Chapter 453

Meta-Analysis of Two Means

Introduction

This module performs a meta-analysis on a set of two-group, continuous-scale, studies. These studies have a treatment group and a control group. Each study's result may be summarized by the sample size, mean, and standard deviation of each of the two groups. The program provides a complete set of numeric reports and plots to allow the investigation and presentation of the studies. Diagnostic plots include the *forest plot, radial plot, L'Abbe plot,* and *funnel plot.* Both fixed- and random-effects models are available for analysis with a variety of settings.

Meta-Analysis refers to methods for the systematic review of a set of individual studies with the aim to combine their results. Meta-analysis has become popular for a number of reasons:

1. The adoption of evidence-based medicine, which requires that all reliable information is considered.
2. The desire to avoid narrative reviews which are often misleading.
3. The desire to interpret the large number of studies that may have been conducted about a specific treatment.
4. The desire to increase the statistical power of the results by combining many small-size studies.

The goals of meta-analysis may be summarized as follows. A meta-analysis seeks to systematically review all pertinent evidence, provide quantitative summaries, integrate results across studies, and provide an overall interpretation of these studies.

We have found many books and articles on meta-analysis. In this chapter, we briefly summarize the information in Schwarzer et al (2015), Chen and Peace (2021), Harrer et al (2022), Sutton et al (2000), Thompson (1998), and Palmer and Sterne (2016). Refer to those sources for more details about the meta-analysis formulas used.

For a great introduction we recommend Chapter 10 of the Cochrane Handbook. The reference for this chapter is Deeks, Higgins, and Altman (2021).

Treatment Effects

Suppose you have obtained the results for *k* studies, labeled *i = 1, ..., k*. Each study consists of a treatment group (1) and a control group (2). The results of each study are summarized by six statistics:

n_{T_i} the number of subjects in the treatment group.

n_{C_i} the number of subjects in the control group.

$\bar{x}_{T_i}$ the sample mean of the treatment group which estimates the treatment mean μ_{Ti}.

$\bar{x}_{C_i}$ the sample mean of the control group which estimates the control mean μ_{Ci}.

s_{T_i} the sample standard deviation of the treatment group.

s_{C_i} the sample standard deviation of the control group.

Mean Difference

The scales (e.g., blood pressure, pulse rate, volume, etc.) of all studies must be the same. If the logarithm has been taken in one study, it must be taken in all studies. You cannot combine studies with different scales using this procedure!

The measure of treatment effect for study *i* is

$$\theta_i = \mu_{T_i} - \mu_{C_i}$$

which is estimated by

$$\hat{\theta}_i = \bar{x}_{T_i} - \bar{x}_{C_i}$$

Unpooled Variance Estimate

If the two group standard deviations are assumed to be different, the standard deviation of the mean difference is

$$V(\hat{\theta}_i) = \left(\frac{\sigma_{T_i}^2}{n_{T_i}} + \frac{\sigma_{C_i}^2}{n_{C_i}}\right)$$

The value of $V(\hat{\theta}_i)$ is estimated by

$$\hat{V}(\hat{\theta}_i) = \left(\frac{s_{T_i}^2}{n_{T_i}} + \frac{s_{C_i}^2}{n_{C_i}}\right)$$

Confidence intervals based on the *normal* distribution are defined for θ_i as follows.

$$\hat{\theta}_i \pm z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\theta}_i)}$$

Pooled Variance Estimate

If the two group standard deviations are assumed to be equal, the standard deviation of the mean difference is

$$V(\hat{\theta}_i) = \sigma_i^2 \left(\frac{1}{n_{T_i}} + \frac{1}{n_{C_i}} \right)$$

The value of σ_i^2 is estimated by the pooled sample standard deviation given by

$$s_i^2 = \frac{(n_{T_i} - 1)s_{T_i}^2 + (n_{C_i} - 1)s_{C_i}^2}{n_{T_i} + n_{C_i}}$$

so that

$$\hat{V}(\hat{\theta}_i) = s_i^2 \left(\frac{1}{n_{T_i}} + \frac{1}{n_{C_i}} \right)$$

Confidence intervals based on the *t* distribution are defined for θ_i in the usual manner.

$$\hat{\theta}_i \pm t_{n_{T_i}+n_{C_i}-2,1-\alpha/2} \sqrt{\hat{V}(\hat{\theta}_i)}$$

Hypothesis Tests for Combined Studies

Several hypothesis tests have been developed to test various aspects about the variation in the effects. This variation is referred to as heterogeneity. Three statistical tests have been devised to test the overall null hypothesis that all treatment effects (mean differences) are the same across studies.

Test Name	Null Hypothesis	Alternative Hypothesis	Test Statistic	Distribution
Nondirectional	$H_0 : \theta_i = 0 \;\; i = 1, \cdots, k$	$H_1 : \theta_i \neq 0 \;\;$ for some i	$X_{ND} = \sum_{i=1}^{k} w_i \hat{\theta}_i^2$	χ_k^2
Directional	$H_0 : \theta_i = \theta \;\; i = 1, \cdots, k$	$H_1 : \theta_i \neq \theta \neq 0 \;\;$ for some i	$X_D = \frac{\left(\sum_{i=1}^{k} w_i \hat{\theta}_i\right)^2}{\sum_{i=1}^{k} w_i}$	χ_1^2
Cochran's Q	$H_0 : \theta_i = \theta \;\; i = 1, \cdots, k$	$H_1 : \theta_i \neq \theta \;\; \neq 0$ for some i	$Q = \sum_{i=1}^{k} w_i \left(\hat{\theta}_i - \hat{\theta}\right)^2$	χ_{k-1}^2
			$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i}$	

Cochran's Q is sometimes called the *heterogeneity test*.

Measures that Quantify Heterogeneity

Several statistics have been proposed to quantify and interpret the heterogeneity that is found in a particular meta-analysis.

H Index

Higgins and Thompson (2002) (see also Chen and Peace (2021) page 157) present a heterogeneity measure which is based on the fact that $E(Q) = k - 1$. This leads to

$$H = \sqrt{\frac{\mathrm{Q}}{k-1}}$$

A reference interval for H is

$$LL_H = \exp\left\{\ln(H) - \left|z_{\frac{\alpha}{2}}\right| SE_{\ln(H)}\right\}$$

$$UL_H = \exp\left\{\ln(H) + \left|z_{\frac{\alpha}{2}}\right| SE_{\ln(H)}\right\}$$

where

$$SE_{\ln(H)} = \begin{cases} \dfrac{\ln(Q) - \ln(k-1)}{2\left(\sqrt{2Q - \sqrt{2k-3}}\right)} & \text{if } Q > k \\ \sqrt{\left(\dfrac{1}{2k-4}\right)\left(1 - \dfrac{1}{3(k-2)^2}\right)} & \text{if } Q \le k \end{cases}$$

If $LL_H > 1$, H is statistically significant.

I^2 Index

Higgins and Thompson (2002) (see also Chen and Peace (2021) page 158) present an *inconsistency* measure called I^2

$$I^2 = 100\% \frac{\hat{\tau}^2}{\hat{\tau}^2 + s^2}$$

where s^2 is a 'typical' estimate of the within-study variance given by Bowden *et al*. (2011) page 3 as

$$s^2 = \frac{(k-1)\sum w_i}{(\sum w_i)^2 - \sum w_i^2}$$

This measure is related to H^2 as follows

$$I^2 = 100\% \frac{(H^2 - 1)}{H^2}$$

Thus, a reference interval for I^2 can be derived from the reference interval for H^2 as

$$LL_{I^2} = 100\% \left(\frac{LL_H^2 - 1}{LL_H^2} \right)$$

$$UL_{I^2} = 100\% \left(\frac{UL_H^2 - 1}{UL_H^2} \right)$$

Higgins *et al.* (2003) explain that this index does not depend on the number of studies as do other heterogeneity measures. They suggest that 25%, 50%, and 75% are reasonable cutoffs for low, medium, and high values of the index.

A generalized version of the I^2 index, which we call I_{PM}^2, along with its reference interval, can then be found by inserting τ_{PM}^2 and its two reference limits into the equation

$$I^2 = 100\% \left(\frac{\hat{\tau}^2}{\hat{\tau}^2 + s^2} \right)$$

which was given above. Similar to results hold for H^2.

Effects Model

Two types of models have been adopted to allow the testing of various hypotheses and creating confidence intervals for the studies in the meta-analysis. The simplest model is the *fixed effects model*. It is used when the studies in the analysis all have similar characteristics. A second model, the *random effects model*, is a little more complicated but is usually more realistic in its assumptions. These models are described next.

Fixed Effects Model

The fixed effects model is given by

$$\hat{\theta}_i = \theta + \sigma e_i$$

where $e_i = N(0,1)$ so that $\hat{\theta}_i \sim N(\theta, \sigma^2)$.

If the effects are assumed to be equal (homogeneous) except for a small random error, the *fixed effects model* may be used to construct a combined confidence interval. The fixed effects model assumes homogeneity of study results. That is, it assumes that a common population effect θ for all *i*. This assumption may not be realistic when combining studies with different patient pools, protocols, follow-up strategies, doses, durations, etc.

Fixed Study Weights

During the analysis, weights w_i are assigned to each study. These weights are often created from the variance as follows.

$$v_i = \hat{V}(\hat{\theta}_i)$$

$$w_i = 1/v_i$$

Inverse Variance Method for Combining Studies

If the fixed effects model is adopted, the *inverse variance* method, described by Sutton (2000) page 58, is used to calculate combined confidence intervals and significance tests for θ. The basic assumption that is used for the fixed effect model is that the studies responses all vary about a single population parameter θ.

The basic formula used for combining confidence interval is

$$\hat{\theta} \pm z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\theta})}$$

where $z_{1-\alpha/2}$ is the appropriate percentage point from the standardized normal distribution, and

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i}$$

$$\hat{V}(\hat{\theta}) = \frac{1}{\sum_{i=1}^{k} w_i}$$

Random Effects Model

The random effects model assumes that the individual θ_i are normally distributed with mean θ and between-study variance τ^2. Palmer and Sterne (2016) page 25 indicate that either the Mantel-Haenszel or the Inverse-Variance estimate may be used for the combined estimate.

The random effects model is given by

$$\hat{\theta}_i = \theta + t_i + \sigma_i e_i$$

where t_i and e_i are independent. Note that $t_i \sim N(0, \tau^2)$ and $e_i = N(0,1)$ so that $\hat{\theta}_i \sim N(\theta, \tau^2 + \sigma^2)$.

Random Study Weights

The individual study weights for the random effects model are

$$w_i^* = \frac{1}{V(\hat{\theta}_i) + \tau^2}$$

where τ^2 is estimated using one of the techniques shown n.

τ^2 Estimation

A very important quantity in meta-analysis is the between-study variance τ^2. **NCSS** provides two estimators of this quantity, one called the DerSimonian and Laird (DL) estimator and the other called the Paule and Mandel (PM) estimator. The PM estimator has been recommended by several recent articles.

DerSimonian and Laird Estimator of τ^2

The between-study variance τ^2 can be estimated by the DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$ (see Higgins and Thompson (2002) page 1543 and Veroniki *et al.* (2016) pages 60 - 61).

$$\hat{\tau}^2_{DL} = \max\left[0, \frac{Q-(k-1)}{\sum w_i - \left(\frac{\sum w_i^2}{\sum w_i}\right)}\right]$$

Note that this formula uses the fixed effect weights w_i.

Using the definition of I^2 in terms of $\hat{\tau}^2$ (see below), a simple confidence interval for τ^2 can be derived by plugging in the I^2 reference limits LL_{I^2} and UL_{I^2} into the rearranged formula

$$\hat{\tau}^2 = s^2\left(\frac{I^2}{100 - I^2}\right)$$

Random effect weights are

$$w_i^* = \frac{1}{V(\hat{\theta}_i) + \hat{\tau}^2_{DL}}$$

The combined estimate of the effect measure is

$$\hat{\theta}_{DL} = \frac{\sum w_i^* \hat{\theta}_i}{w_i^*}$$

The variance of this estimate is

$$V(\hat{\theta}_{DL}) = \frac{1}{\sum w_i^*}$$

The formulas used for the confidence interval are of the form

$$\hat{\theta}_{DL} \pm z_{1-\alpha/2}\sqrt{V(\hat{\theta}_{DL})}$$

where $z_{1-\alpha/2}$ is the appropriate percentage point from the standardized normal distribution.

The formula used for the one degree-of-freedom heterogeneity test is

$$z = \frac{\hat{\theta}_{DL}}{\sqrt{V(\hat{\theta}_{DL})}}$$

Paule and Mandel Estimator of τ²

The between-study variance τ^2 can be estimated by the Paule and Mandel Q-profile estimator $\hat{\tau}^2_{PM}$ (see Bowden *et al.* (2011) page 3 and Veroniki *et al.* (2016) pages 63 - 64). Paule and Mandel proposed that the generalized Q-statistic be iterated for various values of τ^2 until $Q_{gen}(\tau^2) = k - 1$, its expected value.

$$Q_{gen}(\tau^2) = \sum_{i=1}^{k} u_i\left(\hat{\theta}_i - \tilde{\theta}_{PM}\right)^2$$

where

$$u_i = \frac{1}{V\left(\hat{\theta}_i\right) + \tau^2}$$

$$\tilde{\theta}_{PM} = \frac{\sum u_i \hat{\theta}_i}{\sum u_i}$$

The variance of this estimate is

$$V\left(\tilde{\theta}_{PM}\right) = \frac{1}{\sum u_i}$$

Viechtbauer (2007) proposed that an α-level confidence set for τ^2_{PM} is found by iterating the same generalized Q-statistic, this time solving for the values of τ^2 which make $Q_{gen}(\tau^2) = \chi^2_{k-1,1-\alpha/2}$ and $Q_{gen}(\tau^2) = \chi^2_{k-1,\alpha/2}$. The two resulting values of τ^2 form a non-symmetric confidence interval for τ^2_{PM}.

Knapp-Hartung Adjustment

The Knapp-Hartung adjustment, Knapp and Hartung (2003), is recommended when using a random effects model. See, for example, IntHout *et al.* (2014). This procedure requires the following two adjustments.

1. A Student's-t distribution with k-1 degrees of freedom replaces the normal distribution in the confidence intervals and significance tests.
2. The variance of the estimated effect measure is now

$$\mathrm{V}\left(\hat{\theta}_{KH}\right) = \frac{\sum w_i\left(\hat{\theta}_i - \hat{\theta}_{DL}\right)^2}{(k-1)\sum w_i}$$

Graphical Displays

A number of plots have been devised to display the information in a meta-analysis. These include the forest plot, the radial plot, the funnel plot, and the L'Abbe plot. More will be said about each of these plots in the Output section.

Data Structure

The data are entered into a dataset using one row per study. Six variables are required to hold the sample size, mean, and standard deviation of each study. In addition to these, an additional variable is usually used to hold a short (3 or 4 character) label. Another variable may be used to hold a grouping variable.

As an example, we will use a slightly modified version of the data referred to in Sutton (2000) page 30 as the dental dataset. This dataset reviews nine randomized clinical trials that were conducted to study the effects of sodium fluoride (NaF) with sodium monofluorophosphate (SMFP). These nine studies were all on the same continuous scale, so their results could be analyzed using the meta-analysis techniques presented in this chapter. These data are contained in the *SUTTON30 modified* database. You should load this database to see how the data are arranged.

Example 1 – Meta-Analysis of Means

This section presents an example of how to analyze the data contained in the Sutton30 modified dataset. This dataset contains data for 9 randomized clinical trials.

Setup

To run this example, complete the following steps:

1 Open the Sutton30 modified example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Sutton30 modified** and click **OK**.

2 Specify the Meta-Analysis of Two Means procedure options

- Find and open the **Meta-Analysis of Two Means** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Treatment - N Variables	**1**
Treatment - Mean Variables	**2**
Treatment - S.D. Variables	**3**
Control - N Variables	**4**
Control - Mean Variables	**5**
Control - S.D. Variables	**6**
Row Label Variable	**7**
Subgroup Variable	**<Empty>**
Group Variances Are	**Unequal**
Knapp-Hartung Adjustment	**Used**
τ^2 Method	**Paule and Mandel (PM)**
I^2 and H^2 Method	**DerSimonian and Laird (DL)**

Reports Tab

Run Summary	**Checked**
Data Summary	**Checked**
Analysis (Combined Only)	**Unchecked**
Analysis (Combined, Individual)	**Checked**
Q Heterogeneity Reports	**Checked**
τ^2 Between-Study Variation	**Checked**
s^2 Within-Study Variation	**Checked**
I^2 Inconsistency Index	**Checked**
H^2 Relative Excess in Heterogeneity	**Checked**
Confidence Level	**95**

Plots Tab

Forest Plot	**Checked**
Radial Plot	**Checked**
L'Abbe Plot	**Checked**
Horizontal Funnel Plot	**Checked**
Vertical Funnel Plot	**Unchecked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary

Run Summary Section

Parameter	Variable	Parameter	Value
Treatment N	NaFN	Rows Processed	9
Treatment Mean	NaFMean	Rows Filtered Out	0
Treatment SD	NaFSD	Rows with Missing Values*	0
Control N	SMFPN	Rows Analyzed	9
Control Mean	SMFPMean	Number of Observations	7089
Control SD	SMFPSD		
Subgroup	None	Number of Subgroups	1
Row Label	Study		

* Rows are excluded when they have missing values, nonpositive N's or SD's, or text for numeric values.

This report records the variables that were used and the number of rows that were processed.

Data Summary

Data Summary

Group Variances Are: Unequal

	Treatment			Control			Total		
Study	N	Mean	SD	N	Mean	SD	N	Difference	SE*
S1	134	5.960	4.240	113	6.820	4.720	247	-0.860	0.576
S2	175	4.740	4.640	151	5.070	5.380	326	-0.330	0.561
S3	137	2.040	2.590	140	2.510	3.220	277	-0.470	0.351
S4	184	2.700	2.320	179	1.200	2.460	363	1.500	0.251
S5	174	6.090	4.860	169	5.810	5.140	343	0.280	0.540
S6	754	4.720	5.330	736	8.760	5.290	1490	-4.040	0.275
S7	209	10.100	8.100	209	10.900	7.900	418	-0.800	0.783
S8	1151	2.820	3.050	1122	3.010	3.320	2273	-0.190	0.134
S9	679	3.880	8.850	673	4.370	5.370	1352	-0.490	0.398

* The standard error of the difference, SE, uses the unpooled variance formula.

This report shows the input data.

Treatment and Control Input Values

These six columns give the values of N, Mean, and Standard Deviation for each of the two groups: treatment and control.

Total Difference

This is the difference between the two group means.

SE

This is the estimated standard error of the estimated difference for each study.

Mean Difference Analysis

Mean Difference Analysis

Group Variances Are:	Unequal
τ^2 Method:	Paule and Mandel
Knapp-Hartung Adjustment:	Used

Study	Mean Difference	Standard Error	95% CI Limits for the Mean Difference		Test (Diff = 0)		Percent Weights	
			Lower	Upper	Value	P-Value	Fixed	Random
Combined								
Fixed	-0.462	0.095	-0.648	-0.276	-4.859	0.0000	100.0	
Random	-0.603	0.505	-1.768	0.563	-1.192	0.2673		100.0
Studies								
S1	-0.860	0.576	-1.988	0.268	-1.494	0.1352	2.7	10.7
S2	-0.330	0.561	-1.430	0.770	-0.588	0.5564	2.9	10.8
S3	-0.470	0.351	-1.157	0.217	-1.340	0.1802	7.3	11.4
S4	1.500	0.251	1.008	1.992	5.973	0.0000	14.3	11.7
S5	0.280	0.540	-0.779	1.339	0.518	0.6044	3.1	10.8
S6	-4.040	0.275	-4.579	-3.501	-14.684	0.0000	11.9	11.6
S7	-0.800	0.783	-2.334	0.734	-1.022	0.3067	1.5	9.9
S8	-0.190	0.134	-0.452	0.072	-1.420	0.1556	50.5	11.8
S9	-0.490	0.398	-1.270	0.290	-1.232	0.2180	5.7	11.3

This report displays results for the mean difference. The report gives you the combined results for the fixed and random models first, followed by the results for each study.

Mean Difference

This is the mean difference for each study. The first two lines are the combined results for all studies using either the Fixed-Effects Model or the Random Effects Model. Note that the combined values are different because the weights (shown in the last two columns) are different.

Standard Error

This column gives the combined standard errors for the fixed and random models followed by the standard errors for each of the studies.

95% CI Limits for the Mean Difference

These are the lower and upper confidence limits for the mean difference. The first two lines are the combined results for all studies using either the Fixed-Effects Model and the Random Effects Model.

Test (Diff = 0)

These two columns give the result of a hypothesis test of whether the mean difference is zero.

Percent Weight

These two columns give the percentage of the total of the individual weight that was used by this row. You can see how the model choice impacts the distribution of the weights.

Heterogeneity Tests

Heterogeneity Tests

	Cochran's Test			Directional Test			Nondirectional Test		
Rows	Q	DF	P-Value	χ^2_D	DF	P-Value	χ^2_{ND}	DF	P-Value
Combined	236.904	8	0.0000	23.6	1	0.0000	260.5	9	0.0000

Q This statistic (sometimes called 'The heterogeneity test') tests the null hypothesis that all effects are equal (homogeneous effects) versus the alternative that at least one effect had a different effect (heterogeneous effects).
χ^2_D This statistic tests the null hypothesis that all effects are equal (homogeneous effects) versus the alternative that at least one effect had a different effect (heterogeneous effects).
χ^2_{ND} This statistic tests the null hypothesis that all effects are zero versus the alternative that at least one effect is non-zero.

This reports the results of three chi-square tests designed to test whether all treatment effects are equal.

Cochran's Test

Q tests the null hypothesis that all effects are equal versus the alternative that at least one study had a different value (heterogeneous effects) than the rest. This is the computed chi-square value for this test.

Directional Test

This statistic tests the null hypothesis that all effects are equal versus the alternative that at least one study had a different effect.

Nondirectional Test

This statistic tests the null hypothesis that all effects are zero versus the alternative that at least one effect is non-zero.

τ² Between-Study Variation

τ² Between-Study Variation

τ² Method: Paule and Mandel

Variance			Standard Deviation		
	95% CI Limits			95% CI Limits	
τ²	Lower	Upper	τ	Lower	Upper
2.091	0.910	7.896	1.446	0.954	2.810

This report shows the estimated between-study variance with a confidence interval. This value might be compared with the estimated within-study variance given in the next report.

s² Within-Study Variation

s² Within-Study Variation

Variance s²	Standard Deviation s
0.103	0.322

This report shows the estimated within-study variance. This value might be compared with the estimated between-study variance given in the last report.

I² Inconsistency Index

I² Inconsistency Index

I² Method: DerSimonian and Laird

I² Index		
	95% CI Limits	
I²	Lower	Upper
96.623	95.106	97.670

The I^2 index may be interpreted as the proportion of total variation of treatment effects that is due to heterogeneity between studies.

H² Relative Excess in Heterogeneity

H² Relative Excess in Heterogeneity

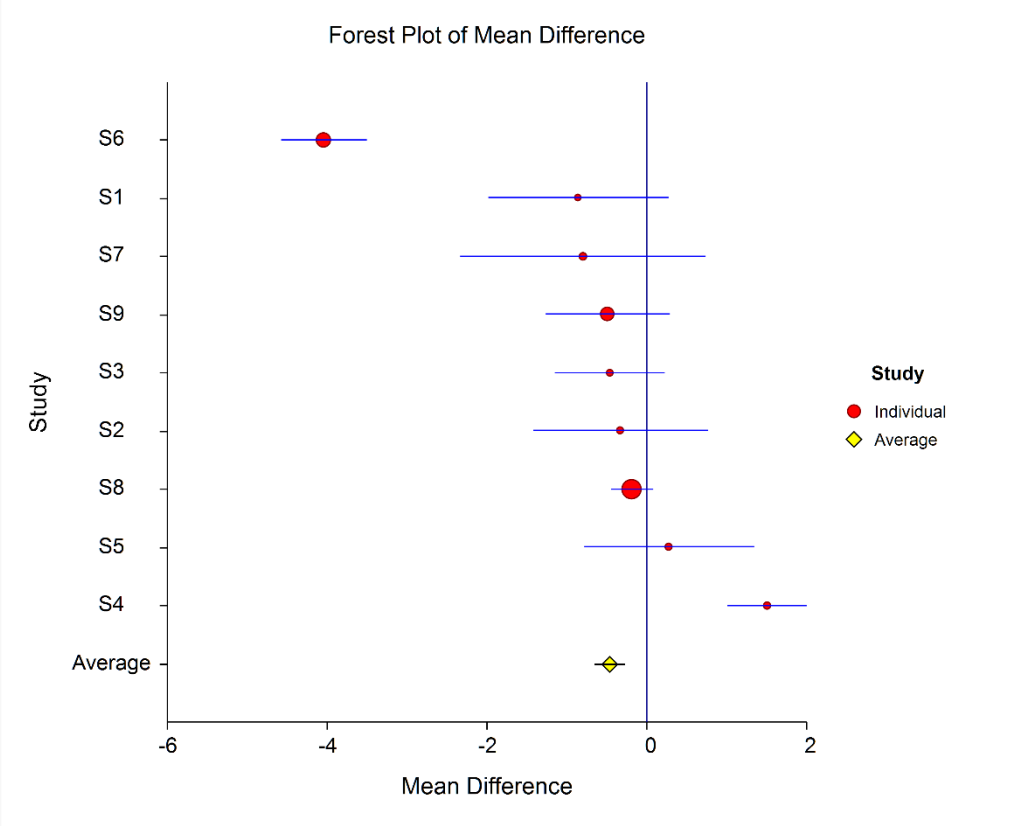
H² Method: DerSimonian and Laird

H² Index			H Index		
	95% CI Limits			95% CI Limits	
H²	Lower	Upper	H	Lower	Upper
29.613	20.432	42.920	5.442	4.520	6.551

The index H^2 is the ratio of Q (the heterogeneity) with its expected value. The minimum value of this index is one which occurs when there is no heterogeneity. A test heterogeneity is obtained by determining if the lower limit of the confidence interval is greater than one.

Forest Plot

Forest Plot

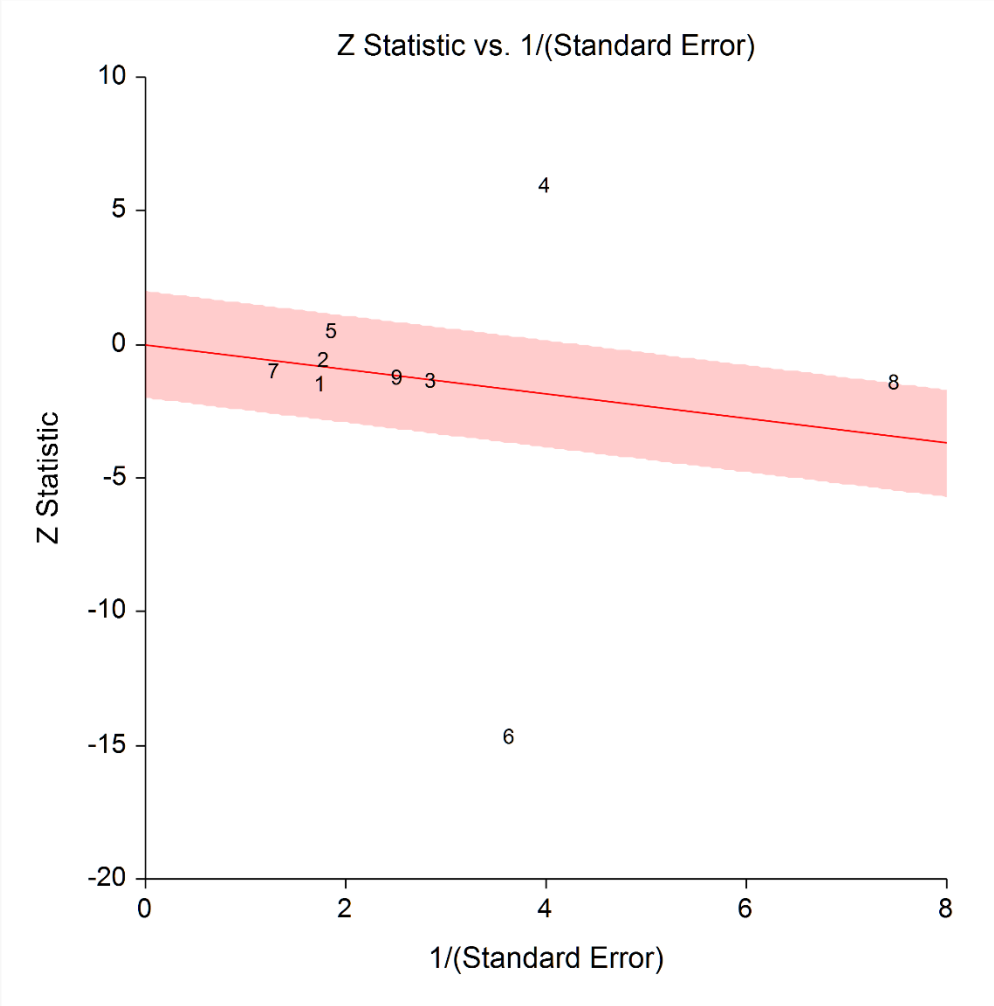


This plot presents the results for each study on one plot. The size of the plot symbol is proportional to the sample size of the study. The points on the plot are sorted mean difference. The lines represent the confidence intervals about the mean difference. Note that the narrower the confidence limits, the better.

By studying this plot, you can determine the main conclusions that can be drawn from the set of studies. For example, you can determine how many studies were significant (the confidence limits do not intersect the vertical line at 0.0).

Radial Plot

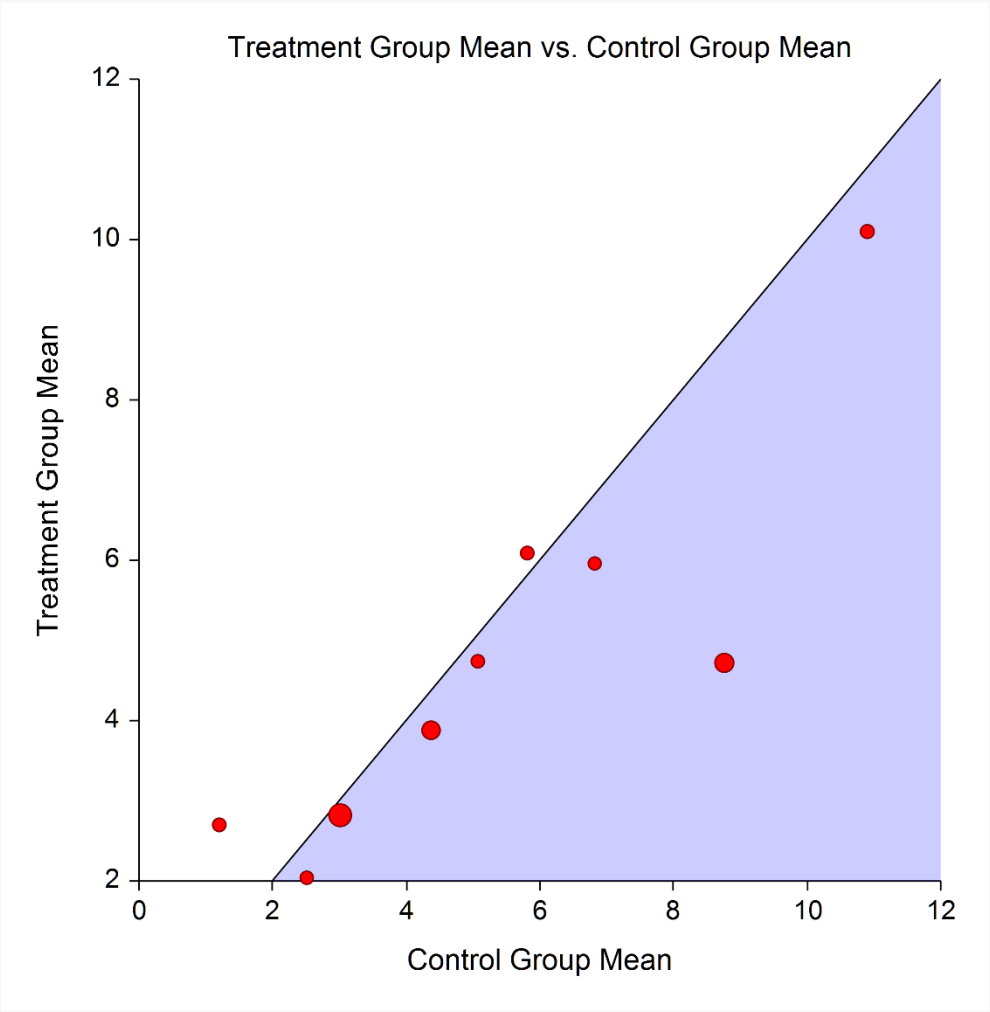
Radial Plot



The radial (or Galbraith) plot shows the z-statistic (outcome divided by standard error) on the vertical axis and a measure of weight on the horizontal axis. Studies that have the largest weight are closest to the Y axis. Studies within the limits are interpreted as homogeneous. Studies outside the limits may be outliers.

L'Abbe Plot

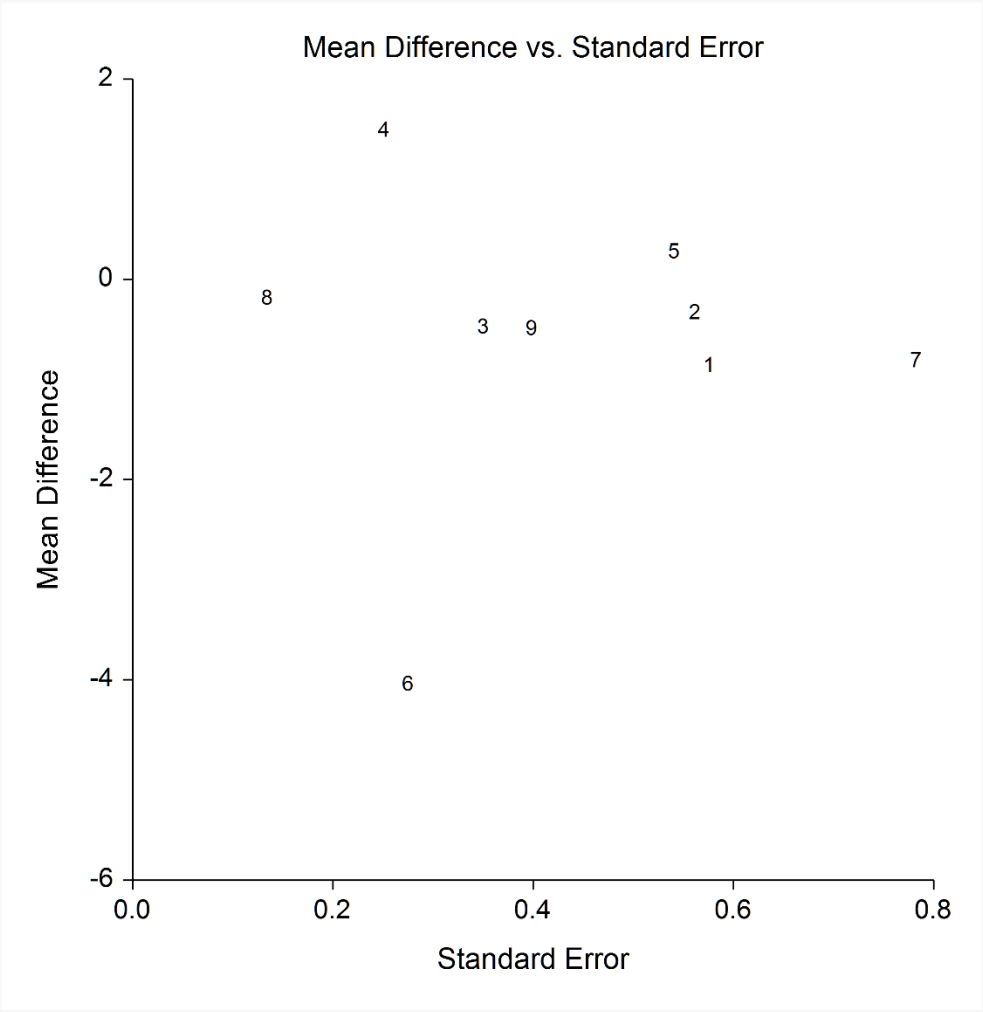
L'Abbe Plot



The L'Abbe plot displays the treatment mean on vertical axis versus the control mean on the horizontal axis. Homogenous studies will be arranged along the diagonal line. This plot is especially useful in determining if the relationship between the treatment group and the control group is the same for all values of the control group mean.

Horizontal Funnel Plot

Horizontal Funnel Plot



The funnel plot is often recommended to assess the publication bias in the meta-analysis. A plot with no bias opens out like a funnel to the right. Bias will appear as no studies in the upper right.

Several authors discouraged the use of the familiar triangle-shaped reference lines (not shown here) because they actually distract from the diagnosis. Also, the funnel plot isn't as useful when there are only a small number of studies as there are here.

We chose the horizontal orientation because it allows you to add various reference lines to the plot such as prediction limits and loess curve. However, the question of bias is probably most easily answered from the simple scatter plot shown here.

Example 2 – Meta-Analysis of Means with Subgroups

This section presents an example of how to analyze the data contained in the Sutton30 modified dataset, including groups.

Setup

To run this example, complete the following steps:

1 Open the Sutton30 modified example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Sutton30 modified** and click **OK**.

2 Specify the Meta-Analysis of Two Means procedure options

- Find and open the **Meta-Analysis of Two Means** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Treatment - N Variables	**1**
Treatment - Mean Variables	**2**
Treatment - S.D. Variables	**3**
Control - N Variables	**4**
Control - Mean Variables	**5**
Control - S.D. Variables	**6**
Row Label Variable	**7**
Subgroup Variable	**8**
Group Variances Are	**Unequal**
Knapp-Hartung Adjustment	**Used**
τ^2 Method	**Paule and Mandel (PM)**
I^2 and H^2 Method	**DerSimonian and Laird (DL)**

Reports Tab

Run Summary	**Checked**
Data Summary	**Checked**
Analysis (Combined Only)	**Unchecked**
Analysis (Combined, Individual)	**Checked**
Q Heterogeneity Reports	**Checked**
τ^2 Between-Study Variation	**Checked**
s^2 Within-Study Variation	**Checked**
I^2 Inconsistency Index	**Checked**
H^2 Relative Excess in Heterogeneity	**Checked**
Confidence Level	**95**

Plots Tab

Forest Plot ..... **Checked**
Radial Plot ..... **Checked**
L'Abbe Plot ..... **Checked**
Horizontal Funnel Plot ..... **Checked**
Vertical Funnel Plot..... **Unchecked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary

Run Summary Section

Parameter	Variable	Parameter	Value
Treatment N	NaFN	Rows Processed	9
Treatment Mean	NaFMean	Rows Filtered Out	0
Treatment SD	NaFSD	Rows with Missing Values*	0
Control N	SMFPN	Rows Analyzed	9
Control Mean	SMFPMean	Number of Observations	7089
Control SD	SMFPSD		
Subgroup	Group	Number of Subgroups	2
Row Label	Study		

* Rows are excluded when they have missing values, nonpositive N's or SD's, or text for numeric values.

This report records the variables that were used and the number of rows that were processed.

Data Summary

Data Summary

Group Variances Are: Unequal

[Group] Study	Treatment N	Treatment Mean	Treatment SD	Control N	Control Mean	Control SD	Total N	Total Difference	Total SE*
[A]									
S1	134	5.960	4.240	113	6.820	4.720	247	-0.860	0.576
S2	175	4.740	4.640	151	5.070	5.380	326	-0.330	0.561
S3	137	2.040	2.590	140	2.510	3.220	277	-0.470	0.351
S4	184	2.700	2.320	179	1.200	2.460	363	1.500	0.251
[B]									
S5	174	6.090	4.860	169	5.810	5.140	343	0.280	0.540
S6	754	4.720	5.330	736	8.760	5.290	1490	-4.040	0.275
S7	209	10.100	8.100	209	10.900	7.900	418	-0.800	0.783
S8	1151	2.820	3.050	1122	3.010	3.320	2273	-0.190	0.134
S9	679	3.880	8.850	673	4.370	5.370	1352	-0.490	0.398

* The standard error of the difference, SE, uses the unpooled variance formula.

This report shows the input data.

Treatment and Control Input Values

These six columns give the values of N, Mean, and Standard Deviation for each of the two groups: treatment and control.

Total Difference

This is the difference between the two group means.

SE

This is the estimated standard error of the estimated difference for each study.

Mean Difference Analysis

Mean Difference Analysis

Group Variances Are:	Unequal
τ^2 Method:	Paule and Mandel
Knapp-Hartung Adjustment:	Used

[Group] Study	Mean Difference	Standard Error	95% CI Limits for the Mean Difference		Test (Diff = 0)		Percent Weights	
			Lower	Upper	Value	P-Value	Fixed	Random
[Combined]								
Fixed	-0.462	0.095	-0.648	-0.276	-4.859	0.0000	100.0	
Random	-0.603	0.505	-1.768	0.563	-1.192	0.2673		100.0
[A]								
Fixed Model	0.541	0.182	0.184	0.898	2.971	0.0030	27.3	
Random Model	0.008	0.543	-1.697	1.756	0.055	0.9598		44.6
S1	-0.860	0.576	-1.988	0.268	-1.494	0.1352	2.7	10.7
S2	-0.330	0.561	-1.430	0.770	-0.588	0.5564	2.9	10.8
S3	-0.470	0.351	-1.157	0.217	-1.340	0.1802	7.3	11.4
S4	1.500	0.251	1.008	1.992	5.973	0.0000	14.3	11.7
[B]								
Fixed Model	-0.838	0.112	-1.057	-0.620	-7.518	0.0000	72.7	
Random Model	-1.067	0.783	-3.249	1.099	-1.373	0.2418		55.4
S5	0.280	0.540	-0.779	1.339	0.518	0.6044	3.1	10.8
S6	-4.040	0.275	-4.579	-3.501	-14.684	0.0000	11.9	11.6
S7	-0.800	0.783	-2.334	0.734	-1.022	0.3067	1.5	9.9
S8	-0.190	0.134	-0.452	0.072	-1.420	0.1556	50.5	11.8
S9	-0.490	0.398	-1.270	0.290	-1.232	0.2180	5.7	11.3

This report displays results for the mean difference. The report gives you the combined results for the fixed and random models first, followed by the individuals results of each study.

Mean Difference

This is the mean difference for each study. The first two lines are the combined results for all studies using either the Fixed-Effects Model or the Random Effects Model. Note that the combined values are different because the weights (shown in the last two columns) are different.

Standard Error

This column gives the combined standard errors for the fixed and random models followed by the standard errors for each of the studies.

95% CI Limits for the Mean Difference

These are the lower and upper confidence limits. The first two lines are the combined results for all studies using either the Fixed-Effects Model or the Random Effects Model.

Test (Diff = 0)

These two columns give the result of a hypothesis test of whether the mean difference is zero.

Percent Weight

These two columns give the percentage of the total of the individual weight that was used by this row. You can see how the model choice impacts the distribution of the weights.

Heterogeneity Tests

Heterogeneity Tests

	Cochran's Test			Directional Test			Nondirectional Test		
Group	Q	DF	P-Value	χ^2_D	DF	P-Value	χ^2_{ND}	DF	P-Value
A	31.227	3	0.0000	8.8	1	0.0030	40.1	4	0.0000
B	163.939	4	0.0000	56.5	1	0.0000	220.5	5	0.0000
Combined	236.904	8	0.0000	23.6	1	0.0000	260.5	9	0.0000

The individual group test values do not sum to the combined value because these group values are calculated using only the studies in the corresponding group.

Q This statistic (sometimes called 'The heterogeneity test') tests the null hypothesis that all effects are equal (homogeneous effects) versus the alternative that at least one effect had a different effect (heterogeneous effects).
χ^2_D This statistic tests the null hypothesis that all effects are equal (homogeneous effects) versus the alternative that at least one effect had a different effect (heterogeneous effects).
χ^2_{ND} This statistic tests the null hypothesis that all effects are zero versus the alternative that at least one effect is non-zero.

This reports the results of three chi-square tests designed to test whether all treatment effects are equal.

Cochran's Test

Q tests the null hypothesis that all effects are equal versus the alternative that at least one effect had a different effect (heterogeneous effects) than the rest. This is the computed chi-square value for this test.

Directional Test

This statistic tests the null hypothesis that all effects are equal versus the alternative that at least one effect had a different effect.

Nondirectional Test

This statistic tests the null hypothesis that all effects are zero versus the alternative that at least one effect is non-zero.

τ² Between-Study Variation

τ² Between-Study Variation

τ² Method: Paule and Mandel

[Subgroup] Group	Variance τ²	Variance 95% CI Limits Lower	Variance 95% CI Limits Upper	Standard Deviation τ	Standard Deviation 95% CI Limits Lower	Standard Deviation 95% CI Limits Upper
A	0.981	0.237	15.226	0.990	0.487	3.902
B	2.848	0.956	24.301	1.688	0.978	4.930
Combined	2.091	0.910	7.896	1.446	0.954	2.810

This report shows the estimated between-study variance with a confidence interval. This value might be compared with the estimated within-study variance given in the next report.

s² Within-Study Variation

s² Within-Study Variation

[Subgroup] Group	Variance s²	Standard Deviation s
A	0.158	0.397
B	0.103	0.321
Combined	0.103	0.322

This report shows the estimated within-study variance. This value might be compared with the estimated between-study variance given in the last report.

I² Inconsistency Index

I² Inconsistency Index

I² Method: DerSimonian and Laird

[Subgroup] Group	I² Index		
	I²	95% CI Limits Lower	Upper
A	90.393	78.399	95.727
B	97.560	96.093	98.476
Combined	96.623	95.106	97.670

The I^2 index may be interpreted as the proportion of total variation of treatment effects that is due to heterogeneity between studies.

H² Relative Excess in Heterogeneity

H² Relative Excess in Heterogeneity

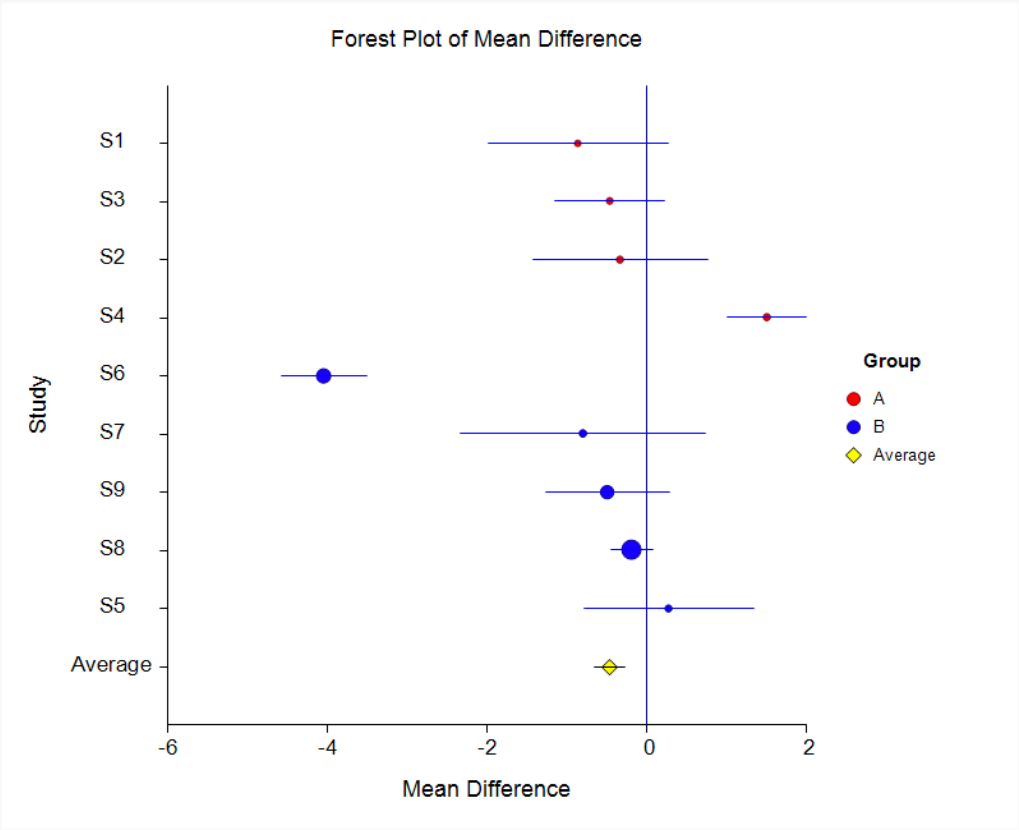
H² Method: DerSimonian and Laird

[Subgroup] Group	H² Index			H Index		
	H²	95% CI Limits Lower	Upper	H	95% CI Limits Lower	Upper
A	10.409	4.629	23.405	3.226	2.152	4.838
B	40.985	25.598	65.621	6.402	5.059	8.101
Combined	29.613	20.432	42.920	5.442	4.520	6.551

The index H^2 is the ratio of Q (the heterogeneity) with its expected value. The minimum value of this index is one which occurs when there is no heterogeneity. A test heterogeneity is obtained by determining if the lower limit of the confidence interval is greater than one.

Forest Plot

Forest Plot

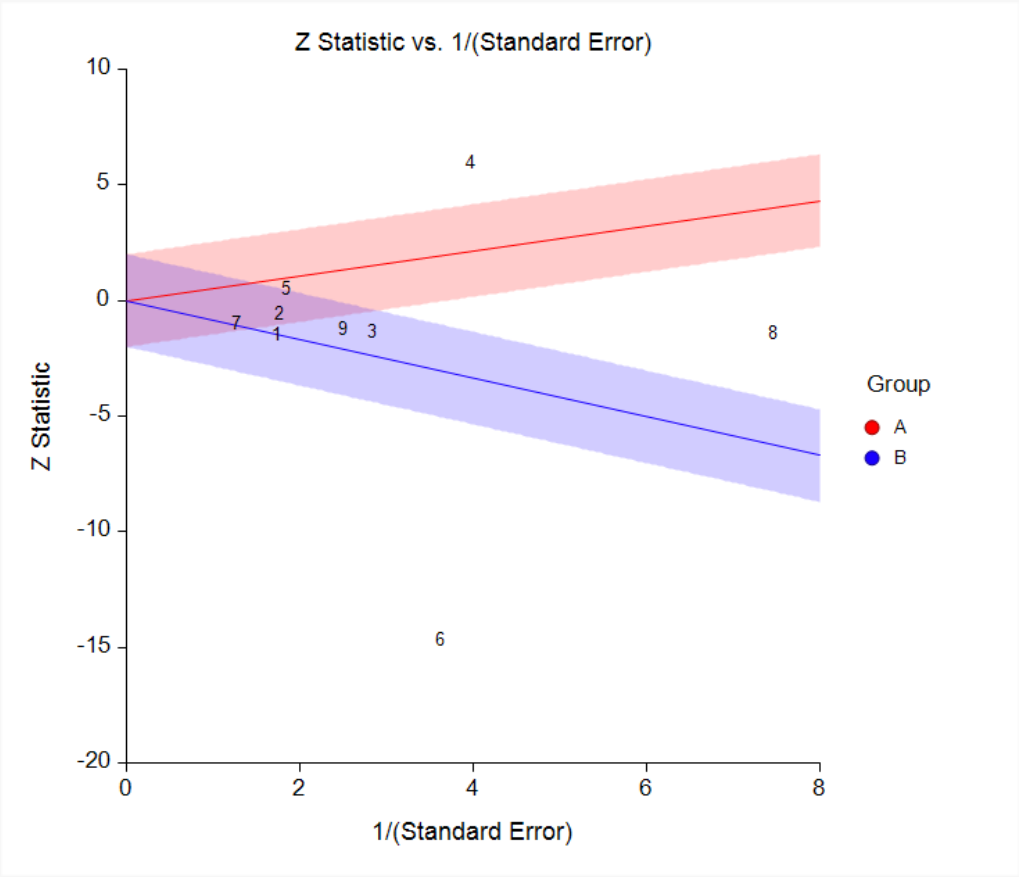


This plot presents the results for each study on one plot. The size of the plot symbol is proportional to the sample size of the study. The points on the plot are sorted mean difference. The lines represent the confidence intervals about the mean difference. Note that the narrower the confidence limits, the better.

By studying this plot, you can determine the main conclusions that can be drawn from the set of studies. For example, you can determine how many studies were significant (the confidence limits do not intersect the vertical line at 0.0).

Radial Plot

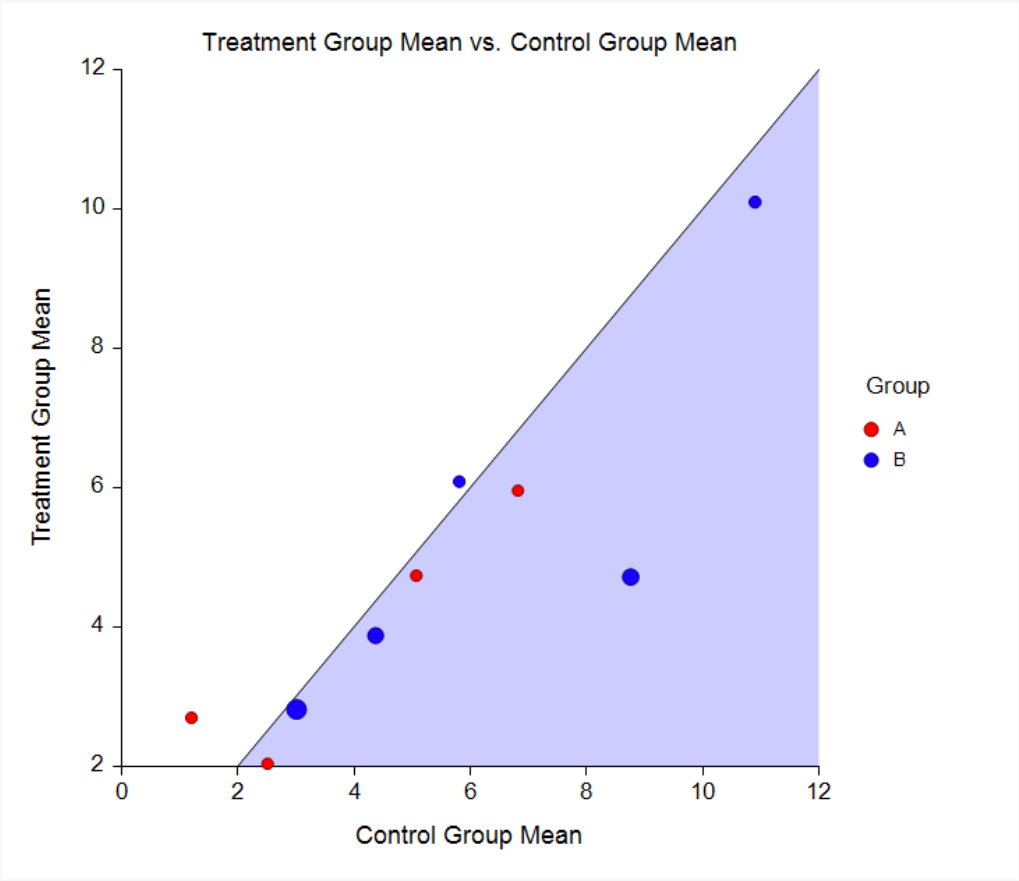
Radial Plot



The radial (or Galbraith) plot shows the z-statistic (outcome divided by standard error) on the vertical axis and a measure of weight on the horizontal axis. Studies that have the largest weight are closest to the Y axis. Studies within the limits are interpreted as homogeneous. Studies outside the limits may be outliers.

L'Abbe Plot

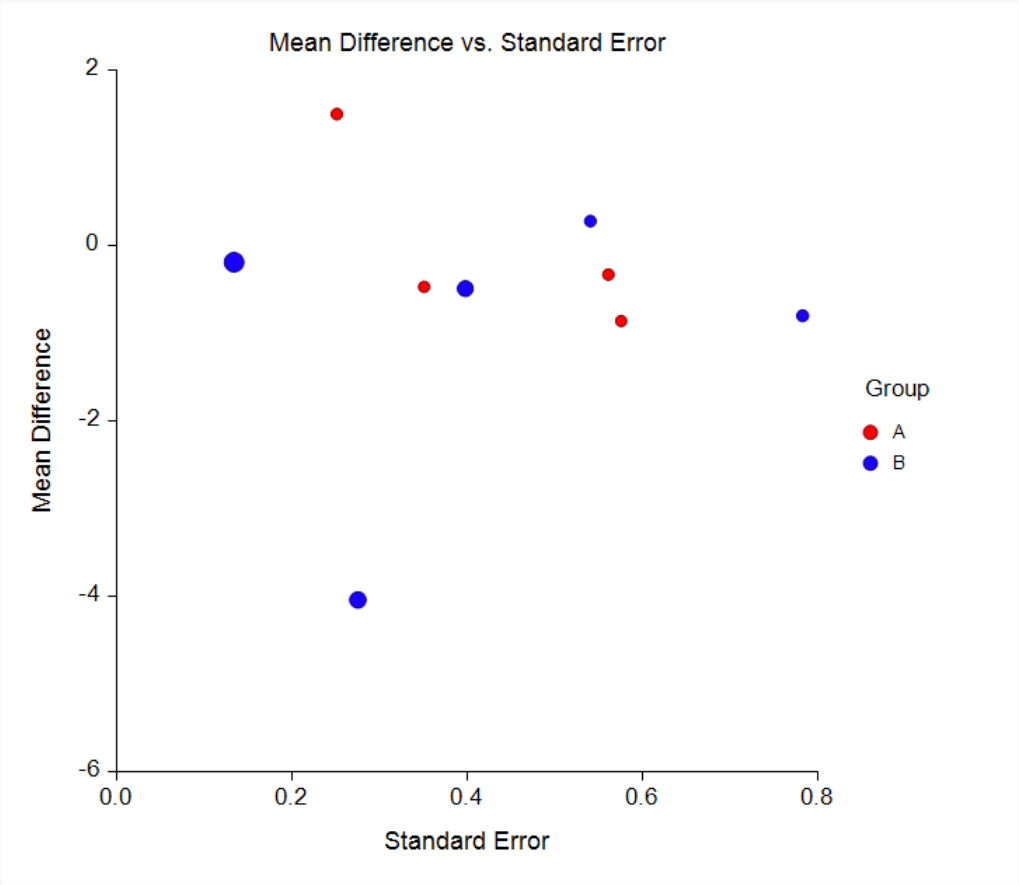
L'Abbe Plot



The L'Abbe plot displays the treatment mean on vertical axis versus the control mean on the horizontal axis. Homogenous studies will be arranged along the diagonal line. This plot is especially useful in determining if the relationship between the treatment group and the control group is the same for all values of the control group risk.

Horizontal Funnel Plot

Horizontal Funnel Plot



The funnel plot is often recommended to assess the publication bias in the meta-analysis. A plot with no bias opens out like a funnel to the right. Bias will appear as no studies in the upper right.

Several authors discouraged the use of the familiar triangle-shaped reference lines (not shown here) because they actually distract from the diagnosis.

We chose the horizontal orientation because it allows you to add various reference lines to the plot such as prediction limits and loess curve. However, the question of bias is probably most easily answered from the plain point plot shown here.