

## Description

Meta-analysis (Glass 1976) is a statistical technique for combining the results from several similar studies. The results of multiple studies that answer similar research questions are often available in the literature. It is natural to want to compare their results and, if sensible, provide one unified conclusion. This is precisely the goal of the meta-analysis, which provides a single estimate of the effect of interest computed as the weighted average of the study-specific effect estimates. When these estimates vary substantially between the studies, meta-analysis may be used to investigate various causes for this variation.

Another important focus of the meta-analysis may be the exploration and impact of small-study effects, which occur when the results of smaller studies differ systematically from the results of larger studies. One of the common reasons for the presence of small-study effects is publication bias, which arises when the results of published studies differ systematically from all the relevant research results.

Comprehensive overview of meta-analysis may be found in Sutton and Higgins (2008); Cooper, Hedges, and Valentine (2019); Borenstein et al. (2009); Higgins and Green (2017); Hedges and Olkin (1985); Sutton et al. (2000a); and Palmer and Sterne (2016). A book dedicated to addressing publication bias was written by Rothstein, Sutton, and Borenstein (2005).

This entry presents a general introduction to meta-analysis and describes relevant statistical terminology used throughout the manual. For how to perform meta-analysis in Stata, see [META] [meta](#).

## Remarks and examples

Remarks are presented under the following headings:

*Brief overview of meta-analysis*

*Meta-analysis models*

*Common-effect (“fixed-effect”) model*

*Fixed-effects model*

*Random-effects model*

*Comparison between the models and interpretation of their results*

*Meta-analysis estimation methods*

*Forest plots*

*Heterogeneity*

*Assessing heterogeneity*

*Addressing heterogeneity*

*Subgroup meta-analysis*

*Meta-regression*

*Publication bias*

*Funnel plots*

*Tests for funnel-plot asymmetry*

*The trim-and-fill method*

*Cumulative meta-analysis*

*Leave-one-out meta-analysis*

*Multivariate meta-regression*

*Multilevel meta-regression*

## Brief overview of meta-analysis

The term meta-analysis refers to the analysis of the data obtained from a collection of studies that answer similar research questions. These studies are known as *primary studies*. Meta-analysis uses statistical methods to produce an overall estimate of an effect, explore between-study heterogeneity, and investigate the impact of publication bias or, more generally, small-study effects on the final results. [Pearson \(1904\)](#) provides the earliest example of what we now call meta-analysis. In that reference, the average of study-specific correlation coefficients was used to estimate an overall effect of vaccination against smallpox on subjects' survival.

There is a lot of information reported by a myriad of studies, which can be intimidating and difficult to absorb. Additionally, these studies may report conflicting results in terms of the magnitudes and even direction of the effects of interest. For example, many studies that investigated the effect of taking aspirin for preventing heart attacks reported contradictory results. Meta-analysis provides a principled approach for consolidating all of this overwhelming information to provide an overall conclusion or reasons for why such a conclusion cannot be reached.

Meta-analysis has been used in many fields of research. See the Cochrane Collaboration (<https://us.cochrane.org/>) for a collection of results from meta-analysis that address various treatments from all areas of healthcare. Meta-analysis has also been used in econometrics (for example, [Dalhuisen et al. \[2003\]](#); [Woodward and Wui \[2001\]](#); [Hay, Knechel, and Wang \[2006\]](#); [Card, Kluge, and Weber \[2010\]](#)); education (for example, [Bernard et al. \[2004\]](#); [Fan and Chen \[2001\]](#)); psychology (for example, [Sin and Lyubomirsky \[2009\]](#); [Barrick and Mount \[1991\]](#); [Harter, Schmidt, and Hayes \[2002\]](#)); psychiatry (for example, [Hanji 2017](#)); criminology (for example, [Gendreau, Little, and Goggin \[1996\]](#); [Pratt and Cullen \[2000\]](#)); and ecology (for example, [Hedges, Gurevitch, and Curtis \[1999\]](#); [Gurevitch, Curtis, and Jones \[2001\]](#); [Winfree et al. \[2009\]](#); [Armqvist and Wooster \[1995\]](#)).

Meta-analysis is the statistical-analysis step of a systematic review. The term systematic review refers to the entire process of integrating the empirical research to achieve unified and potentially more general conclusions. Meta-analysis provides the theoretical underpinning of a systematic review and sets it apart from a narrative review; in the latter, an area expert summarizes the study-specific results and provides final conclusions, which could lead to potentially subjective and difficult-to-replicate findings. The theoretical soundness of meta-analysis made systematic reviews the method of choice for integrating empirical evidence from multiple studies. See [Cooper, Hedges, and Valentine \(2019\)](#) for more information as well as for various stages of a systematic review.

In what follows, we briefly describe the main components of meta-analysis: effect sizes, forest plots, heterogeneity, and publication bias.

**Effect sizes.** Effect sizes (or various measures of outcome) and their standard errors are the two most important components of a meta-analysis. They are obtained from each of the primary studies prior to the meta-analysis. Effect sizes of interest depend on the research objective and type of study. For example, in a meta-analysis comparing two groups, odds ratios and risk ratios are commonly used for binary outcomes and Hedges's  $g$  and Cohen's  $d$  measures for continuous outcomes. For meta-analysis estimating a single proportion (prevalence), the Freeman–Tukey-transformed proportions are typically used. When you deal with correlation data, the Fisher's  $z$ -transformed correlations are often used as the effect size. An overall effect size is computed as a weighted average of study-specific effect sizes, with more precise (larger) studies having larger weights. The weights are determined by the chosen meta-analysis model; see [Meta-analysis models](#). Also see [\[META\] meta esize](#) for how to compute various effect sizes in a meta-analysis.

**Meta-analysis models.** Another important consideration for meta-analysis is that of the underlying model. Three commonly used models are a common-effect, fixed-effects, and random-effects models. The models differ in how they estimate and interpret parameters. See *Meta-analysis models* for details.

**Meta-analysis summary—forest plots.** The results of meta-analysis are typically summarized on a forest plot, which plots the study-specific effect sizes and their corresponding confidence intervals, the combined estimate of the effect size and its confidence interval, and other summary measures such as heterogeneity statistics. See *Forest plots* for details.

**Heterogeneity.** The estimates of effect sizes from individual studies will inherently vary from one study to another. This variation is known as a study heterogeneity. Two types of heterogeneity described by *Deeks, Higgins, and Altman (2017)* are methodological, when the studies differ in design and conduct, and clinical, when the studies differ in participants, treatments, and exposures or outcomes. The authors also define statistical heterogeneity, which exists when the observed effects differ between the studies. It is typically a result of clinical heterogeneity, methodological heterogeneity, or both. There are methods for assessing and addressing heterogeneity that we discuss in detail in *Heterogeneity*.

**Publication bias.** The selection of studies in a meta-analysis is an important step. Ideally, all studies that meet prespecified selection criteria must be included in the analysis. This is rarely achievable in practice. For instance, it may not be possible to have access to some unpublished results. So some of the relevant studies may be omitted from the meta-analysis. This may lead to what is known in statistics as a sample-selection problem. In the context of meta-analysis, this problem is known as publication bias or, more generally, reporting bias. Reporting bias arises when the omitted studies are systematically different from the studies selected in the meta-analysis. For details, see *Publication bias*.

Finally, you may ask, Does it make sense to combine different studies? According to *Borenstein et al. (2009, chap. 40)*, “in the early days of meta-analysis, Robert Rosenthal was asked whether it makes sense to perform a meta-analysis, given that the studies differ in various ways and that the analysis amounts to combining apples and oranges. Rosenthal answered that *combining apples and oranges makes sense if your goal is to produce a fruit salad.*”

Meta-analysis would be of limited use if it could combine the results of identical studies only. The appeal of meta-analysis is that it actually provides a principled way of combining a broader set of studies and can answer broader questions than those originally posed by the included primary studies. The specific goals of the considered meta-analysis should determine which studies can be combined and, more generally, whether a meta-analysis is even applicable.

## Meta-analysis models

The role of a meta-analysis model is important for the computation and interpretation of the meta-analysis results. Different meta-analysis models make different assumptions and, as a result, estimate different parameters of interest. In this section, we describe the available meta-analysis models and point out the differences between them.

Suppose that there are  $K$  independent studies. Each study reports an estimate,  $\hat{\theta}_j$ , of the unknown true effect size  $\theta_j$  and an estimate,  $\hat{\sigma}_j$ , of its standard error,  $j = 1, 2, \dots, K$ . The goal of a meta-analysis is to combine these estimates in a single result to obtain valid inference about the population parameter of interest,  $\theta_{\text{pop}}$ .

Depending on the research objective and assumptions about studies, three approaches are available to model the effect sizes: a common-effect model (historically known as a fixed-effect model—notice the singular “effect”), a fixed-effects model (notice the plural “effects”), and a random-effects model. We briefly define the three models next and describe them in more detail later.

Consider the model

$$\hat{\theta}_j = \theta_j + \epsilon_j \quad j = 1, 2, \dots, K \quad (1)$$

where  $\epsilon_j$ 's are sampling errors and  $\epsilon_j \sim N(0, \sigma_j^2)$ . Although  $\sigma_j^2$ 's are unknown, meta-analysis does not estimate them. Instead, it treats the estimated values,  $\hat{\sigma}_j^2$ 's, of these variances as known and uses them during estimation. In what follows, we will thus write  $\epsilon_j \sim N(0, \hat{\sigma}_j^2)$ .

A common-effect model, as suggested by its name, assumes that all study effect sizes in (1) are the same and equal to the true effect size  $\theta$ ; that is,  $\theta_j = \theta_{j'} = \theta$  for  $j \neq j'$ . The research questions and inference relies heavily on this assumption, which is often violated in practice.

A fixed-effects model assumes that the study effect sizes in (1) are different,  $\theta_j \neq \theta_{j'}$  for  $j \neq j'$ , and “fixed”. That is, the studies included in the meta-analysis define the entire population of interest. So the research questions and inference concern only the specific  $K$  studies included in the meta-analysis.

A random-effects model also assumes that the study effect sizes in (1) are different,  $\theta_j \neq \theta_{j'}$  for  $j \neq j'$ , but that they are “random”. That is, the studies in the meta-analysis represent a sample from a population of interest. The research questions and inference extend beyond the  $K$  studies included in the meta-analysis to the entire population of interest.

The models differ in the population parameter,  $\theta_{\text{pop}}$ , they estimate; see [Comparison between the models and interpretation of their results](#). Nevertheless, they all use the weighted average as the estimator for  $\theta_{\text{pop}}$ :

$$\hat{\theta}_{\text{pop}} = \frac{\sum_{j=1}^K w_j \hat{\theta}_j}{\sum_{j=1}^K w_j} \quad (2)$$

However, they differ in how they define the weights  $w_j$ .

We describe each model and the parameter they estimate in more detail below.

### Common-effect (“fixed-effect”) model

As we mentioned earlier, a common-effect (CE) meta-analysis model ([Hedges 1982](#); [Rosenthal and Rubin 1982](#)) is historically known as a fixed-effect model. The term “fixed-effect model” is easy to confuse with the “fixed-effects model” (plural), so we avoid it in our documentation. The term “common-effect”, as suggested by [Rice, Higgins, and Lumley \(2018\)](#), is also more descriptive of the underlying model assumption. A CE model assumes a common (one true) effect for all studies in (1):

$$\hat{\theta}_j = \theta + \epsilon_j \quad j = 1, 2, \dots, K$$

The target of interest in a CE model is an estimate of a common effect size,  $\theta_{\text{pop}} = \theta$ . The CE model generally uses the weights  $w_j = 1/\hat{\sigma}_j^2$  in (2) to estimate  $\theta$ .

CE models are applicable only when the assumption that the same parameter underlies each study is reasonable, such as with pure replicate studies.

### Fixed-effects model

A fixed-effects (FE) meta-analysis model ([Hedges and Vevea 1998](#); [Rice, Higgins, and Lumley 2018](#)) is defined by (1); it assumes that different studies have different effect sizes ( $\theta_1 \neq \theta_2 \neq \dots \neq \theta_K$ ) and that the effect sizes are fixed quantities. By fixed quantities, we mean that the studies included in the meta-analysis define the entire population of interest. FE models are typically used whenever the analyst wants to make inferences only about the included studies.

The target of interest in an FE model is an estimate of the weighted average of true study-specific effect sizes,

$$\theta_{\text{pop}} = \text{Ave}(\theta_j) = \frac{\sum_{j=1}^K W_j \theta_j}{\sum_{j=1}^K W_j}$$

where  $W_j$ 's represent true, unknown weights, which are defined in [Rice, Higgins, and Lumley \(2018, eq. 3\)](#). The estimated weights,  $w_j = 1/\hat{\sigma}_j^2$ , are generally used in (2) to estimate  $\theta_{\text{pop}}$ .

Based on [Rice, Higgins, and Lumley \(2018\)](#), an FE model answers the question, “What is the magnitude of the average true effects in the set of  $K$  studies included in the meta-analysis?” It is appropriate when the true effects sizes are different across studies and the research interest lies in their average estimate.

## Random-effects model

A random-effects (RE) meta-analysis model ([Hedges 1983](#); [DerSimonian and Laird 1986](#)) assumes that the study effect sizes are different and that the collected studies represent a random sample from a larger population of studies. (The viewpoint of random effect sizes is further explored by Bayesian meta-analysis; see, for example, [Random-effects meta-analysis of clinical trials](#) in [\[BAYES\] bayesmh](#).) The goal of RE meta-analysis is to provide inference for the population of studies based on the sample of studies used in the meta-analysis.

The RE model may be described as

$$\hat{\theta}_j = \theta_j + \epsilon_j = \theta + u_j + \epsilon_j$$

where  $u_j \sim N(0, \tau^2)$  and, as before,  $\epsilon_j \sim N(0, \hat{\sigma}_j^2)$ . Parameter  $\tau^2$  represents the between-study variability and is often referred to as the heterogeneity parameter. It estimates the variability among the studies, beyond the sampling variability. When  $\tau^2 = 0$ , the RE model reduces to the CE model.

Here the target of inference is  $\theta_{\text{pop}} = E(\theta_j)$ , the mean of the distribution of effect sizes  $\theta_j$ 's.  $\theta_{\text{pop}}$  is estimated from (2) with  $w_j = 1/(\hat{\sigma}_j^2 + \hat{\tau}^2)$ .

## Comparison between the models and interpretation of their results

CE and FE models are computationally identical but conceptually different. They differ in their target of inference and the interpretation of the overall effect size. In fact, all three models have important conceptual and interpretation differences. [table 1](#) summarizes the different interpretations of  $\theta_{\text{pop}}$  under the three models.

Table 1. Interpretation of  $\theta_{\text{pop}}$  under various meta-analysis models

Model	Interpretation of $\theta_{\text{pop}}$
common-effect	common effect ( $\theta_1 = \theta_2 = \dots = \theta_K = \theta$ )
fixed-effects	weighted average of the $K$ true study effects
random-effects	mean of the distribution of $\theta_j = \theta + u_j$

A CE meta-analysis model estimates the true effect size under the strong assumption that all studies share the same effect and thus all the variability between the studies is captured by the sampling errors. Under that assumption, the weighted average estimator indeed estimates the true common effect size,  $\theta$ .

In the presence of additional variability unexplained by sampling variations, the interpretation of the results depends on how this variability is accounted for in the analysis.

An FE meta-analysis model uses the same weighted average estimator as a CE model, but the latter now estimates the weighted average of the  $K$  true study-specific effect sizes,  $\text{Ave}(\theta_j)$ .

An RE meta-analysis model assumes that the study contributions,  $u_j$ 's, are random. It decomposes the variability of the effect sizes into the between-study and within-study components. The within-study variances,  $\hat{\sigma}_j^2$ 's, are assumed known by design. The between-study variance,  $\tau^2$ , is estimated from the sample of the effect sizes. Thus, the extra variability attributed to  $\tau^2$  is accounted for during the estimation of the mean effect size,  $E(\theta_j)$ .

So which model should you choose? The literature recommends to start with a random-effects model, which is Stata's default for most meta-analyses. If you are willing to assume that the studies have different true effect sizes and you are interested only in providing inferences about these specific studies, then the FE model is appropriate. If the assumption of study homogeneity is reasonable for your data, a CE model may be considered.

## Meta-analysis estimation methods

Depending on the chosen meta-analysis model, various methods are available to estimate the weights  $w_j$  in (2). The meta-analysis models from the previous sections assumed the inverse-variance estimation method ([Whitehead and Whitehead 1991](#)) under which the weights are inversely related to the variance. The inverse-variance estimation method is applicable to all meta-analysis models and all types of effect sizes. Thus, it can be viewed as the most general approach.

For a two-group comparison of binary outcomes, CE and FE models also support the Mantel–Haenszel estimation method, which can be used to combine odds ratios, risk ratios, and risk differences. The classical Mantel–Haenszel method ([Mantel and Haenszel 1959](#)) is used for odds ratios, and its extension by [Greenland and Robins \(1985\)](#) is used for risk ratios and risk differences. The Mantel–Haenszel method is recommended with sparse data. [Fleiss, Levin, and Paik \(2003\)](#) also suggests that it be used with small studies provided that there are many.

In RE models, the weights are inversely related to the total variance,  $w_j = 1/(\hat{\sigma}_j^2 + \hat{\tau}^2)$ . Different methods are proposed for estimating the between-study variability,  $\tau^2$ , which is used in the expression for the weights. These include the restricted maximum likelihood (REML), maximum likelihood (ML), empirical Bayes (EB), DerSimonian–Laird (DL), Hedges (HE), Sidik–Jonkman (SJ), and Hunter–Schmidt (HS).

REML, ML, and EB are iterative methods, whereas other methods are noniterative (have closed-form expressions). The former estimators produce nonnegative estimates of  $\tau^2$ . The other estimators, except SJ, may produce negative estimates and are thus truncated at zero when this happens. The SJ estimator always produces a positive estimate of  $\tau^2$ .

REML, ML, and EB assume that the distribution of random effects is normal. The other estimators make no distributional assumptions about random effects. Below, we briefly describe the properties of each method. See Sidik and Jonkman (2007), Viechtbauer (2005), and Veroniki et al. (2016) for a detailed discussion and the merit of each estimation method.

The REML method (Raudenbush 2009) produces an unbiased, nonnegative estimate of  $\tau^2$  and is commonly used in practice. (It is the default estimation method in Stata because it performs well in most scenarios.)

When the number of studies is large, the ML method (Hardy and Thompson 1998; Thompson and Sharp 1999) is more efficient than the REML method but may produce biased estimates when the number of studies is small, which is a common case in meta-analysis.

The EB estimator (Berkey et al. 1995), also known as the Paule–Mandel estimator (Paule and Mandel 1982), tends to be less biased than other RE methods, but it is also less efficient than REML or DL (Knapp and Hartung 2003).

The DL method (DerSimonian and Laird 1986), historically, is one of the most popular estimation methods because it does not make any assumptions about the distribution of the random effects and does not require iteration. But it may underestimate  $\tau^2$ , especially when the variability is large and the number of studies is small. However, when the variability is not too large and the studies are of similar sizes, this estimator is more efficient than other noniterative estimators HE and SJ. See Veroniki et al. (2016) for details and relevant references.

The SJ estimator (Sidik and Jonkman 2005), along with the EB estimator, is the best estimator in terms of bias for large  $\tau^2$  (Sidik and Jonkman 2007). This method always produces a positive estimate of  $\tau^2$  and thus does not need truncating at 0, unlike the other noniterative methods.

Like DL, the HE estimator (Hedges 1983) is a method of moments estimator, but, unlike DL, it does not weight effect-size variance estimates (DerSimonian and Laird 1986). Veroniki et al. (2016) note, however, that this method is not widely used in practice.

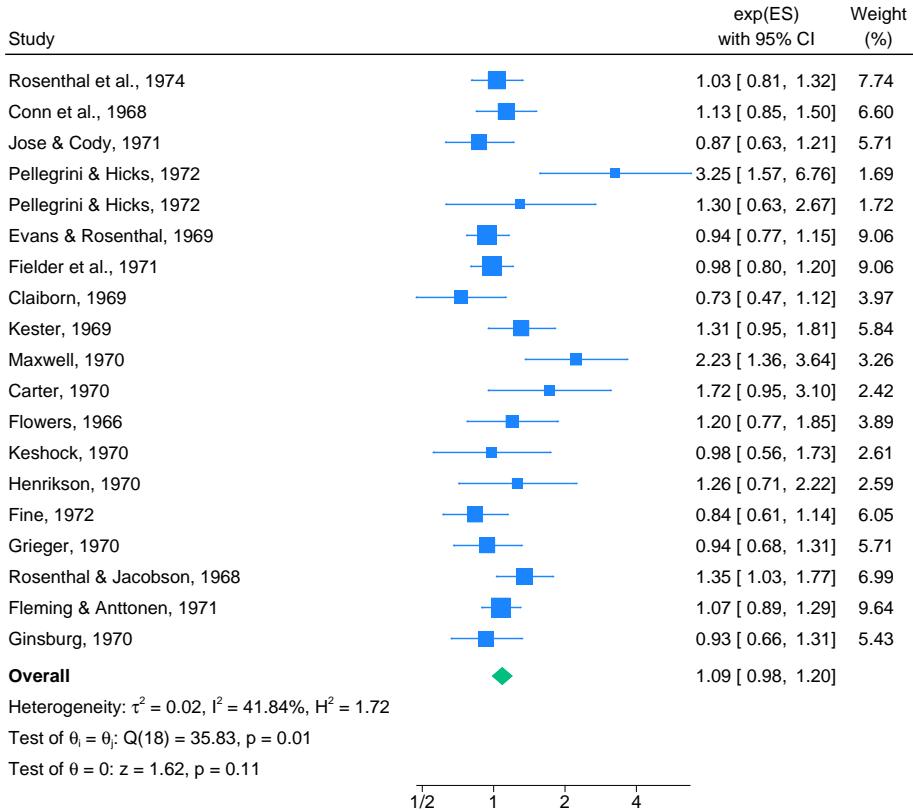
The HS estimator (Schmidt and Hunter 2015) is negatively biased and thus not recommended when unbiasedness is important (Viechtbauer 2005). Otherwise, the mean squared error of HS is similar to that of ML and is smaller than those of HE, DL, and REML.

## Forest plots

Meta-analysis results are often presented using a forest plot (for example, Lewis and Ellis [1982]). A forest plot shows study-specific effect sizes and an overall effect size with their respective confidence intervals. The information about study heterogeneity and the significance of the overall effect size are also

typically presented. This plot provides a convenient way to visually compare the study effect sizes, which can be any summary estimates available from primary studies, such as standardized and unstandardized mean differences, (log) odds ratios, (log) risk ratios, and (log) hazard ratios.

Below is an example of a forest plot.



Random-effects REML model

A blue square is plotted for each study, with the size of the square being proportional to the study weight; that is, larger squares correspond to larger (more precise) studies. Studies' CIs are plotted as whiskers extending from each side of the square and spanning the width of the CI. The estimate of the overall effect size, depicted here by a green diamond, is typically plotted following the individual effect sizes. The diamond is centered at the estimate of the overall effect size and the width of the diamond represents the corresponding CI width. Heterogeneity measures such as the  $I^2$  and  $H^2$  statistics, homogeneity test, and the significance test of the overall effect sizes are also commonly reported.

Three further variations of forest plots are for cumulative, subgroup, and leave-one-out meta-analyses; see [Cumulative meta-analysis](#), [Subgroup meta-analysis](#), and [Leave-one-out meta-analysis](#).

For further details about forest plots, see [META] [meta forestplot](#).

## Heterogeneity

The exposition below is based on [Deeks, Higgins, and Altman \(2017\)](#) and references therein.

It is natural for effect sizes of studies collected in a meta-analysis to vary between the studies because of sampling variability. However, when this variation exceeds the levels that could be explained by sampling variation, it is referred to as the between-study heterogeneity. Between-study heterogeneity may arise for different reasons and is generally divided into two types: clinical and methodological ([Thompson 1994](#); [Deeks, Higgins, and Altman 2017](#)). Clinical heterogeneity is the variability in the intervention strategies, outcomes, and study participants. Methodological heterogeneity is the variability in the study design and conduct. Statistical heterogeneity refers to the cases when the variability between the observed effects cannot be explained by sampling variability alone. It arises when the true effects in each study are different and may be the result of clinical heterogeneity, methodological heterogeneity, or both. In what follows, we refer to statistical heterogeneity simply as heterogeneity.

### Assessing heterogeneity

Forest plots are useful for visual examination of heterogeneity. Its presence can be evaluated by looking at the plotted CIs, which are represented as horizontal lines on the plot. Heterogeneity is suspect if there is a lack of overlap between the CIs.

For many studies, Galbraith plots may be a more visually appealing alternative to forest plots for assessing heterogeneity and presenting meta-analysis results. These plots graph standardized effect sizes against precision for each study with a regression line through the origin with the overall effect size as its slope. Excess variation of the scatter points around the regression line may suggest the presence of heterogeneity. See [\[META\] meta galbraithplot](#).

For a two-group comparison of binary outcomes, L'Abbé plots may be used to assess heterogeneity and compare study-specific event rates in the two groups; see [\[META\] meta labbeplot](#).

You can also test for heterogeneity more formally by using Cochran's homogeneity test. Additionally, various heterogeneity measures such as the  $I^2$  statistic, which estimates the percentage of the between-study variability, are available to quantify heterogeneity.

See [\[META\] meta summarize](#) for details.

### Addressing heterogeneity

There are several strategies to address heterogeneity when it is present. Below, we summarize some of the recommendations from [Deeks, Higgins, and Altman \(2017\)](#):

1. “*Explore heterogeneity*”. [Subgroup analyses](#) and [meta-regression](#) are commonly used to explore heterogeneity. For such analyses to be proper, you must prespecify upfront (before your meta-analysis) the study attributes you would like to explore. Often, meta-analysts are already familiar with the studies, so the genuine prestudy specification may not be possible. In that case, you should use caution when interpreting the results. Once heterogeneity is established, its exploration after the fact is viewed as data snooping and should be avoided.
2. “*Perform an RE meta-analysis*”. After careful consideration of subgroup analysis and meta-regression, you may consider an RE meta-analysis to account for the remaining unexplained between-study heterogeneity. See [Deeks, Higgins, and Altman \(2017, sec. 9.5.4\)](#) for details.

3. “*Exclude studies*”. Generally, you should avoid excluding studies from a meta-analysis because this may lead to bias. You may consider doing this in the presence of a few outlying studies when the reasons for the outlying results are well understood and are unlikely to interfere with your research objectives. Even then, you still need to perform sensitivity analysis and report both the results with and without the outlying studies.
4. “*Do not perform a meta-analysis*”. In the presence of substantial variation that cannot be explained, you may have to abandon the meta-analysis altogether. In this case, it will be misleading to report a single overall estimate of an effect, especially if there is a disagreement among the studies about the direction of the effect.

Below, we discuss ways of exploring heterogeneity via subgroup meta-analysis and meta-regression.

## Subgroup meta-analysis

It is not uncommon for the studies in a meta-analysis to report varying effect-size estimates. But it is important to understand and account for such variation during the meta-analysis to obtain reliable results (Thompson 1994; Berlin 1995). In the presence of substantial between-study variability, meta-analysis may be used to explore the relationship between the effect sizes and study-level covariates of interest, known in the meta-analysis literature as moderators. For example, the effect of a particular vaccine may depend on a study location, the effect of a particular drug may depend on the studies’ dosages, and so on.

Depending on the type of covariates, subgroup meta-analysis or meta-regression may be used to explore the between-study heterogeneity. Subgroup meta-analysis is commonly used with categorical covariates, whereas meta-regression is used when at least one of the covariates is continuous.

In subgroup meta-analysis or simply subgroup analysis, the studies are grouped based on study or participants’ characteristics, and an overall effect-size estimate is computed for each group. The goal of subgroup analysis is to compare these overall estimates across groups and determine whether the considered grouping helps explain some of the observed between-study heterogeneity. Note that subgroup analysis can be viewed as a special case of a meta-regression with only one categorical moderator.

For more details about subgroup analysis, see the `subgroup()` option in [META] [meta summarize](#) and [META] [meta forestplot](#).

## Meta-regression

Meta-regression explores a relationship between the study-specific effect sizes and the study-level covariates, such as a latitude of a study location or a dosage of a drug. These covariates are often referred to as moderators. See, for instance, Greenland (1987), Berkey et al. (1995), Thompson and Sharp (1999), Thompson and Higgins (2002), and Viechtbauer et al. (2015) for more information about meta-regression.

Two types of meta-regression are commonly considered in the meta-analysis literature: fixed-effects meta-regression and random-effects meta-regression.

An FE meta-regression (Greenland 1987) assumes that all heterogeneity between the study outcomes can be accounted for by the specified moderators. Let  $\mathbf{x}_j$  be a  $1 \times p$  vector of moderators with the corresponding unknown  $p \times 1$  coefficient vector  $\beta$ . An FE meta-regression is given by

$$\hat{\theta}_j = \mathbf{x}_j\beta + \epsilon_j \quad \text{weighted by } w_j = \frac{1}{\hat{\sigma}_j^2}, \text{ where } \epsilon_j \sim N(0, \hat{\sigma}_j^2)$$

A traditional FE meta-regression does not model residual heterogeneity, but it can be incorporated by multiplying each of the variances,  $\hat{\sigma}_j^2$ , by a common factor. This model is known as an FE meta-regression with a multiplicative dispersion parameter or a multiplicative FE meta-regression (Thompson and Sharp 1999).

An RE meta-regression (Berkey et al. 1995) can be viewed as a meta-regression that incorporates the residual heterogeneity via an additive error term, which is represented in a model by a study-specific random effect. These random effects are assumed to be normal with mean zero and variance  $\tau^2$ , which estimates the remaining between-study heterogeneity that is unexplained by the considered moderators. An RE meta-regression is

$$\hat{\theta}_j = \mathbf{x}_j\beta + u_j + \epsilon_j \quad \text{weighted by } w_j^* = \frac{1}{\hat{\sigma}_j^2 + \hat{\tau}^2}, \text{ where } u_j \sim N(0, \tau^2) \text{ and } \epsilon_j \sim N(0, \hat{\sigma}_j^2)$$

For more details about meta-regression, see [META] meta regress and [META] meta regress postestimation.

## Publication bias

Publication bias or, more generally, reporting bias occurs when the studies selected for a scientific review are systematically different from all available relevant studies. Specifically, publication bias is known in the meta-analysis literature as an association between the likelihood of a publication and the statistical significance of a study result. The rise of systematic reviews for summarizing the results of scientific studies elevated the importance of acknowledging and addressing publication bias in research. Publication bias typically arises when nonsignificant results are being underreported in the literature (for example, Rosenthal [1979]; Iyengar and Greenhouse [1988]; Begg and Berlin [1988]; Hedges [1992]; Stern and Simes [1997]; Givens, Smith, and Tweedie [1997]; Sutton et al. [2000b]; and Kicinski, Springate, and Kontopantelis [2015]).

Suppose that we are missing some of the studies in our meta-analysis. If these studies are simply a random sample of all the studies that are relevant to our research question, our meta-analytic results will remain valid but will not be as precise. That is, we will likely obtain wider confidence intervals and less powerful tests. However, if the missing studies differ systematically from our observed studies, such as when smaller studies with nonsignificant findings are suppressed from publication, our meta-analytic results will be biased toward a significant result. Any health-policy or clinical decisions based on them will be invalid.

Dickersin (2005) notes that to avoid potentially serious consequences of publication bias, many researchers (for example, Simes [1986]; Dickersin [1988]; Hetherington et al. [1989]; Dickersin and Rennie [2003]; Antes and Chalmers [2003]; and Krakovsky [2004]) called for the registration of clinical trials worldwide at the outset to keep track of the findings, whether or not significant, from all trials. Although this may not necessarily eradicate the problem of publication bias, this will make it more difficult for the results of smaller trials to go undetected. Generally, when one selects the studies for meta-analysis, the review of the literature should be as comprehensive as possible, including searching the grey literature to uncover the relevant unpublished studies.

See Borenstein et al. (2009, chap. 30) for the summary of other factors for publication bias such as language bias and cost bias.

## Funnel plots

The funnel plot (Light and Pillemer 1984) is commonly used to explore publication bias (Sterne, Becker, and Egger 2005). It is a scatterplot of the study-specific effect sizes versus measures of study precision. In the absence of publication bias, the shape of the scatterplot should resemble a symmetric inverted funnel. The funnel-plot asymmetry, however, may be caused by factors other than publication bias such as a presence of a moderator correlated with the study effect and study size or, more generally, the presence of substantial between-study heterogeneity (Egger et al. 1997 ; Peters et al. 2008 ; Sterne et al. 2011 ). The so-called contour-enhanced funnel plots have been proposed to help discriminate between the funnel-plot asymmetry because of publication bias versus other reasons.

See [META] [meta funnelplot](#) for details.

## Tests for funnel-plot asymmetry

Graphical evaluation of funnel plots is useful for data exploration but may be subjective when detecting the asymmetry. Statistical tests provide a more formal evaluation of funnel-plot asymmetry. These tests are also known as tests for small-study effects (Sterne, Gavaghan, and Egger 2000) and, historically, as tests for publication bias. The tests are no longer referred to as “tests for publication bias” because, as we commented earlier, the presence of the funnel-plot asymmetry may not necessarily be attributed to publication bias, particularly in the presence of substantial between-study variability. See Harbord, Harris, and Sterne (2016) for a summary of these tests.

Two types of tests for funnel-plot asymmetry are considered in the literature: regression-based tests (Egger et al. 1997 ; Harbord, Egger, and Sterne 2006; and Peters et al. 2006 ) and a nonparametric rank-based test (Begg and Mazumdar 1994). These tests explore the relationship between the study-specific effect sizes and study precision. The presence of the funnel-plot asymmetry is declared when the association between the two measures is greater than what would have been observed by chance.

For more details regarding the tests of funnel-plot asymmetry, see [META] [meta bias](#).

## The trim-and-fill method

Tests for funnel-plot asymmetry are useful for detecting publication bias but are not able to estimate the impact of this bias on the final meta-analysis results. The nonparametric trim-and-fill method of Duval and Tweedie (2000a, 2000b) provides a way to assess the impact of missing studies because of publication bias on the meta-analysis. It evaluates the amount of potential bias present in meta-analysis and its impact on the final conclusion. This method is typically used as a sensitivity analysis to the presence of publication bias.

See [META] [meta trimfill](#) for more information about the trim-and-fill method.

## Cumulative meta-analysis

Cumulative meta-analysis performs multiple meta-analyses, where each analysis is produced by adding one study at a time. It is useful to identify various trends in the overall effect sizes. For example, when the studies are ordered chronologically, one can determine the point in time of the potential change in the direction or significance of the effect size. A well-known example of a cumulative meta-analysis is presented in *Cumulative meta-analysis* of [META] [meta](#) for the study of the efficacy of streptokinase after a myocardial infarction (Lau et al. 1992). Also see the `cumulative()` option in [META] [meta summarize](#) and [META] [meta forestplot](#).

## Leave-one-out meta-analysis

Just like cumulative meta-analysis, the leave-one-out meta-analysis also performs multiple meta-analyses; however, in this case, each analysis is produced by excluding a single study. It is quite common that studies yield effect sizes that are relatively exaggerated. Their presence in the meta-analysis may distort the overall results, and it is of great importance to identify such studies for further examination. The leave-one-out meta-analysis is a useful tool to investigate the influence of each study on the overall effect size estimate. See the `leaveoneout` option in [META] [meta summarize](#) and [META] [meta forestplot](#) for more information.

## Multivariate meta-regression

Multivariate meta-analysis combines results from studies where multiple dependent effect sizes (outcomes) are reported by each study. Let  $\hat{\theta}_j$  be a  $d \times 1$  vector of estimates of the true population multivariate effect size  $\theta_j$  for study  $j$ . Let  $\mathbf{x}_j$  be a  $1 \times p$  vector of moderators with the corresponding unknown  $p \times 1$  regression coefficient vector  $\beta_i$  for  $i = 1, \dots, d$ .

An FE multivariate meta-regression (Raudenbush, Becker, and Kalaian 1988) is given by

$$\hat{\theta}_j = \mathbf{X}_j \boldsymbol{\beta} + \epsilon_j, \epsilon_j \sim N_d(\mathbf{0}, \boldsymbol{\Lambda}_j)$$

where  $\mathbf{X}_j = \mathbf{x}_j \otimes I_d$  is a  $d \times dp$  matrix and  $\boldsymbol{\beta} = (\beta'_1, \beta'_2, \dots, \beta'_d)'$  is a  $dp \times 1$  vector of unknown regression coefficients;  $\otimes$  is the Kronecker product. The within-study covariance matrices  $\boldsymbol{\Lambda}_j$ 's are assumed known and thus do not require estimation.

The RE multivariate meta-regression (Berkey et al. 1998) can be expressed as

$$\hat{\theta}_j = \mathbf{X}_j \boldsymbol{\beta} + \epsilon_j^* = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{u}_j + \epsilon_j, \text{ where } \epsilon_j^* \sim N_d(\mathbf{0}, \boldsymbol{\Lambda}_j + \boldsymbol{\Sigma})$$

where  $\mathbf{u}_j$  is a  $d \times 1$  vector of random effects corresponding to the  $d$  outcomes.

`meta mvregress` fits multivariate meta-regression; see [META] [meta mvregress](#). By default, a random-effects model is assumed. The goal of multivariate meta-regression is to estimate the regression coefficients  $\boldsymbol{\beta}$  and the random-effects covariance matrix  $\boldsymbol{\Sigma}$ , also known as the between-study covariance matrix. Three estimation methods are available to fit the RE multivariate meta-regression model and multiple covariance structures can be specified to model the between-study covariance  $\boldsymbol{\Sigma}$ . After fitting the multivariate meta-regression model, you can assess heterogeneity; see [META] [estat heterogeneity \(mv\)](#). Various postestimation tools are available such as predicting random effects, computing the linear predictor, residuals, standardized residuals, and more; see [META] [meta mvregress postestimation](#).

## Multilevel meta-regression

Multilevel meta-analysis synthesizes the results from potentially dependent effect sizes that exhibit a hierarchical or nested structure. For example, studies and their corresponding effect sizes may be nested within higher-level groupings such as geographical locations (for example, states or countries) or administrative units (for example, school districts).

When a hierarchical structure is present in the data, the multilevel meta-analysis is preferred over the classical meta-analysis. By properly accounting for the hierarchical structure among the effect sizes, we can obtain more accurate estimates of the overall effect size and better overall statistical inference. We can also decompose the heterogeneity present among the effect sizes across the different hierarchical levels, which can provide valuable insights into the factors that affect our outcome of interest.

The three-level meta-regression model (for example, Goldstein et al. [2000]; Thompson, Turner, and Warn [2001]; and Konstantopoulos [2011]) can be expressed as

$$\hat{\theta}_{jkr} = \mathbf{x}_{jkr}\boldsymbol{\beta} + \mathbf{z}_{jkr}^{(3)}\mathbf{u}_j^{(3)} + \mathbf{z}_{jkr}^{(2)}\mathbf{u}_{jk}^{(2)} + \epsilon_{jkr}$$

where  $j = 1, 2, \dots, M$ ,  $k = 1, 2, \dots, m_j$ , and  $r = 1, 2, \dots, m_{jk}$ . In this case,  $\mathbf{x}_{jkr} = (1, x_{1,jkr}, \dots, x_{p-1,jkr})$  is a  $1 \times p$  vector of moderators and  $\boldsymbol{\beta}$  is the corresponding  $p \times 1$  vector of unknown fixed-effects parameters.  $\mathbf{z}_{jkr}^{(3)}$  is a  $1 \times q_3$  vector of moderators associated with the level-3  $q_3 \times 1$  vector of random effects  $\mathbf{u}_j^{(3)}$  (1 intercept and  $q_3 - 1$  slopes), where  $\mathbf{u}_j^{(3)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(3)})$ . Similarly,  $\mathbf{z}_{jkr}^{(2)}$  is a  $1 \times q_2$  vector of moderators associated with the level-2 (within-level-3)  $q_2 \times 1$  vector of random effects  $\mathbf{u}_{jk}^{(2)}$ , where  $\mathbf{u}_{jk}^{(2)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(2)})$ .  $\epsilon_{jkr} \sim N(0, \hat{\sigma}_{jkr}^2)$ , where  $\hat{\sigma}_{jkr}^2$ 's are known sampling (effect-size) variances.

`meta meregress` fits multilevel meta-regression; see [META] **meta meregress**. If your model contains only random intercepts (no random slopes), you may use the `meta multilevel` command, which has a simpler syntax geared toward random-intercepts multilevel models; see [META] **meta multilevel**. The goal of multilevel meta-regression is to estimate the regression coefficients  $\boldsymbol{\beta}$  and the random-effects covariance matrices  $\boldsymbol{\Sigma}^{(l)}$  for each level  $l > 1$ . By default, the REML estimation method is assumed, but the MLE method is also supported. Multiple covariance structures can be specified to model the  $\boldsymbol{\Sigma}^{(l)}$  matrices. After fitting the multilevel meta-regression model, you can assess multilevel heterogeneity; see [META] **estat heterogeneity (me)**. Various postestimation tools are available, such as predicting random effects and computing the linear predictor, residuals, standardized residuals, and more; see [META] **meta me postestimation**.

## References

- Antes, G., and I. Chalmers. 2003. Under-reporting of clinical trials is unethical. *Lancet* 361: 978–979. [https://doi.org/10.1016/S0140-6736\(03\)12838-3](https://doi.org/10.1016/S0140-6736(03)12838-3).
- Arnqvist, G., and D. Wooster. 1995. Meta-analysis: Synthesizing research findings in ecology and evolution. *Trends in Ecology and Evolution* 10: 236–240. [https://doi.org/10.1016/S0169-5347\(00\)89073-4](https://doi.org/10.1016/S0169-5347(00)89073-4).
- Barrick, M. R., and M. K. Mount. 1991. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology* 44: 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>.
- Begg, C. B., and J. A. Berlin. 1988. Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society, A ser.*, 151: 419–463. <https://doi.org/10.2307/2982993>.
- Begg, C. B., and M. Mazumdar. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50: 1088–1101. <https://doi.org/10.2307/2533446>.
- Berkey, C. S., D. C. Hoaglin, F. Mosteller, and G. A. Colditz. 1995. A random-effects regression model for meta-analysis. *Statistics in Medicine* 14: 395–411. <https://doi.org/10.1002/sim.4780140406>.
- . 1998. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 17: 2537–2550. [https://doi.org/10.1002/\(SICI\)1097-0258\(19981130\)17:22%3C2537::AID-SIM953%3E3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0258(19981130)17:22%3C2537::AID-SIM953%3E3.0.CO;2-C).
- Berlin, J. A. 1995. Invited commentary: Benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *American Journal of Epidemiology* 142: 383–387. <https://doi.org/10.1093/oxfordjournals.aje.a117645>.
- Bernard, R. M., P. C. Abrami, Y. Lou, E. Borokhovski, A. Wade, L. Wozney, P. A. Wallet, M. Fiset, and B. Huang. 2004. How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research* 74: 379–439. <https://doi.org/10.3102/00346543074003379>.
- Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester, UK: Wiley.
- Card, D., J. Kluve, and A. Weber. 2010. Active labour market policy evaluations: A meta-analysis. *Economic Journal* 120: F452–F477. <https://doi.org/10.1111/j.1468-0297.2010.02387.x>.

- Cooper, H., L. V. Hedges, and J. C. Valentine, eds. 2019. *The Handbook of Research Synthesis and Meta-Analysis*. 3rd ed. New York: Russell Sage Foundation.
- Dalhuisen, J. M., R. J. G. M. Florax, H. L. F. de Groot, and P. Nijkamp. 2003. Price and income elasticities of residential water demand: A meta-analysis. *Land Economics* 79: 292–308. <https://doi.org/10.2307/3146872>.
- Deeks, J. J., J. P. T. Higgins, and D. G. Altman. 2017. “Analysing data and undertaking meta-analyses”. In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0*, edited by J. P. T. Higgins and S. Green. London: The Cochrane Collaboration. <https://training.cochrane.org/handbook>.
- DerSimonian, R., and N. M. Laird. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
- Dickersin, K. 1988. Report from the panel on the case for registers of clinical trials at the eighth annual meeting of the society for clinical trials. *Controlled Clinical Trials* 9: 76–80. [https://doi.org/10.1016/0197-2456\(88\)90010-4](https://doi.org/10.1016/0197-2456(88)90010-4).
- . 2005. “Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm”. In *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, edited by H. R. Rothstein, A. J. Sutton, and M. Borenstein, 9–33. Chichester, UK: Wiley. <https://doi.org/10.1002/0470870168.ch2>.
- Dickersin, K., and D. Rennie. 2003. Registering clinical trials. *Journal of the American Medical Association* 290: 516–523. <https://doi.org/10.1001/jama.290.4.516>.
- Duval, S., and R. L. Tweedie. 2000a. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56: 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- . 2000b. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 95: 89–98. <https://doi.org/10.1080/01621459.2000.10473905>.
- Egger, M., G. Davey Smith, M. Schneider, and C. Minder. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315: 629–634. <https://doi.org/10.1136/bmj.315.7109.629>.
- Fan, X., and M. Chen. 2001. Parental involvement and students’ academic achievement: A meta-analysis. *Educational Psychology Review* 13: 1–22. <https://doi.org/10.1023/A:1009048817385>.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: Wiley. <https://doi.org/10.1002/0471445428>.
- Gendreau, P., T. Little, and C. Goggin. 1996. A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology* 34: 575–608. <https://doi.org/10.1111/j.1745-9125.1996.tb01220.x>.
- Givens, G. H., D. D. Smith, and R. L. Tweedie. 1997. Publication bias in meta-analysis: A Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science* 12: 221–240. <https://doi.org/10.1214/ss/1030037958>.
- Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* 5: 3–8. <https://doi.org/10.2307/1174772>.
- Goldstein, H., M. Yang, R. Z. Omar, R. M. Turner, and S. G. Thompson. 2000. Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society, C ser.*, 49: 399–412. <https://doi.org/10.1111/1467-9876.00200>.
- Greenland, S. 1987. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* 9: 1–30. <https://doi.org/10.1093/oxfordjournals.epirev.a036298>.
- Greenland, S., and J. M. Robins. 1985. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 41: 55–68. <https://doi.org/10.2307/2530643>.
- Gurevitch, J., P. S. Curtis, and M. H. Jones. 2001. Meta-analysis in ecology. *Advances in Ecology* 32: 199–247. [https://doi.org/10.1016/S0065-2504\(01\)32013-5](https://doi.org/10.1016/S0065-2504(01)32013-5).
- Hanji, M. B. 2017. *Meta-Analysis in Psychiatry Research: Fundamental and Advanced Methods*. Waretown, NJ: Apple Academic Press. <https://doi.org/10.1201/9781315366234>.
- Harbord, R. M., M. Egger, and J. A. C. Sterne. 2006. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 25: 3443–3457. <https://doi.org/10.1002/sim.2380>.
- Harbord, R. M., R. J. Harris, and J. A. C. Sterne. 2016. “Updated tests for small-study effects in meta-analyses”. In *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*, edited by T. M. Palmer and J. A. C. Sterne, 153–165. 2nd ed. College Station, TX: Stata Press.

- Hardy, R. J., and S. G. Thompson. 1998. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 17: 841–856. [https://doi.org/10.1002/\(sici\)1097-0258\(19980430\)17:8<841::aid-sim781>3.0.co;2-d](https://doi.org/10.1002/(sici)1097-0258(19980430)17:8<841::aid-sim781>3.0.co;2-d).
- Harter, J. K., F. L. Schmidt, and T. L. Hayes. 2002. Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Applied Psychology* 87: 268–279. <https://doi.org/10.1037/0021-9010.87.2.268>.
- Hay, D. C., W. R. Knechel, and N. Wang. 2006. Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research* 23: 141–191. <https://doi.org/10.1506/4XR4-KT5V-E8CN-91GX>.
- Hedges, L. V. 1982. Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 92: 490–499. <https://doi.org/10.1037/0033-2909.92.2.490>.
- . 1983. A random effects model for effect sizes. *Psychological Bulletin* 93: 388–395. <https://doi.org/10.1037/0033-2909.93.2.388>.
- . 1992. Modeling publication selection effects in meta-analysis. *Statistical Science* 7: 246–255. <https://doi.org/10.1214/ss/1177011364>.
- Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. *Ecology* 80: 1150–1156. <https://doi.org/10.2307/177062>.
- Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., and J. L. Vevea. 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* 3: 486–504. <https://doi.org/10.1037/1082-989X.3.4.486>.
- Hetherington, J., K. Dickersin, I. Chalmers, and C. L. Meinert. 1989. Retrospective and prospective identification of unpublished controlled trials: Lessons from a survey of obstetricians and pediatricians. *Pediatrics* 84: 374–380.
- Higgins, J. P. T., and S. Green, eds. 2017. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0*. London: The Cochrane Collaboration. <https://training.cochrane.org/handbook>.
- Iyengar, S., and J. B. Greenhouse. 1988. Selection models and the file drawer problem. *Statistical Science* 3: 109–117. <https://doi.org/10.1214/ss/1177013012>.
- Kicinski, M., D. A. Springate, and E. Kontopantelis. 2015. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. *Statistics in Medicine* 34: 2781–2793. <https://doi.org/10.1002/sim.6525>.
- Knapp, G., and J. Hartung. 2003. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* 22: 2693–2710. <https://doi.org/10.1002/sim.1482>.
- Konstantopoulos, S. 2011. Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods* 2: 61–76. <https://doi.org/10.1002/jrsm.35>.
- Kontopantelis, E., and D. Reeves. 2020. Pairwise meta-analysis of aggregate data using metaan in Stata. *Stata Journal* 20: 680–705.
- Kravovsky, M. 2004. Register or perish: Looking to make the downside of therapies known. *Scientific American* 6: 18–20.
- Lau, J., E. M. Antman, J. Jimenez-Silva, B. Kupelnick, F. Mosteller, and T. C. Chalmers. 1992. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine* 327: 248–254. <https://doi.org/10.1056/NEJM199207233270406>.
- Lewis, J. A., and S. H. Ellis. 1982. A statistical appraisal of post-infarction beta-blocker trials. *Primary Cardiology Suppl.* 1: 31–37.
- Light, R. J., and D. B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press. <https://doi.org/10.2307/j.ctvk12px9>.
- Linden, A. 2022. Computing the fragility index for randomized trials and meta-analyses using Stata. *Stata Journal* 22: 77–88.
- Mantel, N., and W. Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719–748. Reprinted in *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods*, ed. S. Greenland, pp. 112–141. Newton Lower Falls, MA: Epidemiology Resources.
- Palmer, T. M., and J. A. C. Sterne, eds. 2016. *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*. 2nd ed. College Station, TX: Stata Press.
- Paule, R. C., and J. Mandel. 1982. Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards* 87: 377–385. <https://doi.org/10.6028/jres.087.022>.

- Pearson, K. 1904. Report on certain enteric fever inoculation statistics. *BMJ* 2: 1243–1246. <https://doi.org/10.1136/bmj.2.2288.1243>.
- Peters, J. L., A. J. Sutton, D. R. Jones, K. R. Abrams, and L. Rushton. 2006. Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association* 295: 676–680. <https://doi.org/10.1001/jama.295.6.676>.
- . 2008. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology* 61: 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>.
- Pratt, T. C., and F. T. Cullen. 2000. The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology* 38: 931–964. <https://doi.org/10.1111/j.1745-9125.2000.tb00911.x>.
- Raudenbush, S. W. 2009. “Analyzing effect sizes: Random-effects models”. In *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. Cooper, L. V. Hedges, and J. C. Valentine, 295–316. 2nd ed. New York: Russell Sage Foundation.
- Raudenbush, S. W., B. J. Becker, and H. Kalaian. 1988. Modeling multivariate effect sizes. *Psychological Bulletin* 103: 111–120. <https://doi.org/10.1037/0033-2909.103.1.111>.
- Rice, K., J. P. T. Higgins, and T. S. Lumley. 2018. A re-evaluation of fixed effect(s) meta-analysis. *Journal of the Royal Statistical Society, A ser.*, 181: 205–227. <https://doi.org/10.1111/rssa.12275>.
- Rosenthal, R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86: 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>.
- Rosenthal, R., and D. B. Rubin. 1982. Comparing effect sizes of independent studies. *Psychological Bulletin* 92: 500–504. <https://doi.org/10.1037/0033-2909.92.2.500>.
- Rothstein, H. R., A. J. Sutton, and M. Borenstein, eds. 2005. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester, UK: Wiley. <https://doi.org/10.1002/0470870168>.
- Schmidt, F. L., and J. E. Hunter. 2015. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 3rd ed. Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781483398105>.
- Sidik, K., and J. N. Jonkman. 2005. A note on variance estimation in random effects meta-regression. *Journal of Biopharmaceutical Statistics* 15: 823–838. <https://doi.org/10.1081/BIP-200067915>.
- . 2007. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 26: 1964–1981. <https://doi.org/10.1002/sim.2688>.
- Simes, R. J. 1986. Publication bias: The case for an international registry of clinical trials. *Journal of Clinical Oncology* 4: 1529–1541. <https://doi.org/10.1200/JCO.1986.4.10.1529>.
- Sin, N. L., and S. Lyubomirsky. 2009. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis. *Journal of Clinical Psychology* 65: 467–487. <https://doi.org/10.1002/jclp.20593>.
- Stern, J. M., and R. J. Simes. 1997. Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 315: 640–645. <https://doi.org/10.1136/bmj.315.7109.640>.
- Sterne, J. A. C., B. J. Becker, and M. Egger. 2005. “The funnel plot”. In *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, edited by H. R. Rothstein, A. J. Sutton, and M. Borenstein, 73–98. Chichester, UK: Wiley. <https://doi.org/10.1002/0470870168.ch5>.
- Sterne, J. A. C., D. Gavaghan, and M. Egger. 2000. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 53: 1119–1129. [https://doi.org/10.1016/S0895-4356\(00\)00242-0](https://doi.org/10.1016/S0895-4356(00)00242-0).
- Sterne, J. A. C., A. J. Sutton, J. P. A. Ioannidis, N. Terrin, D. R. Jones, J. Lau, J. R. Carpenter, G. Rücker, R. M. Harbord, C. H. Schmid, J. Tetzlaff, J. J. Deeks, J. L. Peters, P. Macaskill, G. Schwarzer, S. Duval, D. G. Altman, D. Moher, and J. P. T. Higgins. 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343: d4002. <https://doi.org/10.1136/bmj.d4002>.
- Sutton, A. J., K. R. Abrams, D. R. Jones, T. A. Sheldon, and F. Song. 2000a. *Methods for Meta-Analysis in Medical Research*. New York: Wiley.
- Sutton, A. J., S. Duval, R. L. Tweedie, K. R. Abrams, and D. R. Jones. 2000b. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 320: 1574–1577. <https://doi.org/10.1136/bmj.320.7249.1574>.

- Sutton, A. J., and J. P. T. Higgins. 2008. Recent developments in meta-analysis. *Statistics in Medicine* 27: 625–650. <https://doi.org/10.1002/sim.2934>.
- Thompson, S. G. 1994. Systematic review: Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 309: 1351–1355. <https://doi.org/10.1136/bmj.309.6965.1351>.
- Thompson, S. G., and J. P. T. Higgins. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 21: 1559–1573. <https://doi.org/10.1002/sim.1187>.
- Thompson, S. G., and S. J. Sharp. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* 18: 2693–2708. [https://doi.org/10.1002/\(sici\)1097-0258\(19991030\)18:20<2693::aid-sim235>3.0.co;2-v](https://doi.org/10.1002/(sici)1097-0258(19991030)18:20<2693::aid-sim235>3.0.co;2-v).
- Thompson, S. G., R. M. Turner, and D. E. Warn. 2001. Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research* 10: 375–392. <https://doi.org/10.1177/096228020101000602>.
- Veroniki, A. A., D. Jackson, W. Viechtbauer, R. Bender, J. Bowden, G. Knapp, O. Kuss, J. P. T. Higgins, D. Langan, and G. Salanti. 2016. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods* 7: 55–79. <https://doi.org/10.1002/jrsm.1164>.
- Viechtbauer, W. 2005. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* 30: 261–293. <https://doi.org/10.3102/10769986030003261>.
- Viechtbauer, W., J. A. López-López, J. Sánchez-Meca, and F. Marín-Martínez. 2015. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods* 20: 360–374. <https://doi.org/10.1037/met0000023>.
- Whitehead, A., and J. Whitehead. 1991. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 10: 1665–1677. <https://doi.org/10.1002/sim.4780101105>.
- Winfree, R., R. Aguilar, D. P. Vázquez, G. LeBuhn, and M. A. Aizen. 2009. A meta-analysis of bees' responses to anthropogenic disturbance. *Ecology* 90: 2068–2076. <https://doi.org/10.1890/08-1245.1>.
- Woodward, R. T., and Y.-S. Wui. 2001. The economic value of wetland services: A meta-analysis. *Ecological Economics* 37: 257–270. [https://doi.org/10.1016/S0921-8009\(00\)00276-7](https://doi.org/10.1016/S0921-8009(00)00276-7).

## Also see

[META] [meta](#) — Introduction to meta

[META] [Glossary](#)

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).