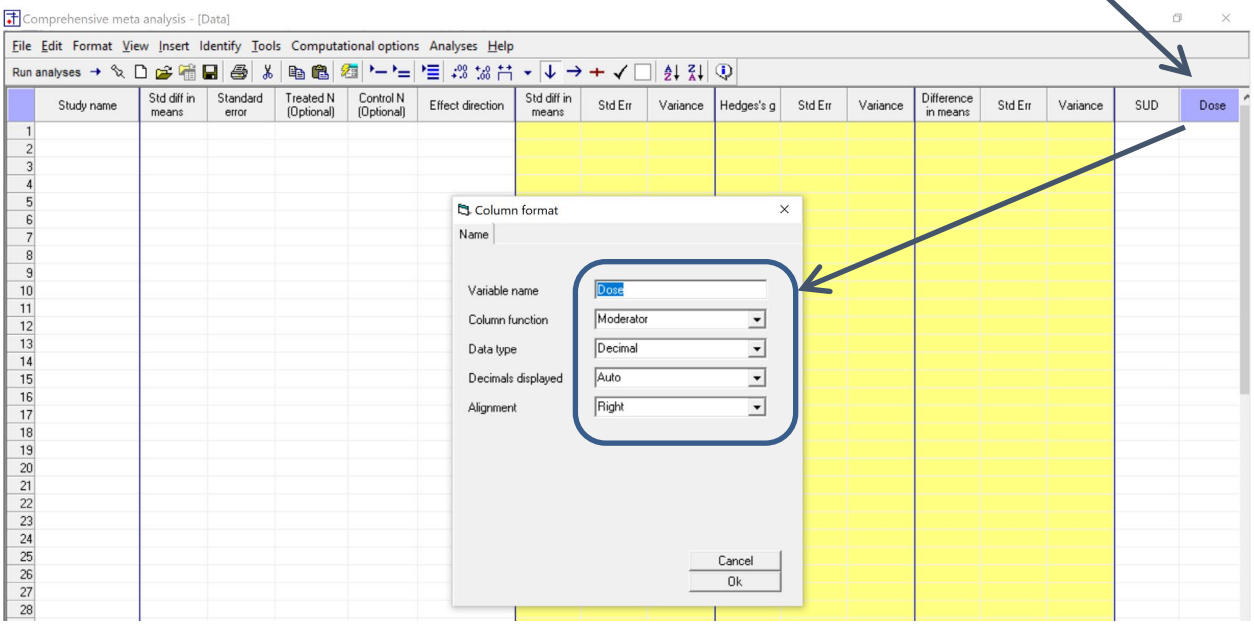
- Set the variable name to SUD
- Set the column function to Moderator
- Set the data type to Categorical
- Click Ok



We want to create a column to hold the moderator “Dose”, the mean dose of methylphenidate used in each study.

Right-click on the first open column (Q) and select Column Properties

- Set the variable name to Dose
- Set the column function to Moderator
- Set the data type to Decimal
- Click Ok

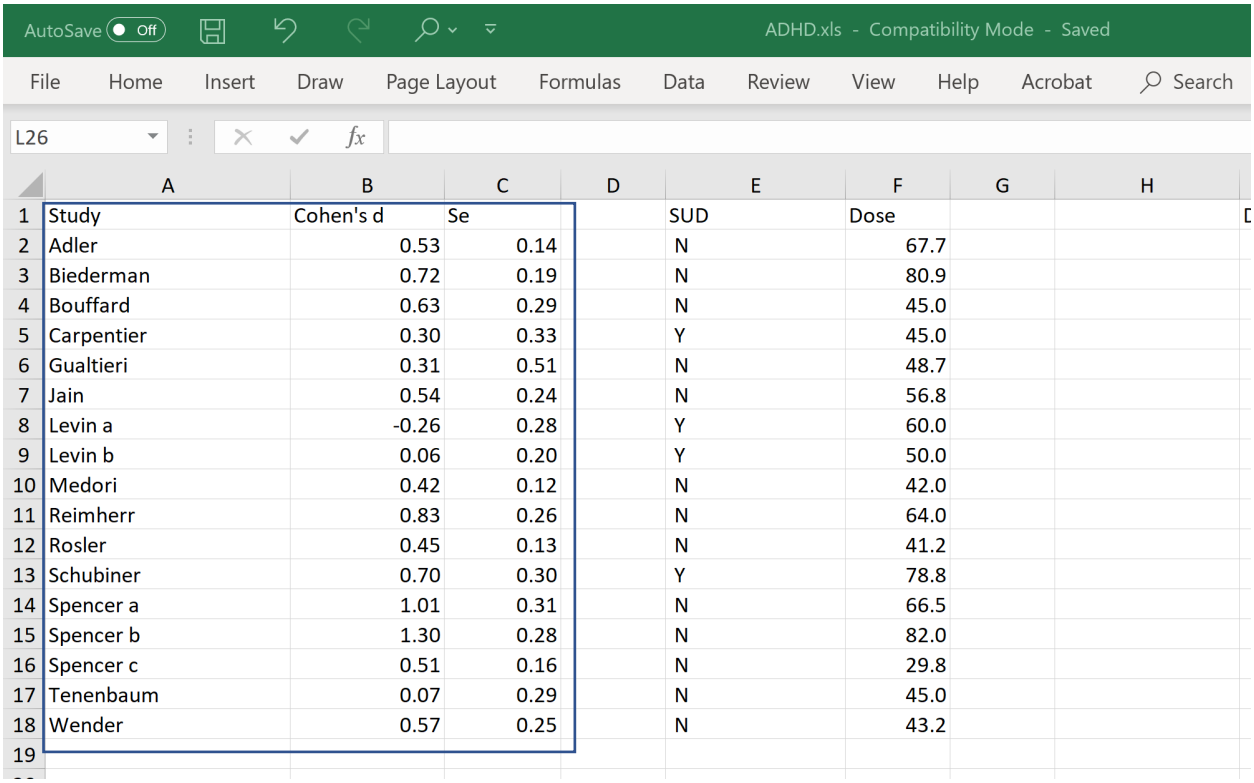


There are three options at this point

- Enter the data directly into CMA
- – or – Open the CMA data file “ADHD.cma”
- – or – Copy the data from Excel “ADHD.xls”

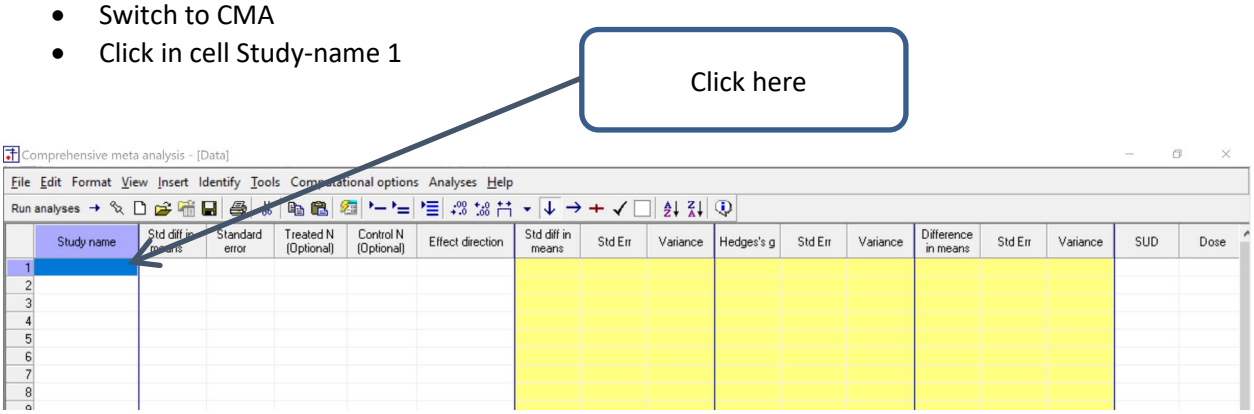
Here, we’ll show how to copy the data from Excel

- Switch to Excel and open the file
- Highlight the rows and columns as shown (Columns A to C), and press CTRL-C to copy to clipboard

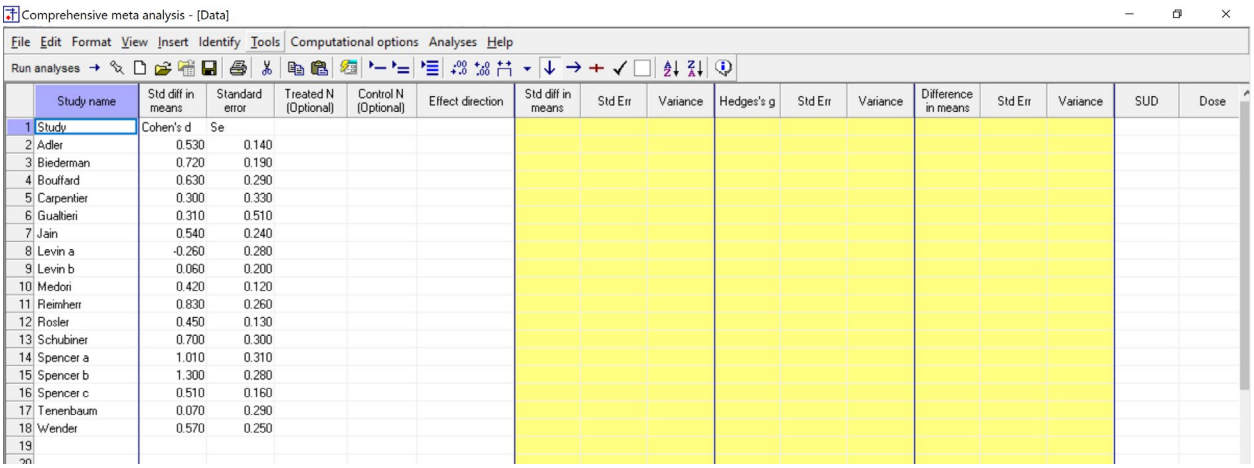


	A	B	C	D	E	F	G	H
1	Study	Cohen's d	Se		SUD	Dose		
2	Adler	0.53	0.14		N	67.7		
3	Biederman	0.72	0.19		N	80.9		
4	Bouffard	0.63	0.29		N	45.0		
5	Carpentier	0.30	0.33		Y	45.0		
6	Gualtieri	0.31	0.51		N	48.7		
7	Jain	0.54	0.24		N	56.8		
8	Levin a	-0.26	0.28		Y	60.0		
9	Levin b	0.06	0.20		Y	50.0		
10	Medori	0.42	0.12		N	42.0		
11	Reimherr	0.83	0.26		N	64.0		
12	Rosler	0.45	0.13		N	41.2		
13	Schubiner	0.70	0.30		Y	78.8		
14	Spencer a	1.01	0.31		N	66.5		
15	Spencer b	1.30	0.28		N	82.0		
16	Spencer c	0.51	0.16		N	29.8		
17	Tenenbaum	0.07	0.29		N	45.0		
18	Wender	0.57	0.25		N	43.2		
19								

- Switch to CMA
- Click in cell Study-name 1



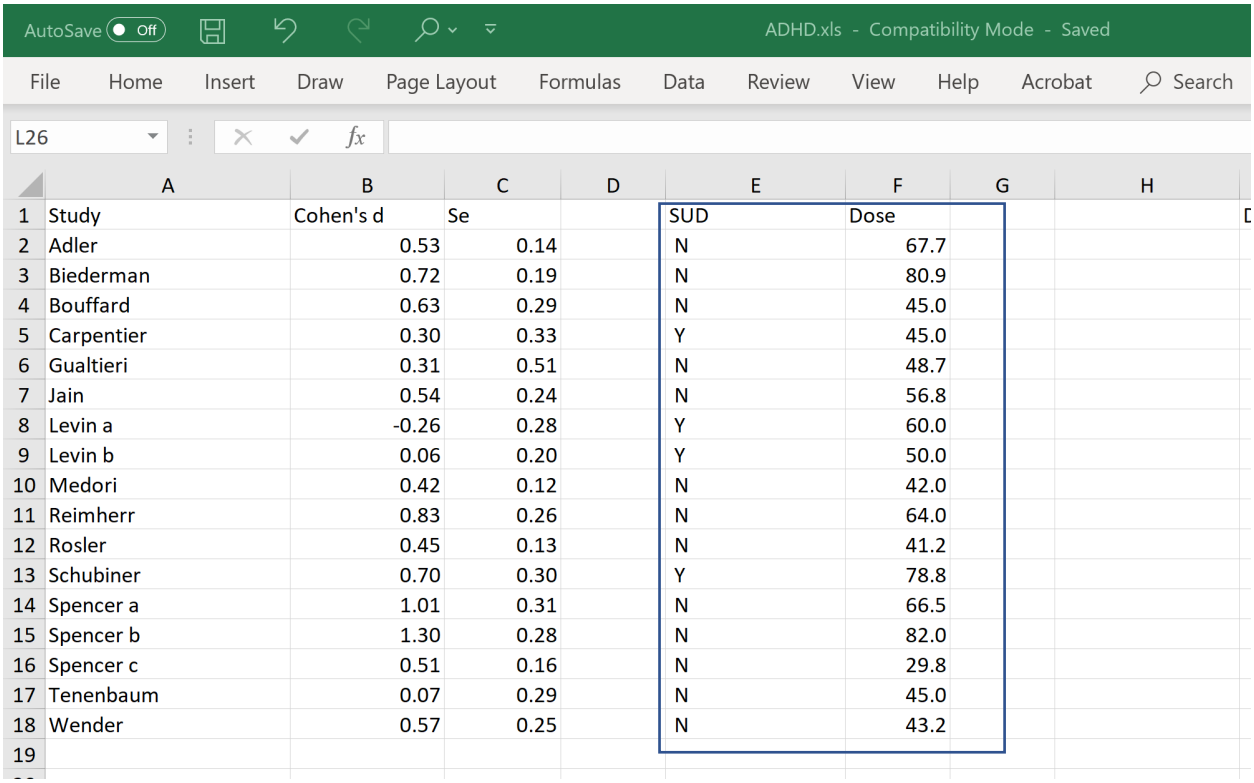
- Press [CTRL-V] to paste the data
- The screen should look like this



Return to Excel

Highlight the columns for SUD and Dose.

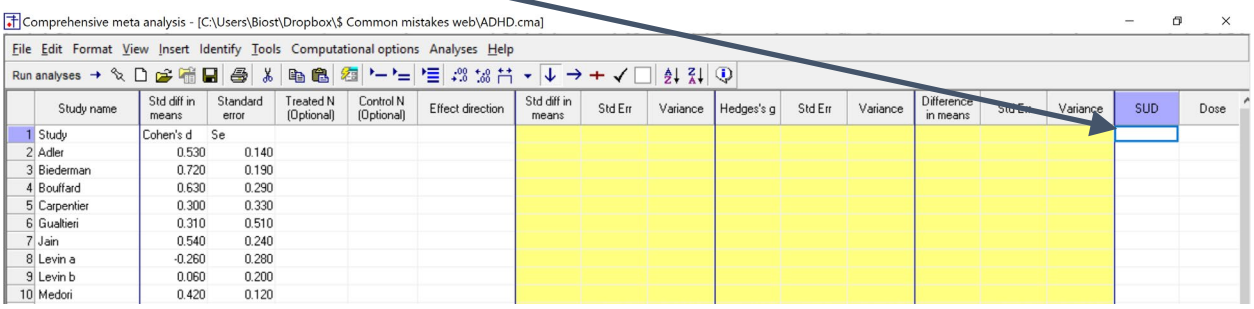
Press CTRL <C> to copy to the clipboard



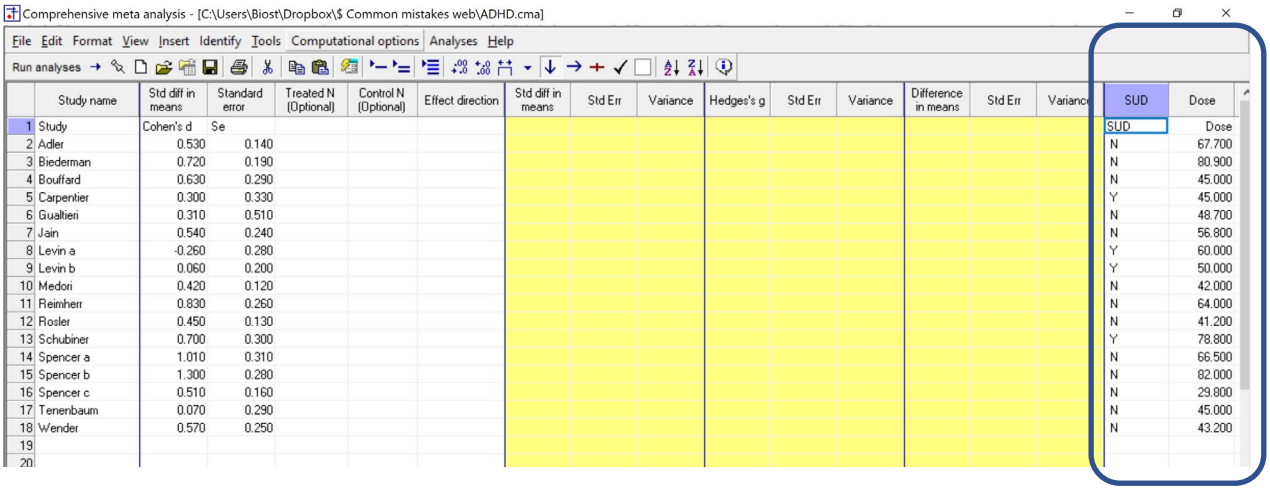
	A	B	C	D	E	F	G	H
1	Study	Cohen's d	Se		SUD	Dose		
2	Adler	0.53	0.14		N	67.7		
3	Biederman	0.72	0.19		N	80.9		
4	Bouffard	0.63	0.29		N	45.0		
5	Carpentier	0.30	0.33		Y	45.0		
6	Gualtieri	0.31	0.51		N	48.7		
7	Jain	0.54	0.24		N	56.8		
8	Levin a	-0.26	0.28		Y	60.0		
9	Levin b	0.06	0.20		Y	50.0		
10	Medori	0.42	0.12		N	42.0		
11	Reimherr	0.83	0.26		N	64.0		
12	Rosler	0.45	0.13		N	41.2		
13	Schubiner	0.70	0.30		Y	78.8		
14	Spencer a	1.01	0.31		N	66.5		
15	Spencer b	1.30	0.28		N	82.0		
16	Spencer c	0.51	0.16		N	29.8		
17	Tenenbaum	0.07	0.29		N	45.0		
18	Wender	0.57	0.25		N	43.2		
19								

Return to CMA

Click in the SUD column, Row 1



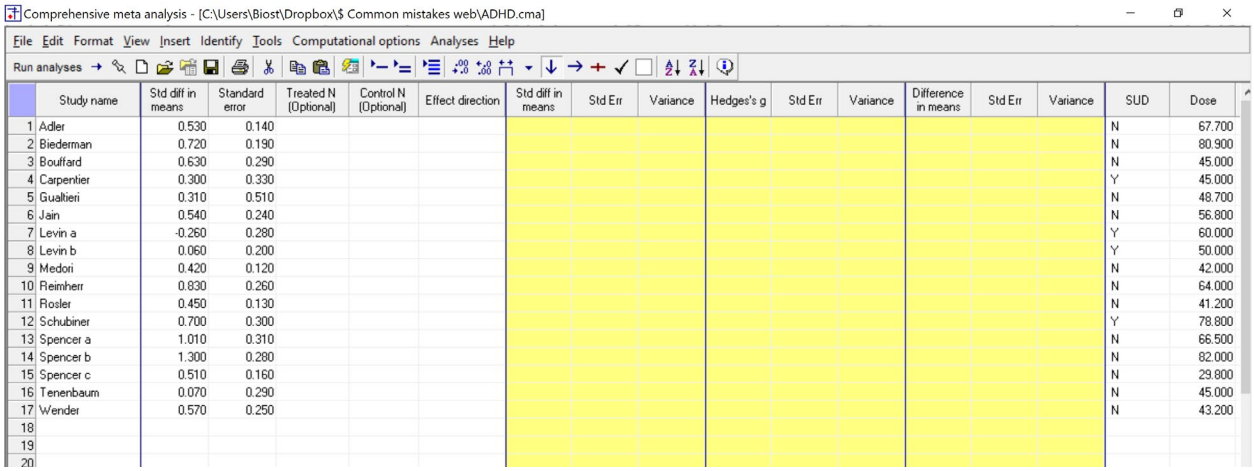
Click CTRL <V> to paste the data



The screen should look like this

After checking that the data has been copied correctly, we can delete Row 1

- Click anywhere in Row 1
- Select Edit > Delete row, and confirm
- Your screen should look like this

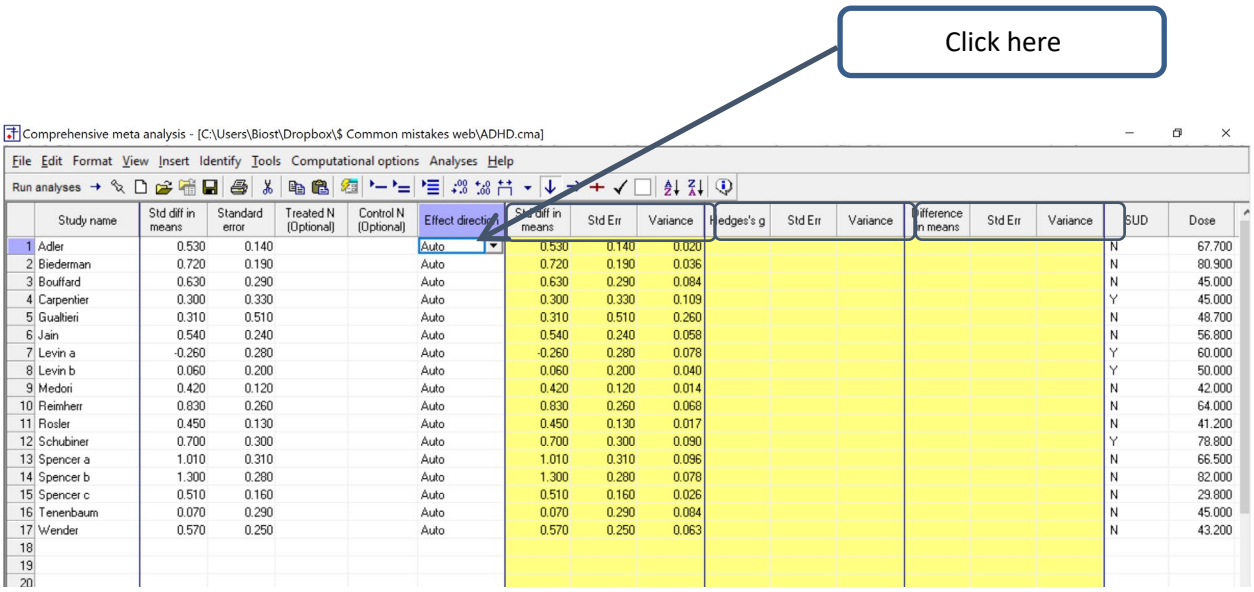


	Study name	Std diff in means	Standard error	Treated N (Optional)	Control N (Optional)	Effect direction	Std diff in means	Std Err	Variance	Hedges's g	Std Err	Variance	Difference in means	Std Err	Variance	SUD	Dose
1	Adler	0.530	0.140													N	67.700
2	Biederman	0.720	0.190													N	80.900
3	Bouffard	0.630	0.290													N	45.000
4	Carpentier	0.300	0.330													Y	45.000
5	Gualtieri	0.310	0.510													N	48.700
6	Jain	0.540	0.240													N	56.800
7	Levin a	-0.260	0.280													Y	60.000
8	Levin b	0.060	0.200													Y	50.000
9	Medori	0.420	0.120													N	42.000
10	Reimherr	0.830	0.260													N	64.000
11	Rosler	0.450	0.130													N	41.200
12	Schubiner	0.700	0.300													Y	78.800
13	Spencer a	1.010	0.310													N	66.500
14	Spencer b	1.300	0.280													N	82.000
15	Spencer c	0.510	0.160													N	29.800
16	Tenenbaum	0.070	0.290													N	45.000
17	Wender	0.570	0.250													N	43.200
18																	
19																	
20																	

We need to enter a value for “Effect Direction”

Enter “Auto” for each study (or enter Auto for one study and copy to the others

<Auto> tells the program to take the first mean minus the second. Since all studies used scales that worked in the same direction (a higher score better function) this is appropriate here.



	Study name	Std diff in means	Standard error	Treated N (Optional)	Control N (Optional)	Effect direction	Std diff in means	Std Err	Variance	Hedges's g	Std Err	Variance	Difference in means	Std Err	Variance	SUD	Dose
1	Adler	0.530	0.140			Auto	0.530	0.140	0.020							N	67.700
2	Biederman	0.720	0.190			Auto	0.720	0.190	0.036							N	80.900
3	Bouffard	0.630	0.290			Auto	0.630	0.290	0.084							N	45.000
4	Carpentier	0.300	0.330			Auto	0.300	0.330	0.109							Y	45.000
5	Gualtieri	0.310	0.510			Auto	0.310	0.510	0.260							N	48.700
6	Jain	0.540	0.240			Auto	0.540	0.240	0.058							N	56.800
7	Levin a	-0.260	0.280			Auto	-0.260	0.280	0.078							Y	60.000
8	Levin b	0.060	0.200			Auto	0.060	0.200	0.040							Y	50.000
9	Medori	0.420	0.120			Auto	0.420	0.120	0.014							N	42.000
10	Reimherr	0.830	0.260			Auto	0.830	0.260	0.068							N	64.000
11	Rosler	0.450	0.130			Auto	0.450	0.130	0.017							N	41.200
12	Schubiner	0.700	0.300			Auto	0.700	0.300	0.090							Y	78.800
13	Spencer a	1.010	0.310			Auto	1.010	0.310	0.096							N	66.500
14	Spencer b	1.300	0.280			Auto	1.300	0.280	0.078							N	82.000
15	Spencer c	0.510	0.160			Auto	0.510	0.160	0.026							N	29.800
16	Tenenbaum	0.070	0.290			Auto	0.070	0.290	0.084							N	45.000
17	Wender	0.570	0.250			Auto	0.570	0.250	0.063							N	43.200
18																	
19																	
20																	

We do not have the number of patients for either group, so we leave those columns blank.

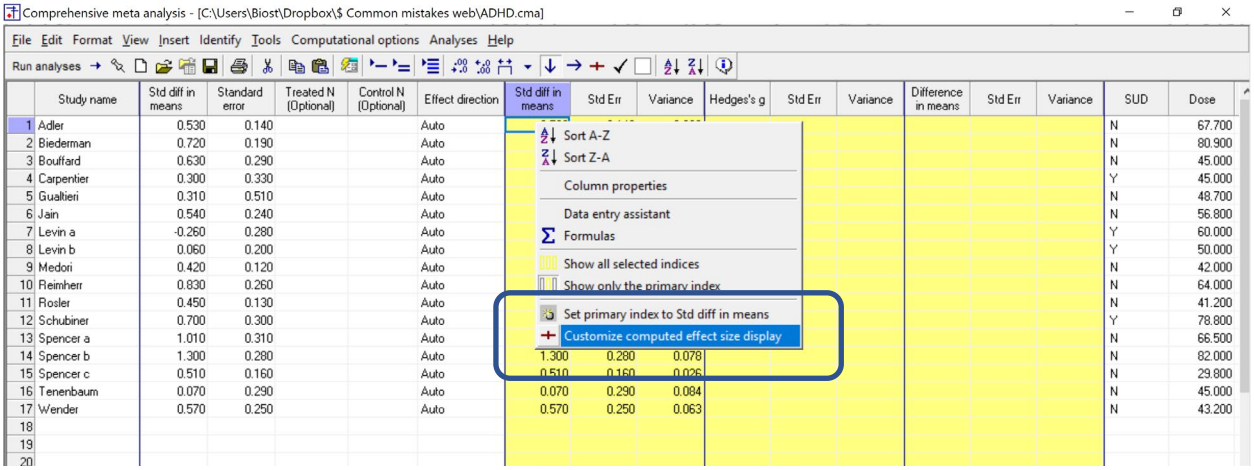
The program displays a set of columns for three effect sizes – the standardized difference in means (*d*), the standardized difference in means (*g*) and the raw difference in means (D). Based on the data we entered, the program is able to compute only one of these (*d*) and so the others are blank.

Std Diff in means

This is the standardized difference in means, sometimes called Cohen’s *d*. It is defined as the raw difference in means divided by the standard deviation (computed within groups and pooled). This index transforms the outcomes for all studies onto a common metric, so they can be used in the same analysis.

In this analysis we will be using Cohen's *d*. To save space on the screen we will hide the raw difference and *g*, leaving only *d* on the display

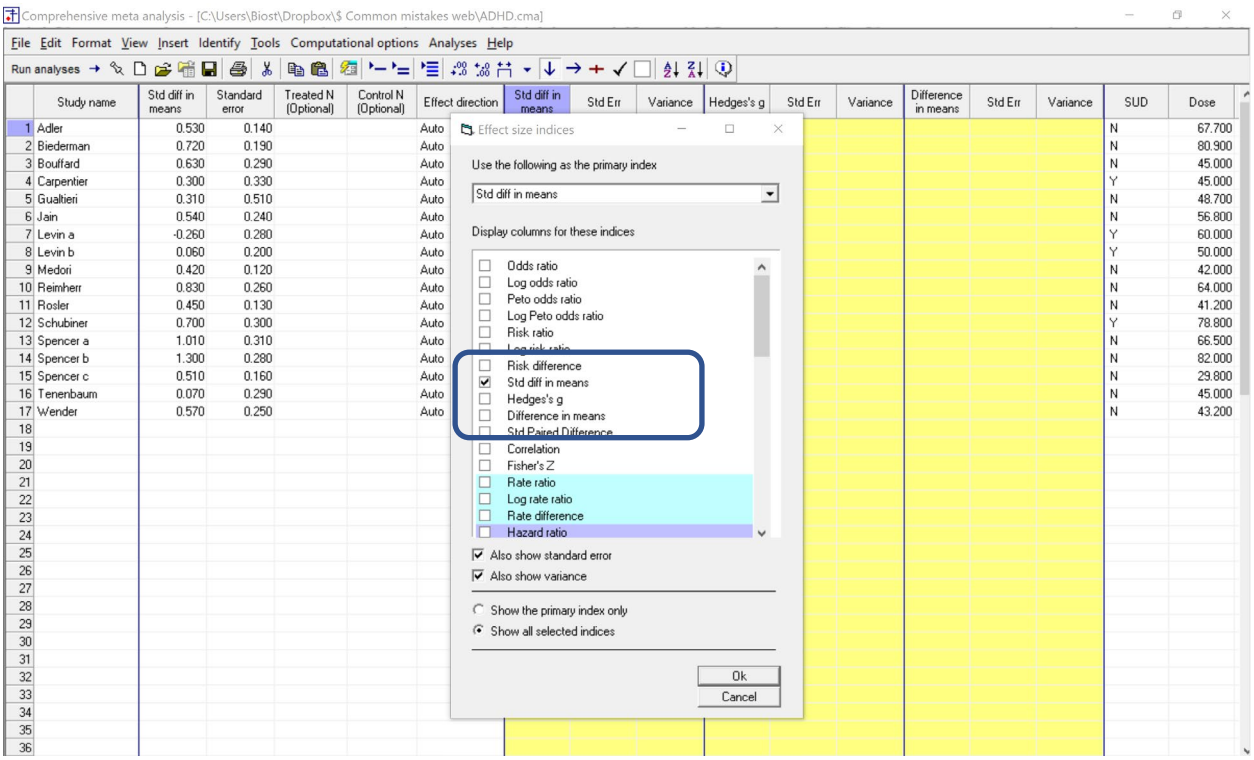
- Right-click in any yellow column
- Click "Customize computed effect size display"



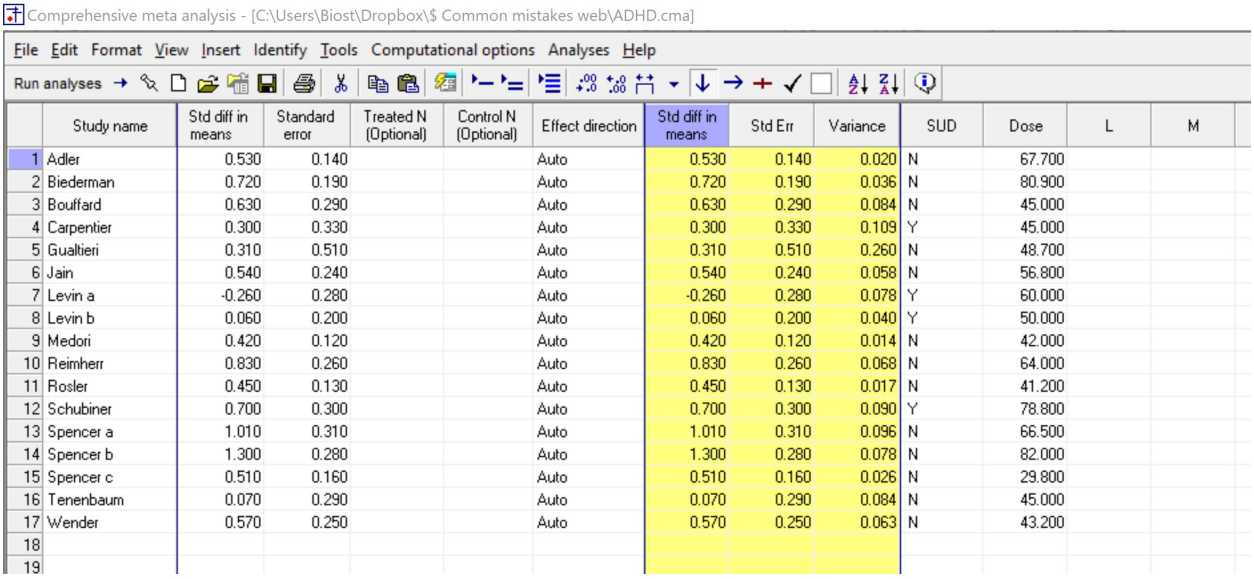
	Study name	Std diff in means	Standard error	Treated N (Optional)	Control N (Optional)	Effect direction	Std diff in means	Std Err	Variance	Hedges's g	Std Err	Variance	Difference in means	Std Err	Variance	SUD	Dose
1	Adler	0.530	0.140			Auto										N	67.700
2	Biederman	0.720	0.190			Auto										N	80.900
3	Bouffard	0.630	0.290			Auto										N	45.000
4	Carpentier	0.300	0.330			Auto										Y	45.000
5	Gualtieri	0.310	0.510			Auto										N	48.700
6	Jain	0.540	0.240			Auto										N	56.800
7	Levin a	-0.260	0.280			Auto										Y	60.000
8	Levin b	0.060	0.200			Auto										Y	50.000
9	Medori	0.420	0.120			Auto										N	42.000
10	Reimherr	0.830	0.260			Auto										N	64.000
11	Rosler	0.450	0.130			Auto										N	41.200
12	Schubiner	0.700	0.300			Auto										Y	78.800
13	Spencer a	1.010	0.310			Auto										N	66.500
14	Spencer b	1.300	0.280			Auto	1.300	0.280	0.078							N	82.000
15	Spencer c	0.510	0.160			Auto	0.510	0.160	0.026							N	29.800
16	Tenenbaum	0.070	0.290			Auto	0.070	0.290	0.084							N	45.000
17	Wender	0.570	0.250			Auto	0.570	0.250	0.063							N	43.200
18																	
19																	
20																	

In the wizard,

- Un-check “Hedges’ *g*”
- Un-check “Difference in means”
- Put a check-mark adjacent to “Also show standard error”
- Put a check-mark adjacent to “Also show variance”
- Click [Ok]



The screen should look like this



	Study name	Std diff in means	Standard error	Treated N (Optional)	Control N (Optional)	Effect direction	Std diff in means	Std Err	Variance	SUD	Dose	L	M
1	Adler	0.530	0.140			Auto	0.530	0.140	0.020	N	67.700		
2	Biederman	0.720	0.190			Auto	0.720	0.190	0.036	N	80.900		
3	Bouffard	0.630	0.290			Auto	0.630	0.290	0.084	N	45.000		
4	Carpentier	0.300	0.330			Auto	0.300	0.330	0.109	Y	45.000		
5	Gualtieri	0.310	0.510			Auto	0.310	0.510	0.260	N	48.700		
6	Jain	0.540	0.240			Auto	0.540	0.240	0.058	N	56.800		
7	Levin a	-0.260	0.280			Auto	-0.260	0.280	0.078	Y	60.000		
8	Levin b	0.060	0.200			Auto	0.060	0.200	0.040	Y	50.000		
9	Medori	0.420	0.120			Auto	0.420	0.120	0.014	N	42.000		
10	Reimherr	0.830	0.260			Auto	0.830	0.260	0.068	N	64.000		
11	Rosler	0.450	0.130			Auto	0.450	0.130	0.017	N	41.200		
12	Schubiner	0.700	0.300			Auto	0.700	0.300	0.090	Y	78.800		
13	Spencer a	1.010	0.310			Auto	1.010	0.310	0.096	N	66.500		
14	Spencer b	1.300	0.280			Auto	1.300	0.280	0.078	N	82.000		
15	Spencer c	0.510	0.160			Auto	0.510	0.160	0.026	N	29.800		
16	Tenenbaum	0.070	0.290			Auto	0.070	0.290	0.084	N	45.000		
17	Wender	0.570	0.250			Auto	0.570	0.250	0.063	N	43.200		
18													
19													

Click File > Save As and save the file

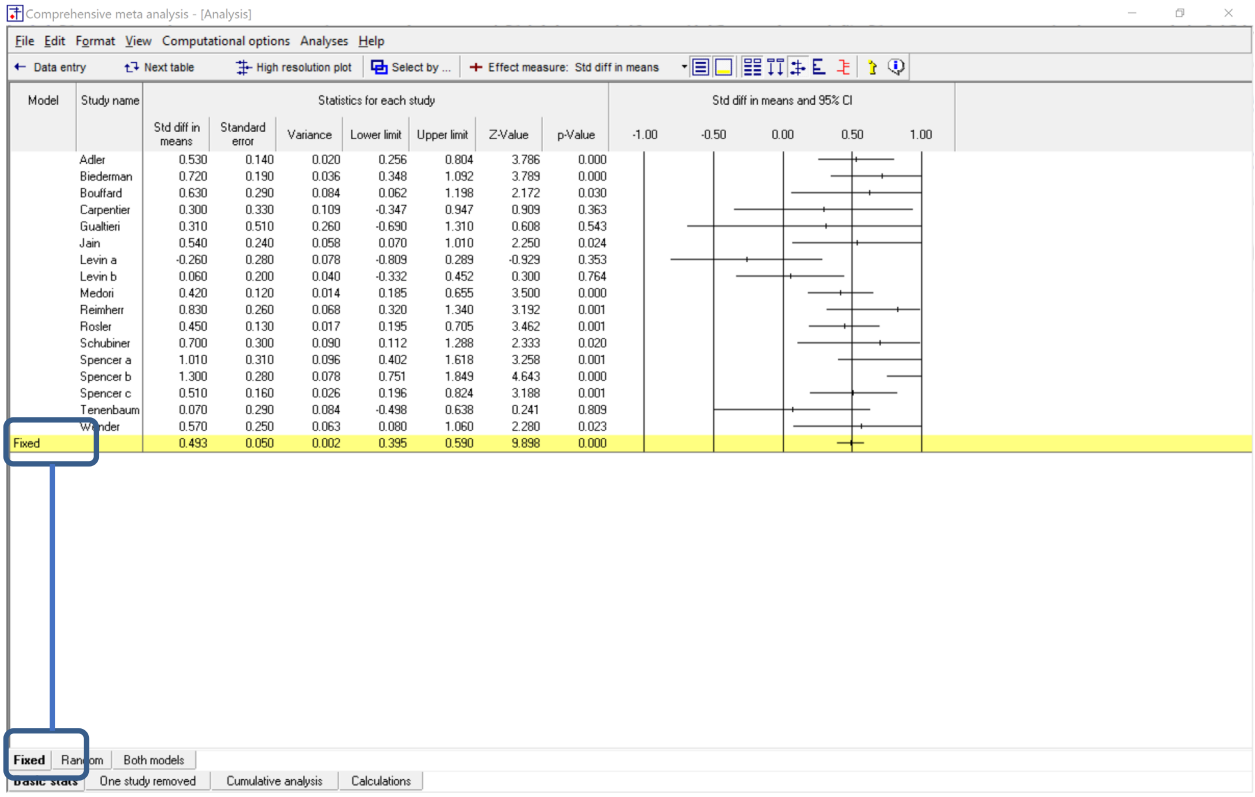
Note that the file name is now in the header.

- [Save] will over-write the prior version of this file without warning
- [Save As…] will allow you to save the file with a new name

Basic Analysis

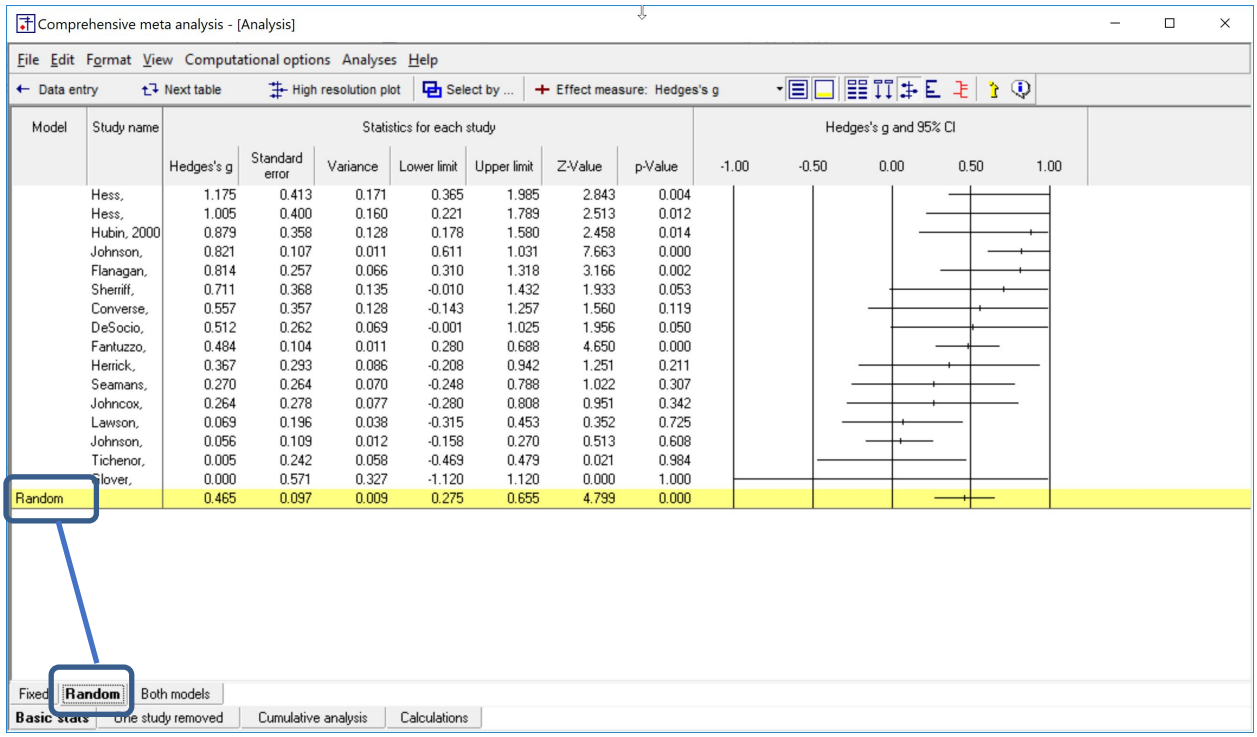
To run the analysis, click [Run analysis]

This is the basic analysis screen



Initially, the program displays the fixed-effect analysis. This is indicated by the tab at the bottom and the label in the plot.

The fixed-effect model is appropriate when all studies are based on the same population, and are identical in all material respects. That's rarely the case, and is clearly not the case here.



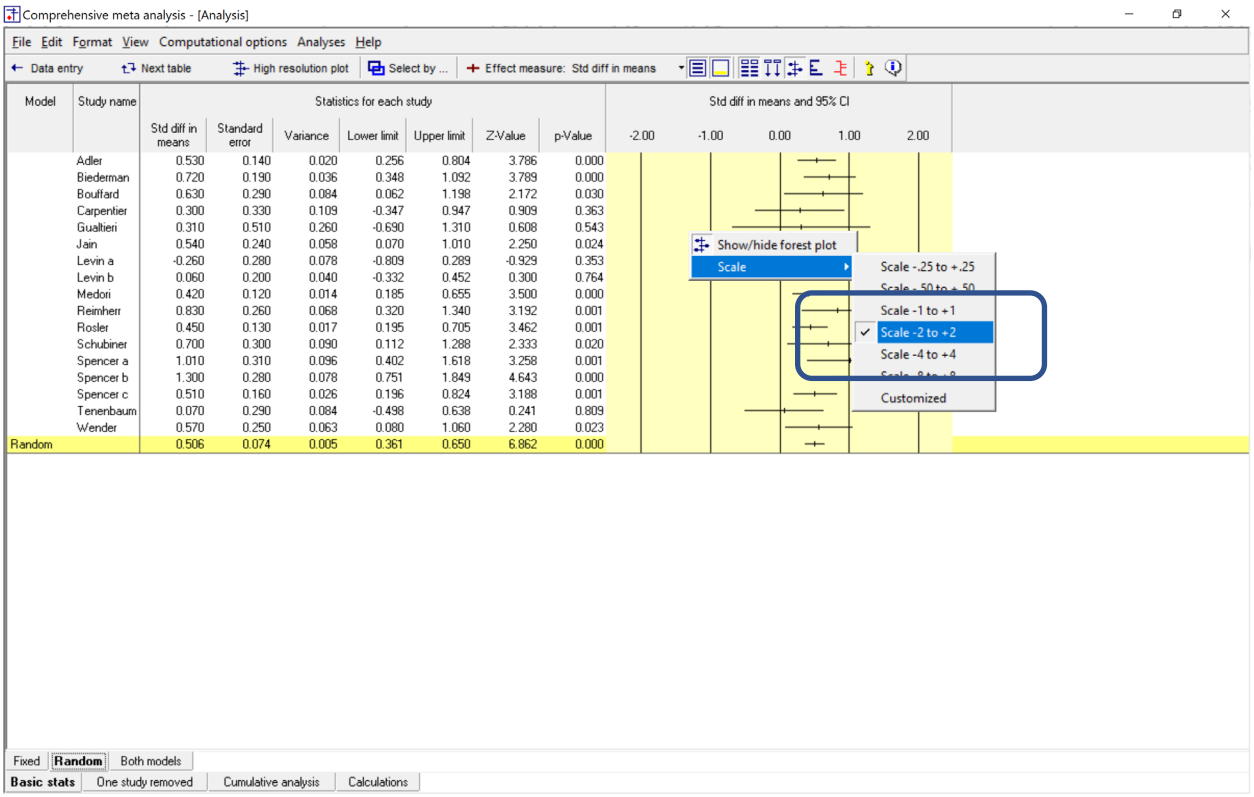
Rather, we want to use the random-effects model. This model assume that the studies in the analysis have been selected from a universe of studies that meet some criteria (as detailed by the description of Populations, Interventions, Outcomes, and Comparison groups in the full paper). The model also assumes that the effect sizes in these studies are a random selection of the effect sizes in the universe.

To switch to the random-effects model, click “Random” at the bottom of the screen.

Next, we may want to modify the scale for the forest plot

At the moment, the scale goes from -1 to +1 and some values are being truncated at the right-end.

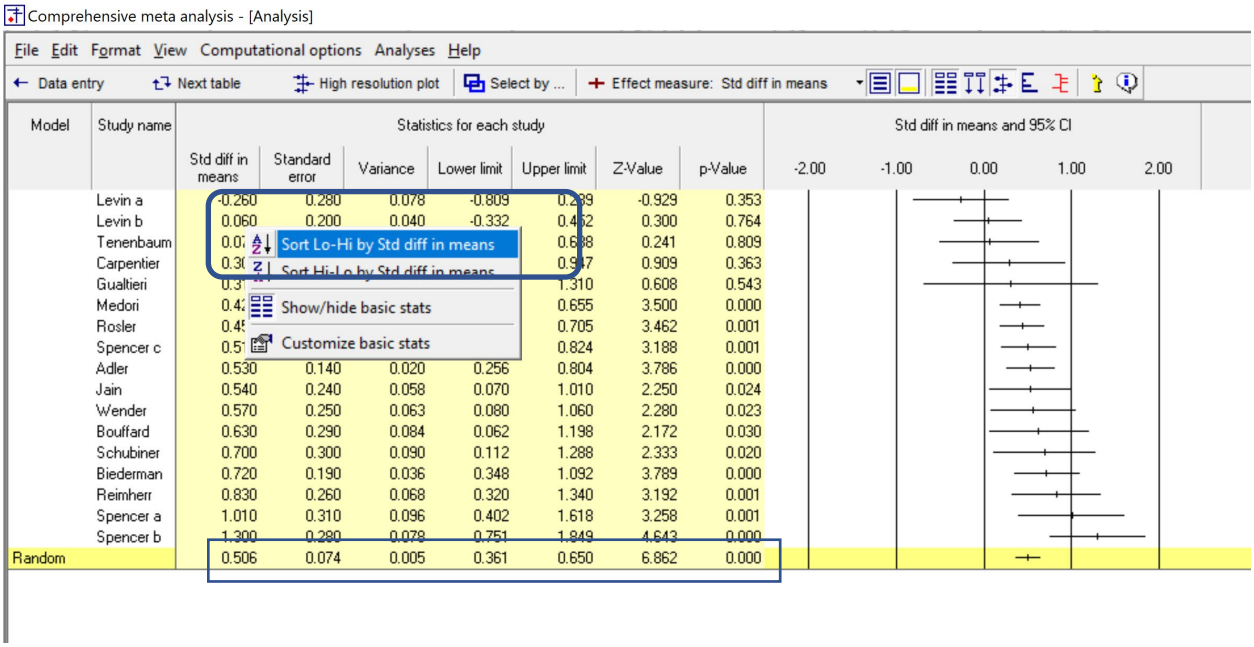
Right-click on the forest plot > Scale > Scale -2 to +2



To get a better sense of the dispersion we can sort the studies based on their effect sizes.

To sort by any column, right-click *on that column*

Here, right-click on Std diff in means, and click <Sort Lo-Hi by Std diff in means>



The mean effect size and confidence interval

The statistics on this row address the mean.

The mean effect size is 0.506.

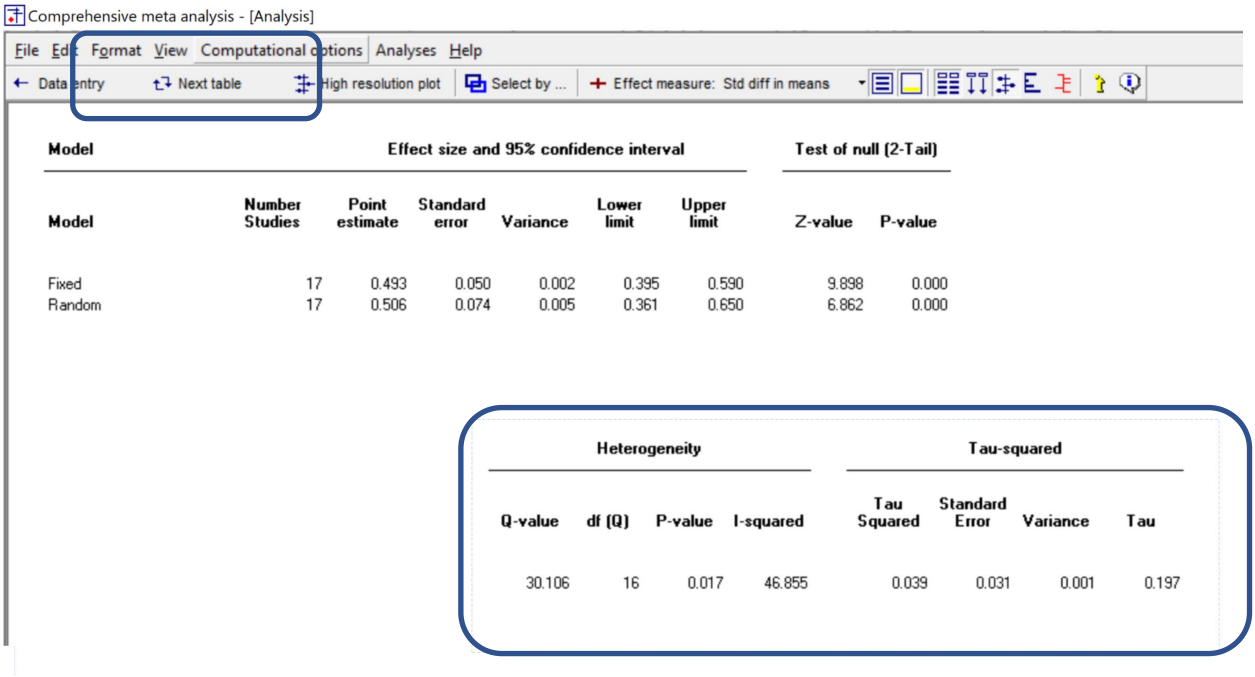
The confidence interval of 0.361 to 0.650 speaks to the precision of the mean. The mean of 0.506 is our estimate of the true mean in the universe of comparable studies. But, it is only an estimate. The true mean in that universe probably falls in the range of 0.361 to 0.650. (More correctly, in 95% of all meta-analyses, the confidence interval will include the true mean for the universe of comparable studies).

In addition to estimating the mean, we can test the null hypothesis that in the universe of comparable studies, the true mean effect is zero. The Z-value for this test is 6.862 and the corresponding p-value is < 0.001. We can reject this null hypothesis, and conclude that *on average* in the universe of studies comparable to those in the analysis, the treatment increases cognitive function.

Heterogeneity

To this point we've addressed the *mean* effect size. We need to also consider how much the effects vary from study to study.

To see the relevant statistics, click Next Table



The statistics displayed at the top address the mean effect size. The statistics displayed at the bottom address the heterogeneity in effect size.

This is a modified version of the actual screen. In practice, these statistics are all displayed across the top of the screen. In this screen-shot I moved the heterogeneity statistics to the bottom for readability.

The relevant statistics are as follows. *Q* is 30.106 with 16 degrees of freedom and a p-value of 0.017. I^2 is 46.855%. Tau-squared (T^2) is 0.039, and Tau (T) is 0.197. I'll start by presenting these statistics in the way that they are typically presented. Then I'll explain why these statistics are largely irrelevant, and suggest a better approach.

The *Q*-value is the sum of squared deviations of all effect-sizes from the mean effect size, on a standardized scale. Under the null hypothesis that the true effect size is precisely the same in all studies (and that all of the variation in observed effects is due to sampling error), the expected value of *Q* would be equal to the degrees of freedom (the number of studies minus 1).

In this analysis, *Q* is 30.106 with 16 degrees of freedom. Under the null hypothesis that the treatment has precisely the same impact in all studies, *Q* would follow a chi-squared distribution with 16 df. So by referring to a table of chi-square we can determine that the p-value for a test of this null hypothesis is

0.017. We reject the null hypothesis, and conclude that the treatment is more effective in some populations than in others.

There is a widespread belief that the I^2 statistic tells us how much the effect size varies across studies. This is a fundamental misunderstanding of what this statistic means.

In fact, I^2 tells us what proportion of the variance in observed effects is due to variance in true effects rather than sampling error. As such, it provides some context for understanding the forest plot. If I^2 is low (near zero), then most of the variance in the forest plot is due to sampling error. If we could somehow plot the variance of true effects (which is what we are about), most of the variance would disappear. Conversely, if I^2 is high (say, more than 75%) then most of the variance in the forest plot is due to variance in true effects. If we could somehow plot the variance of true effects (which is what we are about), most of the variance would remain.

In this case, I^2 is around 47%, so the plot of observed effects provides a useful (if slightly exaggerated) idea of how the true effects are distributed.

The variance of true effects (T^2) is 0.039, and the standard deviation of true effects (T) is 0.197.

All of these statistics are accurate. But, with the exception of T, they all are largely irrelevant. When we ask about heterogeneity, we intend to ask "Over what range do the effects vary", and none of these statistics directly addresses this question. Concretely –

- The *p*-value provides a test of the null hypothesis that the effects are all the same. There is not really any reason to test this hypothesis, since common sense tells us that the intervention will not have precisely the same impact in all populations. In any event, the *p*-value only tells us that the effects vary – it does not tell us how much they vary.

- Similarly, I^2 tells us that most of the variation that we see in the observed effects reflects variation in true effects, rather than sampling error. This (along with the plot) provides some idea of the variation, but it's a rough idea, at best. Based on I^2 and the plot, it's unlikely that any two people will arrive at the same estimate of the dispersion.

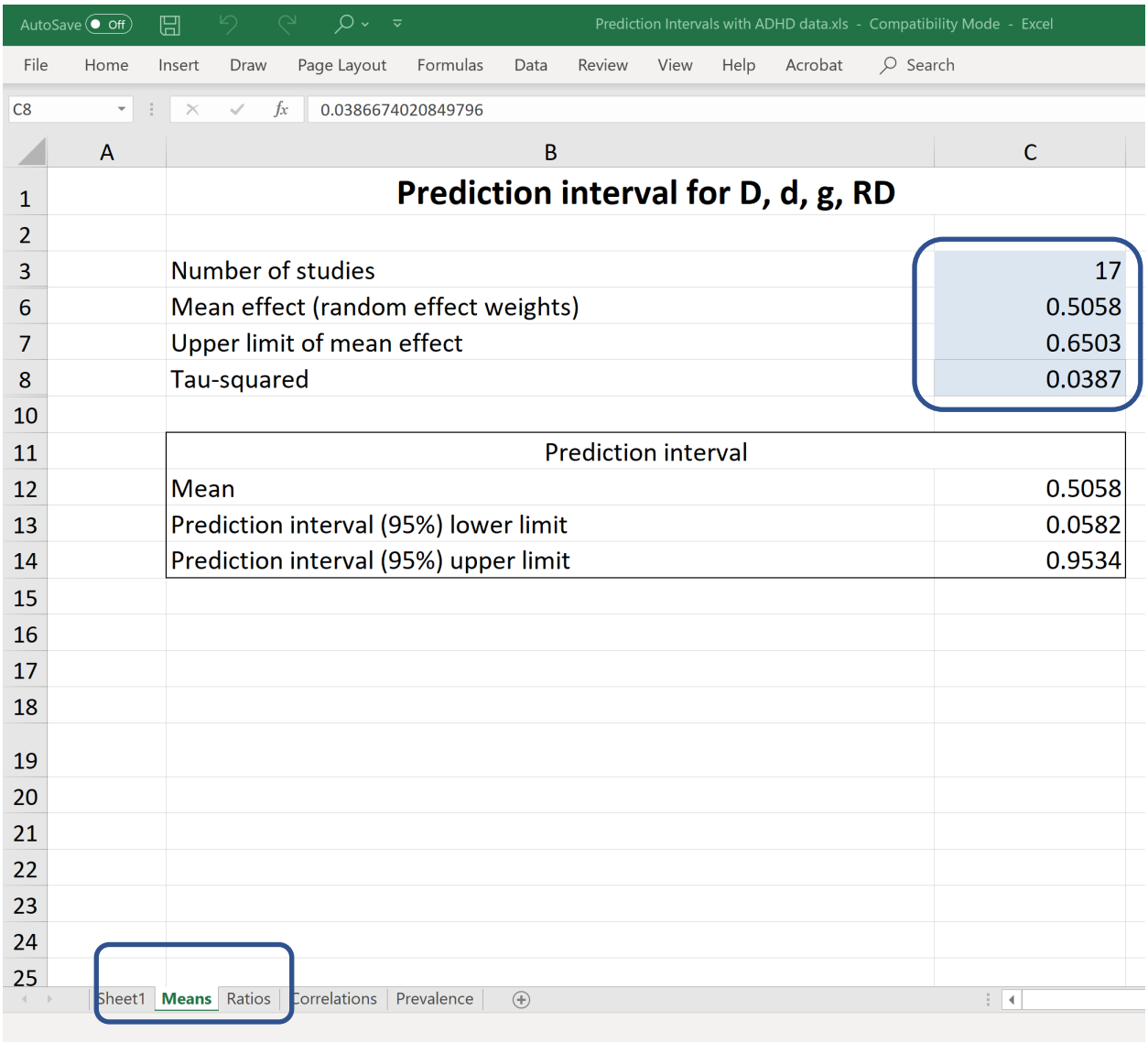
- T^2 is the variance of true effects. As such, it plays the same role as the variance of scores in a primary study. It's an important statistic, but does not directly address the question that we care about.

Finally, we have T, the standard deviation of true effects. This statistic is the one that most directly speaks to the dispersion of effects. If we had a precise estimate of the mean effect size and a precise estimate of T, we could assume that the true effect size in most studies falls within some two T on either side of the mean.

As a practical matter, the estimate of the mean and the estimate of T are subject to error. Therefore, we apply a modified version of this formula to compute the prediction interval. This formula is programmed into the Excel file Prediction_Intervals

Open the Excel file Prediction_Intervals.xls

- At the bottom, select the tab for [Means]
- Copy the four values needed for the blue cells, as shown here (The file saved in the ADHD folder has the data already entered)



The Excel file displays the prediction interval as 0.0582 to 0.9534.

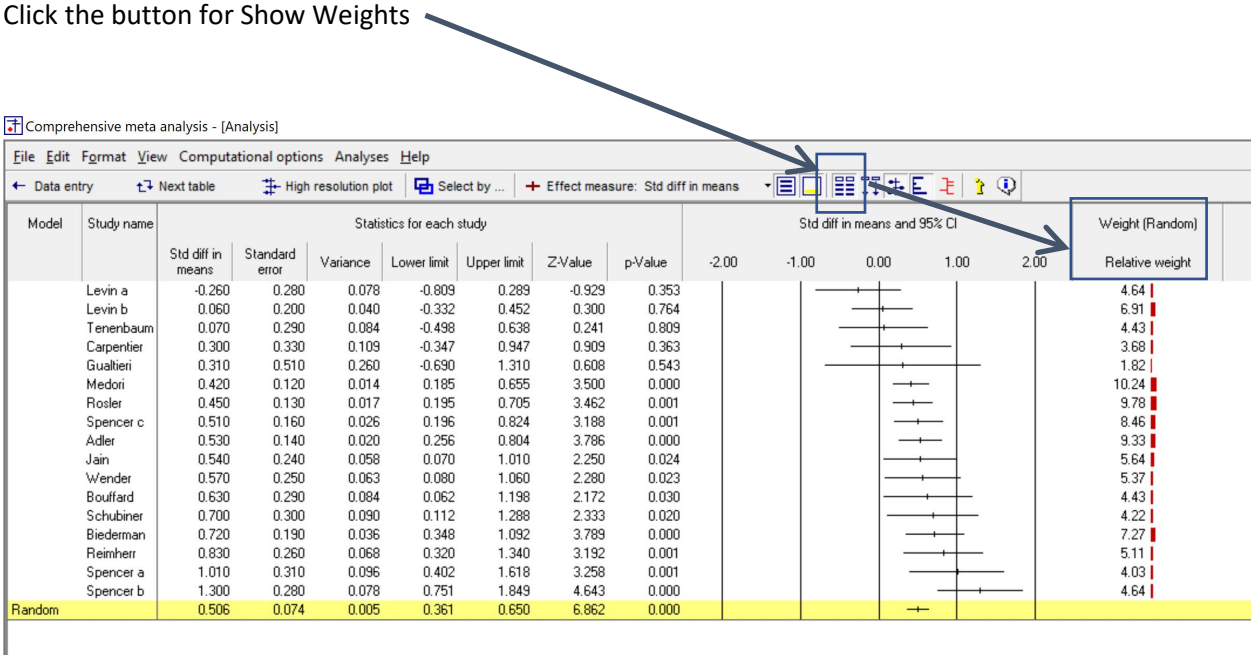
If we can assume that the effects are normally, distributed, then the impact of the intervention will be as low as 0.05 in some populations, to as high as 0.95 in others (in round numbers).

This screen shows where to locate the four values needed. You can type the values into Excel. Or, for more accuracy, use Copy and Paste to copy each to Excel. This will copy all significant digits, and not only the digits that are displayed.

Explore the study weights

We might want to explore the study weights, to see how much weight each study contributed to the mean effect size.

Click the button for Show Weights



Model	Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Weight (Random) Relative weight
	Levin a	-0.260	0.280	0.078	-0.809	0.289	-0.929	0.353	4.64
	Levin b	0.060	0.200	0.040	-0.332	0.452	0.300	0.764	6.91
	Tenenbaum	0.070	0.290	0.084	-0.498	0.638	0.241	0.809	4.43
	Carpentier	0.300	0.330	0.109	-0.347	0.947	0.909	0.363	3.68
	Gualtieri	0.310	0.510	0.260	-0.690	1.310	0.608	0.543	1.82
	Medori	0.420	0.120	0.014	0.185	0.655	3.500	0.000	10.24
	Rosler	0.450	0.130	0.017	0.195	0.705	3.462	0.001	9.78
	Spencer c	0.510	0.160	0.026	0.196	0.824	3.188	0.001	8.46
	Adler	0.530	0.140	0.020	0.256	0.804	3.786	0.000	9.33
	Jain	0.540	0.240	0.058	0.070	1.010	2.250	0.024	5.64
	Wender	0.570	0.250	0.063	0.080	1.060	2.280	0.023	5.37
	Bouffard	0.630	0.290	0.084	0.062	1.198	2.172	0.030	4.43
	Schubiner	0.700	0.300	0.090	0.112	1.288	2.333	0.020	4.22
	Biederman	0.720	0.190	0.036	0.348	1.092	3.789	0.000	7.27
	Reimherr	0.830	0.260	0.068	0.320	1.340	3.192	0.001	5.11
	Spencer a	1.010	0.310	0.096	0.402	1.618	3.258	0.001	4.03
	Spencer b	1.300	0.280	0.078	0.751	1.849	4.643	0.000	4.64
Random		0.506	0.074	0.005	0.361	0.650	6.862	0.000	

Std diff in means and 95% CI: -2.00 -1.00 0.00 1.00 2.00

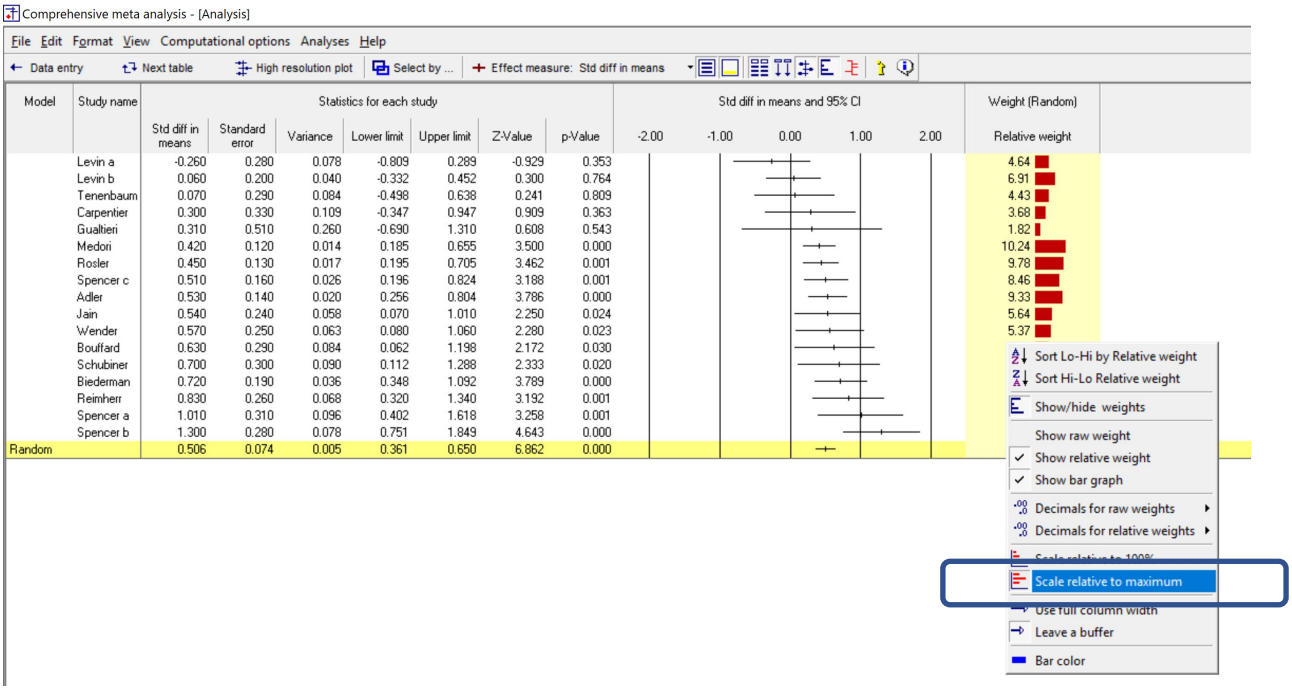
The program displays the relative weight assigned to each study, along with a bar graph on a scale of 0 to 1 corresponding to the percentage of weight (these sum to 1.0).

Right-click on the column of weights to customize the display.

- You may elect to display the raw weight rather than the relative weight.
- You may elect to scale relative to the maximum weight as shown here. The scale will then go (in this case) from 0 to 0.07 rather than 0 to 1, which makes it easier to distinguish among the study weights.

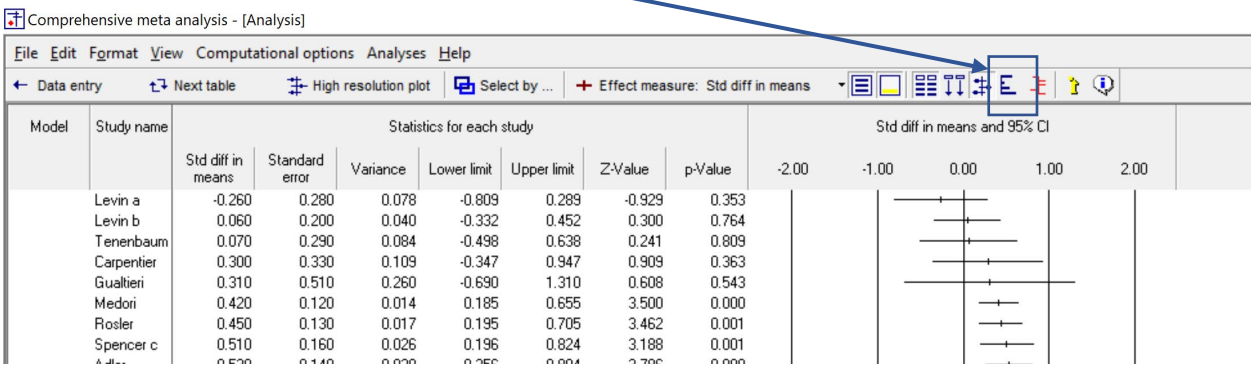
The Excel file displays the prediction interval as 0.0582 to 0.9534.

If we can assume that the effects are normally, distributed, then the impact of the intervention will be as low as 0.05 in some populations, to as high as 0.95 in others (in round numbers).

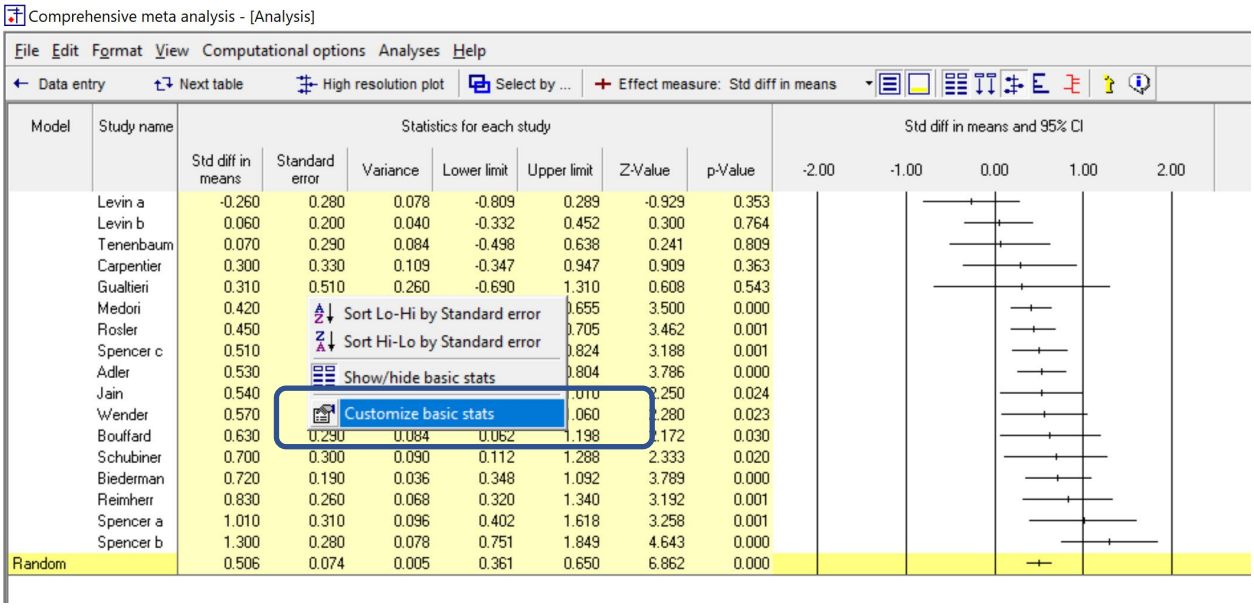


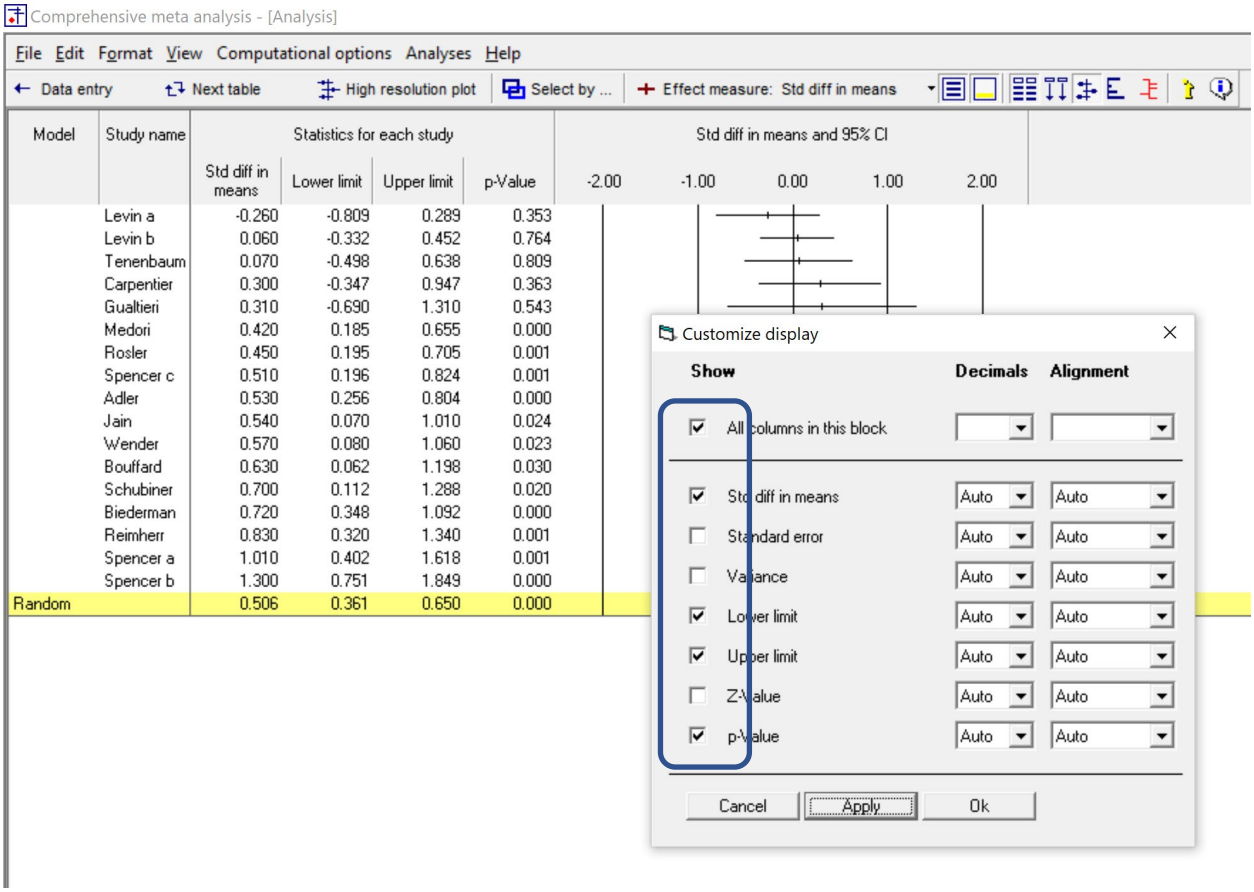
We want to create a high-resolution plot

The basic idea for the plot is to only keep the columns that we really need. If we can remove other columns, the plot will be more useful. On the high-resolution plot we will be using the size of each box to reflect the study weights, so we don't need the column of study weights. Click here to hide the column of weights



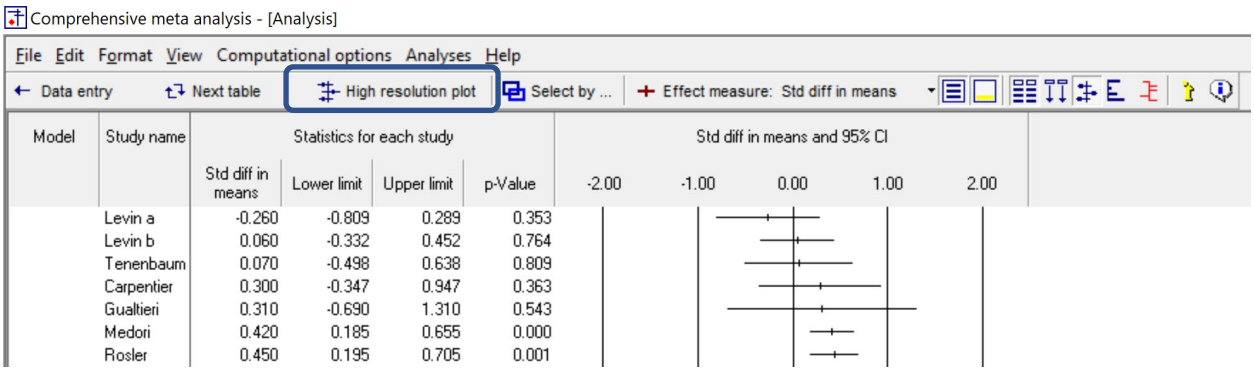
Right-click on the statistics, then select Customize basic stats. This will allow us to choose which columns to include in the high-resolution plot





I've chosen to keep the mean, lower and upper limit, and p-value. I've removed the standard error, variance, and Z-value. This will make the plot more readable.

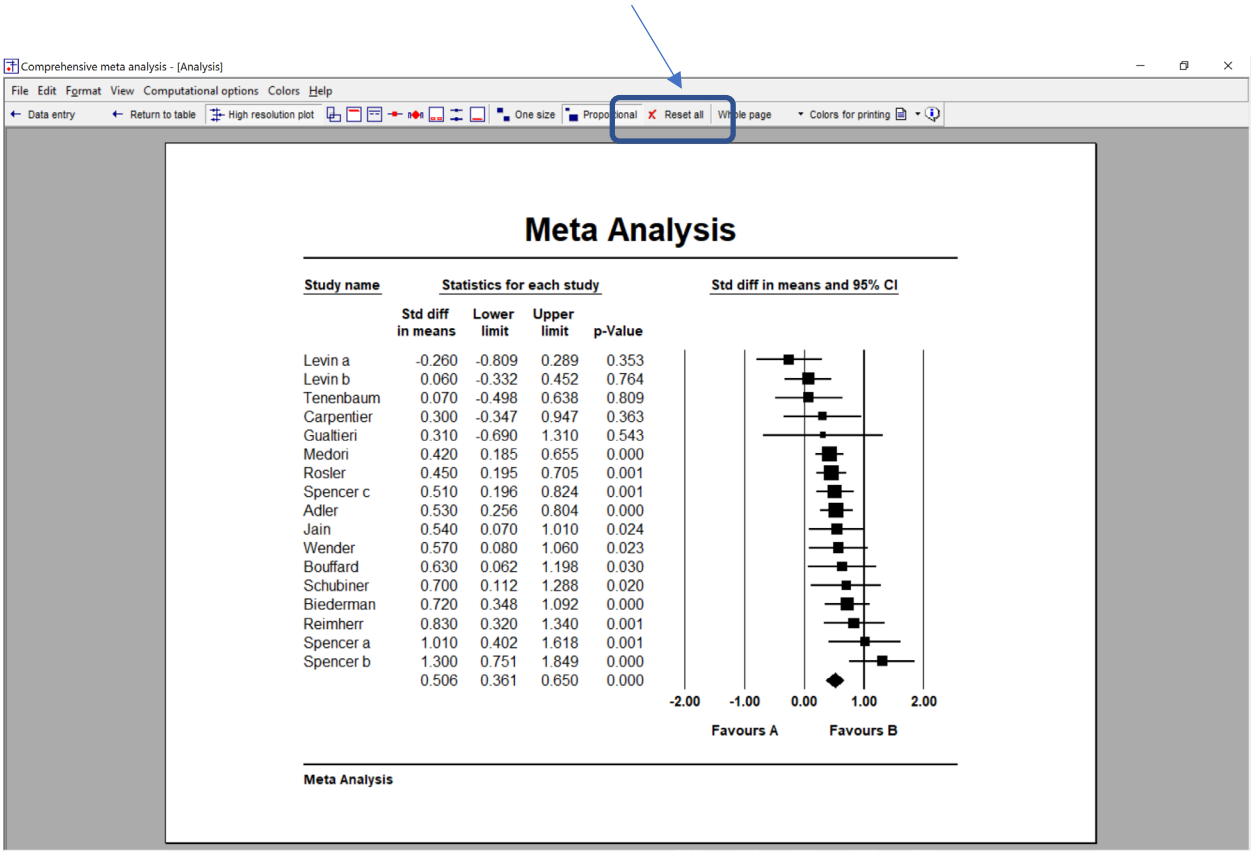
Click <High resolution plot>



The program displays the plot.

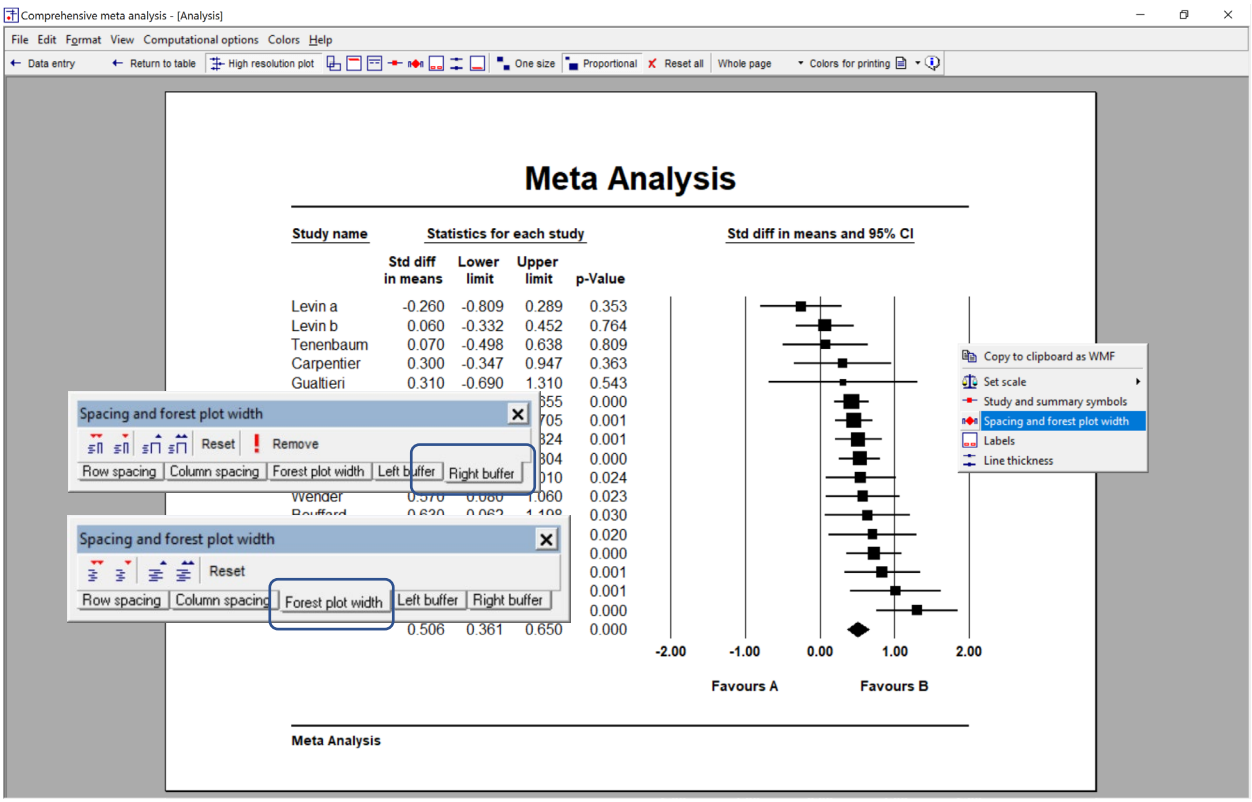
Initialize the plot

Click Reset All to set all the options to their defaults



Modify the spacing

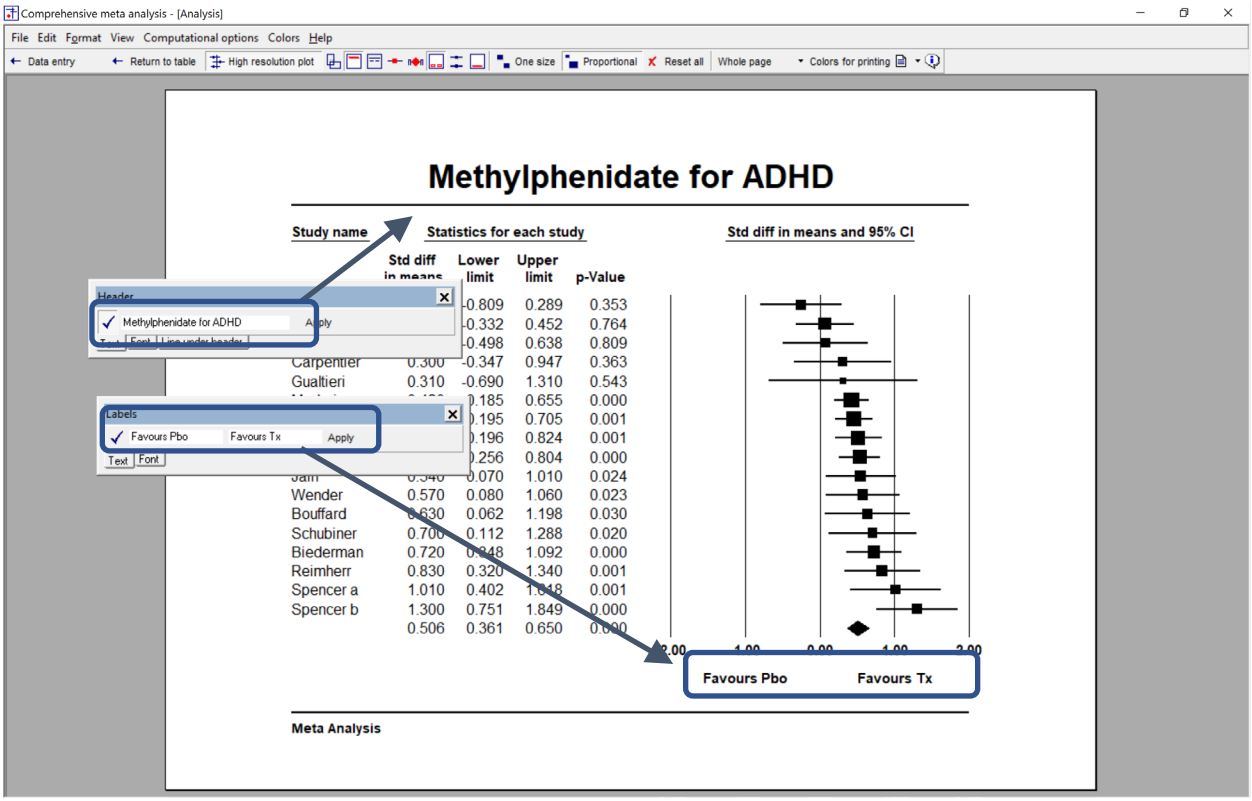
- Right-click on the forest plot
- Select Spacing and forest plot width
- Select Right buffer > Remove. This will remove space to the right of the plot
- Select Forest plot width > [UP] [UP] [UP] [UP]



Edit the labels and Title

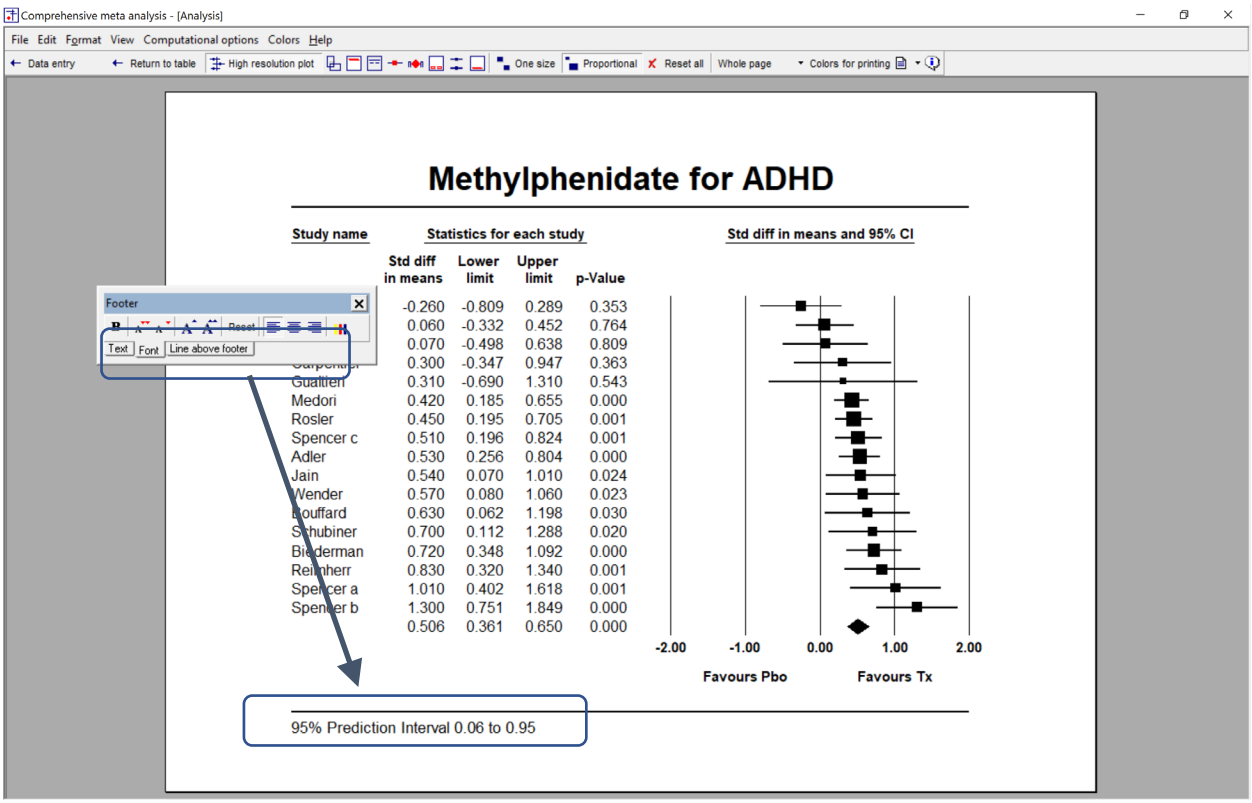
- Right-click on the Title (or click Edit > Header on the menu).
- Enter a new title
- Click Font and increase the font size by one <Up>

- Right-click on the Labels (or click Edit > Labels on the menu).
- Enter new labels
- Click Font and increase the font size by one <Up>



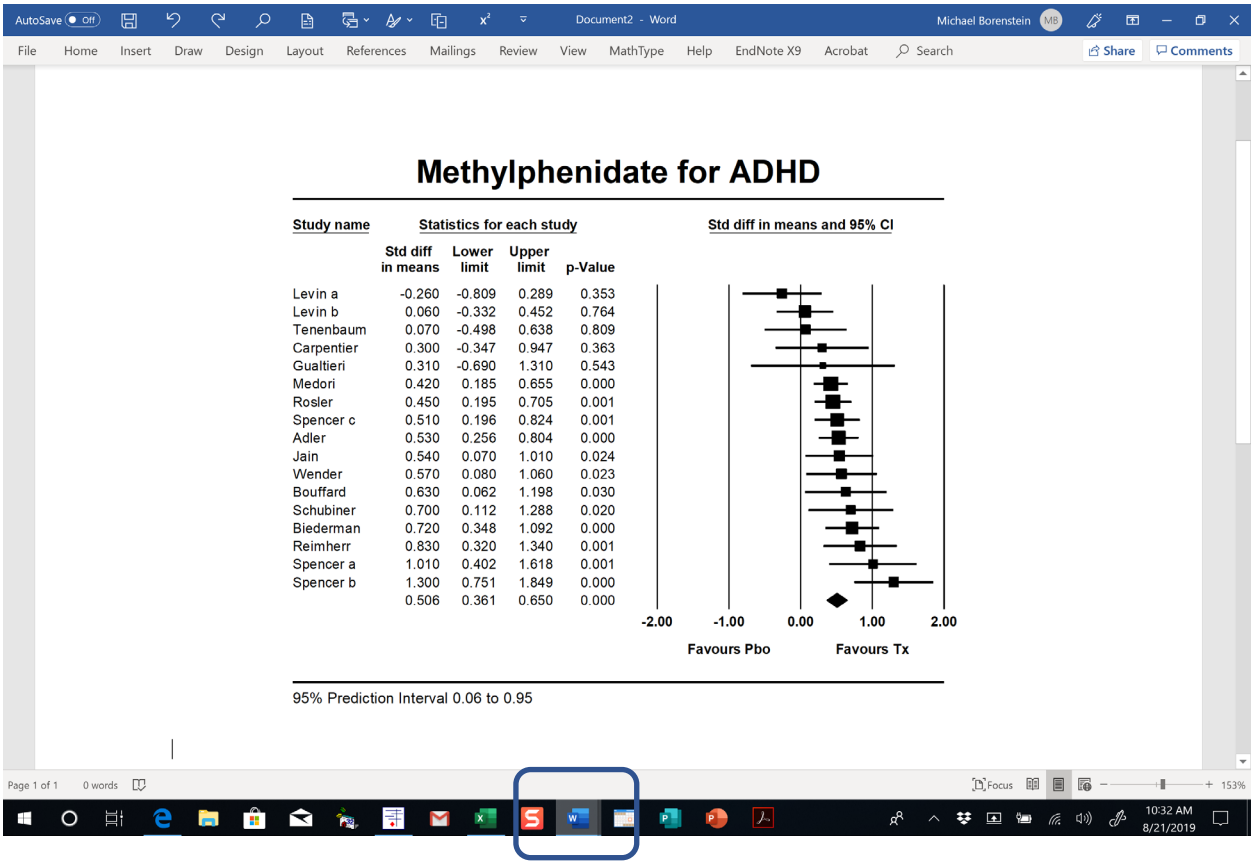
Edit the footer

- Right-click on the footer (or click Edit > Footer on the menu)
- Change the text to (95% prediction interval 0.06 to 0.95)



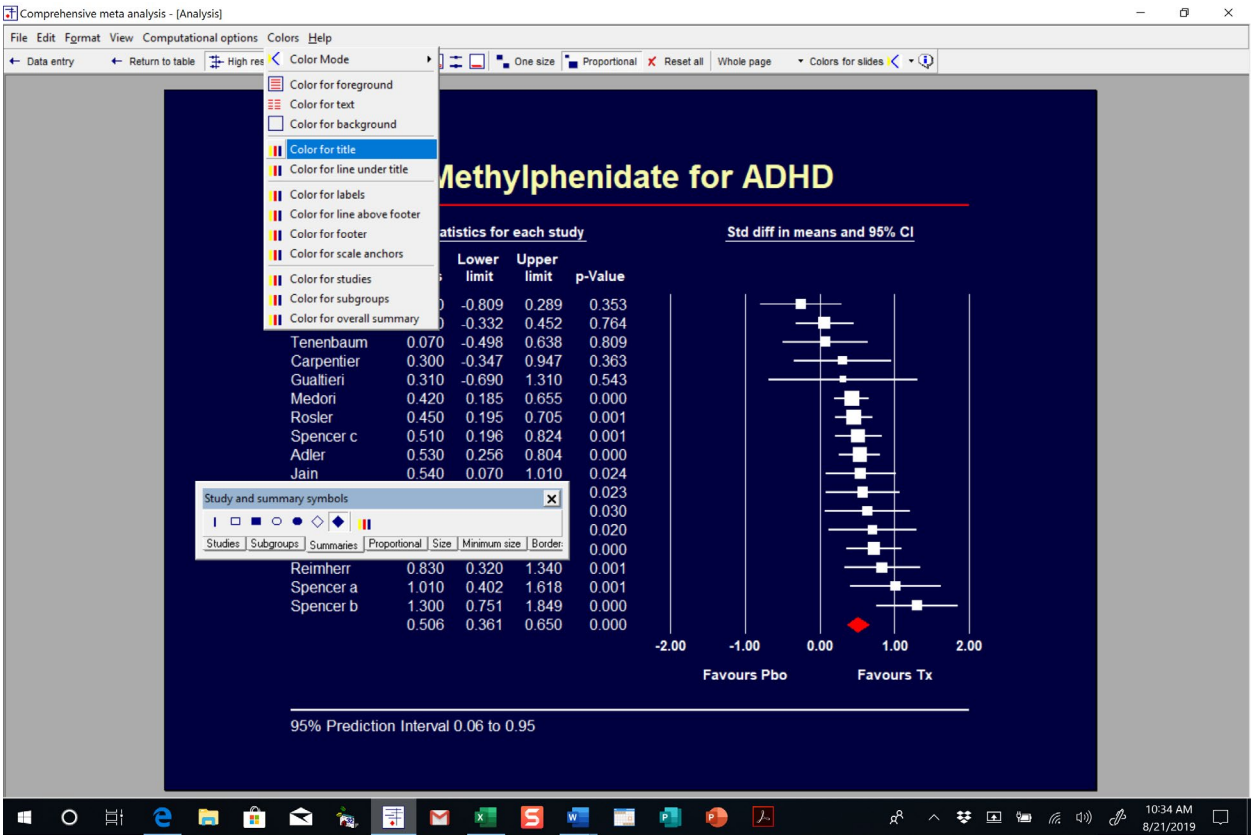
Export to Word

Click File > Export to Word
The program created a Word file and inserts the plot
Look for the Word icon on your taskbar

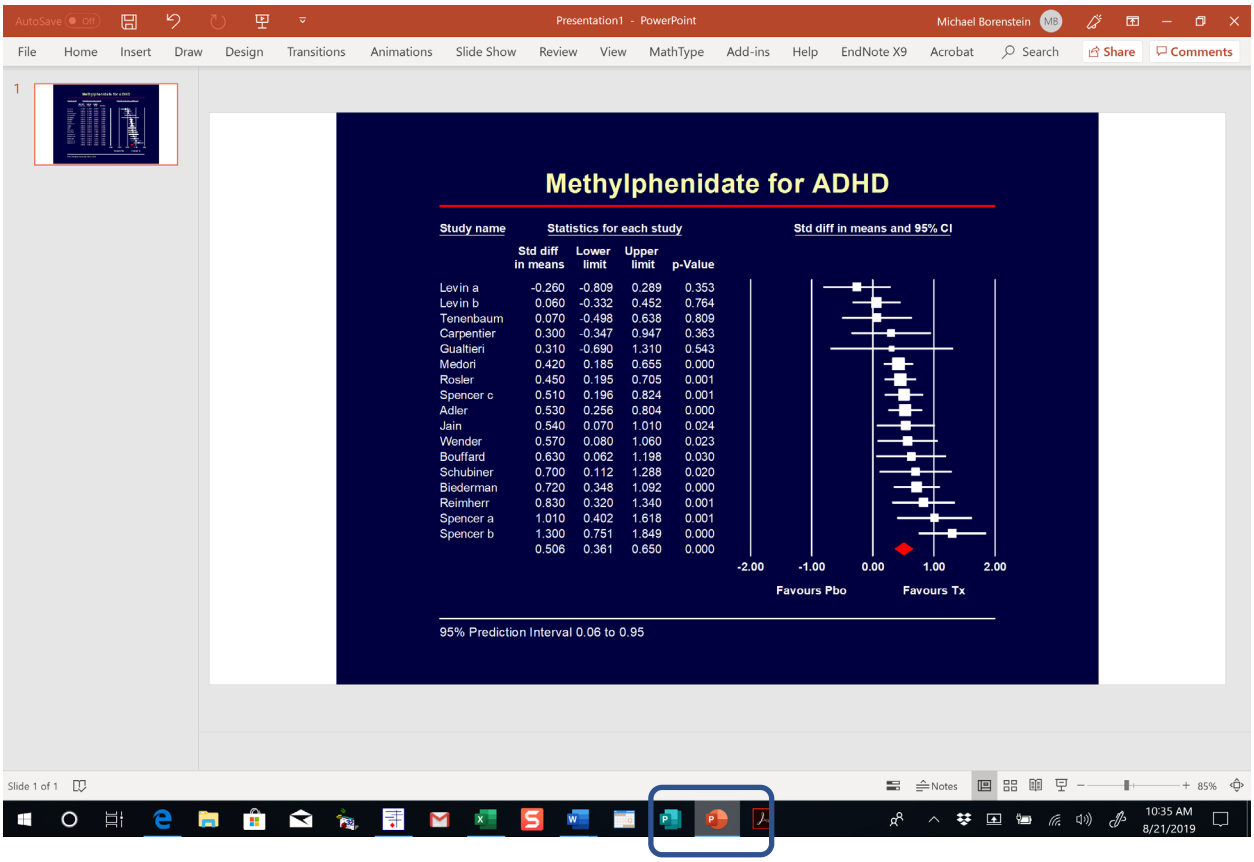


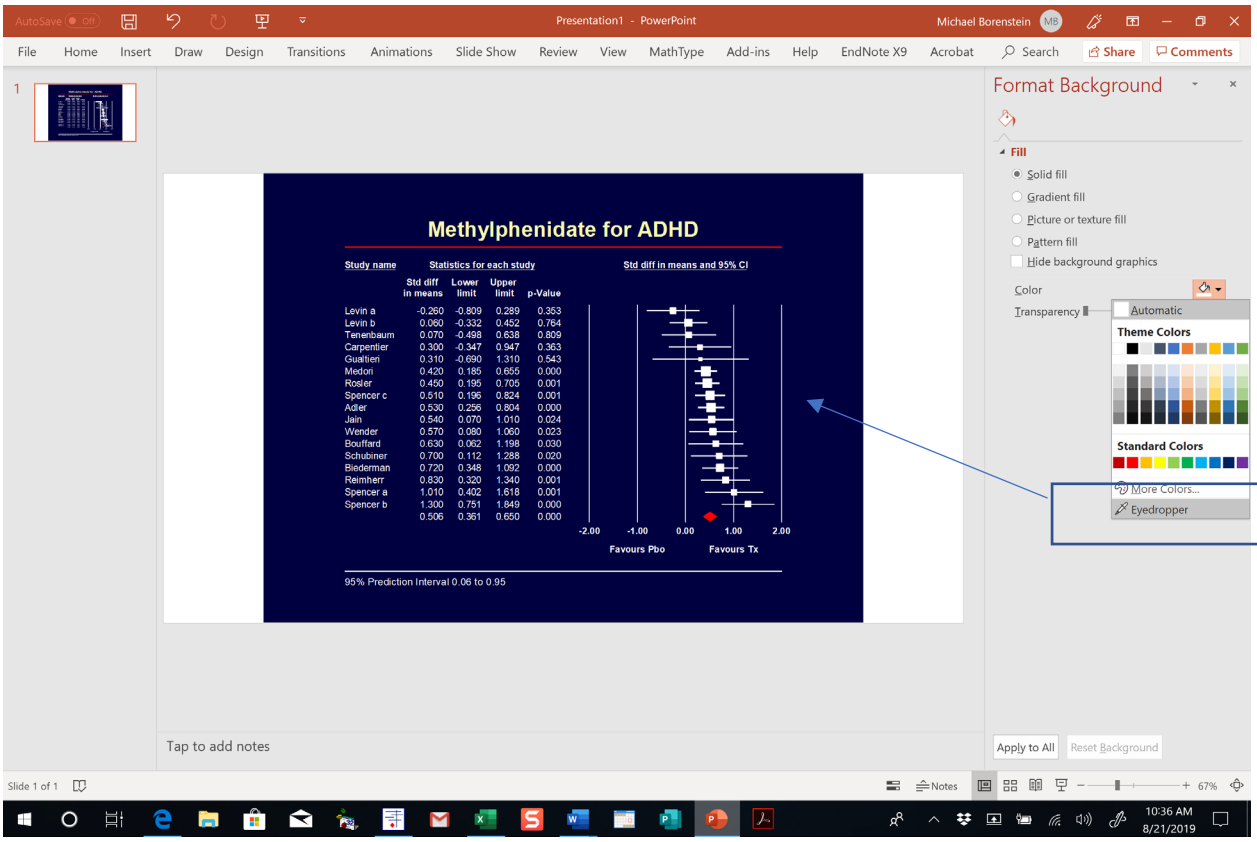
Export to PowerPoint

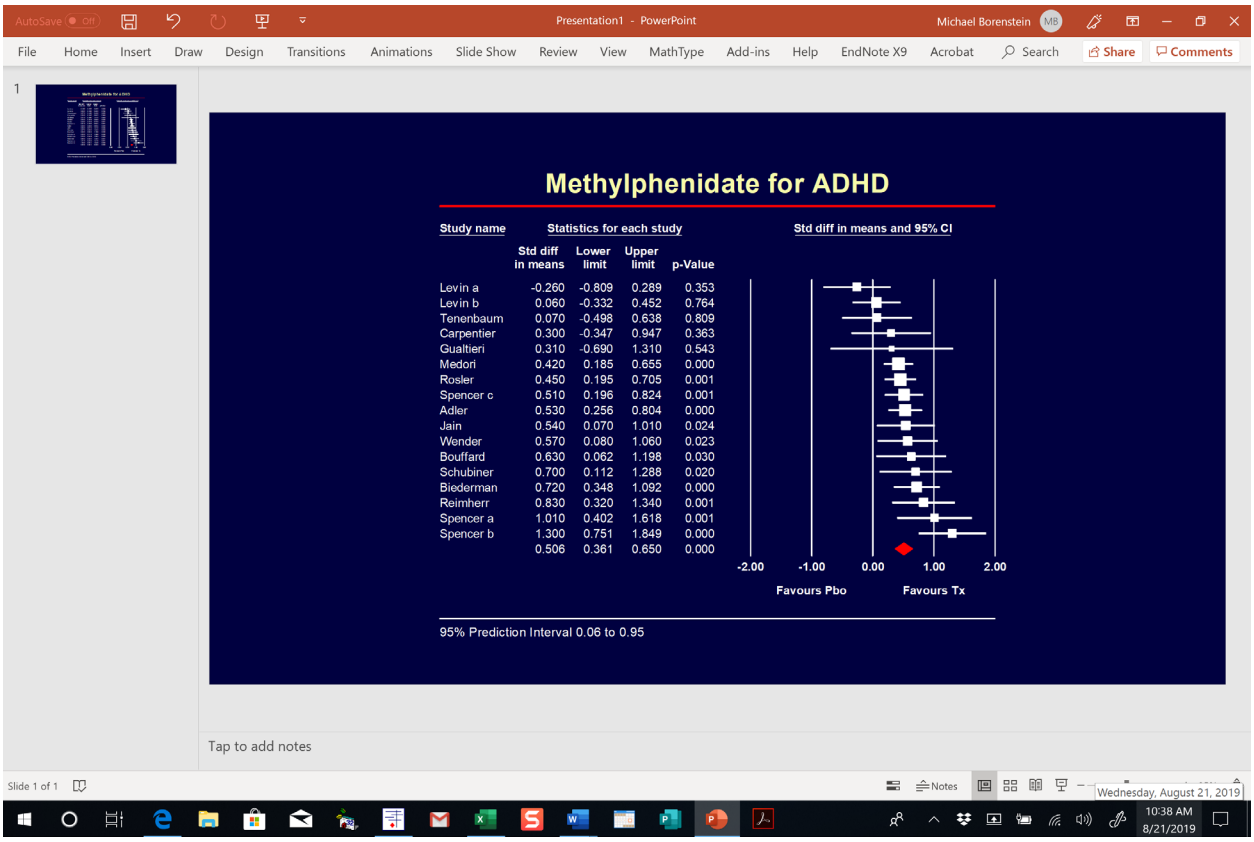
- If you're in Word, return to CMA
- Select colors for slides
- Click File > Export to PowerPoint



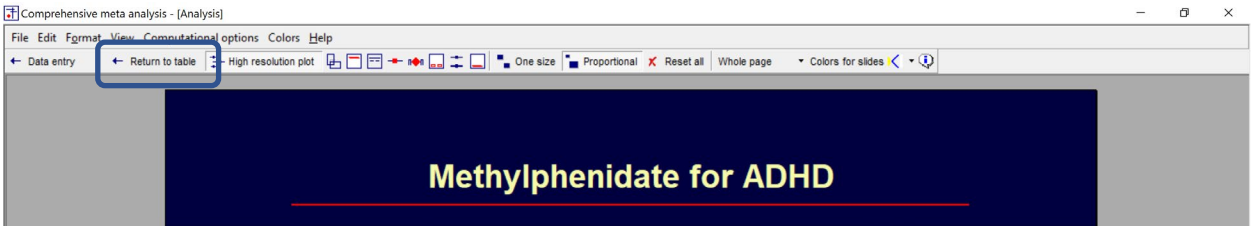
To modify the colors for any element of the plot, click Colors on the menu


Presentation1 - PowerPoint
Michael Borenstein
File Home Insert Draw Design Transitions Animations Slide Show Review View MathType Add-ins Help EndNote X9 Acrobat Search Share Comments
Methylphenidate for ADHD
Study name
Statistics for each study
Std diff in means and 95% CI
Std diff in means
Lower limit
Upper limit
p-Value
Levin a -0.260 -0.809 0.289 0.353
Levin b 0.060 -0.332 0.452 0.764
Tenenbaum 0.070 -0.498 0.638 0.809
Carpentier 0.300 -0.347 0.947 0.363
Gualtieri 0.310 -0.690 1.310 0.543
Medori 0.420 0.185 0.655 0.000
Rosler 0.450 0.195 0.705 0.001
Spencer c 0.510 0.196 0.824 0.001
Adler 0.530 0.256 0.804 0.000
Jain 0.540 0.070 1.010 0.024
Wender 0.570 0.080 1.060 0.023
Bouffard 0.630 0.062 1.198 0.030
Schubiner 0.700 0.112 1.288 0.020
Biederman 0.720 0.348 1.092 0.000
Reimherr 0.830 0.320 1.340 0.001
Spencer a 1.010 0.402 1.618 0.001
Spencer b 1.300 0.751 1.849 0.000
0.506 0.361 0.650 0.000
-2.00 -1.00 0.00 1.00 2.00
Favours Pbo
Favours Tx
95% Prediction Interval 0.06 to 0.95
Slide 1 of 1
Notes
85%
10:35 AM
8/21/2019


Presentation1 - PowerPoint
Michael Borenstein
File Home Insert Draw Design Transitions Animations Slide Show Review View MathType Add-ins Help EndNote X9 Acrobat Search Share Comments
Format Background
Fill
Solid fill
Gradient fill
Picture or texture fill
Pattern fill
Hide background graphics
Color
Transparency
Automatic
Theme Colors
Standard Colors
More Colors...
Eyedropper
Apply to All
Reset Background
Methylphenidate for ADHD
Study name
Statistics for each study
Std diff in means and 95% CI
Std diff in means
Lower limit
Upper limit
p-Value
Levin a -0.260 -0.809 0.289 0.353
Levin b 0.060 -0.332 0.452 0.764
Tenenbaum 0.070 -0.498 0.638 0.809
Carpentier 0.300 -0.347 0.947 0.363
Gualtieri 0.310 -0.690 1.310 0.543
Medori 0.420 0.185 0.655 0.000
Rosler 0.450 0.195 0.705 0.001
Spencer c 0.510 0.196 0.824 0.001
Adler 0.530 0.256 0.804 0.000
Jain 0.540 0.070 1.010 0.024
Wender 0.570 0.080 1.060 0.023
Bouffard 0.630 0.062 1.198 0.030
Schubiner 0.700 0.112 1.288 0.020
Biederman 0.720 0.348 1.092 0.000
Reimherr 0.830 0.320 1.340 0.001
Spencer a 1.010 0.402 1.618 0.001
Spencer b 1.300 0.751 1.849 0.000
0.506 0.361 0.650 0.000
-2.00 -1.00 0.00 1.00 2.00
Favours Pbo
Favours Tx
95% Prediction Interval 0.06 to 0.95
Tap to add notes
Slide 1 of 1
Notes
67%
10:36 AM
8/21/2019



Return to CMA
Click Return to table



Subgroups

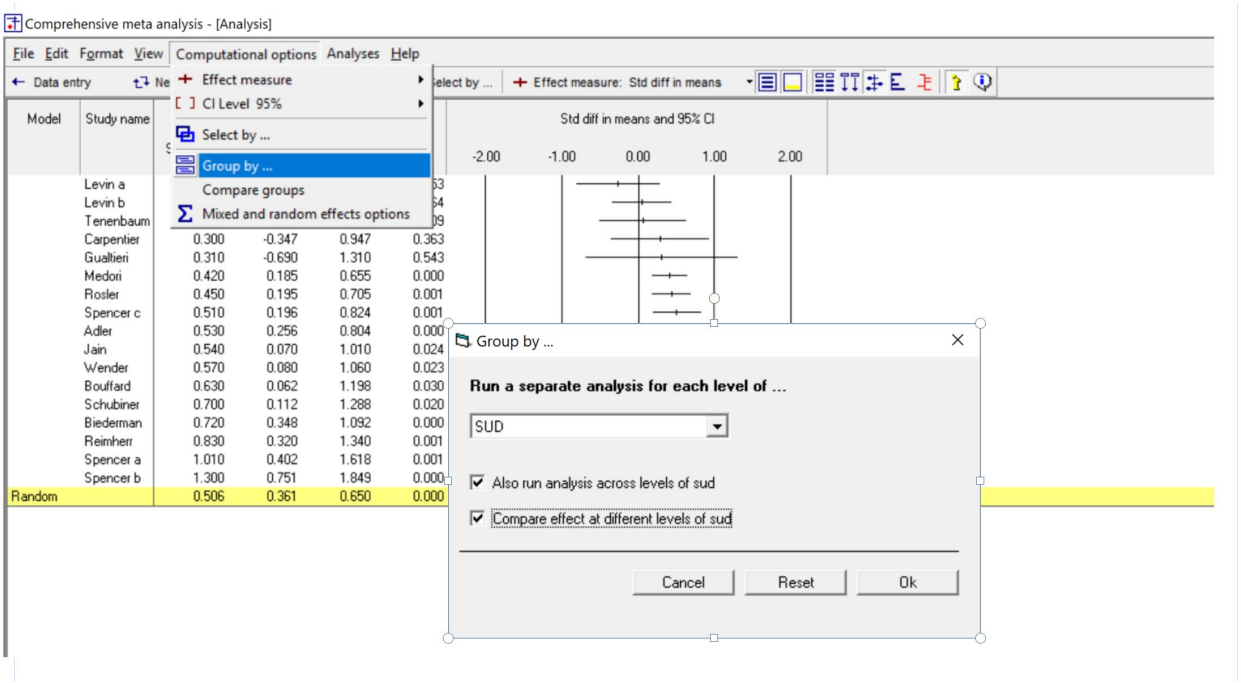
To this point we have established that the impact of methyphenidate is relatively weak in some studies and substantially stronger in others.

Our next goal is to see if we can identify the moderators that are associated with this variation in effect size. Specifically, we will look at SUD and Dose.

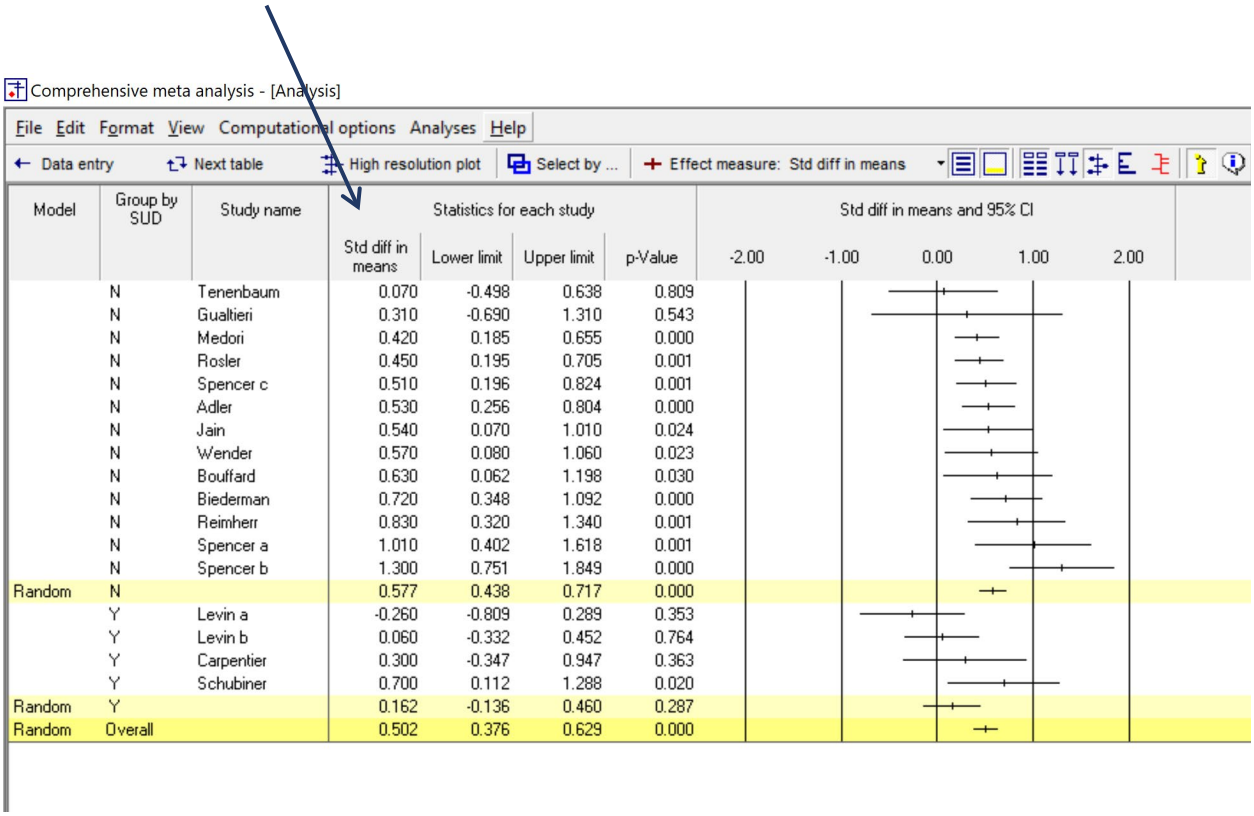
Recall that when we entered the data, we included a column for SUD (substance abuse disorder), to identify studies as including patients who abused drugs (Y) vs. studies that excluded these patients (N).

To address this question we will run the analysis separately for each subgroup of studies.

- Click Computational options > Group By >
- Select SUD
- Check the two boxes on the wizard

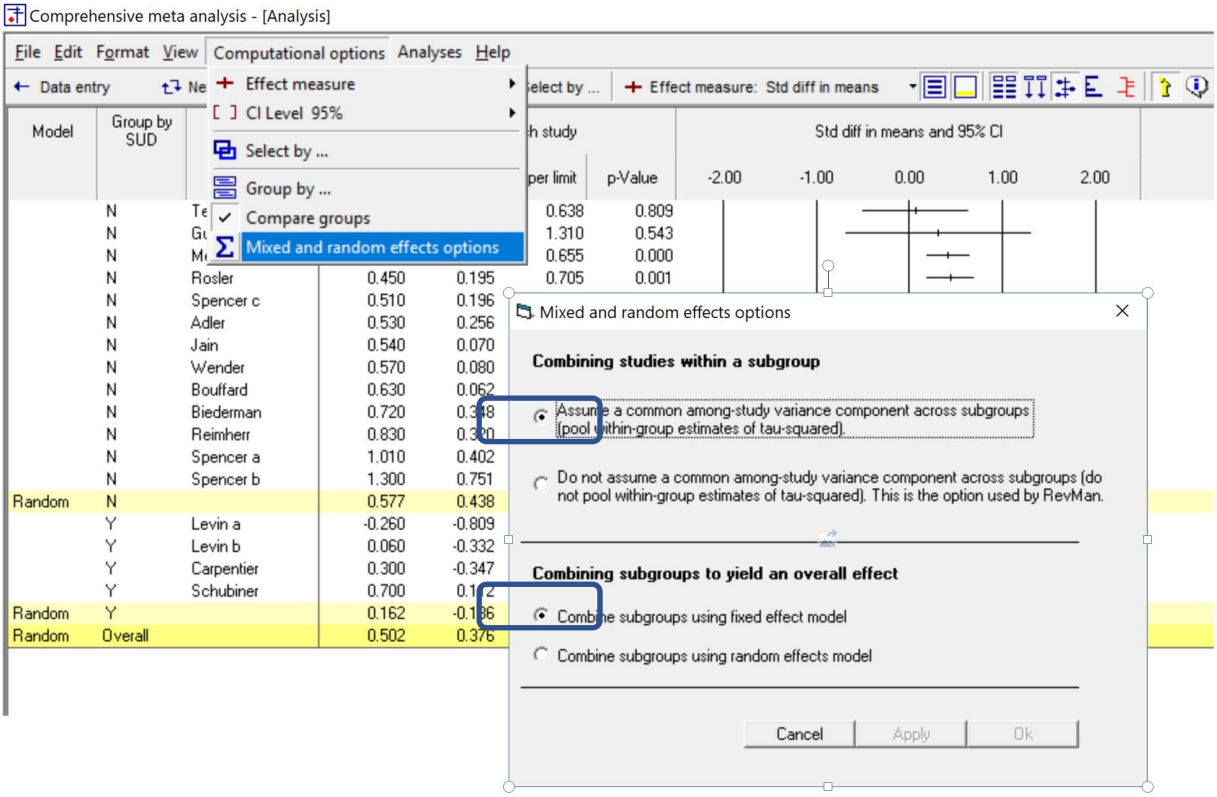


Right-click on Std Diff in Means and Sort by effect size



Model	Group by SUD	Study name	Std diff in means	Lower limit	Upper limit	p-Value
	N	Tenenbaum	0.070	-0.498	0.638	0.809
	N	Gualtieri	0.310	-0.690	1.310	0.543
	N	Medori	0.420	0.185	0.655	0.000
	N	Rosler	0.450	0.195	0.705	0.001
	N	Spencer c	0.510	0.196	0.824	0.001
	N	Adler	0.530	0.256	0.804	0.000
	N	Jain	0.540	0.070	1.010	0.024
	N	Wender	0.570	0.080	1.060	0.023
	N	Bouffard	0.630	0.062	1.198	0.030
	N	Biederman	0.720	0.348	1.092	0.000
	N	Reimherr	0.830	0.320	1.340	0.001
	N	Spencer a	1.010	0.402	1.618	0.001
	N	Spencer b	1.300	0.751	1.849	0.000
Random	N		0.577	0.438	0.717	0.000
	Y	Levin a	-0.260	-0.809	0.289	0.353
	Y	Levin b	0.060	-0.332	0.452	0.764
	Y	Carpentier	0.300	-0.347	0.947	0.363
	Y	Schubiner	0.700	0.112	1.288	0.020
Random	Y		0.162	-0.136	0.460	0.287
Random	Overall		0.502	0.376	0.629	0.000

Click Computational Options > Mixed and Random effects options



This opens a dialog box

At the top, click "Assume a common among-study variance component across subgroups (pool within-group estimates of tau-squared).

The issue here is as follows. We estimate T^2 separately for each subgroup, so we get one estimate for SUD | N and another for SUD | Y. Then we have the option to apply each estimate to its own subgroup, or to pool the estimates and use the pooled value for both subgroups. In general, it's a good idea to pool. The reason is that unless we have a decent amount of studies within each subgroup (there is no consensus, but I would use 20 as a number), we gain more from pooling (and working with more data) than we lose by combining data from subgroups where the true value of T^2 is probably not the same.

At the bottom, click "Combine subgroups using the fixed-effect model". Use this option even if (as you should) you are using the random-effects model in the analysis.

The issue here is as follows. When we group by subgroups, there are two levels of sampling – we need to select subgroups, and we need to select studies within each subgroup.

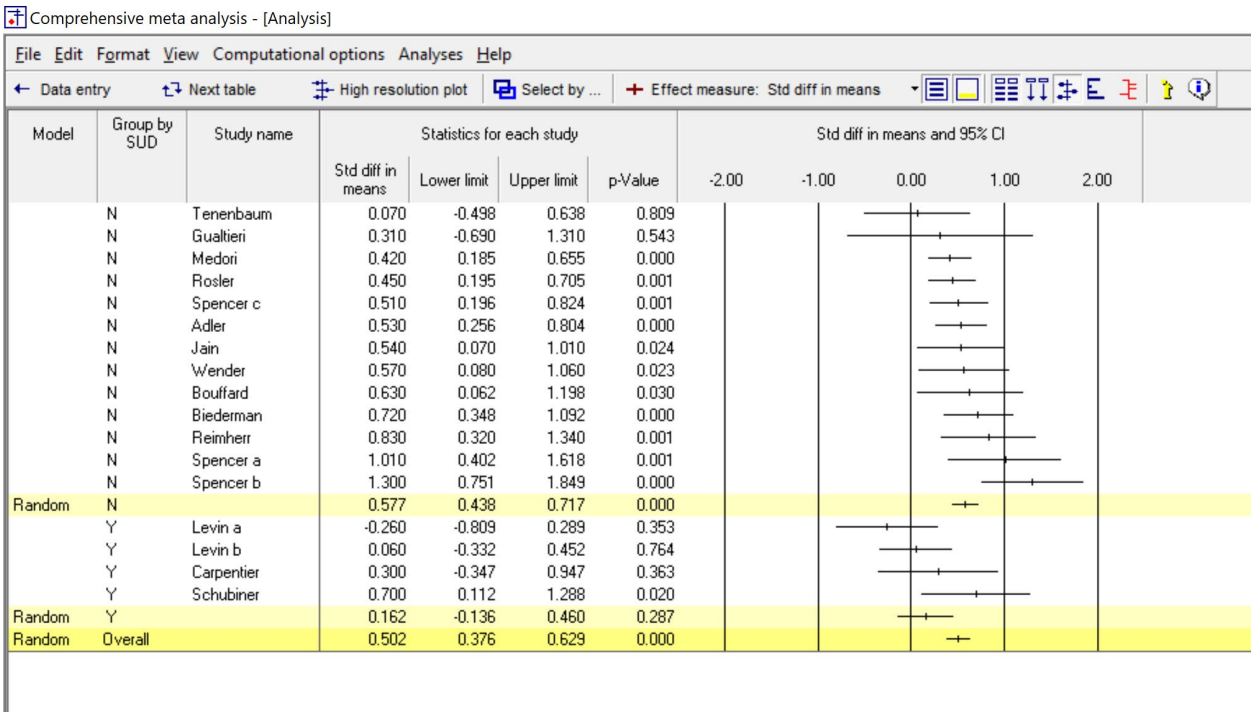
The subgroups that we choose are "Fixed" – we care specifically about SUD| N and SUD | Y. We are making an inference about these two subgroups, and will not be extrapolating from them to a larger set of subgroups. The option here is to combine SUBGROUPS using the fixed-effect model (note that the term fixed here means that these are chosen, not that they are the same).

Within the subgroups, the studies are sampled at random. We happen to have specific studies within each subgroup, but we don't care specifically about these studies – they are being used to estimate the mean for all SUD |N studies and all the mean for all SUD | Y studies. For this reason we select the Random tab at the bottom of the screen.

To reiterate – the tab at the bottom of the screen is Random since the studies within a subgroup are assumed to be a random sample of all comparable studies. The option in the bottom section of the wizard is Fixed since we are making an inference to these two subgroups only, and not extrapolating from these to other subgroups.

	Fixed (Chosen by name)	Random (Sampled randomly)	
Subgroups	Yes		In Wizard choose Fixed
Studies within subgroups		Yes	On plot, Random tab

Since we are using the Fixed-effect model at one level, and the random-effects model at another level, this is called a Mixed-effect model.



Model	Group by SUD	Study name	Std diff in means	Lower limit	Upper limit	p-Value
	N	Tenenbaum	0.070	-0.498	0.638	0.809
	N	Gualtieri	0.310	-0.690	1.310	0.543
	N	Medori	0.420	0.185	0.655	0.000
	N	Rosler	0.450	0.195	0.705	0.001
	N	Spencer c	0.510	0.196	0.824	0.001
	N	Adler	0.530	0.256	0.804	0.000
	N	Jain	0.540	0.070	1.010	0.024
	N	Wender	0.570	0.080	1.060	0.023
	N	Bouffard	0.630	0.062	1.198	0.030
	N	Biederman	0.720	0.348	1.092	0.000
	N	Reimherr	0.830	0.320	1.340	0.001
	N	Spencer a	1.010	0.402	1.618	0.001
	N	Spencer b	1.300	0.751	1.849	0.000
Random	N		0.577	0.438	0.717	0.000
	Y	Levin a	-0.260	-0.809	0.289	0.353
	Y	Levin b	0.060	-0.332	0.452	0.764
	Y	Carpentier	0.300	-0.347	0.947	0.363
	Y	Schubiner	0.700	0.112	1.288	0.020
Random	Y		0.162	-0.136	0.460	0.287
Random	Overall		0.502	0.376	0.629	0.000

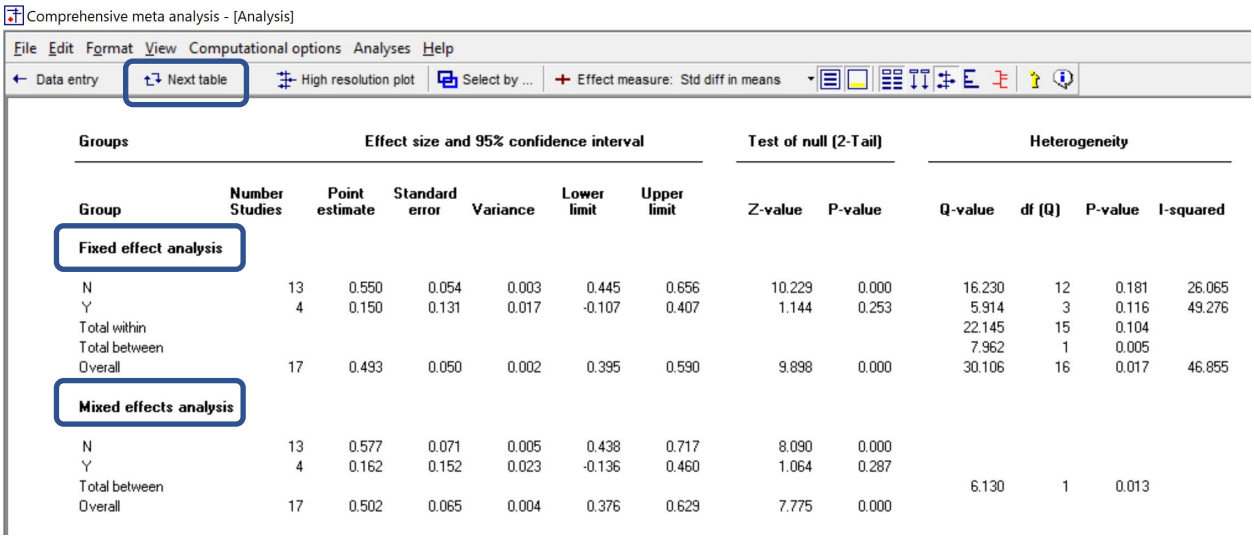
The program displays the analysis for the studies that excluded drug abusers at the top, and for those that included drug abusers at the bottom.

For studies that excluded drug abusers the mean effect size is 0.577 with a 95% confidence interval of 0.438 to 0.717. A test of the null hypothesis that the mean effect is zero yields a p-value of < 0.001.

For studies that excluded drug abusers the mean effect size is 0.162 with a 95% confidence interval of −0.136 to 0.460. A test of the null hypothesis that the mean effect is zero yields a p-value of 0.287.

There us also a line that shows results based on all studies. The mean effect size is 0.502 with a 95% confidence interval of 0.376 to 0.629. A test of the null hypothesis that the mean effect is zero yields a p-value of < 0.001.

Click on Next table



Groups		Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity			
Group	Number Studies	Point estimate	Standard error	Variance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed effect analysis												
N	13	0.550	0.054	0.003	0.445	0.656	10.229	0.000	16.230	12	0.181	26.065
Y	4	0.150	0.131	0.017	-0.107	0.407	1.144	0.253	5.914	3	0.116	49.276
Total within									22.145	15	0.104	
Total between									7.962	1	0.005	
Overall	17	0.493	0.050	0.002	0.395	0.590	9.898	0.000	30.106	16	0.017	46.855
Mixed effects analysis												
N	13	0.577	0.071	0.005	0.438	0.717	8.090	0.000				
Y	4	0.162	0.152	0.023	-0.136	0.460	1.064	0.287				
Total between									6.130	1	0.013	
Overall	17	0.502	0.065	0.004	0.376	0.629	7.775	0.000				

The top of this screen shows the analysis based on a fixed-effect model (fixed at both levels). This is what we would use if all studies within a subgroup were drawn from the same population

The bottom of the screen shows the analysis based on a mixed-effect model. The subgroups are fixed (we chose these two study designs specifically) but within each subgroup, the studies are selected at random. This is the section we will use.

The mean for the two subgroups are 0.577 vs 0.162. We can test the difference for statistical significance using a Q test with 1 degree of freedom. The Q value of 6.130 with 1 degree of freedom yields a p-value of 0.013. We reject the null hypothesis that the mean effect is the same in both subgroups, and conclude that methylphenidate has more of an effect in studies that exclude SUD patients than it does in studies that include SUD patients.

It is important to understand that this finding is observational, not causal. That is,

A. We can conclude that the treatment is more effective in studies that exclude SUD patients.
B. We cannot conclude that the drug is more effective in non-SUD patients.

While it might appear that A and B are the same, there is an important difference between them. We do know that the drug is more effective in the studies that excluded SUD patients. This may be because the drug is more effective in patients who are non-SUD. But is also may be because the studies that excluded these patients also tended to use a lower dose of the drug (for example). It could be that the drug was less effective in these studies because of the lower dose, and not because of the SUD status.

Meta-regression

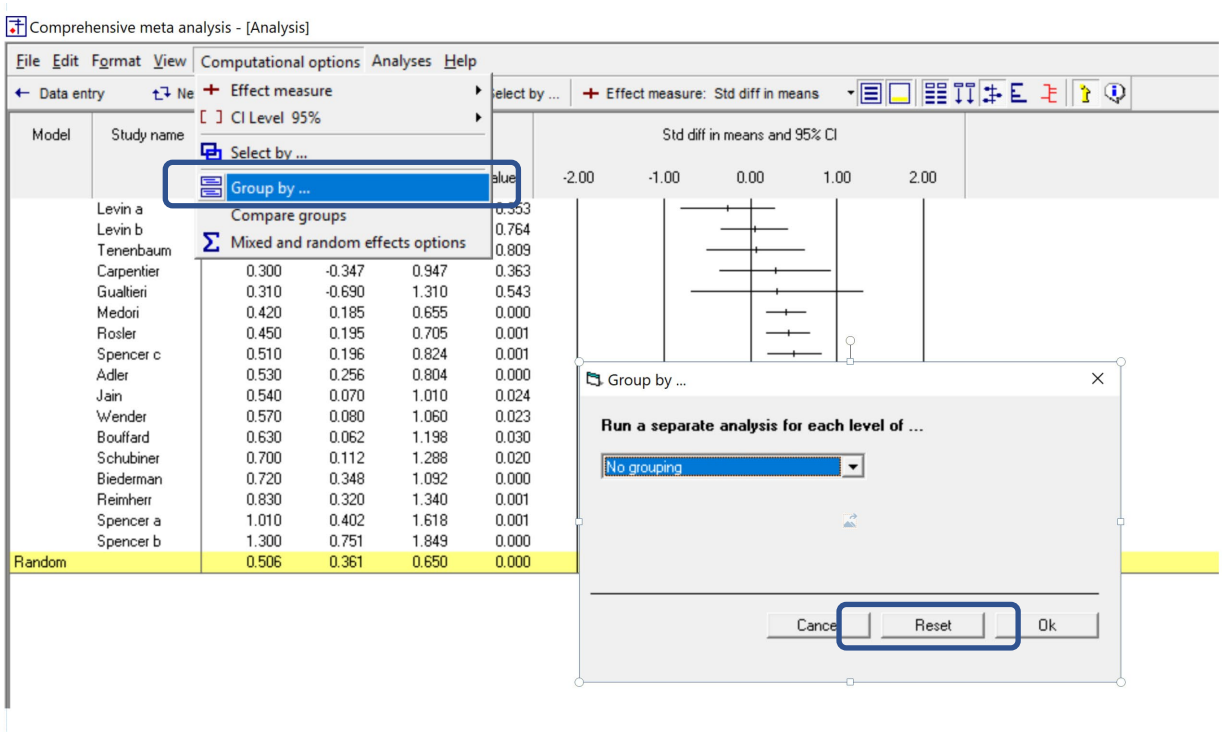
When we looked for a relationship between SUD and effect we used a subgroups analysis. This works when the moderator is categorical (with discrete categories such as Yes and No).

Now, we want to look for a relationship between Dose and effect. Since Dose is a continuous variable, we will use Meta-regression rather than subgroups analysis.

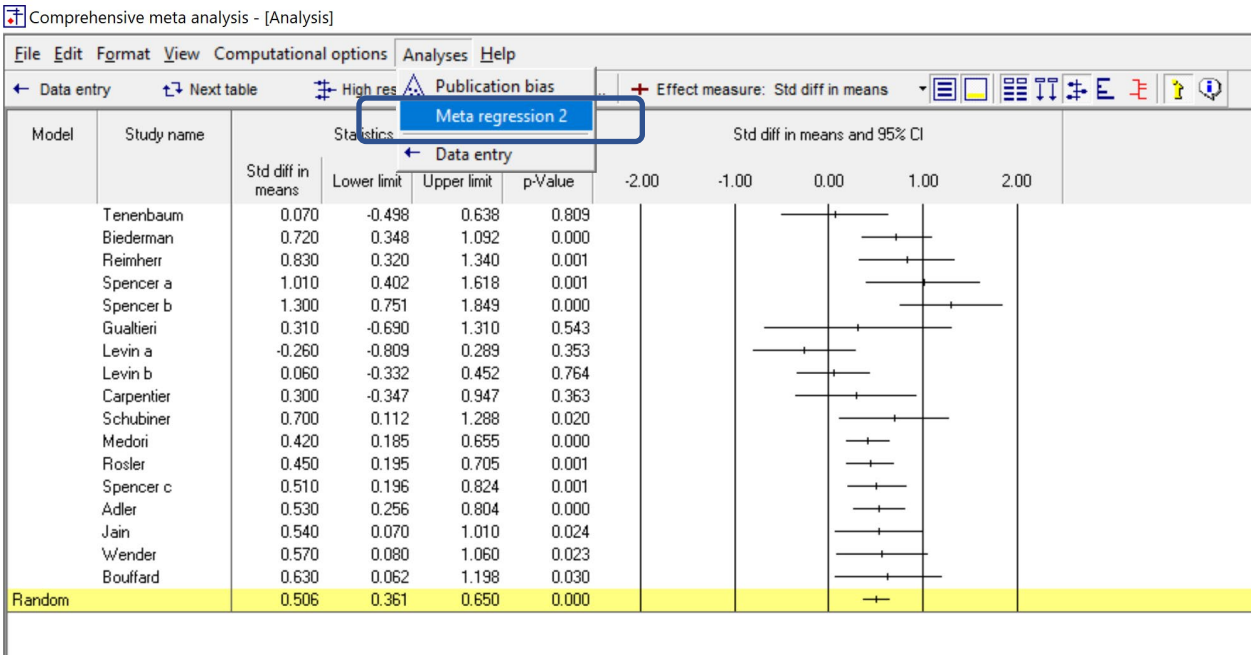
First, we need to remove the grouping

Click Computational options > Group by

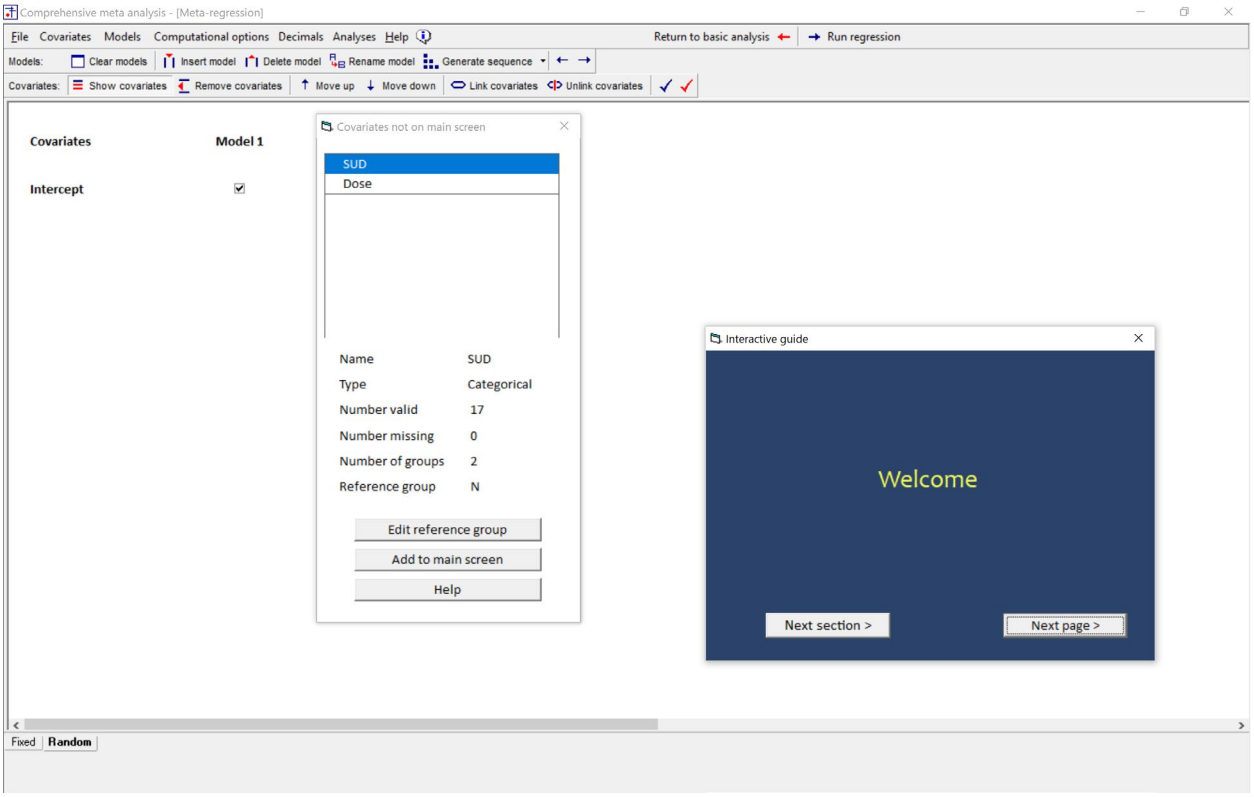
Click Reset



Click Analyses > Meta regression 2

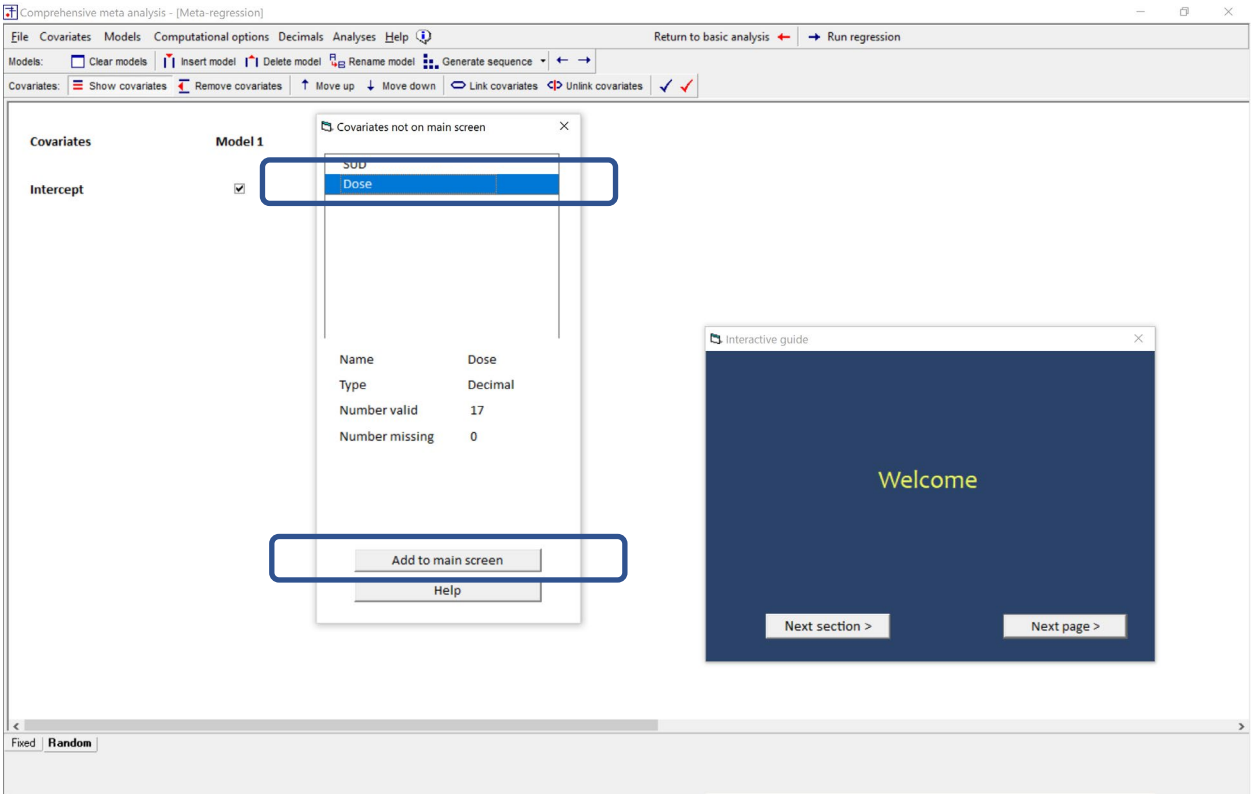


The program displays this screen

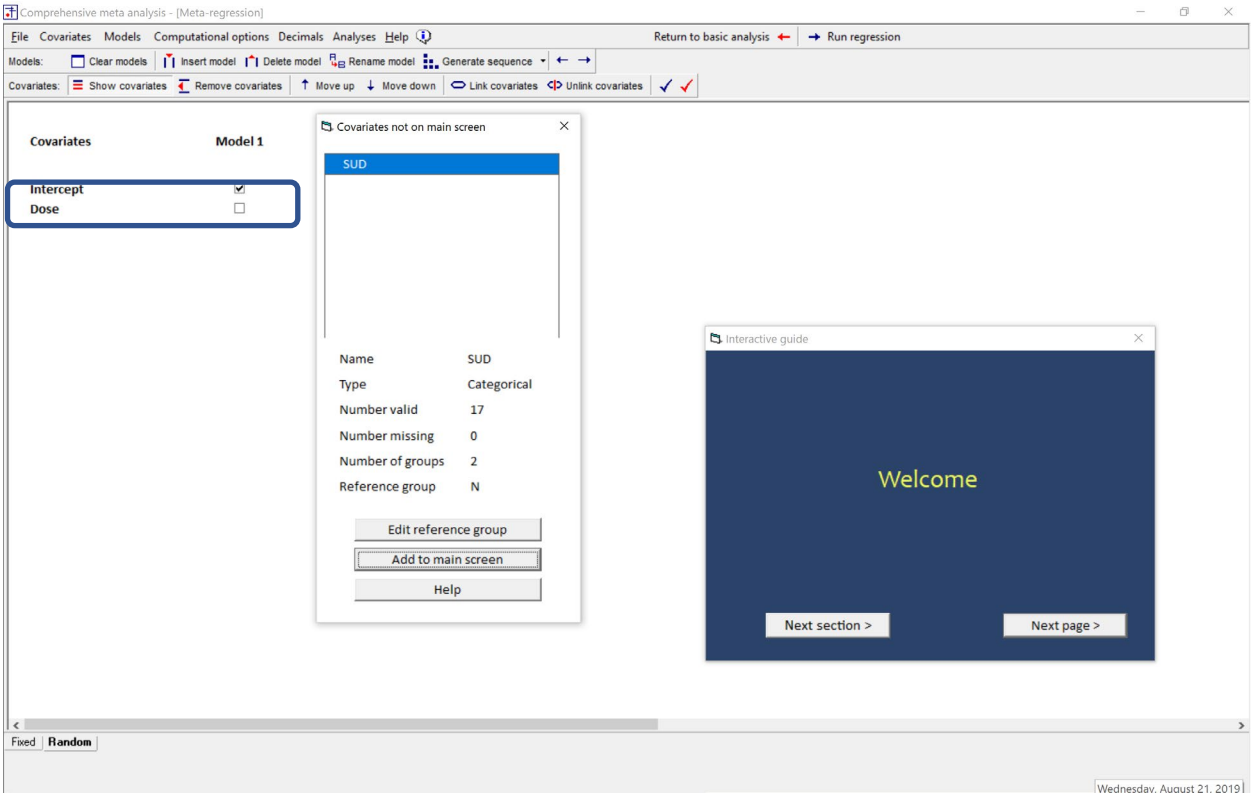


Click on Dose

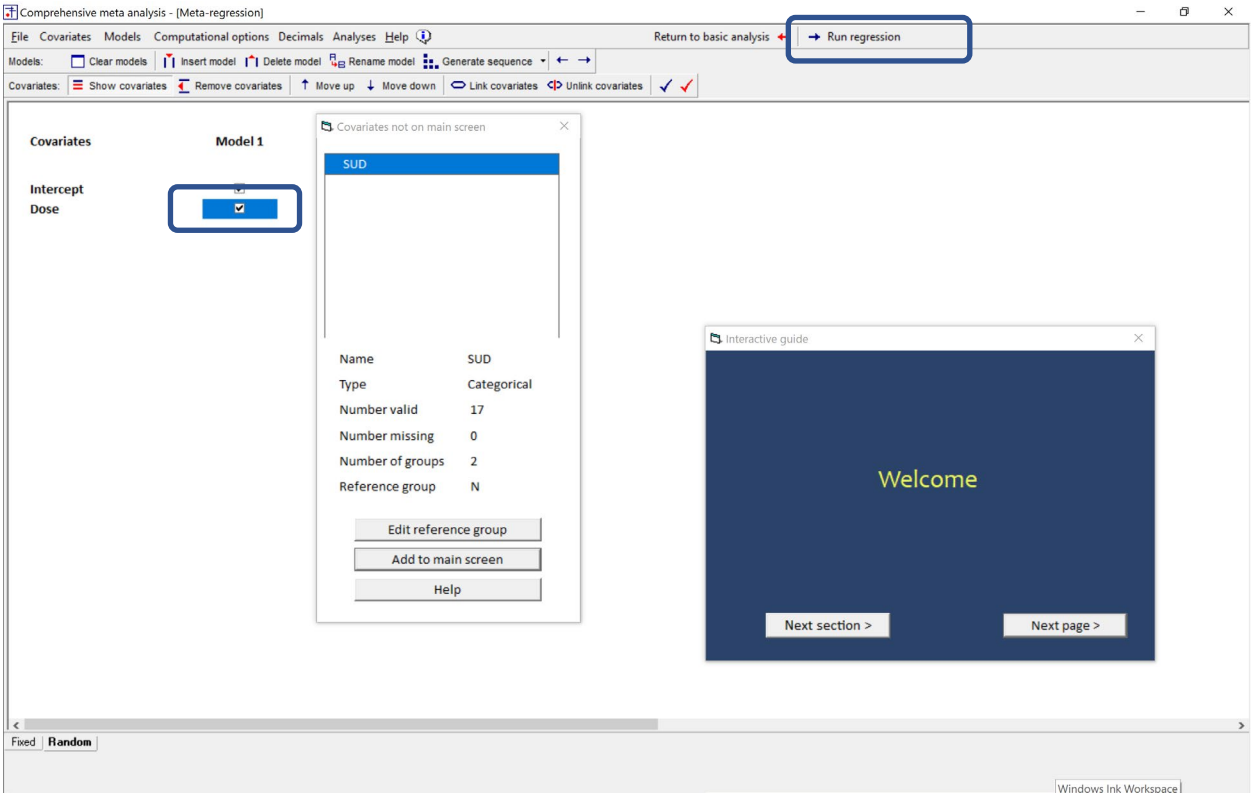
Click on Add to main screen



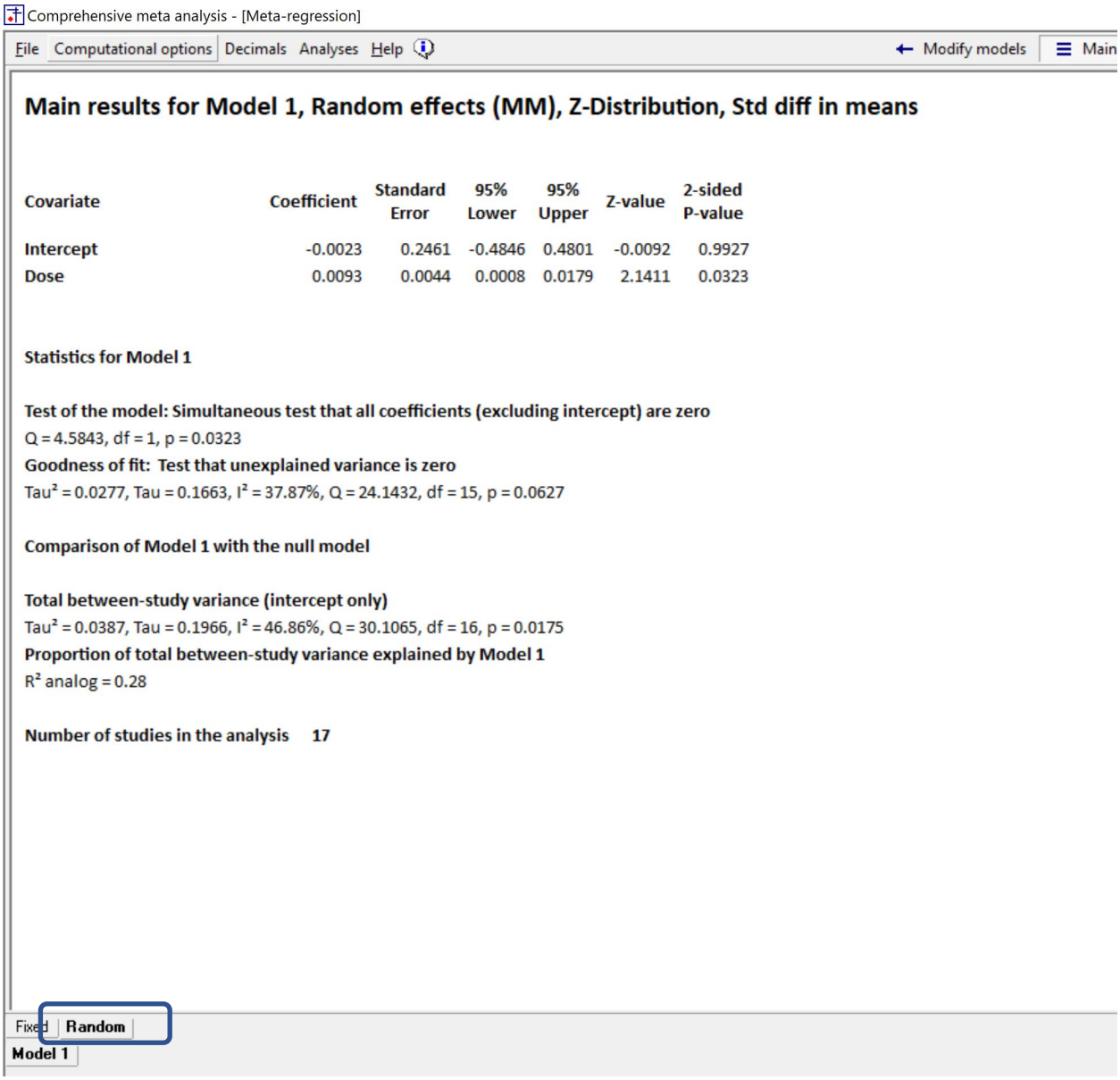
The covariate dose is moved from the list of available covariates onto the list that will be used in the analysis



- Check the box for Dose
- This tells the program to include this covariate in the analysis
- Click Run regression
-



Select the tab for Random at the bottom



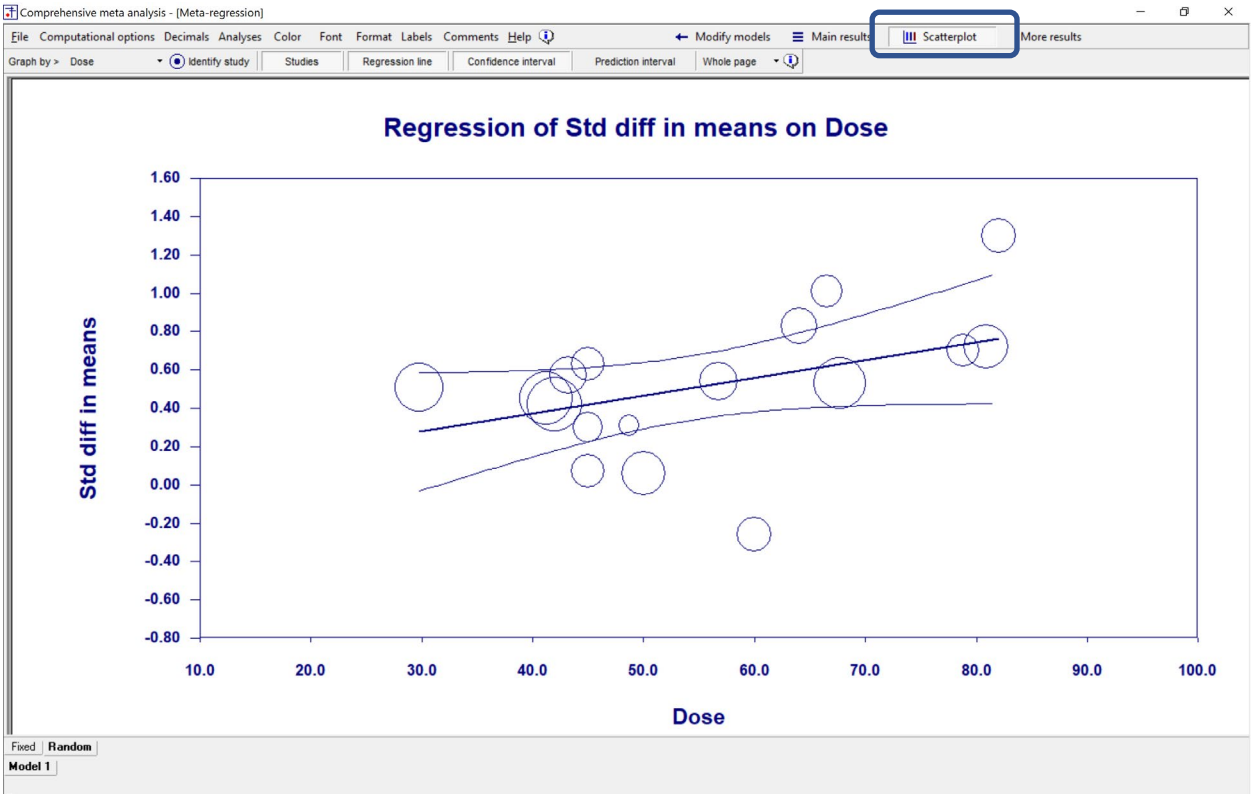
The table at the top shows the relationship between each covariate and the effect when all other covariates are held constant. Here, there is only one covariate, so this table shows the relationship between Dose and effect when no other covariates are held constant.

The coefficient is 0.0093, which means that for every one unit increase in dose the effect size will increase by 0.0093. In round numbers, as dose goes up by one unit, the effect size goes up by 0.01. The confidence interval for the coefficient is 0.0008 to 0.0179.

We can test the null hypothesis that the true coefficient is zero. That test yields a Z-value of 2.1411 and a p-value of 0.323.

To see what this looks like we can plot it

Click Scatterplot



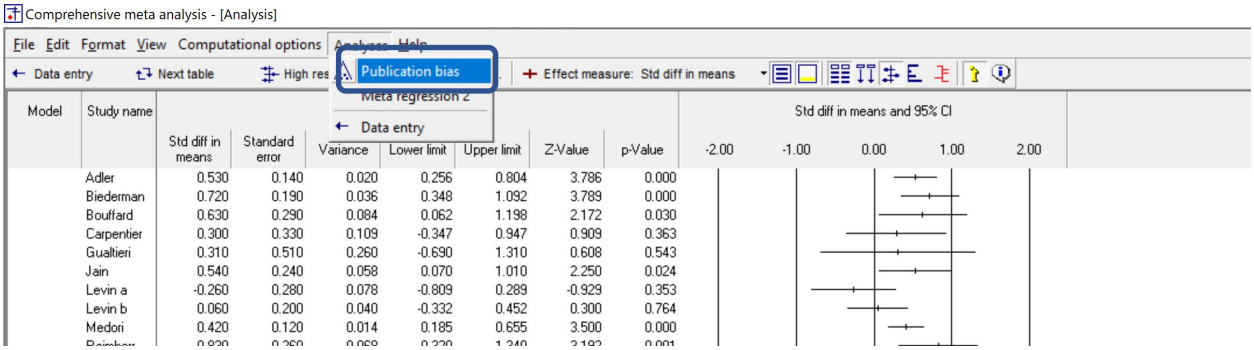
We can see that as Dose increases from around 30 units to 80 units, the effect size increases from around 0.30 to 0.80.

These pages provide only a brief introduction to the regression module.

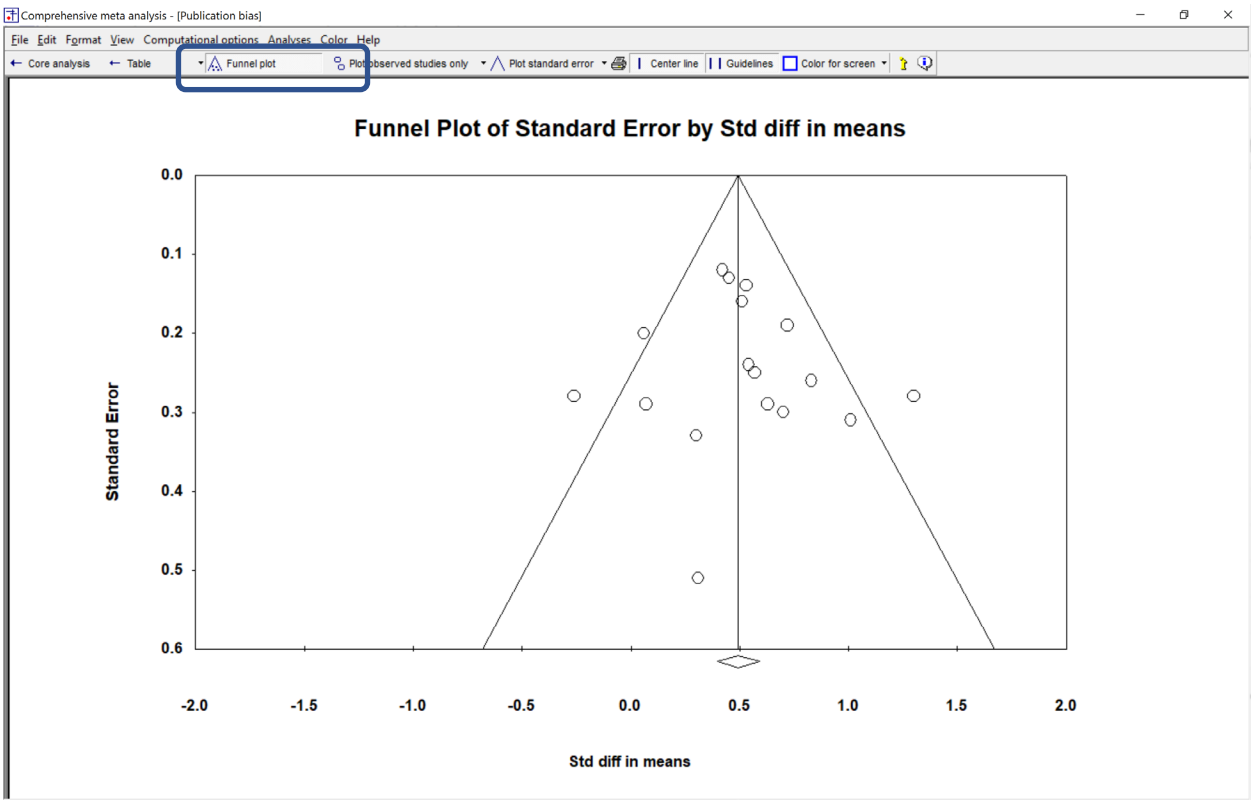
For more detail on this module, see our book on Meta-regression. This is available at no cost as a PDF. Send a note to Biostat100@GMail.com for a copy.

Publication bias

To address the potential impact of publication bias we can select “Analyses > Publication bias” on the main analysis screen.



The program displays this screen

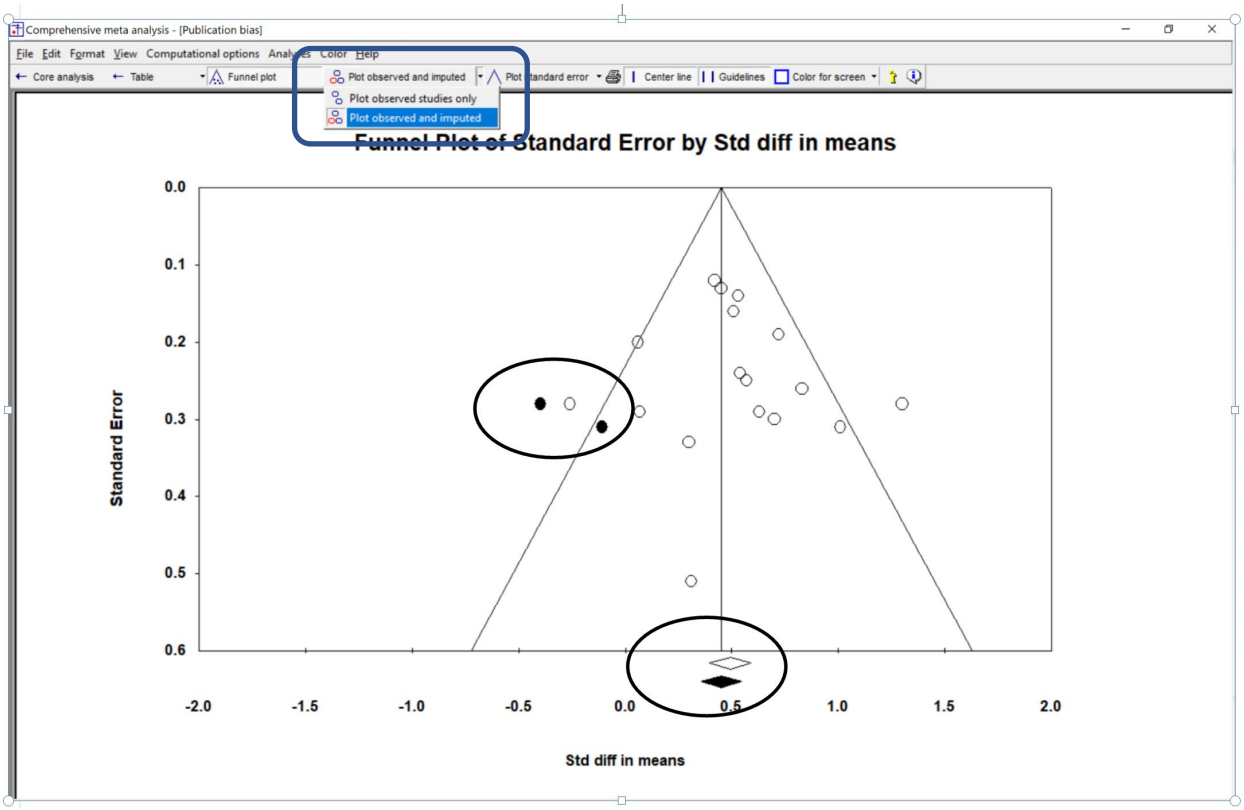


This is a funnel plot, which shows the effect size (on the X-axis) by the standard error (on the Y-axis). The large studies appear at the top, and the smaller studies appear toward the bottom.

The sample size in most studies falls within a narrow range of sample sizes, and so it is not likely that the procedures normally employed to assess publication bias would be effective. For purposes of this exercise we can nevertheless apply the Trim and Fill procedure, as shown here.

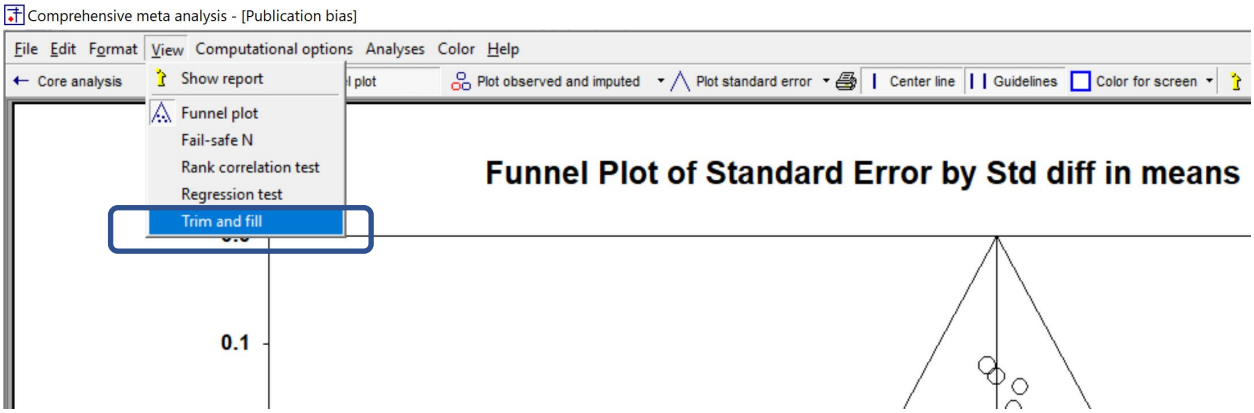
The plot shows the studies that are actually included in the analysis. A vertical line denotes the average effect size of approximately 0.50. If the effects are normally distributed, we would expect half the studies to fall on either side of the line. The Trim and Fill method looks to see if the effects are indeed symmetrically distributed on either side of the mean.

To see the results of this method, click Plot > Plot observed and imputed

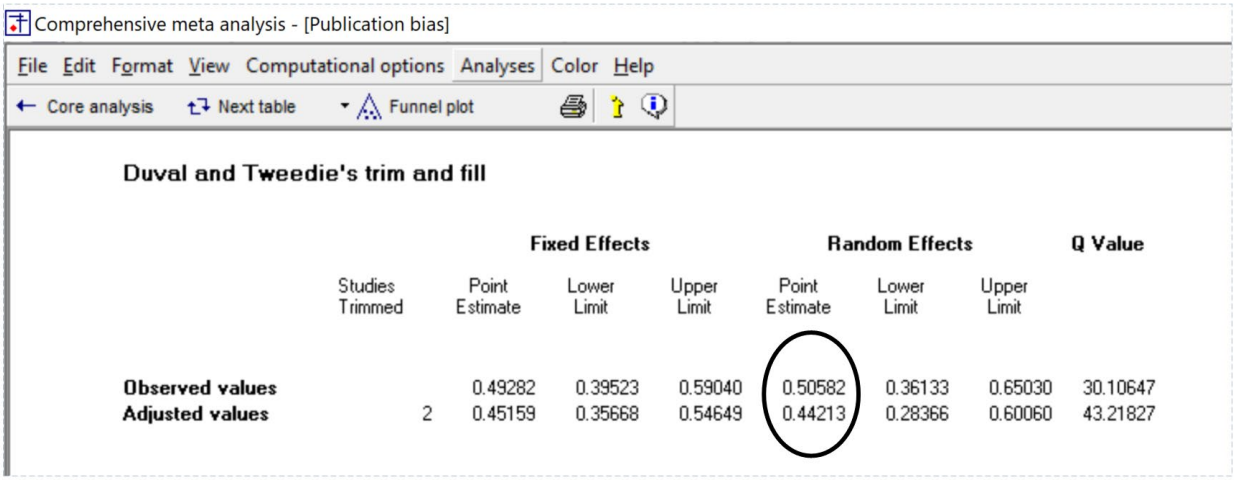


The Trim and Fill algorithm suggests that there may be two studies missing from the left. It imputes those studies and adds them to the analysis (shown as filled circles). At the bottom, the mean effect (filled diamond) is pulled slightly to the left by these two new studies.

To see the actual values associated with the original analysis and the adjusted analysis, click View > Trim and Fill



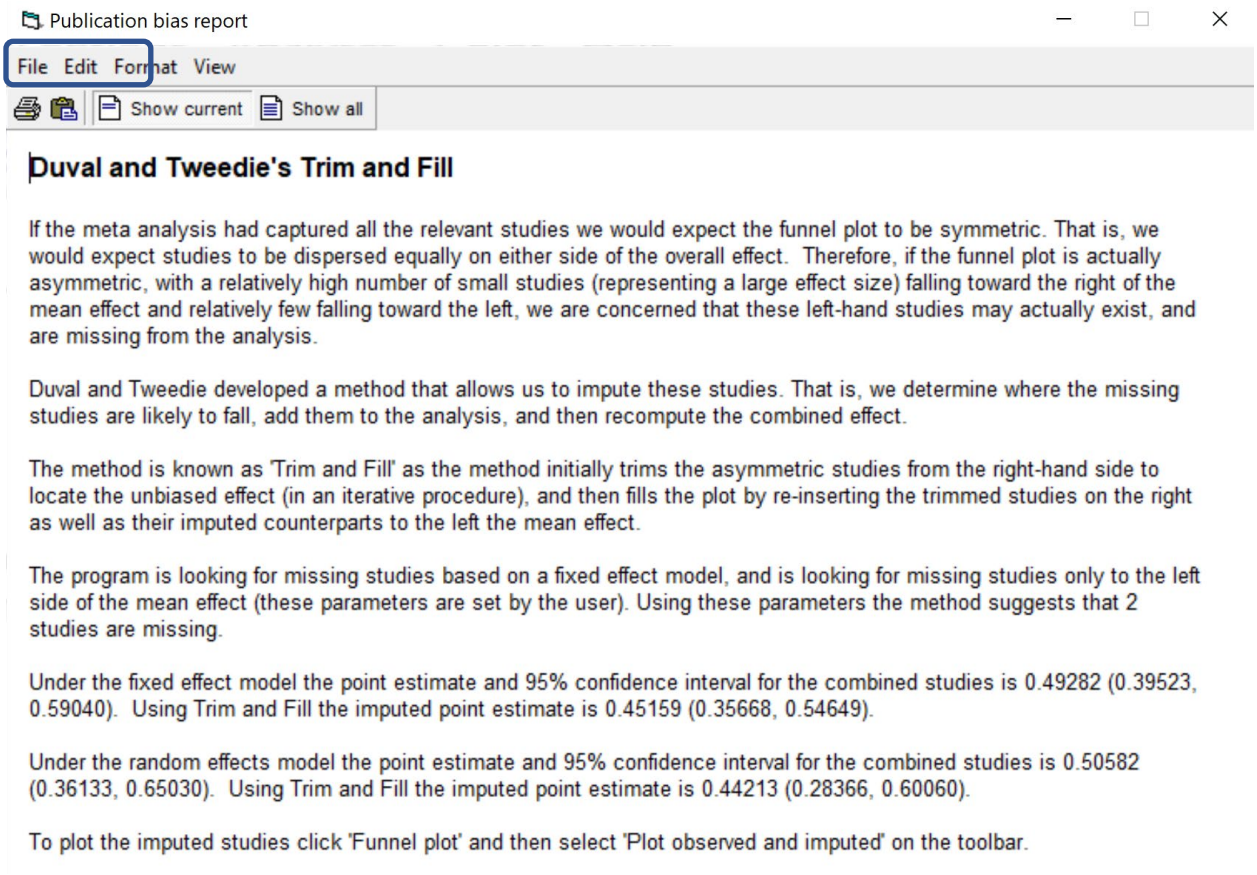
The original effect size using random-effects weights was 0.508. The adjusted estimate is 0.442.



		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0.49282	0.39523	0.59040	0.50582	0.36133	0.65030	30.10647
Adjusted values	2	0.45159	0.35668	0.54649	0.44213	0.28366	0.60060	43.21827

There are many possible reasons for the asymmetry, with publication bias being one of them. If the asymmetry really was due to publication bias, the adjusted estimate is likely to be a better estimate of the true mean than the original estimate. In this case, the adjustment is minor, in the sense that a mean effect size of 0.442 has essentially the same clinical import as a mean effect size of 0.506. So we can say that the basic conclusion (that the mean effect size is moderate) is robust.

The program can generate a summary of this finding. Click View > Show report and the program opens this screen, which can be copied and pasted into another program.



CMA also features other methods that are typically used to test and/or adjust for publication bias. These include the Egger test of the intercept, the Begg and Mazumdar rank correlation test, and Rosenthal's Fail-safe N.

The report below is an example of how to report this analysis

Overview

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 placebo and the researchers assessed the patients' cognitive function at the conclusion of treatment. The effect size 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 co-workers 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. 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 [D]. 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.

Substance abuse disorder (SUD)

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 this comparison is observational, and cannot prove a causal relationship.

Publication bias

There is a relatively narrow range of sample sizes for the studies in this analysis, and therefore tests for publication bias are not likely to be useful. The Trim and Fill method suggests that there might be two studies missing from the analysis. If we were to impute these studies and include them in the analysis, the mean effect size would shift from 0.508 to 0.442, which is a trivial shift in this context.