# A Guide to Common Effect Sizes and Forest Plots

This material is critical to your ability to read, understand, and appraise intervention or program outcome research. While you should be able to critically appraise any research article, intervention research will become particularly important to you. In some ways, this type of research is at the heart of evidence-based practice – using research results to formulate interventions or programs that have a high probability of success for your client(s). This content is especially important to your effective reading of systematic reviews and meta-analyses.

Objectives:

- Describe effect size coefficients for categorical outcomes in an intervention study – risk difference, risk ratio, and odds ratio
- Define effect size coefficients for continuous outcomes in an intervention study – mean difference, standardized mean difference.
- Interpret a forest plot.

Recall that a meta-analysis is a statistical technique for combining the findings from a set of independent studies. In summary,

- Meta-analysis is most often used to assess the clinical effectiveness of social, behavioral, and healthcare interventions; it does this by combining data from two or more randomized controlled or well-designed quasi-experimental trials.
- Meta-analysis of trials provides a precise estimate of treatment effect of each trial and an overall estimate of effect for the set of studies giving due weight to the size of the different studies included.
- Good meta-analyses aim for complete coverage of all relevant studies, look for the presence of heterogeneity, and explore the robustness of the main findings using sensitivity analysis.

You may find yourself confused about the difference between a systematic review and a meta-analysis. Generally, you can think of a systematic review as the overall strategy used to identify and collect a set of studies that address a common outcome. Systematic reviews often include a meta-analysis (or a set of meta-analyses) where findings from the studies are compared, combined, and analyzed. You also can find articles advertised as free standing meta-analyses, that is, that are not a part of a systematic review. For the most part, these free-standing meta-analyses use the same methods as systematic reviews to locate the studies that are included in the meta-analysis. Don't be confused by discussions about systematic reviews, if there is a statistical analysis it is a meta-analysis.

This discussion focuses primarily on meta-analysis. Remember that the goal of a meta-analysis is to provide both a comparative view of a set of studies and an overall summary of the effects of the study set. We don't want to lie here; a meta-analysis can be bewildering when first encountered. Our goal this week is to provide suggestions about the important components for any meta-analysis - if you can get these figured out you will start to be a good meta-analysis consumer. Also, the more you concentrate and practice on the various topics we present here, the more they will start to make sense – meta-analyses included. We can honestly say that many of our students have ended up being skillful in reading meta-analytic studies (after telling us they were having severe heart-burn when we started the discussion).

## Components of a Meta-Analysis

In general, when you read a meta-analysis you should look for these things:

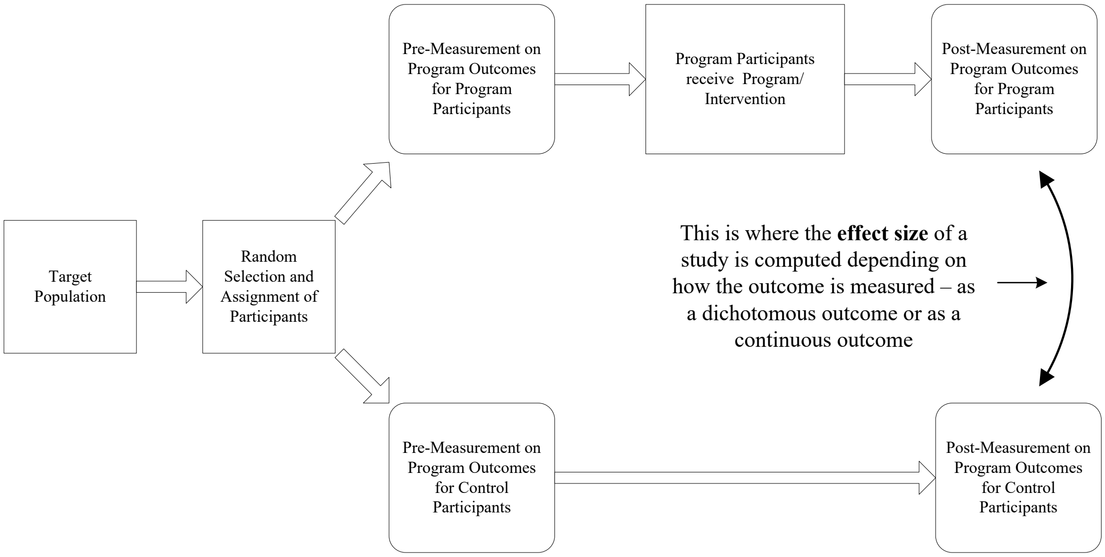
- The outcomes addressed by the studies included in the analysis
- The types of studies included in the analysis
- The effect sizes used to compare studies

Let's discuss each of these areas in turn. First, for the most part you should be able to readily determine the outcomes addressed by the studies included in the analysis. The outcomes are what got you to the systematic review or meta-analysis in the first place. Remember that the outcomes are related to the client problem you are addressing. For example, the outcome for a self-harming client would be the reduction of self-harming behavior. Your interest would be in a meta-analysis of studies that have reported on interventions designed to address self-harming behavior. This logic applies to virtually any client problem translated into an outcome.

Next, you need to identify what types of studies are included in the meta-analysis. Most of the meta-analysis studies you will find report on studies that have been conducted using a randomized controlled trial design. Recall that RCTs are the gold standard for testing the effectiveness of an intervention. That notwithstanding, you may also find a meta-analysis or two that examines a set of quasi-experimental design trials. Don't discount quasi-experimental designs – well-done studies using quasi-experimental methods can produce important evidence about an intervention.

Finally, you need to identify what effect sizes are being used to compare studies. Effect sizes are at the heart of a meta-analysis - they are measures of the effect an intervention had on an outcome compared to a control or other comparison condition. We discuss effect sizes in detail below.

Before getting to effect sizes, however, let's do a quick review of what the design of the studies included in a meta-analysis look like. The following is the diagram for a randomized two-group pre-post research design (an RCT). Take a cognitive snap-shot of this diagram and remember that when you are reading a meta-analysis of a set of RCT studies, each study in the analysis was conducted using this design. The diagram also works for a set of quasi-experimental studies the only difference being that these studies did not use random selection and random assignment procedures.


Target Population
Random Selection and Assignment of Participants
Pre-Measurement on Program Outcomes for Program Participants
Program Participants receive Program/ Intervention
Post-Measurement on Program Outcomes for Program Participants
This is where the **effect size** of a study is computed depending on how the outcome is measured – as a dichotomous outcome or as a continuous outcome
Pre-Measurement on Program Outcomes for Control Participants
Post-Measurement on Program Outcomes for Control Participants

What is really is the ‘heart of the matter’ for a study and for the analysis of a set of studies is the **effect size** represented by the two-headed arrow. To restate, an effect size is a measure that quantifies the difference between people in the study who experienced the intervention and people who were in the control group. In an intervention study, the hypothesis tested is that people in the intervention group will be better or improved as a result of the intervention compared to people who did not get the intervention and the effect size indicates how much better or improved they actually are.

<u>Common Intervention Research and Meta-Analysis Effect Sizes</u>

Before we get into this discussion, we want to make sure that you don’t talk yourself out of understanding the material out-of-hand. We do run into students who claim they just don’t get these technical things and shut down. Avoid that self-defeating behavior at all costs - take your time and look things over with an open mind. Also, keep in mind that this will not be the last conversation you have about effect sizes and meta-analysis – other courses in the sequence will touch on these topics, as well. All we ask is that you do your best…

So let’s get started. There are two general families of effect sizes used in intervention and meta-analysis studies. While the language varies a bit, there are two types of outcomes (remember these are represented by the two-headed in the RCT diagram) at post-test:

- In some studies outcomes are measured at two-levels. These outcomes are considered as **dichotomous**. For example, in a study a dichotomous  outcome could be considered as
  - Successful vs unsuccessful
  - Something present vs something absent
  - Positive benefits vs no benefits
- In other studies outcomes are measured on a **continuous** scale. Continuous here means that the outcome can take on a number of values along some measurement

scale. In a study where the outcome is measured on a continuous scale, it could be expressed as

- On the average, program participants are more successful than control participants on the outcome of interest

Effect Sizes for Dichotomous Outcome Studies

The three effect sizes based on dichotomous outcome data are:

- Risk difference
- Relative risk or risk ratio
- Odds ratio

We have found that the best way to teach how these effect sizes get computed and interpreted is to dig into the details (we hope you agree). A good place to start is to think about the results of a study cast in a 2x2 table (we talked about this type of table in our discussion about chi-square in Week 5). A 2x2 table is called a cross-tabulation. The general format for an intervention study with two groups and a dichotomous outcome is shown below.

| | Positive Effect | No Effect | Row Total |
|---|---|---|---|
| Experimental Group | A | B | A+B |
| Control Group | C | D | C+D |
| Column Total | A+C | B+D | N |

The cross-tabulation is created by the experimental and control groups in the study (the rows) and the dichotomous outcome of positive and no effect (columns). The crossing of these two variables creates a set of cells which are defined as follows:

- Cell A = people in the experimental group who experienced a positive effect
- Cell B = people in the experimental group who experienced no effect
- Cell C = people in the control group who experienced a positive effect
- Cell D = people in the experimental group who experienced no effect
- Cell A+B = total number of people in the experimental group
- Cell C+D = total number of people in the control group
- Cell A+C = total number of people who experienced a positive effect
- Cell B+D = total number of people who experienced no effect
- Cell A+B+C+D = total number of people in the study

The three popular effect sizes for dichotomous outcomes we identified above are simple functions of these cell values (once again **do not** ‘freak out’ here).

- Risk Difference = A/(A+B) - C/(C+D)

  - What this simple equation says is that the risk difference is the difference of the proportion of people in the experimental group who experienced a positive effect (A/(A+B)) minus the proportion of people in the control group who experienced a positive effect (C/(C+D))
  - If you think about it for a second, this makes sense. In a successful study, we would find a higher proportion of positive effects in the experimental than we would in the control group
  - Technically a risk difference is a difference in proportions. The risk language is a holdover from epidemiology studies that use these types of tables to look at risk factors and disease – juts remember a risk difference is a proportion difference

- Relative Risk or Risk Ratio = A/(A+B) / C/(C+D)
  - This equation is an extension of the risk difference equation. It is a ratio defined as the proportion of people in the experimental group who experienced a positive effect (A/(A+B)) divided the proportion of people in the control group who experienced a positive effect (C/(C+D))
  - In a successful study, we would find a higher proportion of positive effects in the experimental than we would in the control group and would express that as being so many times higher (say, 3 times higher) that the control proportion
  - Again, technically a risk ratio is a ratio of proportions.
  - You will find either relative risk or risk ratio used in studies – they are interchangeable

- Odds Ratio = (A/B) / (C/D)
  - Odds ratios are very popular in outcome research. They are also a little harder to figure out. First, an odds ratio is not based on proportions – it is based on odds. The definition of an odds is thus – an odds is the chance that something will happen vs the chance that something won't happen (good so far?). So, for example, in our study table we can compute the odds that people in the experimental group experienced a positive effect as A/B – the chance that they experienced a positive effect divided by the chance they experienced no effect. We would do the same for the control group as C/D. The odds ratio is just the ratio of those two values ((A/B) / (C/D)).
  - In a successful study, we might find that the odds that the experimental group experienced a positive effect is 2.5 times higher than the odds the control group experienced a positive effect.

An example of these effect sizes might help here. Let's say we are reading a study about an intervention designed to help mothers experiencing postnatal depression. The outcome of the study is expressed as a dichotomy – at post-test moms were diagnosed either as having no depressions symptoms (PD No in the table below) or as having depression symptoms (PD Yes in the table below).

| Group by Postnatal Depression Status at Post-test | | | |
|---|---|---|---|
| | PD No | PD Yes | Total |
| Experimental Group | 48 | 12 | 60 |
| Control Group | 21 | 39 | 60 |
| Total | 69 | 51 | 120 |

- Cell A = 48 moms in the experimental group had no depression symptoms
- Cell B = 12 moms in the experimental group had depression symptoms
- Cell C = 21 moms in the control group had no depression symptoms
- Cell D = 39 moms in the control group had depression symptoms
- Cell A+B = 60 total moms in the experimental group
- Cell C+D = 60 total moms in the control group
- Cell A+C = 69 total moms had no depression symptoms
- Cell B+D = 51 total moms had depression symptoms
- Cell A+B+C+D = 120 total moms in the study

Using the equations discussed previously, we find the following effect sizes:

- Risk Difference = (48/60) – (21/60) = .45
- Relative Risk = (48/60) / (21/60) = 2.29
- Odds Ratio = (48/12) / (21/39) = 7.43

We interpret these effect sizes as follows:

- For the risk difference value, we can say that the *proportion* of experimental moms not having postnatal depression symptoms is .45 points higher than the *proportion* of control moms not having postnatal depression symptoms
- For risk ratio value, we can say that the ratio of the *proportion* of experimental moms not having postnatal depression symptoms is 2.3 times higher than the *proportion* of control moms not having postnatal depression symptoms
- For odds ratio value, we can say that the *odds* of experimental moms not having postnatal depression symptoms is 7.4 times higher than the *odds* of control moms not having postnatal depression symptoms

All-in-all our conclusion is that the intervention was effective in impacting postnatal depression.

Effect Sizes for Continuous Outcome Studies

The three effect sizes based on continuous outcome data are:

- Mean difference
- Standardized mean difference

The layout for a study using a continuous outcome is a little simpler that the cross-tabulation we discussed for dichotomous outcomes. Means for the experimental and control groups on the continuous outcome measure represented as shown in the following table.

| | Post-test Mean |
|---|---|
| Experimental Group | $M_1$ |
| Control Group | $M_2$ |

If  a study uses a continuous outcome on some scale or metric that has a commonly understood meaning (weight, income in dollars, behavioral counts, body mass index, minutes spent exercising, etc.), it is often convenient to express the effect size as a simple difference between the two group means (**mean difference**) using the following:

$$MD = M_1 - M_2$$

If a study uses a continuous outcome on some scale metric that does not necessarily have commonly understood meaning (most of our areas of interest), researchers use a standardized form called a **standardized mean difference**. Standardized mean difference effect sizes are computed using the following:

$$SMD = \frac{M_1 - M_2}{SD}$$

You can see in this equation that the standardized mean difference is computed by using the mean difference we noted above in the numerators and a measure of a standard deviation (CD) in the denominator. Thus, standardized mean difference effect sizes are expressed in standard deviation units – they are like the z-scores you may remember from your statistics class.

There are two important parts of a standardized mean difference effect size

- The effect size value itself
  - The value of an effect size indicates how much difference there was in the study between the experimental and control group – larger values reflect larger differences
- The sign of the effect size value
  - The sign of the effect size indicates direction of the difference – in meta-analyses results are typically coded so that a positive sign means the experimental group improved (but not always)

Finally, you might see one of three standardized mean difference effect sizes in an intervention study or a meta-analysis: **Cohen's *d*, Glass's *delta*, and Hedge's *g***. Without getting into great detail, these different standardized mean differences vary by how the standard deviation gets defined. Cohen's *d* uses a standard deviation in the denominator that uses information from both the experimental and control group, Glass's *delta* uses the standard deviation from only the control group, and Hedge's *g* uses the same pooled standard deviation as Cohen's *d* but adds a small correction factor. If the standard deviation in each group in a study is similar, all of these versions will be very close in value.

$$d = \frac{M_1 - M_2}{SD_{pooled}} \qquad delta = \frac{M_1 - M_2}{SD_{control}} \qquad g = \frac{M_1 - M_2}{SD*_{pooled}}$$

An example of how continuous outcome effect sizes look in a study might help here. Let's say we are reading a study about an intervention designed to help child welfare social workers in an agency deal with stress. The outcome of the study is a 'feeling stressed' scale. This is a small N study with 10 people in the experimental group and 10 people in the control group.

You find the following descriptive information about study results. A lower score on the feeling stress scale is a better score so the does appear to be a positive program impact. The standard deviations indicate there was more variability in the experimental group scores than the control group scores.

| Group | N | Mean Stress Score | Standard Deviation |
|---|---|---|---|
| Experimental | 10 | 7.8 | 2.15 |
| Control | 10 | 9.4 | 1.58 |

The authors appropriately conducted an independent samples t-test (remember from Week 5?). It is reported in the following table. The actual mean difference between the groups is -1.6 (the sign is arbitrary depending how the researchers represented group membership in their statistical program). We know that it is in favor of the experimental group. What we don't know is how to interpret it based on the stress scale metric. We also see that the t-test is not statistically significant at the .05 level. The p-Value is larger than .05.

| Mean Difference | t | df | p-Value |
|---|---|---|---|
| -1.6 | -1.89 | 16.51 | 0.075 |

This is where things get interesting. What we need is our standardized mean difference to see if the result of the study is *practically* significant. Studies based on small samples are regularly not statistically significant.

The authors further report the following standardized mean difference values:

- Cohen's *d* = -.84
- Hedge's g = -.81
- Glass's delta = -1.01

A very influential effect size author, Jacob Cohen (of Cohen's *d* fame), suggested that standardized mean difference values can be interpreted using the following rules-of-thumb: a value of .20 would be considered a small effect size, a value of .50 would be considered a medium effect size, and a value of .80 would be considered a large effect size. Based on these cut-offs, we agree with the authors that the intervention had a large effect on stress reduction for the experimental group child welfare workers. Study values for each of the standardized mean difference effect sizes exceeded the .80 threshold for a large effect. Effect sizes are used to gauge the practical significance of a study. In this case, we would conclude that while the study findings were not statistically significant, they were practically significant.

## How Effect Sizes are Reported in a Meta-analysis

We have found that it doesn't work well to dive right into meta-analysis without a conversation about where effect sizes come from in a study – where they are in the design (experimental-control group post-test comparison), how they link to the type of outcome measure (dichotomous or continuous), and how they get computed. You now have all the information you need to start interpreting a meta-analysis.

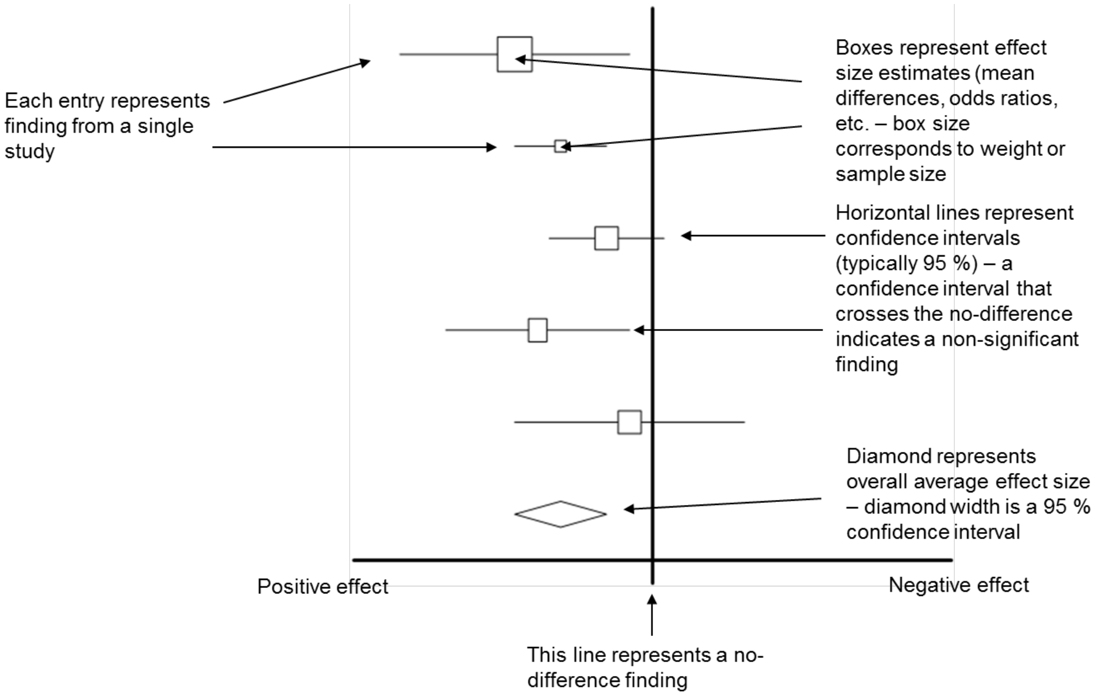
While there is some variability in how effect size data is presented and analyzed in a meta-analysis, for the most part effect size data is reported in graphs called forest plots. While they may look intimidating at first, once you know how they are designed they can be very helpful. Before we venture into our discussion of forest plots, please spend five minutes looking at the following video. It is a nice overview of forest plots. Don't worry that it is a physician doing the video, much of the technical details of evidence-based practice comes from medicine. It should not be a problem to translate into your own evidence-based framework.

https://www.youtube.com/watch?v=py-L8DvJmDc

Now that you have watched the video, let's do a quick review. The next diagram highlights the major features of a forest plot:

- Each line represents the effect size from one study. In the diagram there are five studies.
- The boxes represent the location of the effect size on the effect size scale. Sometimes these locations are represented by a tick mark rather than a box. Their distance away from the no difference or no effect perpendicular line indicates how big the effect size is – the further away it is the larger the effect size. The size of the box in some meta-analyses actually is an indicator of how much weight the effect is given in the aggregate analysis.
- The diamond represents that aggregate effect or overall average effect size – each individual study makes a weighted contribution to that estimate.

- The whiskers on either side of the effect size boxes are confidence intervals. They are indicators of the precision of the effect size estimate in a study. Small confidence intervals are better than large confidence – they indicate a more precise estimate.
- If either tail of a confidence interval crosses the no difference or no effect line, it indicates that that study was not statistically significant. Two studies in the diagram are not significant.

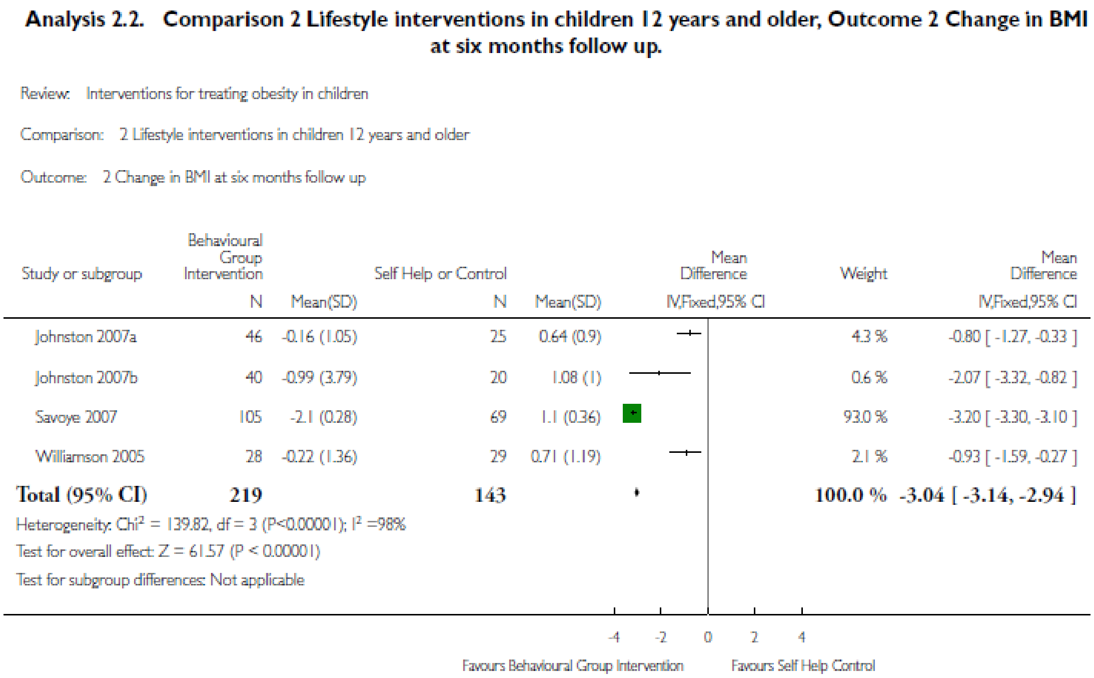

Each entry represents finding from a single study
Boxes represent effect size estimates (mean differences, odds ratios, etc. – box size corresponds to weight or sample size
Horizontal lines represent confidence intervals (typically 95 %) – a confidence interval that crosses the no-difference indicates a non-significant finding
Diamond represents overall average effect size – diamond width is a 95 % confidence interval
Positive effect
Negative effect
This line represents a no-difference finding

The figure below is an actual forest plot from a Cochrane systematic review. This review looked at interventions design to impact obesity in children. Let's take a look at the features of this forest plot:

- The title of the plot indicates that it summarizes studies that tested lifestyle interventions for children 12 years or older and that the outcome of interest was a change in Body Mass Index change scores at six months follow up. All of the studies were randomized controlled trials.
- There are four studies in the plot.
- The effect size used to compare studies is a mean difference (it is shown just above the no effect perpendicular line).
- In a Cochrane review, the location of the effect size on the effect size scale along the bottom is shown by a tick mark.
- Cochrane uses boxes to represent the weight of each study in the summary. In this plot, one study, Savoye, is large compared to the other three studies so it has a noticeably larger box than the others (there really are boxes in the other three studies, they are just very small). You can also see how disproportionally large the Savoye study is by looking in the weight column.
- The confidence intervals vary in width around each effect size estimate. All of the studies are statistically significant. The Savoye study does have confidence

intervals and, while they are not clear in the plot, you can check their values in the far right column.

- The summary diamond at the bottom of the four studies is small and is far away from the no effect line. This summary says that taken as a group, this set of studies provides evidence that lifestyle interventions are effective in impacting BMI scores for children 12 and older. The summary mean difference is -3.04 BMI units and the confidence interval is narrow ranging from -3.14 to -2.94.
- The various statistical values shown in the lower left portion of the plot indicate how variable the estimates of the set of studies are. We will be looking at this aspect of forest plots in future classes.
- The implication of this analysis for practice is that there is evidence that lifestyle interventions are successful in impacting childhood obesity. We would be interested in looking at any of these studies in more detail to see how they were designed and conducted. We would be especially interested in looking at the Savoye study since it was large and very successful.



**Analysis 2.2. Comparison 2 Lifestyle interventions in children 12 years and older, Outcome 2 Change in BMI at six months follow up.**

Review: Interventions for treating obesity in children

Comparison: 2 Lifestyle interventions in children 12 years and older

Outcome: 2 Change in BMI at six months follow up

| Study or subgroup | Behavioural Group Intervention | | Self Help or Control | | Mean Difference | Weight | Mean Difference |
|---|---|---|---|---|---|---|---|
| | N | Mean(SD) | N | Mean(SD) | IV,Fixed,95% CI | | IV,Fixed,95% CI |
| Johnston 2007a | 46 | -0.16 (1.05) | 25 | 0.64 (0.9) | | 4.3 % | -0.80 [ -1.27, -0.33 ] |
| Johnston 2007b | 40 | -0.99 (3.79) | 20 | 1.08 (1) | | 0.6 % | -2.07 [ -3.32, -0.82 ] |
| Savoye 2007 | 105 | -2.1 (0.28) | 69 | 1.1 (0.36) | | 93.0 % | -3.20 [ -3.30, -3.10 ] |
| Williamson 2005 | 28 | -0.22 (1.36) | 29 | 0.71 (1.19) | | 2.1 % | -0.93 [ -1.59, -0.27 ] |
| **Total (95% CI)** | **219** | | **143** | | | **100.0 %** | **-3.04 [ -3.14, -2.94 ]** |

Heterogeneity: Chi² = 139.82, df = 3 (P<0.00001); I² =98%

Test for overall effect: Z = 61.57 (P < 0.00001)

Test for subgroup differences: Not applicable

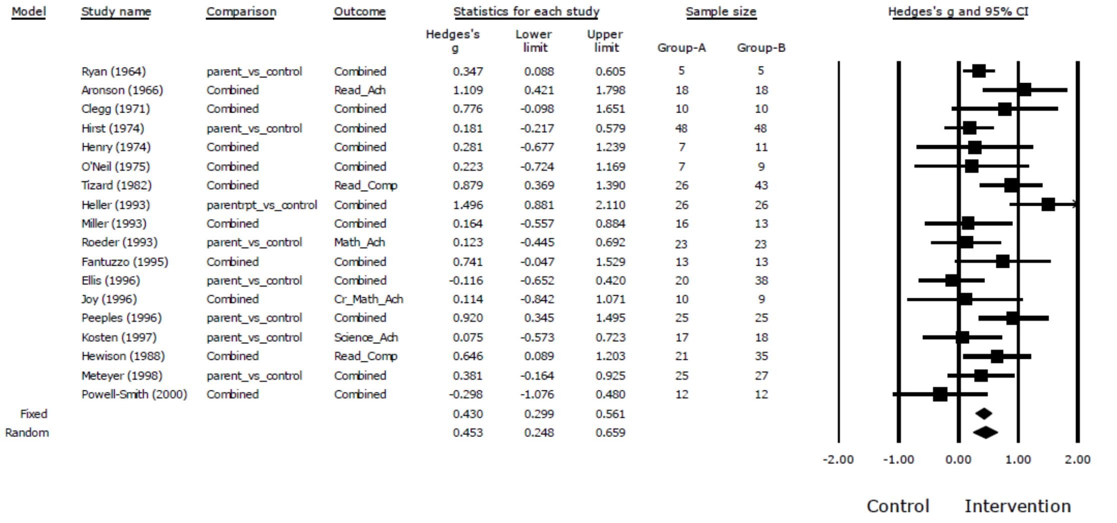
-4
-2
0
2
4
Favours Behavioural Group Intervention
Favours Self Help Control

Forest plots come in many forms so you need to orient yourself to the plot format before you dive into the details. For example, the following plot is from a Campbell systematic review. As you can see, it compares a set of studies that tested parent involvement strategies and their impact on academic performance. We won't spend time on the details presented in the plot but will highlight some differences between this plot and the Cochrane plot:

- The effect size used in this plot was a standardized mean difference – specifically it was a Hedge's *g* (see if you can find it in the plot – you need to do this for every plot you examine – it is a little like finding Waldo)
- The horizontal axis in this plot is the opposite of the Cochrane plot – intervention is favored on the right side of the plot

- It does not appear that box size indicates how much weight the study was given in the summary. The boxes show the location of a study effect size.
- Finally, there are two diamonds at the bottom of the study – one for a fixed effects model and the second for a random effects model (there will be more to come on this). In any event, both of the diamonds are statistically significant.

Figure 1. Effect of Parent Involvement on Children's Academic Performance



| Model | Study name | Comparison | Outcome | Statistics for each study | | | Sample size | |
|---|---|---|---|---|---|---|---|---|
| | | | | Hedges's g | Lower limit | Upper limit | Group-A | Group-B |
| | Ryan (1964) | parent_vs_control | Combined | 0.347 | 0.088 | 0.605 | 5 | 5 |
| | Aronson (1966) | Combined | Read_Ach | 1.109 | 0.421 | 1.798 | 18 | 18 |
| | Clegg (1971) | Combined | Combined | 0.776 | -0.098 | 1.651 | 10 | 10 |
| | Hirst (1974) | parent_vs_control | Combined | 0.181 | -0.217 | 0.579 | 48 | 48 |
| | Henry (1974) | Combined | Combined | 0.281 | -0.677 | 1.239 | 7 | 11 |
| | O'Neil (1975) | Combined | Combined | 0.223 | -0.724 | 1.169 | 7 | 9 |
| | Tizard (1982) | Combined | Read_Comp | 0.879 | 0.369 | 1.390 | 26 | 43 |
| | Heller (1993) | parentrpt_vs_control | Combined | 1.496 | 0.881 | 2.110 | 26 | 26 |
| | Miller (1993) | Combined | Combined | 0.164 | -0.557 | 0.884 | 16 | 13 |
| | Roeder (1993) | parent_vs_control | Math_Ach | 0.123 | -0.445 | 0.692 | 23 | 23 |
| | Fantuzzo (1995) | Combined | Combined | 0.741 | -0.047 | 1.529 | 13 | 13 |
| | Ellis (1996) | parent_vs_control | Combined | -0.116 | -0.652 | 0.420 | 20 | 38 |
| | Joy (1996) | Combined | Cr_Math_Ach | 0.114 | -0.842 | 1.071 | 10 | 9 |
| | Peeples (1996) | parent_vs_control | Combined | 0.920 | 0.345 | 1.495 | 25 | 25 |
| | Kosten (1997) | parent_vs_control | Science_Ach | 0.075 | -0.573 | 0.723 | 17 | 18 |
| | Hewison (1988) | Combined | Read_Comp | 0.646 | 0.089 | 1.203 | 21 | 35 |
| | Meteyer (1998) | parent_vs_control | Combined | 0.381 | -0.164 | 0.925 | 25 | 27 |
| | Powell-Smith (2000) | Combined | Combined | -0.298 | -1.076 | 0.480 | 12 | 12 |
| Fixed | | | | 0.430 | 0.299 | 0.561 | | |
| Random | | | | 0.453 | 0.248 | 0.659 | | |

Heterogeneity Statistics for a Fixed Effects Model: Q =35.6, df = 17, p = 0.005, and I squared = 52.3.

Finally, a meta-analysis may not always be presented in a forest plot. Sometimes studies are summarized in tables or in graphs other that a forest plot. As you gain practice in reading meta-analytic studies, you will be able to find the information you are looking for in any presentation.

The following article is a nice discussion about reading and interpreting forest plots. Read it over carefully and make a copy for future reference.

Ried - Interpreting Meta-Analysis Graphs

These two blogs provide a light-hearted discussion of meta-analysis.

http://blogs.plos.org/absolutely-maybe/2014/01/20/5-key-things-to-know-about-meta-analysis/

http://blogs.plos.org/absolutely-maybe/2015/06/30/another-5-things-to-know-about-meta-analysis/

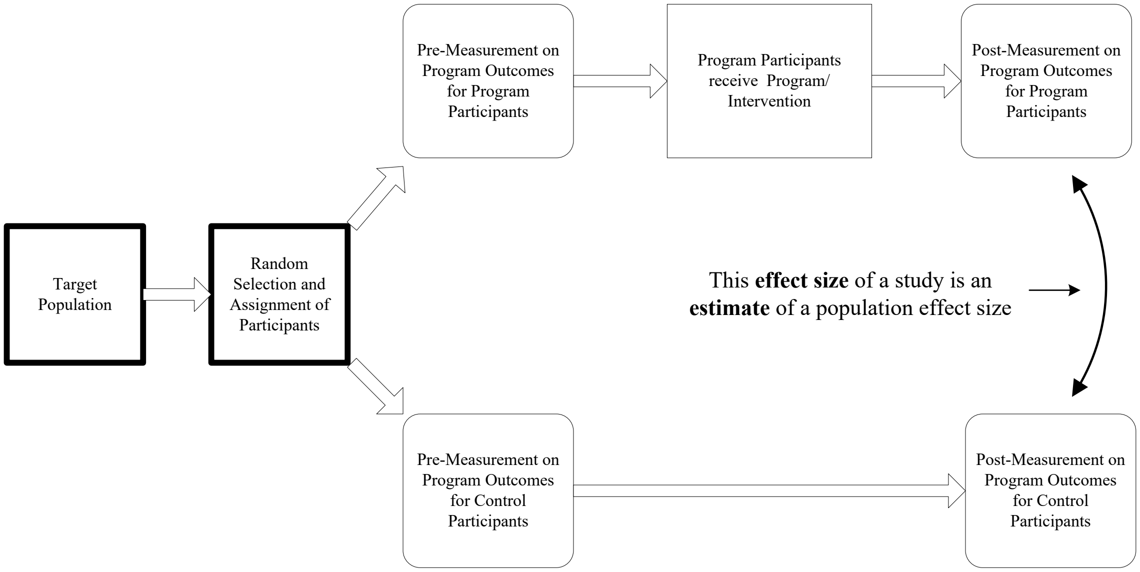
A Brief Discussion About Confidence Intervals

Many of our students have voiced confusion or uncertainty about the role of confidence intervals in a meta-analysis. We did a very brief discussion about confidence interval in the discussion we had about inferential statistics in Week 5. In that discussion, we indicted that inferential statistics are used to make inferences about a larger group of persons using data obtained from a smaller subset of persons who are members of that large group. The larger group of persons is called a **population.** The subset of persons who provide data is called a **sample**. In order to use data obtained from a sample to make inferences about the larger population researchers need to make sure that the sample is **representative** of the population, that is, that the sample accurate reflects population characteristics. The best way to ensure that a sample represents a population is to use a **random selection** process. To be successful, random selection follows some very precise rules using random numbers.

In Week 4, we also discussed the fact that a randomized controlled trial is a form of an inferential study. You can see in the following diagram that a randomized controlled trial starts with a target population from which a sample of study participants is selected using a random selection methodology. This sample of participant is then assigned (or allocated) to the intervention group or the control group using a random assignment methodology. The effect size computed at post-measurement is considered to be an estimate of the effect size in the target population. We can never know the true effect size so we have to rely on this estimate as a best guess.

The confidence intervals we see in forest plots tell us quite a bit about this estimate. A wide confidence interval indicates that the estimate is not very precise – that there is substantial variability in the possible estimates we might see. A narrow confidence interval indicates that the effect size estimate is precise – that there is not as much variability in the possible estimates we might see.

Finally, confidence intervals are useful in determining whether or not a study finding was statistically significant. A study that has a confidence interval that cross the line of no effect (the perpendicular line in a forest plot) would not be statistically significant.


Pre-Measurement on Program Outcomes for Program Participants
Program Participants receive Program/ Intervention
Post-Measurement on Program Outcomes for Program Participants
Target Population
Random Selection and Assignment of Participants
This **effect size** of a study is an **estimate** of a population effect size
Pre-Measurement on Program Outcomes for Control Participants
Post-Measurement on Program Outcomes for Control Participants

The following video presents an understandable discussion about confidence intervals.

https://www.youtube.com/watch?v=llXEGuxvh28