# Title

stata.com

**meta summarize —** Summarize meta-analysis data



## Description

`meta summarize` summarizes `meta` data. It reports individual effect sizes and the overall effect size (ES), their confidence intervals (CIs), heterogeneity statistics, and more. `meta summarize` can perform random-effects (RE), common-effect (CE), and fixed-effects (FE) meta-analyses. It can also perform subgroup, cumulative, and sensitivity meta-analyses. For graphical display of meta-analysis summaries, see [META] **meta forestplot**.

## Quick start

Perform meta-analysis and summarize `meta` data, which were declared by either `meta set` or `meta esize`

```
meta summarize
```

Same as above, but summarize meta-analysis results using the empirical Bayes RE method instead of the declared method

```
meta summarize, random(ebayes)
```

Same as above, but report transformed effect sizes and CIs using the hyperbolic tangent function

```
meta summarize, random(ebayes) transform(tanh)
```

Perform subgroup meta-analysis based on the categorical variable `x1`

```
meta summarize, subgroup(x1)
```

Perform subgroup analysis based on the categorical variables `x1`, `x2`, and `x3`

```
meta summarize, subgroup(x1 x2 x3)
```

Perform cumulative meta-analysis (CMA), where studies are included in the CMA based on the ascending order of observations in variable `x4`

```
meta summarize, cumulative(x4)
```

Same as above, but stratify the results of the CMA based on groups of the categorical variable `x5`

```
meta summarize, cumulative(x4, by(x5))
```

Perform leave-one-out meta-analysis

```
meta summarize, leaveoneout
```

Perform sensitivity meta-analysis by assuming a fixed value of 0.2 for the between-study heterogeneity parameter $\tau^2$, assuming that the declared model is RE

```
meta summarize, tau2(.2)
```

# Menu

Statistics > Meta-analysis

# Syntax

*Meta-analysis as declared with meta set or meta esize*

```
meta summarize [if] [in] [, options reopts]
```

*Random-effects meta-analysis*

```
meta summarize [if] [in], random[(remethod)] [options reopts]
```

*Common-effect meta-analysis*

```
meta summarize [if] [in], common[(cefemethod)] [options]
```

*Fixed-effects meta-analysis*

```
meta summarize [if] [in], fixed[(cefemethod)] [options]
```

| *options* | Description |
|---|---|
| **Main** | |
| `subgroup(`*varlist*`)` | subgroup meta-analysis for each variable in *varlist* |
| `cumulative(`*cumulspec*`)` | cumulative meta-analysis |
| `leaveoneout` | leave-one-out meta-analysis |
| **Options** | |
| `level(#)` | set confidence level; default is as declared for meta-analysis |
| `citype(`*citype*`)` | specify the type of study CI (for meta-analysis of a single proportion) |
| *eform_option* | report exponentiated results |
| `transform(`*transfspec*`)` | report transformed results |
| `sort(`*varlist*`[ , ... ])` | sort studies according to *varlist* |
| `tdistribution` | report $t$ test instead of $z$ test for the overall effect size |
| `proportion` | report proportions (for meta-analysis of a single proportion) |
| `prevalence` | synonym for `proportion` but labels the effect sizes as `Prevalence` in the output |
| `nostudies` | suppress output for individual studies |
| `noheader` | suppress output header |
| `[no]metashow` | display or suppress meta settings in the output |
| *display_options* | control column formats |
| **Maximization** | |
| *maximize_options* | control the maximization process; seldom used |

`collect` is allowed; see **[U] 11.1.10 Prefix commands**.

| *remethod* | Description |
|---|---|
| reml | restricted maximum likelihood; the default |
| mle | maximum likelihood |
| ebayes | empirical Bayes |
| dlaird | DerSimonian–Laird |
| sjonkman | Sidik–Jonkman |
| hedges | Hedges |
| hschmidt | Hunter–Schmidt |

| *cefemethod* | Description |
|---|---|
| mhaenszel | Mantel–Haenszel |
| invvariance | inverse variance |
| ivariance | synonym for invvariance |

| *reopts* | Description |
|---|---|
| tau2(#) | sensitivity meta-analysis using a fixed value of between-study variance $\tau^2$ |
| i2(#) | sensitivity meta-analysis using a fixed value of heterogeneity statistic $I^2$ |
| predinterval[(#)] | report prediction interval for the overall effect size |
| se(*seadj*) | adjust standard error of the overall effect size |

# Options

Main

Options random(), common(), and fixed(), when specified with meta summarize, temporarily override the global model declared by meta set or meta esize during the computation. Options random(), common(), and fixed() may not be combined. If these options are omitted, the declared meta-analysis model is assumed; see *Declaring a meta-analysis model* in [META] **meta data**. Also see *Meta-analysis models* in [META] **Intro**.

random and random(*remethod*) specify that a random-effects model be assumed for meta-analysis; see *Random-effects model* in [META] **Intro**.

*remethod* specifies the type of estimator for the between-study variance $\tau^2$. *remethod* is one of reml, mle, ebayes, dlaird, sjonkman, hedges, or hschmidt. random is a synonym for random(reml). See *Options* in [META] **meta esize** for more information.

common and common(*cefemethod*) specify that a common-effect model be assumed for meta-analysis; see *Common-effect ("fixed-effect") model* in [META] **Intro**. Also see the discussion in [META] **meta data** about common-effect versus fixed-effects models.

common implies common(mhaenszel) for effect sizes lnoratio, lnrratio, and rdiff and common(invvariance) for all other effect sizes. common(mhaenszel) is supported only with effect sizes lnoratio, lnrratio, and rdiff.

*cefemethod* is one of mhaenszel or invvariance (synonym ivariance). See *Options* in [META] **meta esize** for more information.

`fixed` and `fixed(`*cefemethod*`)` specify that a fixed-effects model be assumed for meta-analysis; see *Fixed-effects model* in [META] **Intro**. Also see the discussion in [META] **meta data** about fixed-effects versus common-effect models.

`fixed` implies `fixed(mhaenszel)` for effect sizes `lnoratio`, `lnrratio`, and `rdiff` and `fixed(invvariance)` for all other effect sizes. `fixed(mhaenszel)` is supported only with effect sizes `lnoratio`, `lnrratio`, and `rdiff`.

*cefemethod* is one of `mhaenszel` or `invvariance` (synonym `ivariance`); see *Options* in [META] **meta esize** for more information.

`subgroup(`*varlist*`)` specifies that a subgroup meta-analysis (subgroup analysis) be performed for each variable in *varlist*. Subgroup analysis performs meta-analysis separately for each variable in *varlist* and for each group as defined by that variable. The specified meta-analysis model is assumed for each subgroup. This analysis is useful when the results of all studies are too heterogeneous to be combined into one estimate but the results are similar within certain groups of studies. The specified variables can be numeric or string variables. When multiple variables are specified, only the subgroup results are displayed; that is, the results from individual studies are suppressed for brevity. This option may not be combined with `cumulative()` or `leaveoneout`.

`cumulative(`*ordervar*`[ , ascending | descending by(`*byvar*`) ])` performs a cumulative meta-analysis (CMA). CMA performs multiple meta-analyses and accumulates the results by adding one study at a time to each subsequent analysis. It is useful for monitoring the results of the studies as new studies become available. The studies enter the CMA based on the ordered values of variable *ordervar*. *ordervar* must be a numeric variable. By default, ascending order is assumed unless the suboption `descending` is specified; only one of `ascending` or `descending` is allowed. The `by(`*byvar*`)` option specifies that the CMA be stratified by variable *byvar*. This option may not be combined with `subgroup()` or `leaveoneout`.

`leaveoneout` performs a leave-one-out meta-analysis. For each study, the corresponding leave-one-out meta-analysis is a meta-analysis of all the studies except that study. It is useful for assessing the effect of a single study on the meta-analysis results and for identifying outliers if they exist. This option may not be combined with `subgroup()` or `cumulative()`.

*reopts* are `tau2(#)`, `i2(#)`, `predinterval[ (#) ]`, and `se(khartung[ , truncated ])`. These options are used with random-effects meta-analysis.

`tau2(#)` specifies the value of the between-study variance parameter, $\tau^2$, to use for the random-effects meta-analysis. This option is useful for exploring the sensitivity of the results to different levels of between-study heterogeneity. Only one of `tau2()` or `i2()` may be specified. This option is not allowed in combination with `subgroup()`, `cumulative()`, or `leaveoneout`.

`i2(#)` specifies the value of the heterogeneity statistic $I^2$ (as a percentage) to use for the random-effects meta-analysis. This option is useful for exploring the sensitivity of the results to different levels of between-study heterogeneity. Only one of `i2()` or `tau2()` may be specified. This option is not allowed in combination with `subgroup()`, `cumulative()`, or `leaveoneout`.

`predinterval` and `predinterval(#)` specify that the 95% or #% prediction interval be reported for the overall effect size in addition to the confidence interval. # specifies the confidence level of the prediction interval. The prediction interval provides plausible ranges for the effect size in a future, new study. This option is not allowed in combination with `subgroup()` when specified with more than one variable, `cumulative()`, or `leaveoneout`.

`se(`*seadj*`)` specifies that the adjustment *seadj* be applied to the standard error of the overall effect size. Additionally, the test of significance of the overall effect size is based on a Student's $t$ distribution instead of the normal distribution.

*seadj* is `khartung`[`, truncated`]. Adjustment `khartung` specifies that the Knapp–Hartung adjustment (Hartung and Knapp 2001a, 2001b; Knapp and Hartung 2003), also known as the Sidik–Jonkman adjustment (Sidik and Jonkman 2002), be applied to the standard error of the overall effect size. `hknapp` and `sjonkman` are synonyms for `khartung`. `truncated` specifies that the truncated Knapp–Hartung adjustment (Knapp and Hartung 2003), also known as the modified Knapp–Hartung adjustment, be used.

## Options

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is as declared for the meta-analysis session; see *Declaring a confidence level for meta-analysis* in [META] **meta data**. Also see option `level()` in [META] **meta set**.

`citype(`*citype*`)` specifies the type of CI to be reported for meta-analysis of a single proportion. *citype* is one of `wald` (the default), `exact`, `wilson`, `agresti`, or `jeffreys`. For more details, see *Binomial proportion* in [R] **ci**. This option affects only individual study CIs and not the CI for the overall effect size. Thus, it may not be combined with options `cumulative()`, `leaveoneout`, and `subgroup()` with more than one variable.

*eform_option* is one of `eform`, `eform(`*string*`)`, `or`, or `rr`. It reports exponentiated effect sizes and transforms their respective confidence intervals, whenever applicable. By default, the results are displayed in the metric declared with `meta set` or `meta esize` such as log odds-ratios and log risk-ratios. *eform_option* affects how results are displayed, not how they are estimated and stored.

`eform(`*string*`)` labels the exponentiated effect sizes as *string*; the other options use default labels. The default label is specific to the chosen effect size. For example, option `eform` uses `Odds ratio` when used with log odds-ratios declared with `meta esize` or `Risk ratio` when used with the declared log risk-ratios. Option `or` is a synonym for `eform` when log odds-ratio is declared, and option `rr` is a synonym for `eform` when log risk-ratio is declared. If option `eslabel(`*eslab*`)` is specified during declaration, then `eform` will use the `exp(`*eslab*`)` label or, if *eslab* is too long, the `exp(ES)` label.

`transform(`[*label*`:`] *transf_name*`)` reports transformed effect sizes and CIs. *transf_name* is one of `corr`, `efficacy`, `exp`, `invlogit`, `tanh`, or `invftukey`[`,` *invftopts*]. When *label* is specified, the transformed effect sizes are labeled as *label* instead of using the default label. This option may not be combined with *eform_option*.

`corr` transforms effect sizes (and CIs) specified as Fisher's $z$ values into correlations and, by default, labels them as `Correlation`; that is, `transform(corr)` is a synonym for `transform(Correlation: tanh)`.

`efficacy` transforms the effect sizes and CIs using the $1 - $ `exp()` function (or more precisely, the `-expm1()` function) and labels them as `Efficacy`. This transformation is used, for example, when the effect sizes are log risk-ratios so that the transformed effect sizes can be interpreted as treatment efficacies, 1 − risk ratios.

`exp` exponentiates effect sizes and CIs and, by default, labels them as `exp(ES)`. This transformation is used, for example, when the effect sizes are log risk-ratios, log odds-ratios, and log hazard-ratios so that the transformed effect sizes can be interpreted as risk ratios, odds ratios, and hazard ratios. If the declared effect sizes are log odds-ratios or log risk-ratios, the default label is `Odds ratio` or `Risk ratio`, respectively.

invlogit transforms the effect sizes and CIs using the inverse-logit function, invlogit(), and, by default, labels them as invlogit(ES). This transformation is used, for example, when the effect sizes are logit of proportions so that the transformed effect sizes can be interpreted as proportions.

tanh applies the hyperbolic tangent transformation, tanh(), to the effect sizes and CIs and, by default, labels them as tanh(ES). This transformation is used, for example, when the effect sizes are Fisher's $z$ values so that the transformed effect sizes can be interpreted as correlations.

invftukey[ , *invftopts*] is relevant to meta-analysis of a single proportion. It applies the inverse Freeman–Tukey double arcsine transformation to the effect sizes and CIs and, by default, labels them as Proportion. This transformation is used only when pooling proportions (prevalences) with the default effect size esize(ftukeyprop). See *Inverse Freeman–Tukey transformation* for more details.

*invftopts* are hmean, gmean, amean, ivariance, and scale().

hmean specifies that the harmonic mean of the within-study sample sizes be used to back-transform the overall effect size.

gmean specifies that the geometric mean of the within-study sample sizes be used to back-transform the overall effect size.

amean specifies that the arithmetic mean of the within-study sample sizes be used to back-transform the overall effect size.

ivariance specifies that the inverse of the variance of the overall effect size be used to back-transform the overall effect size.

scale(#) scales the study proportions, the overall proportion, and their CIs by #. This option is relevant when the proportions are very small, in which case it might be preferable to report them as the number of successes per, say, 1,000 or 10,000 observations. # must be an integer greater than 1.

sort(*varlist*[ , ascending | descending]) sorts the studies in ascending or descending order based on values of the variables in *varlist*. This option is useful if you want to sort the studies in the output by effect sizes, sort(_meta_es), or by precision, sort(_meta_se). By default, ascending order is assumed unless the suboption descending is specified; only one of ascending or descending is allowed. *varlist* may contain string and numeric variables. This option is not allowed with cumulative(). When sort() is not specified, the order of the studies in the output is based on the ascending values of variable _meta_id, which is equivalent to sort(_meta_id).

tdistribution reports a $t$ test instead of a $z$ test for the overall effect size. This option may not be combined with option subgroup(), cumulative(), leaveoneout, or se().

proportion reports results as proportions for meta-analysis of a single proportion. By default, the results are displayed in the metric declared with meta esize, such as Freeman–Tukey-transformed proportions or logit-transformed proportions. proportion is a synonym for transform(invftukey, hmean) when the effect size is esize(ftukeyprop) or transform(invlogit) when the effect size is esize(logitprop). This option affects how results are displayed, not how they are estimated or stored.

prevalence is a synonym for proportion but labels the effect sizes as Prevalence instead of Proportion in the output. This option does not appear in the dialog box.

nostudies (synonym nostudy) suppresses the display of information such as effect sizes and their CIs for individual studies from the output table.

noheader suppresses the output header.

metashow and nometashow display or suppress the meta setting information. By default, this information is displayed at the top of the output. You can also specify nometashow with meta update to suppress the meta setting output for the entire meta-analysis session.

*display_options*: cformat(%*fmt*), pformat(%*fmt*), and sformat(%*fmt*); see [R] **Estimation options**. The defaults are cformat(%9.3f), pformat(%5.3f), and sformat(%8.2f).

wgtformat(%*fmt*) specifies how to format the weight column in the output table. The default is wgtformat(%5.2f). The maximum format width is 5.

ordformat(%*fmt*) specifies the format for the values of the order variable, specified in cumulative(*ordervar*). The default is ordformat(%9.0g). The maximum format width is 9.

Maximization

*maximize_options*: iterate(#), tolerance(#), nrtolerance(#), nonrtolerance (see [R] **Maximize**), from(#), and showtrace. These options control the iterative estimation of the between-study variance parameter, $\tau^2$, with random-effects methods reml, mle, and ebayes. These options are seldom used.

from(#) specifies the initial value for $\tau^2$ during estimation. By default, the initial value for $\tau^2$ is the noniterative Hedges estimator.

showtrace displays the iteration log that contains the estimated parameter $\tau^2$, its relative difference with the value from the previous iteration, and the scaled gradient.

# Remarks and examples

stata.com

Remarks are presented under the following headings:



## Introduction

Meta-analysis helps answer research questions based on the results of multiple studies. Does exercise prolong life? Does lack of sleep increase the risk of cancer? Does daylight saving save energy? Or does performing the duck-face technique while taking a selfie increase the number of likes on Facebook? These (except perhaps the last one) and many other research questions have been investigated by multiple studies. These studies may have reported conflicting results: some may have shown effects in one direction, some in the opposite, and others may have shown none that are statistically significant. Meta-analysis uses quantitative methods to explore these conflicting results and, whenever possible, provide a unified conclusion based on the results of the individual studies.

Meta-analysis combines the results of similar multiple studies into a single result. Studies typically report some measures of outcomes, or effect sizes, and their precision (standard errors or CIs). Meta-analysis combines the individual effects sizes to provide various meta-analysis summaries. The main summaries are the overall effect size and its precision. Other meta-analysis summaries include the test of significance of the overall effect size, between-study heterogeneity summaries such as the $I^2$ statistic, and the test of homogeneity between studies. The meta summarize command reports such summaries.

Estimating the overall effect size, $\theta$, and its precision based on the results of multiple studies is at the heart of meta-analysis. There are various methods for estimating $\theta$, which depend on the research goals and model assumptions about the studies. The estimate of the overall (combined) ES is computed as the weighted average of the study-specific effect sizes, with larger weights given to more precise (larger) studies:

$$\widehat{\theta} = \frac{\sum_{j=1}^{K} w_j \widehat{\theta}_j}{\sum_{j=1}^{K} w_j}$$

The weights are determined by the chosen meta-analysis model, estimation method, and potentially the type of effect size; see *Methods and formulas* for details. (In [META] **Intro**, we used $\theta_{\text{pop}}$ to denote the population parameter of interest. For simplicity, here and in the rest of the documentation, we will use $\theta$.)

As we described in *Meta-analysis models* in [META] **Intro**, the choice of a meta-analysis model is important not only for estimation but also for interpretation of $\widehat{\theta}$. `meta summarize` supports random-effects (`random`), fixed-effects (`fixed`), and common-effect (`common`) meta-analysis models. Each meta-analysis model provides various estimation methods such as the random-effects REML method, `random(reml)`, and fixed-effects Mantel–Haenszel method, `fixed(mhaenszel)`. The default model and method are as declared with `meta set` or `meta esize`; see *Declaring a meta-analysis model* in [META] **meta data**. Note that the Mantel–Haenszel method is available only with effect sizes `lnoratio`, `lnrratio`, and `rdiff` declared by using `meta esize`; see [META] **meta esize**.

For random-effects models, you can perform sensitivity meta-analysis to explore the impact of different levels of heterogeneity on the results. You can use the `tau2(#)` option to specify different fixed values for the between-study variance $\tau^2$. Or you can fix the percentage of variation in the effect sizes because of heterogeneity by specifying the values for the $I^2$ statistic in the `i2(#)` option. With random-effects models, you can also compute prediction intervals for $\widehat{\theta}$, `predinterval(#)`, and use the alternative standard-error estimators, `se()`.

You can perform subgroup analysis, `subgroup()`, CMA, `cumulative()`, or leave-one-out meta-analysis, `leaveoneout`; see *Subgroup meta-analysis*, *Cumulative meta-analysis*, and *Leave-one-out meta-analysis* in [META] **Intro**. Also see *Subgroup meta-analysis*, *Cumulative meta-analysis*, and *Leave-one-out meta-analysis* in *Methods and formulas* below.

You can sort the studies based on variables of interest via option `sort()`. For example, use `sort(_meta_es)` or `sort(_meta_weight)` to display the results based on the ascending order of the study effect sizes or study weights, respectively.

You can specify the desired confidence level with `level()`; report exponentiated results by specifying `eform`; report a $t$ test, instead of a $z$ test, for the overall effect size by specifying `tdistribution`; and more.

In the next section, we demonstrate various usages of `meta summarize`.

## Examples of using meta summarize

Recall the pupil IQ data (Raudenbush and Bryk 1985; Raudenbush 1984) described in *Effects of teacher expectancy on pupil IQ (pupiliq.dta)* of [META] **meta**. Here we will use its declared version and will focus on the demonstration of various options of `meta summarize` and explanation of its output.

```
. use https://www.stata-press.com/data/r18/pupiliqset
(Effects of teacher expectancy on pupil IQ; set with -meta set-)
. keep in 1/10
(9 observations deleted)
. meta query, short
-> meta set stdmdiff se , studylabel(studylbl) eslabel(Std. mean diff.)
  Effect-size label: Std. mean diff.
   Effect-size type: Generic
        Effect size: stdmdiff
          Std. err.: se
              Model: Random effects
             Method: REML
```

For brevity, we consider only the first 10 studies. We use `meta query, short` to remind us about the main settings of the declaration step. Our data were declared by using `meta set` with variables `stdmdiff` and `se` specifying the effect sizes and their standard errors, respectively. The declared meta-analysis model is the default random-effects model with the REML estimation method.

Examples are presented under the following headings:

## Example 1: Default random-effects meta-analysis

We type `meta summarize` to obtain a standard meta-analysis summary.

```
. meta summarize
  Effect-size label: Std. mean diff.
        Effect size: stdmdiff
          Std. err.: se
        Study label: studylbl
Meta-analysis summary                           Number of studies =     10
Random-effects model                            Heterogeneity:
Method: REML                                                tau2 =  0.0754
                                                         I2 (%) =   74.98
                                                            H2 =    4.00
                Effect size: Std. mean diff.
------------------------------------------------------------------------------
                       Study |    Effect size    [95% conf. interval]  % weight
-----------------------------+------------------------------------------------
  Rosenthal et al., 1974     |          0.030      -0.215       0.275      12.39
       Conn et al., 1968     |          0.120      -0.168       0.408      11.62
       Jose & Cody, 1971     |         -0.140      -0.467       0.187      10.92
Pellegrini & Hicks, 1972     |          1.180       0.449       1.911       5.25
Pellegrini & Hicks, 1972     |          0.260      -0.463       0.983       5.33
 Evans & Rosenthal, 1969     |         -0.060      -0.262       0.142      13.11
    Fielder et al., 1971     |         -0.020      -0.222       0.182      13.11
          Claiborn, 1969     |         -0.320      -0.751       0.111       9.11
            Kester, 1969     |          0.270      -0.051       0.591      11.02
            Maxwell, 1970    |          0.800       0.308       1.292       8.15
-----------------------------+------------------------------------------------
                       theta |          0.134      -0.075       0.342
------------------------------------------------------------------------------
Test of theta = 0: z = 1.26                     Prob > |z| = 0.2085
Test of homogeneity: Q = chi2(9) = 26.21          Prob > Q = 0.0019
```

As with other `meta` commands, a short information about meta settings is displayed directly following the `meta summarize` command. It can be suppressed with the `nometashow` option; see example 2.

Next, the header reports the information about the meta-analysis model and method, the number of studies (10), and several heterogeneity statistics. The output table reports the effect sizes and their 95% CIs for individual studies and the estimate of the overall, combined ES, labeled as `theta`, and its 95% CI. The test of significance of the overall effect size and the homogeneity test are reported at the bottom of the table.

Because our declared effect-size label, `Std. mean diff.`, was too long to fit as the column header, `meta summarize` used the generic column label `Effect size` but displayed the specified label in the table legend.

The mean effect size in our example is 0.134 with the 95% CI of $[-0.075, 0.342]$. This estimate is computed as the weighted average of the study-specific effect sizes, with the weights representing precision of the studies. The percentages of the total weight for each study are reported in the `% weight` column. The more precise the study is, the larger its weight percentage. For example, studies 6 and 7, with labels `Evans & Rosenthal, 1969` and `Fielder et al., 1971`, have the largest weight percentage among the studies of about 13% (each). Thus, their effect-size estimates, $-0.06$ and $-0.02$, have the largest weights in the weighted-average estimate.

The 95% CI for the overall estimate and the test of $H_0$: $\theta = 0$ with the $z$-test statistic of 1.26 and the $p$-value of 0.2085 suggest that $\theta$ is not statistically significantly different from 0. We should be careful, however, with our conclusions in the presence of between-study heterogeneity.

The heterogeneity statistic $I^2$, reported in the header, is about 75%, which means that 75% of the variability in the effect-size estimates is because of the between-study differences rather than the sampling variation. According to Higgins et al. (2003), this value of $I^2$ corresponds to "large heterogeneity". (The authors suggest that $I^2 = 25\%$ should indicate "small heterogeneity", $I^2 = 50\%$ indicate "medium heterogeneity", and $I^2 = 75\%$ indicate "large heterogeneity".) The between-study variance $\tau^2$ is estimated to be 0.0754. The homogeneity test of $H_0$: $\theta_1 = \theta_2 = \cdots = \theta_{10}$ reports the $Q$ test statistic of 26.21 with a $p$-value of 0.0019.

When there are few studies, which is typical in meta-analysis, the homogeneity test is known to have low power, which means that it may not detect clinically significant heterogeneity (Hedges and Pigott 2001). Thus, you should use caution when interpreting nonsignificant results as "no heterogeneity". In fact, many experts (for example, Berman and Parker [2002]) recommend using a 10% significance level instead of the classical 5% level to determine statistical significance when using this test. On the other hand, when there are many studies, this test is known to have excessive power, which means that it tends to detect heterogeneity that is clinically insignificant (Hardy and Thompson 1998).

In our example, the $p$-value of the homogeneity test is $0.0019 < 0.05 < 0.1$, so there is definitely statistical evidence of the between-study heterogeneity. See example 9 for one way to account for the heterogeneity.

◁

## ▷ Example 2: DerSimonian–Laird random-effects method

Continuing with example 1, let's use the DerSimonian–Laird random-effects method instead of the default (declared) REML method. Let's also suppress the meta setting information displayed at the top of the command output by using the `nometashow` option.

```
. meta summarize, random(dlaird) nometashow
```

```
Meta-analysis summary                          Number of studies =     10
Random-effects model                           Heterogeneity:
Method: DerSimonian-Laird                                  tau2 =  0.0481
                                                        I2 (%) =   65.66
                                                            H2 =    2.91

                Effect size: Std. mean diff.
```

| Study | Effect size | [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Rosenthal et al., 1974 | 0.030 | -0.215 | 0.275 | 13.00 |
| Conn et al., 1968 | 0.120 | -0.168 | 0.408 | 11.88 |
| Jose & Cody, 1971 | -0.140 | -0.467 | 0.187 | 10.90 |
| Pellegrini & Hicks, 1972 | 1.180 | 0.449 | 1.911 | 4.42 |
| Pellegrini & Hicks, 1972 | 0.260 | -0.463 | 0.983 | 4.49 |
| Evans & Rosenthal, 1969 | -0.060 | -0.262 | 0.142 | 14.11 |
| Fielder et al., 1971 | -0.020 | -0.222 | 0.182 | 14.11 |
| Claiborn, 1969 | -0.320 | -0.751 | 0.111 | 8.58 |
| Kester, 1969 | 0.270 | -0.051 | 0.591 | 11.04 |
| Maxwell, 1970 | 0.800 | 0.308 | 1.292 | 7.45 |
| theta | 0.117 | -0.061 | 0.296 | |

```
Test of theta = 0: z = 1.29                    Prob > |z| = 0.1967
Test of homogeneity: Q = chi2(9) = 26.21         Prob > Q = 0.0019
```

The results are now based on the DerSimonian–Laird method, and the header is updated to reflect this. This method is one of the many random-effects methods for estimating the between-study variance $\tau^2$. Its estimate is 0.0481. In random-effects models, the weights depend on $\tau^2$ and thus will differ across different random-effects methods. The mean effect-size estimate under the DerSimonian–Laird method is 0.117 with the 95% CI of $[-0.061, 0.296]$. This estimate is similar to the 0.134 estimate we obtained in example 1. We also arrive at the same inferential conclusion of no statistical significance of the mean effect size as in the previous example.

To shorten the output, let's suppress the meta setting information from the output of `meta summarize` for all remaining examples. We can use `meta update` to update our current meta settings.

```
. quietly meta update, nometashow
```

We specified the `nometashow` option with `meta update` to suppress the display of the meta setting information in all `meta` commands; see *Modifying default meta settings* in [META] **meta data**.

◁

## ▷ Example 3: Fixed-effects meta-analysis

In example 1, we assumed a random-effects meta-analysis model. We can use the `fixed` option to specify a fixed-effects meta-analysis model.

```
. meta summarize, fixed
```

```
Meta-analysis summary                          Number of studies =     10
Fixed-effects model                            Heterogeneity:
Method: Inverse-variance                                 I2 (%) =   65.66
                                                         H2 =     2.91

                Effect size: Std. mean diff.
```

| Study | Effect size | [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Rosenthal et al., 1974 | 0.030 | -0.215 | 0.275 | 15.13 |
| Conn et al., 1968 | 0.120 | -0.168 | 0.408 | 10.94 |
| Jose & Cody, 1971 | -0.140 | -0.467 | 0.187 | 8.48 |
| Pellegrini & Hicks, 1972 | 1.180 | 0.449 | 1.911 | 1.70 |
| Pellegrini & Hicks, 1972 | 0.260 | -0.463 | 0.983 | 1.74 |
| Evans & Rosenthal, 1969 | -0.060 | -0.262 | 0.142 | 22.29 |
| Fielder et al., 1971 | -0.020 | -0.222 | 0.182 | 22.29 |
| Claiborn, 1969 | -0.320 | -0.751 | 0.111 | 4.89 |
| Kester, 1969 | 0.270 | -0.051 | 0.591 | 8.79 |
| Maxwell, 1970 | 0.800 | 0.308 | 1.292 | 3.75 |
| theta | 0.051 | -0.045 | 0.146 | |

```
Test of theta = 0: z = 1.04                      Prob > |z| = 0.2974
Test of homogeneity: Q = chi2(9) = 26.21           Prob > Q = 0.0019
```

As reported in the header, `fixed` implied the inverse-variance estimation method. The between-group variance parameter is not estimated with fixed-effects models, so the heterogeneity summary does not report `tau2`. Under this model, the mean effect-size estimate is 0.051 with the 95% CI of $[-0.045, 0.146]$. As we explain in *Comparison between the models and interpretation of their results* in [META] **Intro**, in a fixed-effects model, `theta` estimates the weighted average of the true study-specific standardized mean differences. Our interpretation is also limited to these 10 studies that we observed in our meta-analysis. That is, the weighted average of the standardized mean differences of these 10 studies is not statistically significantly different from 0.

◁

## Example 4: Common-effect meta-analysis

From example 1 and example 3, we determined that there is substantial between-study variability in these data. Thus, a common-effect model, which assumes that all study-specific effects are the same, is not reasonable for these data. But we will demonstrate it for illustration purposes.

```
. meta summarize, common
Meta-analysis summary                           Number of studies =     10
Common-effect model
Method: Inverse-variance
                Effect size: Std. mean diff.
```

| Study | Effect size | [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Rosenthal et al., 1974 | 0.030 | -0.215 | 0.275 | 15.13 |
| Conn et al., 1968 | 0.120 | -0.168 | 0.408 | 10.94 |
| Jose & Cody, 1971 | -0.140 | -0.467 | 0.187 | 8.48 |
| Pellegrini & Hicks, 1972 | 1.180 | 0.449 | 1.911 | 1.70 |
| Pellegrini & Hicks, 1972 | 0.260 | -0.463 | 0.983 | 1.74 |
| Evans & Rosenthal, 1969 | -0.060 | -0.262 | 0.142 | 22.29 |
| Fielder et al., 1971 | -0.020 | -0.222 | 0.182 | 22.29 |
| Claiborn, 1969 | -0.320 | -0.751 | 0.111 | 4.89 |
| Kester, 1969 | 0.270 | -0.051 | 0.591 | 8.79 |
| Maxwell, 1970 | 0.800 | 0.308 | 1.292 | 3.75 |
| theta | 0.051 | -0.045 | 0.146 | |

```
Test of theta = 0: z = 1.04                     Prob > |z| = 0.2974
```

We use the `common` option to specify a common-effect model. Because this model implies no heterogeneity, the corresponding summaries and the homogeneity test are not reported for this model. As we point out in *Comparison between the models and interpretation of their results* in [META] **Intro**, a common-effect model is computationally the same as a fixed-effects model. So we obtain the exact same results as in example 3. However, the interpretation of our results is different. Here `theta` estimates a single effect, which is common to all studies. Although the two models produce the same results, to encourage proper interpretation, we provide both options, `common` and `fixed`, to distinguish between these models; see *Declaring a meta-analysis model* in [META] **meta data** for details.

## ▷ Example 5: Knapp–Hartung standard-error adjustment

Let's return to our random-effects model from example 1. For random-effects models, `meta summarize` provides several additional options, which we explore in the next three examples.

The Knapp–Hartung adjustment (also known as the Sidik–Jonkman adjustment) to the standard error of the overall effect size (Knapp and Hartung 2003 and Hartung and Knapp 2001a, 2001b) is sometimes used in practice. We can specify it with the `se(khartung)` option. We also specify the `nostudies` option to suppress the output from individual studies because it is unaffected by the `se(khartung)` option.

```
. meta summarize, se(khartung) nostudies
Meta-analysis summary                               Number of studies =     10
Random-effects model                                Heterogeneity:
Method: REML                                                    tau2 =  0.0754
SE adjustment: Knapp-Hartung                                 I2 (%) =   74.98
                                                                H2 =    4.00

        theta: Overall Std. mean diff.
------------------------------------------------------------------------------
             |    Estimate    Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       theta |    .1335309    .1215065     1.10   0.300    -.1413358     .4083976
------------------------------------------------------------------------------
Test of homogeneity: Q = chi2(9) = 26.21                  Prob > Q = 0.0019
```

Without the individual studies, the output table is slightly different. The test of significance is now reported in the output table instead of at the bottom of the output table.

The estimate `theta` is the same as in example 1, 0.134, but it is reported with more digits in this table. The confidence intervals and the test of significance are different. In addition to making an adjustment to the standard error, Knapp and Hartung also use a Student's $t$ distribution as a sampling distribution instead of the normal distribution. Thus, the $t$ statistic is reported in the output table instead of the $z$ statistic. Regardless, we still conclude that our overall effect size is not statistically significant.

Another standard error adjustment, also used in practice, is the so-called truncated or modified Knapp–Hartung adjustment; see *Methods and formulas* for details. This adjustment can be specified with the `se(khartung, truncated)` option.

```
. meta summarize, se(khartung, truncated)
  (output omitted)
```

◁

## ▷ Example 6: Prediction interval

Recall from *Random-effects model* in [META] **Intro** that a random-effects model implies that the observed studies in a meta-analysis represent a sample from a larger population of similar studies. What if we want to estimate the plausible ranges for the overall effect size in a new, future study? We cannot use the confidence interval for the overall effect size because it does not incorporate the uncertainty in estimating the between-study variance, which is important if we want to predict an effect in a new study. We can compute the prediction interval.

```
. meta summarize, predinterval(90) nostudies
Meta-analysis summary                           Number of studies =     10
Random-effects model                            Heterogeneity:
Method: REML                                                tau2 =  0.0754
                                                         I2 (%) =   74.98
                                                             H2 =    4.00

        theta: Overall Std. mean diff.
------------------------------------------------------------------------------
             |    Estimate    Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       theta |     .1335309    .1061617     1.26   0.208    -.0745422     .3416041
------------------------------------------------------------------------------
90% prediction interval for theta: [-0.414, 0.681]
Test of homogeneity: Q = chi2(9) = 26.21                 Prob > Q = 0.0019
```

We specified `predinterval(90)` to compute the 90% prediction interval for the mean effect size; use `predinterval` to compute the 95% interval. Following example 5, we also used `nostudies` to suppress individual studies.

The 90% prediction interval, reported at the bottom of the table, is $[-0.414, 0.681]$. The prediction interval will be wider than the confidence interval because it additionally accounts for the uncertainty in the between-study variability.

◁

## ▷ Example 7: Sensitivity meta-analysis

For random-effects models, we can perform sensitivity analysis to explore various levels of heterogeneity between studies. Let's see how our results change for different values of the between-study variance $\tau^2$ and the heterogeneity statistic $I^2$.

Let's compute the results assuming that $\tau^2$ equals 0.25.

```
. meta summarize, tau2(0.25) nostudies
Sensitivity meta-analysis summary               Number of studies =     10
Random-effects model                            Heterogeneity:
Method: User-specified tau2                                 tau2 =  0.2500
                                                         I2 (%) =   90.86
                                                             H2 =   10.94

        theta: Overall Std. mean diff.
------------------------------------------------------------------------------
             |    Estimate    Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       theta |      .173588     .171407     1.01   0.311    -.1623636     .5095395
------------------------------------------------------------------------------
Test of homogeneity: Q = chi2(9) = 26.21                 Prob > Q = 0.0019
```

Our estimate of the mean effect size is 0.174 with the 95% CI of $[-0.162, 0.51]$ compared with 0.134 with the 95% CI of $[-0.075, 0.342]$ from example 1.

The specified value of $\tau^2$ corresponds to an $I^2$ of about 91%. Let's now compute the results assuming $I^2$ of 10%.

```
. meta summarize, i2(10) nostudies
Sensitivity meta-analysis summary                   Number of studies =     10
Random-effects model                                Heterogeneity:
Method: User-specified I2                                        tau2 =  0.0028
                                                              I2 (%) =   10.00
                                                                  H2 =    1.11

        theta: Overall Std. mean diff.
------------------------------------------------------------------------------
             |    Estimate    Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       theta |    .0589369    .0527232     1.12   0.264    -.0443987    .1622724
------------------------------------------------------------------------------
Test of homogeneity: Q = chi2(9) = 26.21                 Prob > Q = 0.0019
```

The estimate of the mean effect size is 0.059 with the 95% CI of $[-0.044, 0.162]$. The corresponding $\tau^2$ value is 0.0028.

In both cases above, the mean effect size is not statistically significant.

◁

## ▷ Example 8: Other options: CI level, t distribution, sort, eform

`meta summarize` provides other options such as `level()` to temporarily change the declared confidence level and `tdistribution` to use a Student's $t$ distribution as the sampling distribution instead of the default normal distribution.

```
. meta summarize, level(90) tdistribution
Meta-analysis summary                               Number of studies =     10
Random-effects model                                Heterogeneity:
Method: REML                                                     tau2 =  0.0754
                                                              I2 (%) =   74.98
                                                                  H2 =    4.00

                Effect size: Std. mean diff.
------------------------------------------------------------------------------
                       Study |    Effect size    [90% conf. interval]  % weight
-----------------------------+------------------------------------------------
  Rosenthal et al., 1974     |          0.030          -0.176       0.236     12.39
       Conn et al., 1968     |          0.120          -0.122       0.362     11.62
       Jose & Cody, 1971     |         -0.140          -0.415       0.135     10.92
Pellegrini & Hicks, 1972     |          1.180           0.566       1.794      5.25
Pellegrini & Hicks, 1972     |          0.260          -0.347       0.867      5.33
 Evans & Rosenthal, 1969     |         -0.060          -0.229       0.109     13.11
    Fielder et al., 1971     |         -0.020          -0.189       0.149     13.11
          Claiborn, 1969     |         -0.320          -0.682       0.042      9.11
            Kester, 1969     |          0.270           0.000       0.540     11.02
          Maxwell, 1970      |          0.800           0.387       1.213      8.15
-----------------------------+------------------------------------------------
                     theta   |          0.134          -0.061       0.328
------------------------------------------------------------------------------
Test of theta = 0: t(9) = 1.26                           Prob > |t| = 0.2401
Test of homogeneity: Q = chi2(9) = 26.21                   Prob > Q = 0.0019
```

Notice that all CIs, including those for the individual studies, now correspond to the 90% confidence level, compared with example 1. Also, the significance test now uses the Student's $t$ distribution with 9 degrees of freedom, but the conclusion remains the same—the mean effect size is not statistically significant.

You may also find `meta summarize`'s option `eform` useful when dealing with the effect sizes in the log-transformed metric such as log odds-ratios or log risk-ratios. By default, `meta summarize` reports results in the declared metric, which should be chosen such that the sampling distributions of the effect sizes are well approximated by normal distributions. It may be more convenient, however, to display the final results in the original metric. When you specify the `eform` option, it reports the exponentiated results and the corresponding CIs. Note that the significance tests and other summary measures are still computed based on the nonexponentiated results.

It does not make sense to exponentiate standardized mean differences in our example, but we will do this just to demonstrate the option.

We will also use the `sort()` option to sort our results based on the descending order of study weights, with larger, more precise studies appearing first.

```
. meta summarize, eform sort(_meta_weight, descending)
Meta-analysis summary                           Number of studies =     10
Random-effects model                            Heterogeneity:
Method: REML                                                tau2 =  0.0754
                                                         I2 (%) =   74.98
                                                            H2 =    4.00

                    exp(ES): exp(Std. mean diff.)
------------------------------------------------------------------------------
                       Study |     exp(ES)    [95% conf. interval]  % weight
-----------------------------+------------------------------------------------
 Evans & Rosenthal, 1969     |        0.942       0.770       1.152      13.11
    Fielder et al., 1971     |        0.980       0.801       1.199      13.11
  Rosenthal et al., 1974     |        1.030       0.807       1.317      12.39
       Conn et al., 1968     |        1.127       0.845       1.504      11.62
            Kester, 1969     |        1.310       0.950       1.807      11.02
       Jose & Cody, 1971     |        0.869       0.627       1.206      10.92
          Claiborn, 1969     |        0.726       0.472       1.118       9.11
           Maxwell, 1970     |        2.226       1.361       3.640       8.15
Pellegrini & Hicks, 1972     |        1.297       0.629       2.673       5.33
Pellegrini & Hicks, 1972     |        3.254       1.567       6.760       5.25
-----------------------------+------------------------------------------------
                exp(theta)   |        1.143       0.928       1.407
------------------------------------------------------------------------------
Sorted by: _meta_weight
Test of theta = 0: z = 1.26                     Prob > |z| = 0.2085
Test of homogeneity: Q = chi2(9) = 26.21          Prob > Q = 0.0019
```

`meta summarize, eform` reports exponentiated effect sizes and their corresponding CIs. It labels the effect-size column as `exp(ES)`, but you can change this label to *string* by specifying `eform(`*string*`)`.

Note that the `eform` option worked in our example because `meta set` declared our precomputed effect sizes as generic. They could have been log odds-ratios, in which case `eform` would make perfect sense. However, if you use `meta esize` to compute the standardized mean differences (for example, Hedges's $g$) and try to use `eform` with `meta summarize`, you will receive an error message because `meta summarize` knows that exponentiation is not appropriate with effect sizes that correspond to continuous data. With effect sizes `lnoratio` (or `lnorpeto`) and `lnrratio` computed by `meta esize`, you can also use the respective options `or` and `rr`, which are synonyms for `eform` in those cases. These options (and `eform`) will label your results as `Odds ratio` (`Peto's OR`) and `Risk ratio`.

◁

## ▷ Example 9: Subgroup meta-analysis

In example 1 and example 3, we identified the presence of substantial heterogeneity between the observed studies. Sometimes, the heterogeneity can be explained by some study-level covariates, also known as moderators. With categorical moderators, we can perform subgroup analysis, which performs meta-analysis separately for each category of each moderator.

We have binary variable `week1`, which records whether teachers had prior contact with students for more than 1 week or for 1 week or less. Let's use this variable as the moderator in our subgroup analysis. We specify the variable `week1` in the `subgroup()` option.

```
. meta summarize, subgroup(week1)
Subgroup meta-analysis summary                    Number of studies =     10
Random-effects model
Method: REML
Group: week1
                Effect size: Std. mean diff.
```

| Study | Effect size | [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Group: <= 1 week | | | | |
| Pellegrini & Hicks, 1972 | 1.180 | 0.449 | 1.911 | 5.25 |
| Pellegrini & Hicks, 1972 | 0.260 | -0.463 | 0.983 | 5.33 |
| Kester, 1969 | 0.270 | -0.051 | 0.591 | 11.02 |
| Maxwell, 1970 | 0.800 | 0.308 | 1.292 | 8.15 |
| theta | 0.581 | 0.174 | 0.989 | |
| Group: > 1 week | | | | |
| Rosenthal et al., 1974 | 0.030 | -0.215 | 0.275 | 12.39 |
| Conn et al., 1968 | 0.120 | -0.168 | 0.408 | 11.62 |
| Jose & Cody, 1971 | -0.140 | -0.467 | 0.187 | 10.92 |
| Evans & Rosenthal, 1969 | -0.060 | -0.262 | 0.142 | 13.11 |
| Fielder et al., 1971 | -0.020 | -0.222 | 0.182 | 13.11 |
| Claiborn, 1969 | -0.320 | -0.751 | 0.111 | 9.11 |
| theta | -0.033 | -0.137 | 0.071 | |
| Overall | | | | |
| theta | 0.134 | -0.075 | 0.342 | |

Heterogeneity summary

| Group | df | Q | P > Q | tau2 | % I2 | H2 |
|---|---|---|---|---|---|---|
| <= 1 week | 3 | 7.14 | 0.068 | 0.095 | 57.03 | 2.33 |
| > 1 week | 5 | 3.53 | 0.618 | 0.000 | 0.00 | 1.00 |
| Overall | 9 | 26.21 | 0.002 | 0.075 | 74.98 | 4.00 |

Test of group differences: Q_b = chi2(1) = 8.18 Prob > Q_b = 0.004

We now have two output tables. Our main table now reports results from individual studies separately for each group, in addition to the group-specific overall effect size. The overall effect size computed using all studies is reported at the bottom under `Overall`.

The second table reports the group-specific and overall heterogeneity summaries. The test of group differences is reported at the bottom of this table.

The estimated `theta` for the group with contact `<= 1 week` is 0.581 with the 95% CI of $[0.174, 0.989]$. The mean effect size in this group is statistically significant at the 5% level. The estimated `theta`

for the group with contact > 1 week is $-0.033$ with the 95% CI of $[-0.137, 0.071]$. The mean effect size in this group is not statistically significant at the 5% level.

If we look at the heterogeneity summaries, the <= 1 week group still has some unexplained between-study heterogeneity with an estimated $I^2$ of 57% and a $p$-value of the homogeneity test of $0.068 < 0.1$. There does not appear to be any between-study heterogeneity in the > 1 week group: $I^2$ is essentially 0%, and the homogeneity test $p$-value is 0.618.

We should interpret our results with caution because each subgroup analysis used a few studies, with the <= 1 week group having only 4 studies.

We can specify multiple variables in the subgroup() option. Let's also include variable tester in our subgroup analysis.

```
. meta summarize, subgroup(week1 tester)
Subgroup meta-analysis summary                    Number of studies =     10
Random-effects model
Method: REML
Group: week1 tester
```

| Group | No. of studies | Std. mean diff. | [95% conf. | interval] | p-value |
|---|---|---|---|---|---|
| week1 | | | | | |
| <= 1 week | 4 | 0.581 | 0.174 | 0.989 | 0.005 |
| > 1 week | 6 | -0.033 | -0.137 | 0.071 | 0.535 |
| tester | | | | | |
| Aware | 7 | 0.059 | -0.129 | 0.247 | 0.535 |
| Blind | 3 | 0.316 | -0.206 | 0.837 | 0.235 |
| Overall | | | | | |
| theta | 10 | 0.134 | -0.075 | 0.342 | 0.208 |

Heterogeneity summary

| Group | df | Q | P > Q | tau2 | % I2 | H2 |
|---|---|---|---|---|---|---|
| week1 | | | | | | |
| <= 1 week | 3 | 7.14 | 0.068 | 0.095 | 57.03 | 2.33 |
| > 1 week | 5 | 3.53 | 0.618 | 0.000 | 0.00 | 1.00 |
| tester | | | | | | |
| Aware | 6 | 16.35 | 0.012 | 0.035 | 59.07 | 2.44 |
| Blind | 2 | 9.31 | 0.009 | 0.154 | 75.14 | 4.02 |
| Overall | 9 | 26.21 | 0.002 | 0.075 | 74.98 | 4.00 |

Tests of group differences

| | df | Q_b | P > Q_b |
|---|---|---|---|
| week1 | 1 | 8.18 | 0.004 |
| tester | 1 | 0.82 | 0.365 |

With more than one variable in subgroup(), meta summarize reports three output tables. To conserve space, the main table does not report individual studies but reports the number of studies in each group. It also reports the $p$-values of the corresponding significance tests of the overall effect sizes in each group.

The heterogeneity table reports the group summaries for each variable, in addition to the overall summaries. The new table reports the results of tests of subgroup differences for each variable.

The studies appear to be homogeneous across the levels of the `tester` variable.

◁

## ▷ Example 10: Meta-analysis of correlations and the transform() option

Molloy, O'Carroll, and Ferguson (2013) conducted a meta-analysis to examine to what degree conscientiousness is related to medication adherence. Medication adherence is the extent to which typically chronically ill patients follow medical recommendations as prescribed. Conscientiousness is defined as "*socially prescribed impulse control* that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing and prioritizing tasks" (John and Srivastava 1999, 121).

The dataset contains the variables `studylbl`, `rho`, and `n` to indicate the authors and year of publication of the studies, the correlation coefficient between conscientiousness and medication adherence, and the study sample size, respectively.

```
. use https://www.stata-press.com/data/r18/adherence
(Conscientiousness and medication adherence)
. describe n rho studylbl
Variable      Storage   Display    Value
    name         type    format    label      Variable label
─────────────────────────────────────────────────────────────────────────
n               int     %9.0g                 Sample size of the study
rho             double  %9.0g               * Correlation coefficient
studylbl        str26   %26s                  Study label
```

The correlation coefficient `rho` is measured on the natural scale ($-1 \leq r \leq 1$), so the first step is to transform `rho` using the Fisher'$z$ transformation as follows:

$$\texttt{fisherz} = \frac{1}{2}\log\left(\frac{1+\texttt{rho}}{1-\texttt{rho}}\right) = \texttt{atanh(rho)} \sim N\left(0, \frac{1}{\texttt{n}-3}\right)$$

If the underlying data are bivariate normal, the variance of `fisherz` equals $1/(\texttt{n}-3)$ and depends only on the within-study sample size and not on the correlation parameter itself. Fisher's $z$ transformation is available in Stata using the `atanh()` function. Below, we also generate the `se` variable to contain the values of the (asymptotic) standard errors of the Fisher's $z$ values in each study and use `meta set` to declare our `meta` data.

```
. generate double fisherz = atanh(rho)
. generate double se = sqrt(1/(n-3))
. meta set fisherz se, studylabel(studylbl) nometashow
Meta-analysis setting information
 Study information
    No. of studies: 16
       Study label: studylbl
        Study size: N/A
       Effect size
              Type: <generic>
             Label: Effect size
          Variable: fisherz
         Precision
         Std. err.: se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%
  Model and method
             Model: Random effects
            Method: REML
```

The meta-analysis summary may be obtained as follows:

```
. meta summarize
Meta-analysis summary                          Number of studies =     16
Random-effects model                           Heterogeneity:
Method: REML                                              tau2 =  0.0081
                                                       I2 (%) =   61.73
                                                           H2 =    2.61
```

| Study | Effect size | [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Axelsson et al. (2009) | 0.189 | -0.001 | 0.380 | 5.68 |
| Axelsson et al. (2011) | 0.163 | 0.092 | 0.235 | 10.54 |
| Bruce et al. (2010) | 0.354 | 0.082 | 0.626 | 3.64 |
| Christensen et al. (1999) | 0.332 | 0.139 | 0.524 | 5.62 |
| Christensen & Smith (1995) | 0.277 | 0.041 | 0.513 | 4.41 |
| Cohen et al. (2004) | 0.000 | -0.249 | 0.249 | 4.11 |
| Dobbels et al. (2005) | 0.177 | 0.027 | 0.327 | 7.14 |
| Ediger et al. (2007) | 0.050 | -0.059 | 0.159 | 8.89 |
| Insel et al. (2006) | 0.266 | 0.002 | 0.530 | 3.79 |
| Jerant et al. (2011) | 0.010 | -0.061 | 0.081 | 10.58 |
| Moran et al. (1997) | -0.090 | -0.359 | 0.179 | 3.69 |
| O'Cleirigh et al. (2007) | 0.388 | 0.179 | 0.597 | 5.11 |
| Penedo et al. (2003) | 0.000 | -0.184 | 0.184 | 5.87 |
| Quine et al. (2012) | 0.151 | 0.066 | 0.236 | 9.98 |
| Stilley et al. (2004) | 0.245 | 0.087 | 0.402 | 6.84 |
| Wiebe & Christensen (1997) | 0.040 | -0.209 | 0.289 | 4.11 |
| theta | 0.150 | 0.088 | 0.212 | |

```
Test of theta = 0: z = 4.75                    Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(15) = 38.16        Prob > Q = 0.0009
```

The overall Fisher's $z$ value (transformed correlation coefficient) across the 16 studies is estimated to be 0.150 using the REML RE meta-analysis model.

The interpretation of the results, however, is easier in the natural correlation-coefficient metric, which we can compute using the inverse transformation:

$$\texttt{rho} = \frac{\exp(2 \times \texttt{fisherz}) - 1}{\exp(2 \times \texttt{fisherz}) + 1} = \texttt{tanh(fisherz)}$$

Thus, you may obtain the value of the correlation coefficient and its CI by typing

```
. display tanh(r(theta))
.14880413
. display "[" tanh(r(ci_lb)) ", " tanh(r(ci_ub)) "]"
[.08783366, .20866384]
```

More conveniently, you can use the `transform(corr)` option to report correlations. This option applies the hyperbolic tangent (`tanh()`) transformation to the Fisher's $z$ values and labels the resulting effect sizes as `Correlation`. Notice that specifying `transform(corr)` is equivalent to specifying `transform(Correlation: tanh)`.

```
. meta summarize, transform(corr)
Meta-analysis summary                           Number of studies =     16
Random-effects model                            Heterogeneity:
Method: REML                                                tau2 =  0.0081
                                                         I2 (%) =   61.73
                                                            H2 =    2.61
```

| Study | Correlation | [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Axelsson et al. (2009) | 0.187 | -0.001 | 0.362 | 5.68 |
| Axelsson et al. (2011) | 0.162 | 0.091 | 0.231 | 10.54 |
| Bruce et al. (2010) | 0.340 | 0.082 | 0.555 | 3.64 |
| Christensen et al. (1999) | 0.320 | 0.139 | 0.481 | 5.62 |
| Christensen & Smith (1995) | 0.270 | 0.041 | 0.472 | 4.41 |
| Cohen et al. (2004) | 0.000 | -0.244 | 0.244 | 4.11 |
| Dobbels et al. (2005) | 0.175 | 0.027 | 0.316 | 7.14 |
| Ediger et al. (2007) | 0.050 | -0.059 | 0.158 | 8.89 |
| Insel et al. (2006) | 0.260 | 0.002 | 0.486 | 3.79 |
| Jerant et al. (2011) | 0.010 | -0.061 | 0.081 | 10.58 |
| Moran et al. (1997) | -0.090 | -0.345 | 0.177 | 3.69 |
| O'Cleirigh et al. (2007) | 0.370 | 0.178 | 0.535 | 5.11 |
| Penedo et al. (2003) | 0.000 | -0.182 | 0.182 | 5.87 |
| Quine et al. (2012) | 0.150 | 0.066 | 0.232 | 9.98 |
| Stilley et al. (2004) | 0.240 | 0.087 | 0.382 | 6.84 |
| Wiebe & Christensen (1997) | 0.040 | -0.206 | 0.281 | 4.11 |
| tanh(theta) | 0.149 | 0.088 | 0.209 | |

```
Test of theta = 0: z = 4.75                     Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(15) = 38.16         Prob > Q = 0.0009
```

The overall correlation value is 0.149 with a CI of $[0.088, 0.209]$.

◁

## ▷ Example 11: Meta-analysis of a single proportion

Continuing from the `meta esize ndeaths pensize` setting in example 4 of [META] **meta data**, we produce a meta-analysis summary and compute the overall proportion as follows:

```
. meta summarize, proportion

  Effect-size label: Freeman-Tukey's p
        Effect size: _meta_es
          Std. err.: _meta_se

Meta-analysis summary                     Number of studies =      4
Random-effects model                      Heterogeneity:
Method: REML                                          tau2 =  0.0000
                                                   I2 (%) =    0.00
                                                       H2 =    1.00

------------------------------------------------------------------------------
             Study |      Proportion    [95% conf. interval]  % weight
-------------------+----------------------------------------------------------
           Study 1 |           0.273       0.044       0.579     20.18
           Study 2 |           0.353       0.140       0.598     30.70
           Study 3 |           0.476       0.264       0.693     37.72
           Study 4 |           0.167       0.145       0.586     11.40
-------------------+----------------------------------------------------------
 invftukey(theta)  |           0.360       0.230       0.499
------------------------------------------------------------------------------
Test of theta = 0: z = 7.67                  Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(3) = 2.18       Prob > Q = 0.5368
```

The overall proportion is estimated to be 0.360 with a CI of $[0.230, 0.499]$.

The `proportion` option was used to report proportions instead of the Freeman–Tukey-transformed proportions. This option is equivalent to `transform(invftukey, hmean)`, where `hmean` specifies that the harmonic mean of the study-specific sample sizes be used as $n_\theta$ to back-transform the overall effect size [see (4) in *Inverse Freeman–Tukey transformation* for details]. Instead of the harmonic mean, Barendregt et al. (2013) suggested to use the inverse of the variance of the overall Freeman–Tukey-transformed proportion as an estimate of $n_\theta$. This may be requested via `transform(invftukey, ivariance)`.

```
. meta summarize, transform(invftukey, ivariance)

  Effect-size label: Freeman-Tukey's p
        Effect size: _meta_es
          Std. err.: _meta_se

Meta-analysis summary                     Number of studies =      4
Random-effects model                      Heterogeneity:
Method: REML                                          tau2 =  0.0000
                                                   I2 (%) =    0.00
                                                       H2 =    1.00

------------------------------------------------------------------------------
             Study |      Proportion    [95% conf. interval]  % weight
-------------------+----------------------------------------------------------
           Study 1 |           0.273       0.044       0.579     20.18
           Study 2 |           0.353       0.140       0.598     30.70
           Study 3 |           0.476       0.264       0.693     37.72
           Study 4 |           0.167       0.145       0.586     11.40
-------------------+----------------------------------------------------------
 invftukey(theta)  |           0.369       0.247       0.499
------------------------------------------------------------------------------
Note: Method ivariance is used to compute overall proportion.
Test of theta = 0: z = 8.89                  Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(3) = 2.18       Prob > Q = 0.5368
```

Finally, the CIs for the Freeman–Tukey-transformed proportions are the standard normal-based Wald intervals. These are stored in system variables _meta_cil and _meta_ciu. The CIs displayed in the table above are the corresponding back-transformed (using transform(invftukey)) confidence intervals in the proportion metric, and these are stored in _meta_cil_transf and _meta_ciu_transf.

When you report proportions either via the proportion or transform() option, you can use the citype() option to display other types of CIs for the study proportions. Below, we display Wilson CIs for the study proportions.

```
. meta summarize, transform(invftukey, ivariance) citype(wilson)

  Effect-size label: Freeman-Tukey's p
        Effect size: _meta_es
          Std. err.: _meta_se

Meta-analysis summary                     Number of studies =      4
Random-effects model                      Heterogeneity:
Method: REML                                          tau2 =  0.0000
                                                   I2 (%) =    0.00
                                                       H2 =    1.00
```

| Study | Proportion | Wilson [95% conf. | interval] | % weight |
|---|---|---|---|---|
| Study 1 | 0.273 | 0.097 | 0.566 | 20.18 |
| Study 2 | 0.353 | 0.173 | 0.587 | 30.70 |
| Study 3 | 0.476 | 0.283 | 0.676 | 37.72 |
| Study 4 | 0.167 | 0.030 | 0.564 | 11.40 |
| invftukey(theta) | 0.369 | 0.247 | 0.499 | |

```
Note: Method ivariance is used to compute overall proportion.
Note: Wilson CIs are reported only for individual studies.
Test of theta = 0: z = 8.89                     Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(3) = 2.18           Prob > Q = 0.5368
```

The citype() option applies to the CIs of individual studies only and not to the CI of the overall proportion.

◁

## ▷ Example 12: Cumulative meta-analysis

CMA (Lau et al. 1992; Sterne 2016) performs multiple meta-analyses by accumulating studies one at a time. The studies are first ordered with respect to a variable of interest, the ordering variable. Meta-analysis summaries are then computed for the first study, for the first two studies, for the first three studies, and so on. The last meta-analysis will correspond to the standard meta-analysis using all studies.

CMA is useful, for instance, for identifying the point in time of the potential change in the direction or significance of the effect size when the ordering variable is time. You can use the cumulative() option to perform CMA.

For demonstration purposes, let's continue with the dataset in example 1 and use `year` as our ordering variable.

```
. meta summarize, cumulative(year)
Cumulative meta-analysis summary                    Number of studies =     10
Random-effects model
Method: REML
Order variable: year
               theta: Overall Std. mean diff.
```

| Study | theta | [95% conf. | interval] | p-value | year |
|---|---|---|---|---|---|
| Conn et al., 1968 | 0.120 | -0.168 | 0.408 | 0.414 | 1968 |
| Evans & Rosent~1969 | -0.001 | -0.166 | 0.165 | 0.995 | 1969 |
| Claiborn, 1969 | -0.042 | -0.201 | 0.117 | 0.605 | 1969 |
| Kester, 1969 | 0.022 | -0.177 | 0.221 | 0.830 | 1969 |
| Maxwell, 1970 | 0.140 | -0.178 | 0.459 | 0.389 | 1970 |
| Jose & Cody, 1971 | 0.089 | -0.177 | 0.355 | 0.510 | 1971 |
| Fielder et al., 1~1 | 0.064 | -0.141 | 0.270 | 0.539 | 1971 |
| Pellegrini & H~1972 | 0.161 | -0.117 | 0.438 | 0.257 | 1972 |
| Pellegrini & H~1972 | 0.161 | -0.090 | 0.413 | 0.208 | 1972 |
| Rosenthal et.., 1~4 | 0.134 | -0.075 | 0.342 | 0.208 | 1974 |

The output table reports the overall effect size and its CIs for each cumulative analysis. The `p-value` column contains the $p$-values of the significance tests of the overall effect sizes from these analyses. The last column displays the values of the ordering variable.

In our example, no particular trend is apparent.

We can perform stratified CMA by specifying a categorical variable in `cumulative()`'s option `by()`. To demonstrate, we also specify `cumulative()`'s option `descending` to list results in descending order of `year`.

```
. meta summarize, cumulative(year, by(week1) descending)
Stratified cumulative meta-analysis summary         Number of studies =     10
Random-effects model
Method: REML
Order variable: year (descending)
Stratum: week1
               theta: Overall Std. mean diff.
```

| Study | theta | [95% conf. | interval] | p-value | year |
|---|---|---|---|---|---|
| Group: <= 1 week | | | | | |
| Pellegrini & H~1972 | 0.260 | -0.463 | 0.983 | 0.481 | 1972 |
| Pellegrini & H~1972 | 0.718 | -0.183 | 1.620 | 0.118 | 1972 |
| Maxwell, 1970 | 0.755 | 0.320 | 1.190 | 0.001 | 1970 |
| Kester, 1969 | 0.581 | 0.174 | 0.989 | 0.005 | 1969 |
| Group: > 1 week | | | | | |
| Rosenthal et.., 1~4 | 0.030 | -0.215 | 0.275 | 0.810 | 1974 |
| Fielder et al., 1~1 | 0.000 | -0.156 | 0.156 | 0.998 | 1971 |
| Jose & Cody, 1971 | -0.026 | -0.166 | 0.115 | 0.720 | 1971 |
| Claiborn, 1969 | -0.054 | -0.188 | 0.080 | 0.429 | 1969 |
| Evans & Rosent~1969 | -0.056 | -0.167 | 0.056 | 0.326 | 1969 |
| Conn et al., 1968 | -0.033 | -0.137 | 0.071 | 0.535 | 1968 |

CMA is performed separately for each group of `week1`.

Also see *Cumulative meta-analysis* in [META] **meta**.

◁

## ▷ Example 13: Leave-one-out meta-analysis

For each study in the meta-analysis, the corresponding leave-one-out meta-analysis will omit that study and perform a meta-analysis on the remaining set of studies ($k - 1$ studies). It is useful for exploring the influence of a single study on the overall effect size estimate.

Continuing with example 1, we will use option `leaveoneout` to perform a leave-one-out meta-analysis and sort our results according to variable `se` so that larger studies appear first.

```
. meta summarize, leaveoneout sort(se)
Leave-one-out meta-analysis summary          Number of studies =     10
Random-effects model
Method: REML
```

theta: Overall Std. mean diff.

| Omitted study | theta | [95% conf. | interval] | p-value |
|---|---|---|---|---|
| Evans & Rosenthal, 1969 | 0.172 | -0.073 | 0.418 | 0.169 |
| Fielder et al., 1971 | 0.168 | -0.081 | 0.418 | 0.186 |
| Rosenthal et al., 1974 | 0.161 | -0.090 | 0.413 | 0.208 |
| Conn et al., 1968 | 0.149 | -0.102 | 0.400 | 0.244 |
| Kester, 1969 | 0.127 | -0.115 | 0.368 | 0.304 |
| Jose & Cody, 1971 | 0.174 | -0.060 | 0.408 | 0.146 |
| Claiborn, 1969 | 0.175 | -0.036 | 0.386 | 0.105 |
| Maxwell, 1970 | 0.021 | -0.076 | 0.119 | 0.665 |
| Pellegrini & Hicks, 1972 | 0.132 | -0.095 | 0.358 | 0.254 |
| Pellegrini & Hicks, 1972 | 0.057 | -0.090 | 0.204 | 0.446 |
| theta | 0.134 | -0.075 | 0.342 | 0.208 |

Sorted by: se

The output table reports the overall effect size and its CIs for each leave-one-out analysis. In this example, the first row reports the overall effect size estimate based on all the studies excluding the `Evans & Rosenthal, 1969` study ($10 - 1 = 9$ studies). The `p-value` column contains the $p$-values of the significance tests of the overall effect sizes from these analyses. The last row displays the results based on all 10 studies. It seems that the `Maxwell, 1970` study has a relatively large influence because the 95% CI from the meta-analysis excluding that study, $[-0.076, 0.119]$, does not contain the overall effect size estimate based on all studies, 0.134.

◁

# Stored results

meta summarize stores the following in r():

Scalars

| | |
|---|---|
| r(N) | number of observations |
| r(theta) | overall effect size |
| r(se) | standard error of overall effect size |
| r(ci_lb) | lower CI bound for overall effect size |
| r(ci_ub) | upper CI bound for overall effect size |
| r(tau2) | between-study variance |
| r(I2) | $I^2$ heterogeneity statistic (not for CE model) |
| r(H2) | $H^2$ heterogeneity statistic (not for CE model) |
| r(z) | $z$ statistic for test of significance of overall effect size (when se() not specified) |
| r(t) | $t$ statistic for test of significance of overall effect size (when se() specified) |
| r(df) | degrees of freedom for $t$ distribution |
| r(p) | $p$-value for test of significance of overall effect size |
| r(Q) | Cochran's $Q$ heterogeneity test statistic (not for CE model) |
| r(df_Q) | degrees of freedom for heterogeneity test |
| r(p_Q) | $p$-value for heterogeneity test |
| r(Q_b) | Cochran's $Q$ statistic for test of group differences (for subgroup() with one variable) |
| r(df_Q_b) | degrees of freedom for test of group differences |
| r(p_Q_b) | $p$-value for test of group differences |
| r(seadj) | standard error adjustment |
| r(level) | confidence level for CIs |
| r(pi_lb) | lower bound of prediction interval |
| r(pi_ub) | upper bound of prediction interval |
| r(pilevel) | confidence level for prediction interval |
| r(converged) | 1 if converged, 0 otherwise (with iterative random-effects methods) |

Macros

| | |
|---|---|
| r(model) | meta-analysis model |
| r(method) | meta-analysis estimation method |
| r(citype) | type of CI used in option citype() for meta-analysis of a single proportion |
| r(subgroupvars) | names of subgroup-analysis variables |
| r(ordervar) | name of order variable used in option cumulative() |
| r(byvar) | name of variable used in suboption by() within option cumulative() |
| r(direction) | ascending or descending |
| r(seadjtype) | type of standard error adjustment |

Matrices

| | |
|---|---|
| r(esgroup) | ESs and CIs from subgroup analysis |
| r(hetgroup) | heterogeneity summary from subgroup analysis |
| r(diffgroup) | results for tests of group differences from subgroup analysis |
| r(cumul) | results from cumulative meta-analysis |
| r(leaveoneout) | results from leave-one-out meta-analysis |
| r(pi_info) | prediction intervals from subgroup analysis |

meta summarize also creates a system variable, _meta_weight, which contains study weights. When the transform() option is specified, meta summarize creates system variables _meta_es_transf, _meta_cil_transf, and _meta_ciu_transf, which contain the transformed effect sizes and lower and upper bounds of the corresponding transformed CIs.

Also see *Stored results* in [META] **meta set** and *Stored results* in [META] **meta esize** for other system variables.

# Methods and formulas

Methods and formulas are presented under the following headings:



The formulas and methods below are based on Veroniki et al. (2016), Viechtbauer et al. (2015), Borenstein et al. (2009), Schwarzer, Carpenter, and Rücker (2015), Kontopantelis and Reeves (2016), Fisher (2016), and Bradburn, Deeks, and Altman (2016).

## Fixed-effects and common-effect methods for combining study estimates

Consider the data from $K$ independent studies. Let $\widehat{\theta}_j$ be the estimate of the population effect size $\theta_j$ reported by the $j$th study and $\widehat{\sigma}^2_j$ be the corresponding estimate of the within-study variance, which is equal to the squared standard error of $\widehat{\theta}_j$. $\widehat{\theta}_j$ is one of Hedges's $g_j$, Cohen's $d_j$, $\ln\left(\widehat{\text{OR}}_j\right)$, $\ln\left(\widehat{\text{RR}}_j\right)$, and so on, as defined in *Methods and formulas* of [META] **meta esize**, or a generic (precomputed) effect size as declared by [META] **meta set**.

Consider a fixed-effects model (Hedges and Vevea 1998; Rice, Higgins, and Lumley 2018) from *Meta-analysis models* in [META] **Intro**,

$$\widehat{\theta}_j = \theta_j + \epsilon_j \qquad \epsilon_j \sim N\left(0, \widehat{\sigma}^2_j\right)$$

where $\widehat{\sigma}^2_j$'s are treated as known values that do not require estimation. Under the assumption that $\theta_1 = \theta_2 = \cdots = \theta_K = \theta$, the above fixed-effects model simplifies to a common-effect model (Hedges 1982; Rosenthal and Rubin 1982):

$$\widehat{\theta}_j = \theta + \epsilon_j \qquad \epsilon_j \sim N\left(0, \widehat{\sigma}^2_j\right)$$

The estimation methods we describe below are the same for the two models, but the interpretation of the estimates is different; see *Comparison between the models and interpretation of their results* in [META] **Intro**. The two models estimate different population parameters. A common-effect model estimates the common effect $\theta_{\text{pop}} = \theta$, whereas a fixed-effects model estimates a weighted average of the study-specific effects $\widehat{\theta}_j$'s,

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

where $W_j$'s represent true, unknown weights, which are defined in Rice, Higgins, and Lumley (2018, eq. 3). For simplicity, in what follows, we will use $\theta$ to mean $\theta_{\text{pop}}$.

### Inverse-variance method

Under the inverse-variance method, the MLE of $\theta$ is

$$\widehat{\theta}_{\text{IV}} = \frac{\sum_{j=1}^{K} \widehat{\theta}_j/\widehat{\sigma}_j^2}{\sum_{j=1}^{K} 1/\widehat{\sigma}_j^2} = \frac{\sum_{j=1}^{K} w_j\widehat{\theta}_j}{\sum_{j=1}^{K} w_j}$$

where the weight $w_j = 1/\widehat{\sigma}_j^2$ is used to estimate the true weight $W_j$. The inverse-variance method takes its name from the weights being the reciprocal of the effect-size variances.

The variance estimate of $\widehat{\theta}_{\text{IV}}$

$$\widehat{\text{Var}}\left(\widehat{\theta}_{\text{IV}}\right) = \frac{1}{w_{\cdot}}$$

where $w_{\cdot} = \sum_{j=1}^{K} w_j$.

### Mantel–Haenszel method for two-group comparison of binary outcomes

For meta-analysis that compares two binary outcomes, the Mantel–Haenszel method can be used to combine odds ratios (OR), risk ratios (RR), and risk differences (RD) instead of the inverse-variance method. The classical Mantel–Haenszel method (Mantel and Haenszel 1959) is used for OR, and its extension by Greenland and Robins (1985) is used for RR and RD. The Mantel–Haenszel method may be preferable with sparse data (Emerson 1994). This is the default pooling method in `meta esize` for the effect sizes mentioned above with fixed-effects and common-effect models.

Consider the following $2 \times 2$ table for the $j$th study.

| group | event | no event | size |
|---|---|---|---|
| treatment | $a_j$ | $b_j$ | $n_{1j} = a_j + b_j$ |
| control | $c_j$ | $d_j$ | $n_{2j} = c_j + d_j$ |

The sample size for the $j$th study is denoted by $n_j = n_{1j} + n_{2j}$.

For the overall risk difference, the formula is

$$\widehat{\theta}_{\text{MH}} = \frac{\sum_{j=1}^{K} w_j^{(\text{MH})} \times \widehat{\theta}_j}{\sum_{j=1}^{K} w_j^{(\text{MH})}}$$

where $\widehat{\theta}_j$ is $\widehat{\text{RD}}$ from the $j$th study.

Unlike the inverse-variance method, with log odds-ratios and log risk-ratios, the Mantel–Haenszel method combines the individual effect sizes in the original metric and then takes the log to obtain the final overall log odds-ratio or log risk-ratio estimate,

$$\widehat{\theta}_{\text{MH}} = \ln\left\{\frac{\sum_{j=1}^{K} w_j^{(\text{MH})} \times \exp(\widehat{\theta}_j)}{\sum_{j=1}^{K} w_j^{(\text{MH})}}\right\}$$

where $\widehat{\theta}_j$ is $\ln(\widehat{\text{OR}})$ or $\ln\left(\widehat{\text{RR}}\right)$ from the $j$th study.

The MH weights are defined as follows. In the formula for the overall risk difference, the weight assigned to each study is

$$w_j^{(\mathrm{MH})} = \frac{n_{1j}n_{2j}}{n_j}$$

For the overall log risk-ratio, the $j$th weight is given by

$$w_j^{(\mathrm{MH})} = \frac{n_{1j}c_j}{n_j}$$

And for the overall log odds-ratio, the $j$th weight is given by

$$w_j^{(\mathrm{MH})} = \frac{b_j c_j}{n_j}$$

An estimator of the variance of the overall risk difference $\widehat{\theta}_{\mathrm{MH}} = \widehat{\mathrm{RD}}_{\mathrm{MH}}$ (Greenland and Robins 1985) is

$$\widehat{\mathrm{Var}}\left(\widehat{\mathrm{RD}}_{\mathrm{MH}}\right) = \frac{\sum_{j=1}^{K}\left(a_j b_j n_{2j}^3 + c_j d_j n_{1j}^3\right)/n_{1j}n_{2j}n_j^2}{\left(\sum_{j=1}^{K} n_{1j}n_{2j}/n_j\right)^2}$$

An estimator of the variance of the overall log risk-ratio $\widehat{\theta}_{\mathrm{MH}} = \ln(\widehat{\mathrm{RR}}_{\mathrm{MH}})$ (Greenland and Robins 1985) is

$$\widehat{\mathrm{Var}}\left\{\ln(\widehat{\mathrm{RR}}_{\mathrm{MH}})\right\} = \frac{\sum_{j=1}^{K}\left\{n_{1j}n_{2j}\left(a_j + c_j\right) - a_j c_j n_j\right\}/n_j^2}{\left(\sum_{j=1}^{K} a_j n_{2j}/n_j\right) \times \left(\sum_{j=1}^{K} c_j n_{1j}/n_j\right)}$$

And an estimator of the variance of the overall log odds-ratio $\widehat{\theta}_{\mathrm{MH}} = \ln(\widehat{\mathrm{OR}}_{\mathrm{MH}})$ (Robins, Breslow, and Greenland 1986a; Robins, Greenland, and Breslow 1986b) is

$$\widehat{\mathrm{Var}}\left\{\ln(\widehat{\mathrm{OR}}_{\mathrm{MH}})\right\} = \frac{\sum_{j=1}^{K} P_j R_j}{2\left(\sum_{j=1}^{K} R_j\right)^2} + \frac{\sum_{j=1}^{K}\left(P_j S_j + Q_j R_j\right)}{2\sum_{j=1}^{K} R_j \sum_{j=1}^{K} S_j} + \frac{\sum_{j=1}^{K} Q_j S_j}{2\left(\sum_{j=1}^{K} S_j\right)^2}$$

where

$$P_j = \frac{a_j + d_j}{n_j}, \quad Q_j = \frac{b_j + c_j}{n_j}, \quad R_j = \frac{a_j d_j}{n_j}, \quad \text{and } S_j = \frac{b_j c_j}{n_j}$$

Greenland and Robins (1985) and Robins, Breslow, and Greenland (1986a) demonstrate consistency of all the above variance estimators in the two cases they call a sparse-data limiting model, in which the number of $2 \times 2$ tables (studies) increases but the cell sizes remain fixed, and a large-strata limiting model, in which the number of studies remains fixed but individual cell sizes increase.

### Peto's method for odds ratios

An alternative to the Mantel–Haenszel method for combining odds ratios is the Peto's method (Peto et al. 1977; Yusuf et al. 1985). It is based on the inverse-variance method but uses an alternate way to compute the odds ratios (and consequently the log odds-ratio).

Let $\ln\left(\widehat{\text{OR}}_j^{\text{Peto}}\right)$ be Peto's log odds-ratio for the $j$th study as defined in *Odds ratio* in [META] **meta esize**. Then, Peto's overall log odds-ratio is defined following the inverse-variance method as follows,

$$\widehat{\theta}_{\text{Peto}} = \ln\left(\widehat{\text{OR}}^{\text{Peto}}\right) = \frac{\sum_{j=1}^{K} w_j \ln\left(\widehat{\text{OR}}_j^{\text{Peto}}\right)}{\sum_{j=1}^{K} w_j}$$

where $w_j = 1/\widehat{\sigma}_j^2 = \text{Var}(a_j)$ and $\text{Var}(a_j)$ is as defined in *Methods and formulas* of [META] **meta esize** of [META] **meta esize**.

The variance estimate is

$$\widehat{\text{Var}}\left\{\ln\left(\widehat{\text{OR}}^{\text{Peto}}\right)\right\} = \frac{1}{\sum_{j=1}^{K} w_j}$$

## Random-effects methods for combining study estimates

Suppose that the observed study-specific effect sizes represent a random sample from a population of effect sizes that is normally distributed with mean $\theta$ and variance $\tau^2$.

Consider a random-effects model (Hedges 1983; DerSimonian and Laird 1986) from *Meta-analysis models* in [META] **Intro**,

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

where $\epsilon_j$ and $u_j$ are assumed to be independent with $\epsilon_j \sim N\left(0, \widehat{\sigma}_j^2\right)$ and $u_j \sim N\left(0, \tau^2\right)$.

The overall effect $E(\widehat{\theta}_j) = \theta$ is estimated as the weighted average,

$$\widehat{\theta}^* = \frac{\sum_{j=1}^{K} w_j^* \widehat{\theta}_j}{\sum_{j=1}^{K} w_j^*} \tag{1}$$

where $w_j^* = 1/\left(\widehat{\sigma}_j^2 + \widehat{\tau}^2\right)$. The variance of $\widehat{\theta}^*$ is estimated by

$$\widehat{\text{Var}}\left(\widehat{\theta}^*\right) = \frac{1}{w_\cdot^*}$$

where $w_\cdot^* = \sum_{j=1}^{K} w_j^*$.

Different estimators of the between-study variance, $\tau^2$, lead to different estimators of $\theta$. `meta summarize` supports seven estimation methods of $\tau^2$. Three methods are iterative: the maximum likelihood (ML) estimator (Hardy and Thompson 1996); the restricted maximum-likelihood (REML) estimator (Raudenbush 2009); and the empirical Bayes (EB) estimator (Morris 1983; Berkey et al. 1995), also known as the Paule–Mandel estimator (Paule and Mandel 1982). Four methods are noniterative (have a closed-form expression): DerSimonian–Laird (DL) estimator (DerSimonian and Laird 1986); Hedges estimator (HE) (Hedges 1983; Hedges and Olkin 1985), also known as the Cochran estimator or variance-component estimator; Hunter–Schmidt (HS) estimator (Schmidt and Hunter 2015); and Sidik–Jonkman (SJ) estimator (Sidik and Jonkman 2005).

The formulas for and properties of these estimators have been discussed at length in Veroniki et al. (2016). Expressions for these estimators are given in the more general context of meta-regression in *Methods and formulas* of [META] **meta regress**. Below, we provide the simplified expressions when no covariates (moderators) are included in the regression model. The simplified expressions were obtained by replacing the $\mathbf{X}$ matrix with $K \times 1$ column vector of 1s.

### Iterative methods

The ML method (Hardy and Thompson 1996; Thompson and Sharp 1999) computes the MLE of $\tau^2$ by maximizing the following log-likelihood function,

$$\ln L_{\mathrm{ML}}\left(\tau^2\right) = -\frac{K}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{K}\ln\left(\widehat{\sigma}_j^2+\tau^2\right) - \frac{1}{2}\sum_{j=1}^{K}\frac{\left(\widehat{\theta}_j-\widehat{\theta}^*\right)^2}{\widehat{\sigma}_j^2+\tau^2}$$

with respect to $\tau^2$, where $\widehat{\theta}^*$ is defined in (1) and is based on the current value of $\widehat{\tau}^2$.

The ML method is asymptotically efficient but may produce biased results in small samples. The REML method estimates $\tau^2$ by accounting for the uncertainty in the estimation of $\theta$, which leads to nearly an unbiased estimate of $\tau^2$.

The REML log-likelihood function is

$$\ln L_{\mathrm{REML}}\left(\tau^2\right) = \ln L_{\mathrm{ML}}\left(\tau^2\right) - \frac{1}{2}\ln\left\{\sum_{j=1}^{K}\left(\widehat{\sigma}_j^2+\tau^2\right)^{-1}\right\} + \frac{\ln(2\pi)}{2}$$

The EB estimator and a description of the iterative process for each estimator in this section is presented in the *Methods and formulas* of [META] **meta regress**.

### Noniterative methods

The methods in this section do not make any assumptions about the distribution of the random effects. They also do not require any iteration.

The most popular noniterative estimation method is the DL method. This is a method of moment estimator for $\tau^2$, and it is defined as follows,

$$\widehat{\tau}^2_{\mathrm{DL}} = \frac{Q-(K-1)}{\sum_{j=1}^{K}w_j - \sum_{j=1}^{K}w_j^2/\sum_{j=1}^{K}w_j}$$

where $Q = \sum_{j=1}^{K} w_j\left(\widehat{\theta}_j - \widehat{\theta}_{\mathrm{IV}}\right)^2$ and $w_j = 1/\widehat{\sigma}_j^2$.

Because $\widehat{\tau}^2_{\mathrm{DL}}$ is negative when $Q < K-1$, it is truncated at 0 in practice, and thus $\max\left(0, \widehat{\tau}^2_{\mathrm{DL}}\right)$ is used to estimate the between-study variance:

$$\widehat{\tau}^2_{\mathrm{DL}} = \max\left\{0, \frac{\sum_{j=1}^{K}w_j\left(\widehat{\theta}_j-\widehat{\theta}_{\mathrm{IV}}\right)^2-(K-1)}{\sum_{j=1}^{K}w_j - \sum_{j=1}^{K}w_j^2/\sum_{j=1}^{K}w_j}\right\}$$

The HE estimator is another method of moment estimator defined as follows,

$$\widehat{\tau}^2_{\mathrm{HE}} = \max\left\{0, \frac{1}{K-1}\sum_{j=1}^{K}\left(\widehat{\theta}_j-\overline{\theta}\right)^2 - \frac{1}{K}\sum_{j=1}^{K}\widehat{\sigma}_j^2\right\}$$

where $\overline{\theta} = \left(\sum_{j=1}^{K}\widehat{\theta}_j\right)/K$.

The HS estimator is given by

$$\widehat{\tau}^2_{\text{HS}} = \max\left\{0, \frac{Q-K}{\sum_{j=1}^{K} w_j}\right\}$$

For the SJ estimator, consider an initial estimate of $\tau^2$, given by

$$\widehat{\tau}^2_0 = \frac{\sum_{j=1}^{K}\left(\widehat{\theta}_j - \overline{\theta}\right)^2}{K}$$

Then, the estimator is defined as

$$\widehat{\tau}^2_{\text{SJ}} = \frac{\sum_{j=1}^{K} w_j^{\text{SJ}}\left(\widehat{\theta}_j - \widehat{\theta}^{\text{SJ}}\right)^2}{K-1}$$

where $w_j^{\text{SJ}} = \widehat{\tau}^2_0 / \left(\widehat{\sigma}^2_j + \widehat{\tau}^2_0\right)$ and $\widehat{\theta}^{\text{SJ}} = \sum_{j=1}^{K} w_j^{\text{SJ}}\widehat{\theta}_j / \sum_{j=1}^{K} w_j^{\text{SJ}}$.

### Knapp–Hartung standard-error adjustment

Hartung and Knapp (2001a) and Sidik and Jonkman (2002) proposed an adjustment to the variance of $\widehat{\theta}^*$ to account for the uncertainty in estimating $\tau^2$, which is used in the expression for weights. They proposed to multiply $\widehat{\text{Var}}(\widehat{\theta}^*) = 1/w_\cdot^*$ by the following quadratic form,

$$q_{\text{KH}} = \frac{1}{K-1}\sum_{j=1}^{K} w_j^*\left(\widehat{\theta}_j - \widehat{\theta}^*\right)^2$$

or by $\max(1, q_{\text{KH}})$.

The variance estimator for $\widehat{\theta}^*$ can then be defined as

$$\widehat{\text{Var}}_{\text{HK}}\left(\widehat{\theta}^*\right) = \begin{cases} q_{\text{KH}} \times 1/w_\cdot^* & \text{with option } \texttt{se(khartung)} \\ \max(1, q_{\text{KH}}) \times 1/w_\cdot^* & \text{with option } \texttt{se(khartung, truncated)} \end{cases}$$

Hartung (1999) established that the statistic

$$\frac{\widehat{\theta}^* - \theta}{\sqrt{\widehat{\text{Var}}_{\text{HK}}\left(\widehat{\theta}^*\right)}}$$

has a Student's $t$ distribution with $K-1$ degrees of freedom.

Correspondingly, the $(1-\alpha)\times 100\%$ CI for $\theta$ using the Knapp–Hartung standard error is

$$\widehat{\theta}^{*} \pm t_{K-1,1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}_{\mathrm{HK}}\left(\widehat{\theta}^{*}\right)}$$

where $t_{K-1,1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the Student's $t$ distribution with $K-1$ degrees of freedom.

The test statistic for the significance test of an overall effect, $H_0$: $\theta = 0$, is

$$\frac{\widehat{\theta}^{*}}{\sqrt{\widehat{\mathrm{Var}}_{\mathrm{HK}}\left(\widehat{\theta}^{*}\right)}}$$

and has the Student's $t$ distribution with $K-1$ degrees of freedom.

Also see Sidik and Jonkman (2002, 2003) and Cornell et al. (2014) for more discussion about the Knapp–Hartung adjustment.

### Prediction intervals

In a random-effects model, you can compute a prediction interval (Higgins, Thompson, and Spiegelhalter 2009) that estimates plausible ranges for $\theta$ in a future study. Compared with the CI, a prediction interval incorporates the uncertainty in estimating $\tau^2$ in the computation.

A $(1-\alpha)\times 100\%$ prediction interval is defined as

$$\widehat{\theta}^{*} \pm t_{K-2,1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\theta}^{*}\right) + \widehat{\tau}^{2}}$$

where $t_{K-1,1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the Student's $t$ distribution with $K-2$ degrees of freedom. This prediction interval may be specified with the `predinterval()` option.

## Confidence intervals and significance test

Let $\widehat{\theta}$ be any of the estimators considered in the previous sections such as $\widehat{\theta}_{\mathrm{IV}}$ or $\widehat{\theta}^{*}$. The $(1-\alpha)\times 100\%$ confidence interval for $\theta$ is

$$\widehat{\theta} \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\theta}\right)}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution.

We reject the hypothesis of no treatment effect $H_0$: $\theta = 0$ at level $\alpha$, if

$$\frac{\left|\widehat{\theta}\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\theta}\right)}} > z_{1-\alpha/2}$$

If the `tdistribution` option is specified, the $z_{1-\alpha/2}$ critical value is replaced with the $t_{K-1,1-\alpha/2}$ critical value in the above formulas.

## Heterogeneity measures

The homogeneity test can be used to test whether the study-specific effects are the same; see *Homogeneity test*. But with a small number of studies, this test may have low power (Hedges and Pigott 2001). Also, it does not provide an estimate of the magnitude of the between-study heterogeneity. Some authors (for example, Higgins and Thompson [2002] and Higgins et al. [2003]) suggest examining the heterogeneity statistics rather than relying solely on the homogeneity test.

Higgins and Thompson (2002) proposed two heterogeneity measures: $I^2$ and $H^2$. We define them separately for random-effects and fixed-effects models.

For a random-effects model, the two heterogeneity measures are defined as follows:

$$I^2 = \frac{\widehat{\tau}^2}{\widehat{\tau}^2 + s^2} \times 100\% \tag{2}$$

and

$$H^2 = \frac{\widehat{\tau}^2 + s^2}{s^2} \tag{3}$$

where

$$s^2 = \frac{K-1}{\sum_{j=1}^{K} w_j - \sum_{j=1}^{K} w_j^2 / \sum_{j=1}^{K} w_j}$$

is the within-study variance and $\widehat{\tau}^2$ is an estimator of the between-study variance. The values of $I^2$ and $H^2$ will vary depending on which estimator of $\widehat{\tau}^2$ is specified in the `random()` option.

For a fixed-effects model, the expressions for $I^2$ and $H^2$ are given by

$$I^2 = \left\{ \frac{Q - (K-1)}{Q} \right\} \times 100\%$$

and

$$H^2 = \frac{Q}{K-1}$$

where $Q$ is defined in *Homogeneity test*.

The formulas above for $I^2$ and $H^2$ are equivalent to the corresponding formulas (2) and (3), when the DL method is used to estimate $\tau^2$. $I^2$ is negative when $Q < (K-1)$ and is thus reset to zero in that case.

## Inverse Freeman–Tukey transformation

For each study, let $\widehat{p}_{\mathrm{FT}}$ be the Freeman–Tukey-transformed proportion as defined in *Freeman–Tukey-transformed proportion* in *Methods and formulas* in [META] **meta esize**.

The inverse Freeman–Tukey transformation, which back-transforms $\widehat{p}_{\mathrm{FT}}$ to a proportion (option `transform(invftukey)`), is given by (Miller 1978)

$$\widehat{p} = 0.5 \left\{ 1 - \operatorname{sgn}\left(\cos \widehat{p}_{\mathrm{FT}}\right) \sqrt{1 - \left( \sin \widehat{p}_{\mathrm{FT}} + \frac{\sin \widehat{p}_{\mathrm{FT}} - \frac{1}{\sin \widehat{p}_{\mathrm{FT}}}}{n} \right)^2} \right\}$$

where sgn is the sign operator. The expression depends on the study sample size $n$, which is available for each study but not for the overall (pooled) effect size. To back-transform the overall effect size $\widehat{\theta}$, where $\widehat{\theta}$ is obtained by pooling the study-specific $\widehat{p}_{\mathrm{FT}}$'s, to obtain the overall proportion, Miller (1978) suggested to use $n_\theta$, the harmonic mean (default) of the study-specific sample sizes, in place of $n$ in the above formula. Other estimators for $n_\theta$ include the geometric mean, arithmetic mean, or the inverse of the variance of the overall effect size.

Because $0 \leq e \leq n$, each study's $\widehat{p}_{\mathrm{FT}}$ must be between $\texttt{asin}\{\sqrt{1/(n+1)}\}$ and $\texttt{asin}\{\sqrt{n/(n+1)}\} + \pi/2$ [see (1) in [META] **meta esize**]. Thus, the above back-transformation is valid only if $\texttt{asin}\{\sqrt{1(n_\theta+1)}\} \leq \widehat{\theta} \leq \texttt{asin}\{\sqrt{n_\theta/(n_\theta+1)}\} + (\pi/2)$. Therefore, in practice, the overall proportion, $\widehat{p}_{\mathrm{ov}}$, is computed as follows:

$$\widehat{p}_{\mathrm{ov}} = \begin{cases} 0 & \text{if } \widehat{\theta} < \texttt{asin}\left(\sqrt{\frac{1}{n_\theta+1}}\right) \\ 1 & \text{if } \widehat{\theta} > \texttt{asin}\left(\sqrt{\frac{n_\theta}{n_\theta+1}}\right) + \frac{\pi}{2} \\ 0.5\left\{1 - \mathrm{sgn}\left(\cos\widehat{\theta}\right)\sqrt{1 - \left(\sin\widehat{\theta} + \frac{\sin\widehat{\theta} - \frac{1}{\sin\widehat{\theta}}}{n_\theta}\right)^2}\right\} & \text{otherwise} \end{cases} \tag{4}$$

Because $\widehat{\theta}$ can be bounded away from 0 whenever $\widehat{\theta} > \texttt{asin}\{\sqrt{1/(n_\theta+1)}\}$, the test statistic for $H_0$: $\theta = 0$ is adjusted as follows:

$$\frac{\left|\widehat{\theta} - \texttt{asin}\left(\sqrt{\frac{1}{n_\theta+1}}\right)\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\theta}\right)}}$$

## Homogeneity test

Consider a test of $H_0$: $\theta_1 = \theta_2 = \cdots = \theta_K = \theta$, known as the homogeneity test, that evaluates whether the effect sizes are the same across the studies. It uses the following test statistic,

$$Q = \sum_{j=1}^{K} w_j\left(\widehat{\theta}_j - \widehat{\theta}\right)^2 = \sum_{j=1}^{K} w_j\widehat{\theta}_j^2 - \frac{\left(\sum_{j=1}^{K} w_j\widehat{\theta}_j\right)^2}{\sum_{j=1}^{K} w_j}$$

where $w_j = 1/\widehat{\sigma}_j^2$, and $\widehat{\theta}_j$ and $\widehat{\theta}$ depend on the type of the effect size chosen.

Under the null hypothesis of homogeneity, $Q$ follows a $\chi^2$ distribution with $K-1$ degrees of freedom.

Hedges and Pigott (2001) showed that the test has low power when the number of studies ($K$) is small, which is typical in meta-analysis. This means that the null hypothesis of homogeneity is not rejected as often as it should be. Thus, for the homogeneity test, the meta-analysis literature (for example, Petitti [2001]; Berman and Parker [2002]; Sutton and Higgins [2008]) suggests using the significance level $\alpha = 0.1$ instead of the conventional $\alpha = 0.05$.

The homogeneity test checks for the potential presence of heterogeneity but does not estimate the magnitude of the heterogeneity. Thus, many authors (for example, Higgins and Thompson [2002]; Higgins et al. [2003]) suggest exploring the heterogeneity statistics rather than solely relying on the test. See *Heterogeneity measures*.

## Subgroup meta-analysis

When the `subgroup(`*varname*`)` option is specified, we assume that the $K$ studies are partitioned into $L$ subgroups defined by *varname*. Estimates of the overall effect size and their corresponding standard errors are calculated for each of the $L$ subgroups.

Let $\widehat{\theta}_{jl}$ be the effect-size estimate from study $j$ within subgroup $l$ and $\widehat{\sigma}^2_{jl}$ be the corresponding variance, where $l = 1, 2, \ldots, L$ and $j = 1, 2, \ldots, K_l$.

Below, we describe the formulas separately for fixed-effects and random-effects models. The formulas for the common-effect model are the same as for the fixed-effects model. When you specify a common-effect model with subgroup analysis, this model is assumed within each subgroup $l = 1, 2, \ldots, L$, but not for the entire sample of studies.

### Fixed-effects model

In what follows, we assume the inverse-variance method, but the same principles apply to the Mantel–Haenszel method.

In subgroup analysis, a fixed-effects model may be formulated as

$$\widehat{\theta}_{jl} = \theta_{jl} + \epsilon_{jl}, \quad \epsilon_{jl} \sim N\left(0, \widehat{\sigma}^2_{jl}\right)$$

For the $l$th group, $\widehat{\theta}_{\mathrm{IV},l}$ is a weighted average of the effect sizes $\widehat{\theta}_{jl}$ with weights $w_{jl} = 1/\widehat{\sigma}^2_{jl}$:

$$\widehat{\theta}_{\mathrm{IV},l} = \frac{\sum_{j=1}^{K_l} w_{jl}\widehat{\theta}_{jl}}{\sum_{j=1}^{K_l} w_{jl}}$$

The variance estimate of $\widehat{\theta}_{\mathrm{IV},l}$ is

$$\widehat{\mathrm{Var}}\left(\widehat{\theta}_{\mathrm{IV},l}\right) = \frac{1}{w_{.l}}$$

where $w_{.l} = \sum_{j=1}^{K_l} w_{jl}$.

Other meta-analytic quantities such as $I^2_l$ and $Q_l$ may also be computed for the $l$th subgroup just as we described in the previous sections.

The Cochran's $Q$ statistic can be extended to test for differences between the $L$ subgroups:

$$Q_b = \sum_{l=1}^{L} w_{.l} \left( \widehat{\theta}_{\mathrm{IV},l} - \frac{\sum_{l=1}^{L} w_{.l} \widehat{\theta}_{\mathrm{IV},l}}{\sum_{l=1}^{L} w_{.l}} \right)^2$$

The subscript $b$ in $Q_b$ stands for "between" to emphasize that $Q_b$ tests for "between-group" differences.

Under the null hypothesis of homogeneity between the subgroups ($\theta_{.1} = \theta_{.2} = \cdots = \theta_{.L} = \theta$), the statistic $Q_b$ has a $\chi^2$ distribution with $L - 1$ degrees of freedom.

### Random-effects model

Consider a random-effects model with $L$ subgroups and separate between-study variances $\tau_l^2$:

$$\widehat{\theta}_{jl} = \theta_{.l} + u_{jl} + \epsilon_{jl} \qquad \epsilon_{jl} \sim N\left(0, \widehat{\sigma}_{jl}^2\right) \qquad u_{jl} \sim N\left(0, \tau_l^2\right)$$

The formulas for the random-effects model are the same as for the above fixed-effects model, except we replace the weights with the random-effects weights.

The estimate, $\widehat{\theta}_l^*$, and its variance in the $l$th group are

$$\widehat{\theta}_l^* = \frac{\sum_{j=1}^{K_l} w_{jl}^* \widehat{\theta}_{jl}}{\sum_{j=1}^{K_l} w_{jl}^*}$$

$$\widehat{\mathrm{Var}}\left(\widehat{\theta}_l^*\right) = \frac{1}{w_{.l}^*}$$

where $w_{jl}^* = 1/(\widehat{\sigma}_{jl}^2 + \widehat{\tau}_l^2)$ and $w_{.l}^* = \sum_{j=1}^{K_l} w_{jl}^*$.

The Cochran's statistic for testing differences between the $L$ subgroups is defined as

$$Q_b^* = \sum_{l=1}^{L} w_{.l}^* \left( \widehat{\theta}_l^* - \frac{\sum_{l=1}^{L} w_{.l}^* \widehat{\theta}_l^*}{\sum_{l=1}^{L} w_{.l}^*} \right)^2$$

Under the null hypothesis of homogeneity between the subgroups ($\theta_{.1} = \theta_{.2} = \cdots = \theta_{.L} = \theta$), $Q_b^*$ has a $\chi^2$ distribution with $L - 1$ degrees of freedom.

Also see Borenstein et al. (2009, chap. 19) and Schwarzer, Carpenter, and Rücker (2015).

## Cumulative meta-analysis

To perform CMA, we first sort the studies in ascending order according to the values of the variable specified in the `cumulative()` option. If suboption `descending` is specified within the `cumulative()` option, the order is reversed. Mathematically, this corresponds to sorting the pairs $\left(\widehat{\theta}_j, \widehat{\sigma}_j^2\right)$ in the specified order. Let $\left(\widehat{\theta}_j^s, \widehat{\sigma}_j^{2,s}\right)$ denote the sorted pairs.

CMA estimates $K$ overall effect sizes $\widehat{\theta}_j^c$'s as follows,

$$
\begin{aligned}
\widehat{\theta}_1^c &= \widehat{\theta}_1^s \\
\widehat{\theta}_2^c &= \text{MA}\left(\widehat{\theta}_1^s, \widehat{\theta}_2^s\right) \\
\widehat{\theta}_3^c &= \text{MA}\left(\widehat{\theta}_1^s, \widehat{\theta}_2^s, \widehat{\theta}_3^s\right) \\
&\vdots \\
\widehat{\theta}_j^c &= \text{MA}\left(\widehat{\theta}_1^s, \widehat{\theta}_2^s, \widehat{\theta}_3^s, \ldots, \widehat{\theta}_j^s\right) \\
&\vdots \\
\widehat{\theta}_K^c &= \text{MA}\left(\widehat{\theta}_1^s, \widehat{\theta}_2^s, \widehat{\theta}_3^s, \ldots, \widehat{\theta}_K^s\right)
\end{aligned}
$$

where $\text{MA}\left(\widehat{\theta}_1^s, \widehat{\theta}_2^s, \widehat{\theta}_3^s, \ldots, \widehat{\theta}_j^s\right)$ denotes a meta-analysis applied to the sorted studies 1 through $j$. Note that the meta-analysis also depends on the values $\widehat{\sigma}_j^{2,s}$ but we omitted them from MA() for notational convenience.

If suboption `by(`*byvar*`)` is specified within the `cumulative()` option, the above procedure is repeated for each subgroup defined by variable *byvar*.

## Leave-one-out meta-analysis

Leave-one-out meta-analysis estimates $K$ overall effect sizes $\widehat{\theta}_{-j}$'s as follows,

$$
\begin{aligned}
\widehat{\theta}_{-1} &= \text{MA}\left(\widehat{\theta}_2, \widehat{\theta}_3, \ldots, \widehat{\theta}_K\right) \\
\widehat{\theta}_{-2} &= \text{MA}\left(\widehat{\theta}_1, \widehat{\theta}_3, \ldots, \widehat{\theta}_K\right) \\
&\vdots \\
\widehat{\theta}_{-j} &= \text{MA}\left(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_{j-1}, \widehat{\theta}_{j+1}, \ldots, \widehat{\theta}_K\right) \\
&\vdots \\
\widehat{\theta}_{-K} &= \text{MA}\left(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3, \ldots, \widehat{\theta}_{K-1}\right)
\end{aligned}
$$

where $\text{MA}\left(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_{j-1}, \widehat{\theta}_{j+1}, \ldots, \widehat{\theta}_K\right)$ denotes a meta-analysis applied to all the studies except the $j$th study. Note that the meta-analysis also depends on the values $\widehat{\sigma}_j^2$, but we omitted them from MA() for notational convenience.

# References


Barendregt, J. J., S. A. Doi, Y. Y. Lee, R. E. Norman, and T. Vos. 2013. Meta-analysis of prevalence. *Journal of Epidemiology and Community Health* 67: 974–978. https://doi.org/10.1136/jech-2013-203104.

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.

Berman, N. G., and R. A. Parker. 2002. Meta-analysis: Neither quick nor easy. *BMC Medical Research Methodology* 2: 10. https://doi.org/10.1186/1471-2288-2-10.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester, UK: Wiley.

Bradburn, M. J., J. J. Deeks, and D. G. Altman. 2016. metan—A command for meta-analysis in Stata. In *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*, ed. T. M. Palmer and J. A. C. Sterne, 2nd ed., 4–28. College Station, TX: Stata Press.

Cornell, J. E., C. D. Mulrow, A. R. Localio, C. B. Stack, A. R. Meibohm, E. Guallar, and S. N. Goodman. 2014. Random-effects meta-analysis of inconsistent effects: A time for change. *Annals of Internal Medicine* 160: 267–270. https://doi.org/10.7326/M13-2886.

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.

Emerson, J. D. 1994. Combining estimates of the odds ratio: The state of the art. *Statistical Methods in Medical Research* 3: 157–178. http://doi.org/10.1177/096228029400300204.

Fisher, D. J. 2016. Two-stage individual participant data meta-analysis and generalized forest plots. In *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*, ed. T. M. Palmer and J. A. C. Sterne, 2nd ed., 280–307. College Station, TX: Stata Press.

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.

Hardy, R. J., and S. G. Thompson. 1996. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 15: 619–629. https://doi.org/10.1002/(SICI)1097-0258(19960330)15:6⟨619::AID-SIM188⟩3.0.CO;2-A.

——. 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.

Hartung, J. 1999. An alternative method for meta-analysis. *Biometrical Journal* 41: 901–916. https://doi.org/10.1002/(SICI)1521-4036(199912)41:8⟨901::AID-BIMJ901⟩3.0.CO;2-W.

Hartung, J., and G. Knapp. 2001a. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine* 20: 1771–1782. https://doi.org/10.1002/sim.791.

——. 2001b. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine* 20: 3875–3889. https://doi.org/10.1002/sim.1009.

Hedges, L. V. 1982. Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 92: 490–499. http://doi.org/10.1037/0033-2909.92.2.490.

——. 1983. A random effects model for effect sizes. *Psychological Bulletin* 93: 388–395. http://doi.org/10.1037/0033-2909.93.2.388.

Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Hedges, L. V., and T. D. Pigott. 2001. The power of statistical tests in meta-analysis. *Psychological Methods* 6: 203–217. https://doi.org/10.1037/1082-989X.6.3.203.

Hedges, L. V., and J. L. Vevea. 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* 3: 486–504. http://doi.org/10.1037/1082-989X.3.4.486.

Higgins, J. P. T., and S. G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21: 1539–1558. https://doi.org/10.1002/sim.1186.

Higgins, J. P. T., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. Measuring inconsistency in meta-analyses. *BMJ* 327: 557–560. http://doi.org/10.1136/bmj.327.7414.557.

Higgins, J. P. T., S. G. Thompson, and D. J. Spiegelhalter. 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A* 172: 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x.

John, O. P., and S. Srivastava. 1999. The big five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and Research*, ed. L. A. Pervin and O. P. John, 2nd ed., 102–138. New York: Guilford.

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.

Kontopantelis, E., and D. Reeves. 2016. metaan: Random-effects meta-analysis. In *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*, ed. T. M. Palmer and J. A. C. Sterne, 2nd ed., 55–67. College Station, TX: Stata Press.

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.

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.

Miller, J. J. 1978. The inverse of the Freeman–Tukey double arcsine transformation. *American Statistician* 32: 138. https://doi.org/10.1080/00031305.1978.10479283.

Molloy, G. J., R. E. O'Carroll, and E. Ferguson. 2013. Conscientiousness and medication adherence: A meta-analysis. *Annals of Behavioral Medicine* 47: 92–101. https://doi.org/10.1007/s12160-013-9524-4.

Morris, C. N. 1983. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* 78: 47–55. https://doi.org/10.2307/2287098.

Paule, R. C., and J. Mandel. 1982. Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards* 87: 377–385. http://doi.org/10.6028/jres.087.022.

Petitti, D. B. 2001. Approaches to heterogeneity in meta-analysis. *Statistics in Medicine* 20: 3625–3633. https://doi.org/10.1002/sim.1091.

Peto, R., M. C. Pike, P. Armitage, N. E. Breslow, D. R. Cox, S. V. Howard, N. Mantel, K. McPherson, J. Peto, and P. G. Smith. 1977. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *British Journal of Cancer* 35: 1–39. https://doi.org/10.1038/bjc.1977.1.

Raudenbush, S. W. 1984. Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology* 76: 85–97. http://doi.org/10.1037/0022-0663.76.1.85.

——. 2009. Analyzing effect sizes: Random-effects models. In *The Handbook of Research Synthesis and Meta-Analysis*, ed. H. Cooper, L. V. Hedges, and J. C. Valentine, 2nd ed., 295–316. New York: Russell Sage Foundation.

Raudenbush, S. W., and A. S. Bryk. 1985. Empirical Bayes meta-analysis. *Journal of Educational Statistics* 10: 75–98. https://doi.org/10.2307/1164836.

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, Series A* 181: 205–227. https://doi.org/10.1111/rssa.12275.

Robins, J. M., N. E. Breslow, and S. Greenland. 1986a. Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics* 42: 311–323. https://doi.org/10.2307/2531052.

Robins, J. M., S. Greenland, and N. E. Breslow. 1986b. A general estimator for the variance of the Mantel–Haenszel odds ratio. *American Journal of Epidemiology* 124: 719–723. https://doi.org/10.1093/oxfordjournals.aje.a114447.

Rosenthal, R., and D. B. Rubin. 1982. Comparing effect sizes of independent studies. *Psychological Bulletin* 92: 500–504. http://doi.org/10.1037/0033-2909.92.2.500.

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.

Schwarzer, G., J. R. Carpenter, and G. Rücker. 2015. *Meta-Analysis with R*. New York: Springer.

Sidik, K., and J. N. Jonkman. 2002. A simple confidence interval for meta-analysis. *Statistics in Medicine* 21: 3153–3159. https://doi.org/10.1002/sim.1262.

——. 2003. On constructing confidence intervals for a standardized mean difference in meta-analysis. *Communications in Statistics—Simulation and Computation* 32: 1191–1203. https://doi.org/10.1081/SAC-120023885.

——. 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.

Sterne, J. A. C. 2016. Cumulative meta-analysis. In *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*, ed. T. M. Palmer and J. A. C. Sterne, 2nd ed., 68–77. College Station, TX: Stata Press.

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., 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.

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., 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.

Yusuf, S., R. Peto, J. Lewis, R. Collins, and P. Sleight. 1985. Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases* 27: 335–371. https://doi.org/10.1016/S0033-0620(85)80003-7.

## Also see

[META] **meta data** — Declare meta-analysis data

[META] **meta forestplot** — Forest plots

[META] **meta galbraithplot** — Galbraith plots

[META] **meta regress** — Meta-analysis regression

[META] **meta** — Introduction to meta

[META] **Glossary**

[META] **Intro** — Introduction to meta-analysis