meta psycorr — Correct correlations and declare psychometric meta-analysis data+

+This command is part of StataNow.

Description

`meta psycorr` computes correlations that are corrected for attenuation because of statistical artifacts, such as measurement errors and range restriction, in the context of psychometric meta-analysis. It declares the data in memory as `meta` data, informing Stata of key variables and their roles in a meta-analysis. If correcting for artifacts is not of interest, see [META] **meta esize** for computing effect sizes for binary, continuous, and correlation data, or see [META] **meta set** for working with precomputed effect sizes.

If you need to update some of the meta settings after the data declaration, see [META] **meta update**. To display current meta settings, use `meta query`; see [META] **meta update**.

Quick start

Compute correlations, and their standard errors, that are corrected for measurement errors in random variables X and Y from variables `r` (observed correlations) and `n` (study sample size) by using reliability estimates stored in variables `rxx` and `ryy`

```
meta psycorr r n, xreliability(rxx) yreliability(ryy)
```

Same as above, but also correct for indirect (implied) range restriction in X by using observed-score u-ratios stored in the variable `ux`

```
meta psycorr r n, xreliability(rxx) yreliability(ryy) xuratios(ux)
```

Same as above, but correct for direct range restriction and assume the reliability estimates of X are from the unrestricted sample

```
meta psycorr r n, xreliability(rxx, unrestricted) yreliability(ryy)     ///
    xuratios(ux) direct
```

Same as above, but use 90% confidence level and suppress the display of meta settings for all subsequent meta-analysis commands

```
meta psycorr r n, xreliability(rxx, unrestricted) yreliability(ryy)     ///
    xuratios(ux) direct level(90) nometashow
```

Correct correlations for measurement errors and indirect range restriction in X, but specify true-score u-ratios `ut` (instead of the default observed-score u-ratios) and a mixture of restricted and unrestricted reliability estimates for Y based on the indicator variable `idx`

```
meta psycorr r n, xreliability(rxx) yreliability(ryy, restricted(idx))    ///
    xuratios(ut, true)
```

Correct correlations for dichotomization of variable X, for small-study bias, and for indirect range restriction in Y where the values of `uy` are either observed or true-score u-ratios based on the indicator variable `idx`

```
meta psycorr r n, xdich(px) small yuratios(uy, observed(idx))
```

Correct correlations for measurement errors in both random variables, and impute missing reliability estimates by using bootstrap and specify the random-number seed for reproducibility

```
meta psycorr r n, xreliability(rxx) yreliability(ryy)            ///
   impute(bootstrap, rseed(19))
```

Correct for measurement errors and indirect (implied) bivariate range restriction

```
meta psycorr r n, xreliability(rxxr) yreliability(ryyr) xuratios(ux)    ///
   yuratios(uy)
```

Correct for measurement errors and direct bivariate range restriction

```
meta psycorr r n, xreliability(rxxr) yreliability(ryyr) xuratios(ux)    ///
   yuratios(uy) direct
```

Menu

Statistics > Meta-analysis

Syntax

Compute correlations corrected for attenuation because of statistical artifacts

`meta psycorr` *r n* [*if*] [*in*] [, *options*]

Variables *r* and *n* contain observed correlations and sample sizes, respectively, from individual studies.

options	Description
Main	
`xreliability(`*relspec*`)`	specify reliability estimates for random variable *X* to correct for measurement error in *X*
`yreliability(`*relspec*`)`	specify reliability estimates for random variable *Y* to correct for measurement error in *Y*
rrspec	specify range restriction model
Model	
`xdich(`#|*varname*`)`	proportions of successes or failures after dichotomization of *X*
`ydich(`#|*varname*`)`	proportions of successes or failures after dichotomization of *Y*
`small`	apply small-sample bias-correction factor
`impute(`*imethodspec*`)`	specify imputation method for handling missing artifact values; default is `impute(bootstrap)`
Options	
`studylabel(`*varname*`)`	variable to be used to label studies in all meta-analysis output
`eslabel(`*string*`)`	effect-size label to be used in all meta-analysis output; default is `eslabel(Corrected correlation)`
`level(#)`	confidence level for all subsequent meta-analysis commands
[`no`]`metashow`	display or suppress meta settings in the output

The syntax of *relspec* is

varname [, *relopts*]

where the variable *varname* contains the reliability estimates, and options *relopts* are relevant only in the presence of range restriction.

relopts	Description
`restricted`	specify that reliability estimates are based on the restricted sample; the default in the presence of range restriction
`restricted(`*rest_idvar*`)`	specify that reliability estimates are a mixture from restricted and unrestricted samples based on the indicator variable *rest_idvar*
`unrestricted`	specify that reliability estimates are based on the unrestricted sample
`unrestricted(`*unrest_idvar*`)`	specify that reliability estimates are a mixture from unrestricted and restricted samples based on the indicator variable *unrest_idvar*

The syntax of *rrspec* is one of

`xuratios(`*varname* [, *uopts*]`)` [`indirect` | `direct`]

`yuratios(`*varname* [, *uopts*]`)` [`indirect` | `direct`]

`xuratios(`*varname* [, *uopts*]`)` `yuratios(`*varname* [, *uopts*]`)` [*rropts*]

uopts	Description
`observed`	specify that u-ratios are ratios of observed-score standard deviations; the default
`observed(`*obs_idvar*`)`	specify that u-ratios are a mixture of ratios of observed-score and true-score standard deviations based on the indicator variable *obs_idvar*
`true`	specify that u-ratios are ratios of true-score standard deviations
`true(`*true_idvar*`)`	specify that u-ratios are a mixture of ratios of true-score and observed-score standard deviations based on the indicator variable *true_idvar*

rropts	Description
`indirect`	specify that range restriction is indirect; the default
`direct`	specify that range restriction is direct
`nu(`*varname*`)`	specify the within-study sample size of the unrestricted sample with bivariate indirect range restriction
`signrxz(`*signvar*`)`	specify signs of the correlations between X and suitability variable Z in bivariate indirect range restriction
`signryz(`*signvar*`)`	specify signs of the correlations between Y and suitability variable Z in bivariate indirect range restriction

Options

Main

`xreliability(`*relspec*`)` and `yreliability(`*relspec*`)` specify the reliability estimates for X and Y, respectively, to correct for measurement error in the corresponding variable. When range restriction is also present in the meta-analysis, the specified reliability estimates are assumed to be based on the restricted sample, by default. *relspec* is *varname*[, *relopts*], where *relopts* is one of the following options: `restricted`, `restricted(`*rest_idvar*`)`, `unrestricted`, or `unrestricted(`*unrest_idvar*`)`. *varname* must contain positive values that are not greater than 1, with values closer to 1 meaning more reliable measurements.

`restricted` and `restricted(`*rest_idvar*`)` specify that reliability estimates are based on the restricted sample. This is the default. When the indicator variable *rest_idvar* is specified, the reliability estimates are based on the restricted sample when *rest_idvar* $=1$ and the unrestricted sample when *rest_idvar* $=0$.

`unrestricted` and `unrestricted(`*unrest_idvar*`)` specify that reliability estimates are based on the unrestricted sample. When the indicator variable *unrest_idvar* is specified, the reliability estimates are based on the unrestricted sample when *unrest_idvar* $=1$ or the restricted sample when *unrest_idvar* $=0$.

rrspec specifies the range restriction model in the meta-analysis. *rrspec* is one of the following:

`xuratios(`*varname*[, *uopts*]`)` [`indirect` | `direct`]

`yuratios(`*varname*[, *uopts*]`)` [`indirect` | `direct`]

`xuratios(`*varname*[, *uopts*]`)` `yuratios(`*varname*[, *uopts*]`)` [*rropts*]

`xuratios(`*varname*`)` specifies the range restriction in X, where *varname* represents the u-ratio variable that contains the ratios of the standard deviations in the restricted sample to those in the unrestricted sample.

`yuratios(`*varname*`)` specifies the range restriction in Y, where *varname* represents the u-ratio variable that contains the ratios of standard deviations in the restricted sample to those in the unrestricted sample.

When both `xuratios()` and `yuratios()` are specified, a bivariate range restriction in both X and Y is assumed. *varname* within `xuratios()` or `yuratios()` quantifies the degree of range restriction for the respective variable. *varname* must contain positive values that are not greater than 1, with values closer to 1 indicating weaker effects of range restriction.

uopts is one of the following: `observed`, `observed(`*obs_idvar*`)`, `true`, or `true(`*true_idvar*`)`.

`observed`, the default, and `observed(`*obs_idvar*`)` specify that the u-ratio values are the ratios of the observed-score standard deviations. That is, the u-ratio variable contains $S_{X,r}/S_{X,u}$ for range restriction in X and contains $S_{Y,r}/S_{Y,u}$ for range restriction in Y. When the indicator variable *obs_idvar* $=1$, the u-ratio values are the ratios of the respective observed-score standard deviations, $S_{X,r}/S_{X,u}$ or $S_{Y,r}/S_{Y,u}$; when *obs_idvar* $=0$, the u-ratio values are the ratios of the respective true-score standard deviations, $S_{T,r}/S_{T,u}$ or $S_{P,r}/S_{P,u}$.

`true` and `true(`*true_idvar*`)` specify that the u-ratio values are the ratios of the true-score standard deviations. That is, the u-ratio variable contains $S_{T,r}/S_{T,u}$ for range restriction in X and $S_{P,r}/S_{P,u}$ for range restriction in Y. When the indicator variable *obs_idvar* $=1$, the u-ratio values are the ratios of the respective true-score standard deviations, $S_{T,r}/S_{T,u}$ or $S_{P,r}/S_{P,u}$; when *true_idvar* $=$ 0, the u-ratio values are the ratios of the respective observed-score standard deviations, $S_{X,r}/S_{X,u}$ or $S_{Y,r}/S_{Y,u}$.

rropts are `indirect`, `direct`, `nu(`*varname*`)`, `signrxz(`*signvar*`)`, and `signryz(`*signvar*`)`. `indirect` and `direct` may not be combined. `nu()`, `signrxz()`, and `signryz()` are only relevant with bivariate indirect range restriction.

`indirect` specifies that the range restriction is indirect. This is the default.

`direct` specifies that the range restriction is direct.

`nu(`*varname*`)` specifies the within-study sample sizes of the unrestricted sample. This option is only relevant for bivariate indirect range restriction and is used to compute the sampling variances of the corrected correlations, $\rho^{\text{bvirr}}_{TP,u}$; see *Sampling variances of the corrected correlations* $\rho_{TP,u}$. It is required when at least one of the reliability estimates specified in option `xreliability()` or `yreliability()` is assumed to be from the unrestricted sample.

`signrxz(`*varname*`)` specifies the signs of the correlations ρ_{XZ_u}, where Z is the variable on which selection occurred. By default, positive signs are assumed. This option is only relevant for bivariate indirect range restriction. It is used to compute the corrected correlations, $\rho^{\text{bvirr}}_{TP,u}$; see *Correcting correlations for measurement errors and bivariate range restriction*.

`signryz(`*varname*`)` specifies the signs of the correlations ρ_{YZ_u}, where Z is the variable on which selection occurred. By default, positive signs are assumed. This option is only relevant for bivariate indirect range restriction. It is used to compute the corrected correlations, $\rho^{\text{bvirr}}_{TP,u}$; see *Correcting correlations for measurement errors and bivariate range restriction*.

Model

`xdich(`#|*varname*`)` and `ydich(`#|*varname*`)` specify the proportions of successes or failures after dichotomization of X and Y, respectively. # assumes that the proportions of successes or failures are the same across all studies and are equal to #. # and the values of *varname* must be between 0 and 1.

`small` corrects study correlations for small-study bias by multiplying each correlation with a correction factor of $(2n_j - 2)/(2n_j - 1)$, where n_j is the sample size of the jth study.

`impute(`*methodspec*`)` imputes a single value for missing reliability estimates and u-ratios unless `impute(none)` is used. *methodspec* is one of `bootstrap` [`, rseed(`#`)`], `perfect`, `mean`, `wmean`, and `none`.

`bootstrap` [`, rseed(`#`)`] uses bootstrap (sampling from the available observations with replacement) to replace missing artifact values. This is the default. `rseed(`#`)` sets the random-number seed. This suboption is used to reproduce results; see [R] **set seed**.

`perfect` replaces missing reliability values and u-ratios with 1. This is equivalent to assuming perfect measurement (no measurement error) and no range restriction.

`mean` replaces missing artifact values with the mean of the available observations.

`wmean` replaces missing artifact values with the sample-size weighted mean of the available observations.

`none` ignores studies with missing artifact values and performs no imputation.

Imputed values are stored in the corresponding system variables; see *System variables* in [META] **meta data**.

Options

studylabel(*varname*) specifies a string variable containing labels for the individual studies to be used in all applicable meta-analysis output. The default study labels are Study 1, Study 2, ..., Study K, where K is the total number of studies in the meta-analysis.

eslabel(*string*) specifies that *string* be used as the effect-size label in all relevant meta-analysis output. The default label is Corrected correlation.

level(#) specifies the confidence level, as a percentage, for confidence intervals. It will be used by all subsequent meta-analysis commands when computing confidence intervals. The default is level(95) or as set by set level; see [R] **level**. After the declaration, you can specify level() with meta update to update the confidence level to be used throughout the rest of the meta-analysis session. You can also specify level() directly with the meta commands to modify the confidence level, temporarily, during the execution of the command.

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

Remarks and examples

Remarks are presented under the following headings:

Overview

The goal of meta-analysis is to pool effect sizes across studies to summarize the overall evidence regarding a phenomenon of interest. A fundamental assumption in meta-analysis is that the reported effect sizes are accurate reflections of the true relationships being studied. When this assumption is violated, it can lead to the well-known “garbage in, garbage out” problem, where biased inputs produce unreliable conclusions.

In many research domains, such as education, sociology, and psychology, the variables of interest are not directly observable; instead, they represent latent constructs such as intelligence, job satisfaction, or depression. These constructs are measured by using instruments or tests that are inherently subject to measurement error. For example, consider the correlation between job satisfaction and employee performance. If the instruments used to measure either construct are unreliable, the observed correlation may underestimate the true relationship between these variables.

Additionally, other factors, such as range restriction, can further distort observed effect sizes. Range restriction occurs when the variability (or range) of scores on a variable is artificially reduced because of a selection process. For instance, if a study only includes high-performing employees, the observed relationship between job satisfaction and performance might differ from what would have been observed in a broader population because of the reduced variability or restricted range in performance.

To address these issues, psychometric meta-analysis (also known as validity generalization) corrects effect sizes and their associated standard errors for measurement error, range restriction, and other potential statistical artifacts. The effect sizes are typically correlations (also known as validities). By eliminating the distorting effects of statistical artifacts, we can obtain more accurate estimates of the mean correlation θ and the between-study variance τ^2, leading to more valid and generalizable conclusions about the phenomena under investigation.

For details about psychometric meta-analysis; see Schmidt and Hunter (2015).

Differences between psychometric and traditional meta-analyses

It is natural to ask how psychometric meta-analysis compares with the traditional meta-analysis discussed elsewhere in this manual. Borenstein et al. (2021, chap. 43) summarized the differences between psychometric meta-analysis and traditional meta-analysis in four main aspects. Below, we expand on their list:

1. **Use of raw correlations**. Psychometric meta-analysis uses raw (untransformed) correlations as input, whereas Fisher's z-transformed correlations are usually used in traditional meta-analysis.
2. **Correction for statistical artifacts**. Effect sizes (correlations) and their standard errors are corrected for statistical artifacts before pooling to estimate the overall effect size.
3. **Weighting of effect sizes**. The weights used in computing the overall effect size (the mean corrected correlation) are based on both the sample sizes and a compound attenuation factor (to be defined later), instead of the inverse-variance weights typically used in traditional meta-analysis.
4. **Heterogeneity parameter estimation**. The heterogeneity parameter τ^2 is estimated by using a noniterative procedure that differs from those typically employed in traditional meta-analysis, such as the DerSimonian–Laird method.
5. **Interpretation of heterogeneity measures**. In psychometric meta-analysis, sampling error is not the only artifact affecting the observed variability among the correlations. Consequently, measures like I^2 and H^2 have slightly different interpretations; see *I^2 and H^2 statistics* in *Methods and formulas* of [META] **meta summarize** for details.
6. **Credibility intervals instead of prediction intervals**. Credibility intervals are reported in psychometric meta-analysis to provide a plausible range for the distribution of corrected correlations; see *Credibility intervals* in *Methods and formulas* of [META] **meta summarize** for details.

Statistical artifacts

Below, we describe each of the statistical artifacts most commonly encountered in psychometric meta-analysis and explain its impact on the correlation coefficient when considered as the sole artifact in the meta-analysis. In the section of *Methods and formulas*, we will discuss how to address these artifacts when they occur simultaneously in the meta-analysis.

Sampling error and bare-bones meta-analysis

Sampling error refers to the random variations in effect sizes (for example, correlations) that arise because of differences in sample sizes (or samples in general) across studies. It is considered an unsystematic artifact because it does not follow a consistent pattern and cannot be individually corrected for in each study. However, the impact of sampling error is substantially mitigated in meta-analysis by pooling effect sizes across multiple studies. This aggregation increases the total sample size, providing a more stable and accurate estimate of the overall effect size compared to the individual effect sizes.

Schmidt and Hunter (2015, 94) referred to a meta-analysis that accounts only for sampling error while ignoring other potential artifacts, such as measurement error or range restriction, as a bare-bones meta-analysis. Although bare-bones meta-analysis addresses the distortions caused by sampling error, it does not adjust for other systematic artifacts that may bias effect-size estimates, which can limit the precision and generalizability of the findings.

Measurement error

Measurement error occurs because variables in scientific research are rarely measured with perfect accuracy. This is especially relevant in psychometric applications, where the goal is often to understand the relationship between two constructs (latent variables), such as emotional intelligence and leadership effectiveness or job satisfaction and organizational commitment. If the instruments (for example, tests or surveys) used to measure these constructs are not reliable, then any observed correlation might not reflect the true relationship between the constructs of interest.

In measurement error theory, the observed scores of random variables are modeled as follows (Lord and Novick 1968; Schmidt and Hunter 2015, chap. 3):

$$X = T + \epsilon_X$$
$$Y = P + \epsilon_Y$$

where X and Y are the observed scores and T and P are the true scores of the constructs being measured. True scores are the values that would have been observed had we been able to measure them perfectly. Here, ϵ_X and ϵ_Y are the respective errors of measurement in the random variables X and Y.

Measurement error attenuates the correlation coefficient between the true scores ρ_{TP} such that the correlation between the observed scores ρ_{XY} is less than ρ_{TP}. Under the assumption of independent errors ϵ_X and ϵ_Y and the assumption that the true scores T and P are not correlated with their respective errors ϵ_X and e_Y, we can correct for the attenuation effect of measurement error as follows:

$$\rho_{TP} = \frac{\rho_{XY}}{\sqrt{r_{XX} r_{YY}}}$$

where r_{XX} and r_{YY} are the reliability estimates of X and Y, respectively:

$$r_{XX} = \rho^2_{XT} = \text{Var}(T)/\text{Var}(X) \tag{1}$$
$$r_{YY} = \rho^2_{YP} = \text{Var}(P)/\text{Var}(Y) \tag{2}$$

In other words, reliability is the proportion of variance in observed scores attributable to true scores. Instruments with low reliability introduce greater measurement error, weakening the observed relationships between constructs. Reliability values range between 0 to 1, with values closer to 1 indicating more reliable measurements.

Like sampling error, measurement error is an inherent feature of psychometric research (and arguably of all research) and must be accounted for in any meta-analysis.

Range restriction

Range restriction is most common in educational and employment selection research. It refers to a specific type of sample selection bias in which the variability (or range) of scores on one or more variables of interest is artificially reduced because of the selection process. This typically occurs because only top-scoring individuals—those who exceed a particular cutoff on some selection criterion (for example, high test scores or strong interview performance)—are included in the study. Unlike measurement error, which is prevalent in nearly all psychometric research, range restriction may or may not occur, depending on the specific area of research. For an overview of different range restriction scenarios, see Sackett and Yang (2000).

Consider a study investigating how well high school seniors' SAT scores (X) predict their academic performance in college (Y) within the population of college applicants (unrestricted population). We can only measure academic performance for those who were actually admitted to college (restricted population). In this case, the correlation coefficient, $\rho_{XY,r}$, measures the relationship between SAT score and academic performance in the restricted population, which is denoted by the subscript r. Generally, the goal of psychometric meta-analysis is to estimate the correlation, $\rho_{TP,u}$, in the unrestricted population, where T and P are the true-score variables corresponding to X and Y, respectively, and the subscript u denotes the unrestricted population.

Range restriction can be either direct or indirect. In the above example, if admission to college was based solely on X (SAT score), for example, when applicants with a score above a certain threshold are admitted, we say that the range restriction is direct. Direct range restriction attenuates the relationship between X and Y in the restricted sample. Correcting for the attenuating effect of direct range restriction requires knowledge of the degree of range restriction in variable X, measured by $u_X = S_{X,r}/S_{X,u}$, and the observed correlation in the restricted sample, $\rho_{XY,r}$. $S_{X,r}$ and $S_{X,u}$ are the standard deviations of X in the restricted and unrestricted samples, respectively. The correction can be done by using Thorndike's Case II formula (Pearson 1903, eq. 51; Thorndike 1949):

$$\rho_{XY,u} = \frac{\rho_{XY,r}}{u_X\sqrt{\left(\frac{1}{u_X^2} - 1\right)\rho_{XY,r}^2 + 1}}$$

Correction for direct range restriction can also be done by using Thorndike's Case I formula, which assumes direct selection on X only but requires that the u-ratio of Y, $u_Y = S_{Y,r}/S_{Y,u}$, be known. Because this is a very uncommon scenario, the approach has been rarely used in practice.

In practice, range restriction is usually indirect. Direct range restriction requires an unusual situation in which individuals are selected in a strictly top-down manner based on a single criterion. In the example above, universities typically admit students based on multiple factors, such as GPA, letters of recommendation, admission essays, and extracurricular activities. Therefore, there exists a composite construct Z (referred to as suitability by Hunter, Schmidt, and Le [2006]) that is implicitly used to make admission decisions. Thorndike's Case III provided a correction formula for indirect range restriction, but it required knowledge of the standard deviations of the selection variable, Z, in both the restricted and the unrestricted samples. Hunter, Schmidt, and Le (2006) noted that this information is unavailable in most research settings and derived an alternative correction formula, which we describe below.

The key statistic used to correct for indirect range restriction is $u_T = S_{T,r}/S_{T,u}$, which is the ratio of the true-score standard deviations. The formulas to correct for direct and indirect range restriction are functionally identical, except for using u_T instead of u_X in the case of indirect range restriction (Hunter, Schmidt, and Le 2006):

$$\rho_{XY,u} = \frac{\rho_{XY,r}}{u_T\sqrt{\left(\frac{1}{u_T^2}-1\right)\rho_{XY,r}^2+1}}$$

When u_T is not available, it can be estimated from the observed-score u-ratio, u_X, and the reliability estimate of X in the restricted sample, $r_{XX,r}$, or the unrestricted sample $r_{XX,u}$; see *Estimating the values of artifacts for univariate range restriction* for more details.

The above formulas are derived under three assumptions:

1. The slopes of the simple linear regressions of Y on X in the restricted and unrestricted samples are equal.
2. The standard errors of the slope estimates from the above linear regressions are equal.
3. For indirect range restriction, the effect of selection on a third (suitability) variable Z on the dependent variable Y is fully caused (mediated) by the independent variable true-score T; see Hunter, Schmidt, and Le (2006, fig. 1) for a graphical depiction of this mediating assumption. This is known as the full-mediation assumption in the literature.

Instead of naming correction formulas as Case I, II, etc., which are not informative names, Dahlke and Wiernik (2020) referred to the range restriction presented in the above formulas as univariate direct/indirect range restriction on the independent variable X, because they require only one u-ratio variable to be known (u_X in the case of direct range restriction or u_T in the case of indirect range restriction). Also, univariate range restriction assumes that the selection process is entirely determined by only one of the variables in the correlation being studied. Similar formulas can be derived for univariate direct/indirect range restriction on the dependent variable Y (with associated true-score P) by replacing u_X and u_T with $u_Y = S_{Y,r}/S_{Y,u}$ and $u_P = S_{P,r}/S_{P,u}$, where $S_{Y,r}$, $S_{Y,u}$, $S_{P,r}$, and $S_{P,u}$ are the respective standard deviations of Y and P in the restricted and unrestricted samples; see *Correcting correlations for measurement errors and univariate range restriction* for more details.

When the full-mediation assumption is not met, Bryant and Gokhale (1972) proposed a method for correcting indirect range restriction that does not rely on this assumption. However, this method requires knowledge of the ratio of the standard deviations of Y in the restricted sample to the unrestricted sample, $u_Y = S_{Y,r}/S_{Y,u}$. Obtaining the value of the standard deviation estimate in the unrestricted sample, $S_{Y,u}$, may not be possible in certain research fields. For example, in personnel selection, when Y is job performance, it is nearly impossible to compute $S_{Y,u}$ (and therefore u_Y) in the unrestricted sample consisting of all persons who applied for the job regardless of whether they were hired. The approach of Bryant and Gokhale (1972) was further refined by Alexander (1990) and Dahlke and Wiernik (2020). Because this method requires knowledge of both u_X and u_Y, Dahlke and Wiernik (2020) referred to it as a bivariate indirect range restriction correction method; see *Correcting correlations for measurement errors and bivariate range restriction* for more details.

Selection may (rarely) occur directly on both variables X and Y. Alexander et al. (1987) derived a method to correct for this type of range restriction, which they referred to as bidimensional direct truncation. Their method also requires knowledge of both u_X and u_Y. Dahlke and Wiernik (2020) referred to this type of range restriction as bivariate direct range restriction; see *Correcting correlations for measurement errors and bivariate range restriction* for more details.

Artificial dichotomization and small-sample bias

You may also correct for other simple artifacts. For example, the dichotomization of a continuous variable (X or Y) is common in psychometric research. Dichotomization typically results in a point-biserial correlation (Pearson's correlation between a binary variable and a continuous variable) that is lower than the correlation originally observed when both variables were treated as continuous. Thus, dichotomization has an attenuating effect on correlations.

Suppose that, after dichotomization, a proportion p of the observations is assigned to one of the two levels of the new binary variable. The correction formulas are identical regardless of whether p represents the proportion of successes or failures. You may use the `xdich()` and `ydich()` options to correct correlations for the effect of dichotomizing either of the continuous variables X and Y (Schmidt and Hunter 2015, eq. 2.4, 43). The attenuation factor of dichotomization is given by

$$a_{\text{dich}} = \frac{\phi\left\{\Phi^{-1}(p)\right\}}{\sqrt{p(1-p)}}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the respective density and cumulative distribution functions of the standard normal distribution. The corrected correlation, accounting for the attenuating effect of dichotomization, is calculated as

$$\rho_{XY}^{\text{dich}} = \frac{\rho_{XY}}{a_{\text{dich}}}$$

Sample correlation typically exhibits a small negative bias, which is practically negligible for studies with sample sizes of 20 or more (Schmidt, Le, and Oh 2019, 320). You may also correct for small-study bias (Schmidt and Hunter 2015, eq. 3.23, 140), via the `small` option. The attenuation factor of small-study bias and the corresponding corrected correlation are given by

$$a_{\text{bias}} = \frac{2n-2}{2n-1} \qquad \text{and} \qquad \rho_{XY}^{\text{bias}} = \frac{\rho_{XY}}{a_{\text{bias}}}$$

Consider a study where both X and Y variables have been dichotomized and the sample size is small (less than 20). In such cases, corrections for both dichotomization and small-study bias are necessary. Let a_1, a_2, and a_3 represent the attenuation factors for each of these three artifacts, respectively. The product of these individual attenuation factors, $A = a_1 a_2 a_3$, is referred to as the compound attenuation factor. The corrected correlation, ρ^*_{XY}, accounting for the combined distorting effects of the three statistical artifacts, is calculated as

$$\rho^*_{XY} = \frac{\rho_{XY}}{A} \qquad \text{and} \qquad \text{Var}\left(\rho^*_{XY}\right) = \frac{\text{Var}(\rho_{XY})}{A^2} \tag{3}$$

The sample size will be adjusted to correspond to the value of n that would have generated the sampling variances for the corrected correlations.

$$n^* = \left\lceil \frac{\left\{1-(\rho^*_{XY})^2\right\}^2}{\text{Var}\left(\rho^*_{XY}\right)} + 1 \right\rceil \tag{4}$$

The adjusted sample size is stored in the system variable `_meta_studysize`.

❑ Technical note

Corrections for the distorting effects of the artifacts discussed in this section (that is, dichotomization and small-study bias) are applied before corrections for other artifacts, such as measurement error or range restriction. Consequently, the values of ρ^*_{XY} and n^* serve as the inputs for psychometric meta-analysis. These values will be identical to ρ_{XY} and n originally specified with `meta psycorr` if none of the `xdich()`, `ydich()`, or `small` options is specified.

❑

Using meta psycorr

To perform a psychometric meta-analysis using the `meta psycorr` command, both the observed correlations and their within-study sample sizes must be available. The command computes the corrected (unattenuated) correlations by adjusting the observed correlations for the artifacts specified in the model. In addition, the corresponding standard errors of the corrected correlations are computed. These values are stored in the system variables `_meta_es` (effect size) and `_meta_se` (standard error), respectively. Asymptotic 95% CI variables are computed and stored in variables `_meta_cil` (lower bound) and `_meta_ciu` (upper bound). Other confidence levels may be specified via the `level()` option.

To correct for measurement errors, reliability estimates for the variables X and Y (in measuring true-score variables T and P, respectively) are specified by using the `xreliability()` and `yreliability()` options, respectively. In the presence of range restriction, these reliability estimates might be derived from either restricted or unrestricted samples. By default, `meta psycorr` assumes the reliability estimates are based on the restricted sample for all the studies. You can specify the `unrestricted` option within `xreliability()` or `yreliability()` to specify they are based on the unrestricted sample in all the studies. Alternatively, you can use `restricted(`*rest_idvar*`)` or `unrestricted(`*unrest_idvar*`)` to specify a study-specific mixed pattern of restricted and unrestricted reliability estimates.

Range restriction can be specified by using the `xuratios()` option for range restriction in X, the `yuratios()` option for range restriction in Y, or both. Univariate range restriction is assumed when only one of the `xuratios()` or `yuratios()` options is specified. *varname* within `xuratios()` or `yuratios()` quantifies the degree of range restriction present in each study. It represents the ratios of standard deviations of X or Y in the restricted sample to those in the unrestricted sample. By default, *varname* represents observed-score u-ratios, $S_{X,r}/S_{X,u}$ for `xuratios()` and $S_{Y,r}/S_{Y,u}$ for `yuratios()`. You may specify the `true` suboption within `xuratios()` or `yuratios()` to indicate that u-ratios are true-score ratios of standard deviations, $S_{T,r}/S_{T,u}$ or $S_{P,r}/S_{P,u}$, respectively. Alternatively, you can use the `observed(`*obs_idvar*`)` or `true(`*true_idvar*`)` suboptions within `xuratios()` or `yuratios()` to specify a study-specific mixed pattern of observed-score and true-score u-ratios. When both `xuratios()` and `yuratios()` are specified, a bivariate range restriction is assumed. By default, indirect range restriction is assumed unless the `direct` option is specified. For bivariate indirect range restriction, additional options `signrxz()`, `signryz()`, and `nu()` are available. The first two are used in the computation of the corrected correlations, $\rho_{TP,u}$, and the latter is used for computing the sampling variance of $\rho_{TP,u}$; see *Sampling variances of the corrected correlations* $\rho_{TP,u}$ and example 8.

You may also correct for artificial dichotomization of variables X and Y via the `xdich()` and `ydich()` options, respectively. You must specify the proportion of successes (or failures) after dichotomization. This can be specified either as a single value #, which assumes the same proportion of successes across all studies, or as a variable containing these proportions for each individual study. When some of the studies included in the meta-analysis have a small sample size, you may use the `small` option to apply a small-sample correction factor. This adjustment reduces the negative bias in the corresponding observed correlations; see *Artificial dichotomization and small-sample bias* for details.

Missing reliability estimates or u-ratio values are common in psychometric meta-analysis. Missing values may be imputed using the `impute()` option. By default, bootstrap is used to impute the missing artifact values. You may use the `rseed()` suboption within the `impute()` option for reproducibility when using bootstrap. Other imputation methods (for example, `perfect` or `mean`) can be specified within the `impute()` option; see the `impute(`*methodspec*`)` option for details.

Examples of using meta psycorr

Consider the following fictional meta-analysis dataset:

```
. use https://www.stata-press.com/data/r19/mapsycorr_uvirr
(Fictional data for psychometric meta-analysis)

. describe studylbl rho n rxxr ryyr ux

Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
studylbl        str22   %22s                  Study label
rho             double  %9.0g                 X-Y correlation in restricted
                                                sample
n               int     %9.0g                 Study sample size
rxxr            double  %9.0g                 Reliability estimates for X in
                                                restricted sample
ryyr            double  %9.0g                 Reliability estimates for Y in
                                                restricted sample
ux              double  %9.0g                 Std. dev. ratio
                                                (restricted/unrestricted) of X
```

We will use it to describe various usages of the `meta psycorr` command. This dataset includes both measurement error and range restriction, but we also use it to demonstrate a hypothetical scenario when only measurement error is present.

The `rho` variable contains the observed correlations between two variables X (cognitive ability test scores) and Y (job performance ratings) computed from the restricted sample of hired candidates. Reliability estimates of X and Y in measuring T (true cognitive ability) and P (true job-performance ability), respectively, in the restricted sample are stored in variables `rxxr` and `ryyr`. Suppose only top-scoring candidates on the cognitive ability test are hired. The variability in test scores X is artificially reduced in the restricted sample (employed sample), attenuating the observed correlation between X and Y. This is a classic case of univariate direct range restriction. In reality, the hiring process is based on a different variable, such as a suitability variable Z (a mixture of education level, work experience, reference letters, etc.), which is correlated with cognitive ability X. This creates an indirect range restriction on X because only those with certain suitability levels (and thus higher cognitive ability) are included in the observed sample. Variable `ux` contains the observed-score u-ratio, u_X. It is the ratio of the standard deviation of X in the restricted sample to that in the unrestricted sample, $S_{X,r}/S_{X,u}$.

In the examples below, we show different specifications of `meta psycorr`, starting with ignoring the effects of the statistical artifacts (bare-bones meta-analysis) to then accounting for them under different scenarios.

Examples are presented under the following headings:

▷ Example 1: Bare-bones meta-analysis

We will start by ignoring the distorting effects of statistical artifacts such as measurement error and range restriction and therefore perform a bare-bones meta-analysis that corrects for sampling error only by aggregating the correlations across the studies. This can be done by using `meta psycorr` as follows:

```
. meta psycorr rho n, studylabel(studylbl)
Psychometric meta-analysis setting information

 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Correlation
          Variable: _meta_es

         Precision
          Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Bare-bones meta-analysis
```

Briefly, `meta psycorr` reports that we are performing a bare-bones meta-analysis with 12 studies, that `rho` and `n` are the variables used to declare effect sizes and compute their standard errors, that the default confidence level is 95%, and more. See *Meta settings with meta psycorr (StataNow)* in [META] **meta data** for a detailed description of all settings for this dataset.

We can now use, for example, `meta summarize` to compute the overall effect size, which is the mean correlation in this case, labeled as `theta` in the output below.

```
. meta summarize

  Effect-size label: Correlation
        Effect size: _meta_es
          Std. err.: _meta_se
        Study label: studylbl

Meta-analysis summary                         Number of studies =     12
Random-effects model                          Heterogeneity:
Method: Bare-bones MA                                      tau2 =  0.0088
                                                         I2 (%) =   59.34
                                                             H2 =    2.46

----------------------------------------------------------------------
                 Study |   Correlation    [95% conf. interval]  % weight
-----------------------+----------------------------------------------
Samwell et al. (2012)  |         0.240     -0.002       0.482      3.28
Cressen et al. (2007)  |         0.160      0.033       0.287     11.84
Pycelle et al. (2018)  |         0.210      0.027       0.393      5.74
 Qyburn et al. (2005)  |         0.390      0.098       0.682      2.24
 Wolkan et al. (2016)  |         0.120     -0.011       0.251     11.11
Volarik et al. (2019)  |         0.520      0.272       0.768      3.13
  Luwin et al. (2011)  |         0.030     -0.105       0.165     10.48
 Assaad et al. (2023)  |         0.320      0.119       0.521      4.74
  Aemon et al. (2008)  |         0.050     -0.078       0.178     11.63
Creylen et al. (2013)  |         0.110     -0.004       0.224     14.70
Colemon et al. (2017)  |         0.030     -0.099       0.159     11.47
  Vyman et al. (2015)  |         0.270      0.129       0.411      9.65
-----------------------+----------------------------------------------
                 theta |         0.147      0.079       0.216
----------------------------------------------------------------------
Test of theta = 0: z = 4.20                         Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(11) = 27.06             Prob > Q = 0.0045
```

The mean observed correlation is $\widehat{\theta} = 0.147$ with a 95% CI of $[0.079, 0.216]$. Given that we are not correcting for any statistical artifacts (other than sampling error), the interpretations of $\widehat{\theta}$, $\widehat{\tau}^2$, I^2, and CIs are identical to the case of traditional meta-analysis (defined by `meta set` or `meta esize`). See [META] **meta summarize** for details about this command. One difference is that, here, weights are the within-study sample sizes, n_j (specified by the `n` variable), expressed as percentages $100 \times n_j / \sum_{j=1}^{K} n_j$, whereas in traditional meta-analysis, weights are the inverse variances $1/(\widehat{\sigma}_j^2 + \widehat{\tau}^2)$, expressed as percentages.

◁

▷ Example 2: Measurement error

Continuing with example 1, we know that cognitive ability test scores X might not perfectly capture true cognitive ability T, and job performance ratings Y could be subject to rater bias or inconsistency, and therefore may not accurately measure true job-performance ability P. Therefore, the observed correlations ρ_{XY}'s are attenuated by the measurement error ($\rho_{XY} < \rho_{TP}$), and the observed mean correlation computed in example 1 underestimates the true mean correlation between T and P. To demonstrate how to correct for measurement error only, we ignore the range restriction in this example and assume that the reliability estimates are obtained from the entire sample. We use the `xreliability(rxxr)` and `yreliability(ryyr)` options to specify the reliability estimates.

We compute the correlations ρ_{TP}, corrected for attenuation because of measurement error, as follows:

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr) studylabel(studylbl)
Psychometric meta-analysis setting information
 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n
       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es
         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%
  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis
 Reliability for X
            Values: rxxr
 Reliability for Y
            Values: ryyr
```

Compared with example 1, `Method` now displays `Individual-correction meta-analysis`, and the effect size label is now `Corrected correlation`. This is because we are now correcting for the distorting effects of measurement error. The corrected correlations stored in the `_meta_es` variable are the ρ_{TP} values defined in (5) in *Correcting correlations for measurement errors only*, whereas the sampling standard errors, stored in the `_meta_se` variable, are computed as the square root of the sampling variances, $\widehat{\text{Var}}\,(\rho_{TP})$, given in (11).

The output additionally reports information about the reliability estimates of X and Y. In particular, these estimates are stored in variables `rxxr` (`Values: rxxr`) and `ryyr` (`Values: ryyr`).

As with example 1, we can now perform a psychometric meta-analysis to compute the mean corrected correlation by using `meta summarize` or `meta forestplot`. For details on interpreting the output of these commands when the observed correlations are individually corrected for statistical artifacts, see example 3 and example 4.

◁

▷ Example 3: Measurement error and univariate indirect range restriction

In example 2, we ignored that our data are based on the restricted sample and focused solely on correcting the correlations for measurement error. In this example, we will adjust for both artifacts present in these data: measurement error and range restriction. As before, to account for measurement error, we use the `xreliability(rxxr)` and `yreliability(ryyr)` options to specify the reliability estimates. But in the presence of range restriction, the reliability estimates, by default, are assumed to be obtained from the restricted sample.

We assume that range restriction is indirect in X and follow the full-mediation assumption discussed in *Range restriction*; that is, any range restriction that exists on Y is fully caused (mediated) by the range restriction in X. We use the `xuratios()` option to specify the variable (`ux`) that stores the observed-score u-ratio for X.

We compute the corrected correlations for attenuation because of measurement error and univariate indirect range restriction, $\rho^{\text{uvirr}_x}_{TP,u}$, as follows:

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr)
> xuratios(ux) studylabel(studylbl)
Psychometric meta-analysis setting information
 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n
       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es
         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%
  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis
 Reliability for X
            Values: rxxr
              Type: restricted
 Reliability for Y
            Values: ryyr
              Type: restricted
 Range restriction
        u_X values: ux
          u_X type: observed
              Type: indirect
```

Here, compared with example 2, we are now correcting for the distorting effects of measurement error and range restriction. The corrected correlations stored in the _meta_es variable are the $\rho^{\text{uvirr}_x}_{TP,u}$ values shown in *Correcting correlations for measurement errors and univariate range restriction*, whereas the sampling standard errors are computed as the square root of the sampling variances, $\widehat{\text{Var}}\left(\rho_{TP,u}\right)$, defined in (11).

The output also reports additional information about the reliability estimates of X and Y. In particular, these estimates, which are stored in variables `rxxr` (`Values: rxxr`) and `ryyr` (`Values: ryyr`), are assumed computed from the restricted sample (`Type: restricted`). Information about the range restriction is also shown; in particular, the u-ratio values are stored in the ux variable (`u_X values: ux`), and these values are the observed-score u-ratios (`u_X type: observed`). In other words, the u-ratio values represent the ratio of observed-score standard deviations of X in the restricted sample to those in the unrestricted sample, $S_{X,r}/S_{X,u}$. Finally, the range restriction is assumed indirect (`Type: indirect`).

The expression for the corrected correlations, $\rho^{\text{uvirr}_x}_{TP,u}$, as defined in (8), depends on $r_{XX,r}$, $r_{YY,r}$, and the true-score u-ratios u_T; therefore, `meta psycorr` uses the values of the observed-score u-ratios u_X (stored in the ux variable) to compute u_T as follows:

$$u_T = \sqrt{\frac{r_{XX,r}u_X^2}{1 + r_{XX,r}u_X^2 - u_X^2}}$$

These values are stored in the system variable _meta_ut. See *Estimating the values of artifacts for univariate range restriction* for more details.

Let's now perform a psychometric meta-analysis to compute the mean corrected correlation by using meta summarize.

```
. meta summarize

  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
        Study label: studylbl
     Correcting for: Measurement errors in X and Y.
                     Univariate indirect range restriction in X.

Meta-analysis summary                         Number of studies =     12
Random-effects model                          Heterogeneity:
Method: Individual-correction MA                          tau2 =  0.0312
                                                        I2 (%) =   57.45
                                                            H2 =    2.35

           Effect size: Corrected correlation

                    Study     Effect size    [95% conf. interval]  % weight

  Samwell et al. (2012)         0.446        0.018      0.874       4.25
  Cressen et al. (2007)         0.308        0.068      0.547      14.17
  Pycelle et al. (2018)         0.474        0.077      0.870       5.02
   Qyburn et al. (2005)         0.737        0.257      1.000       2.83
   Wolkan et al. (2016)         0.242       -0.020      0.504      12.06
  Volarik et al. (2019)         0.866        0.633      1.000       5.05
    Luwin et al. (2011)         0.071       -0.248      0.390       8.32
   Assaad et al. (2023)         0.596        0.257      0.935       6.10
    Aemon et al. (2008)         0.106       -0.165      0.376      11.52
  Creylen et al. (2013)         0.279       -0.005      0.562      10.14
  Colemon et al. (2017)         0.067       -0.221      0.355      10.18
    Vyman et al. (2015)         0.548        0.291      0.806      10.36

                    theta       0.326        0.195      0.458

Test of theta = 0: z = 4.85                          Prob > |z| = 0.0000
Test of homogeneity: Q = chi2(11) = 25.85              Prob > Q = 0.0068
```

The output header lists the artifacts used to correct the correlations for attenuation. The second column of the output table shows the individually corrected correlations, whereas the third and fourth columns present their corresponding 95% CIs. For instance, in the first study (Samwell et al. (2012)), the corrected correlation is 0.446 with a 95% CI of [0.018, 0.874].

The last column provides the weights for each study as a percentage. These weights are calculated by using (2) in [META] **meta summarize** and are used in the meta-analysis to compute the mean corrected correlation $\hat{\theta}$. Unlike the weights from the bare-bones meta-analysis in example 1, which depend solely on study sample sizes, these weights also incorporate the degree of attenuation caused by statistical artifacts. Consequently, studies with larger sample sizes and lower levels of attenuation are assigned greater weights in the meta-analysis.

The mean corrected correlation is 0.326 with a 95% CI of [0.195, 0.458]. The heterogeneity parameter is $\hat{\tau}^2 = 0.0312$ and $I^2 = 57.45\%$ [see (3) in [META] **meta summarize** for the computation of I^2], which means that 57.45% of the variability in the corrected correlations is because of the between-study differences rather than variability because of statistical artifacts (measurement error and range restriction) or sampling error.

◁

▷ Example 4: Credibility intervals and forest plots

Continuing with example 3, we will construct an 80% credibility interval by using the `credinterval` option. We specify the `nostudies` option to suppress the individually corrected correlations and their CIs and instead focus on the mean (overall) corrected correlations. We also specify the `tdistribution` option to display a t test and a 95% CI based on the Student's t distribution instead of the default normal distribution.

```
. meta summarize, nostudies credinterval tdistribution

  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
        Study label: studylbl
     Correcting for: Measurement errors in X and Y.
                     Univariate indirect range restriction in X.

Meta-analysis summary                            Number of studies =     12
Random-effects model                             Heterogeneity:
Method: Individual-correction MA                             tau2 =  0.0312
                                                           I2 (%) =   57.45
                                                               H2 =    2.35

        theta: Overall corrected correlation
```

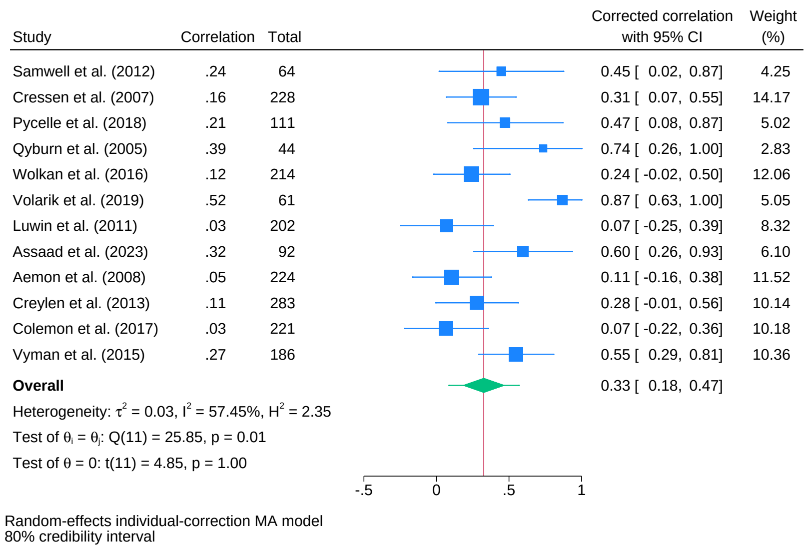
	Estimate	Std. err.	t	P>|t|	[95% conf. interval]	
theta	.3264477	.0672589	4.85	0.001	.1784118	.4744835

```
80% credibility interval for theta: [0.086, 0.567]
Test of homogeneity: Q = chi2(11) = 25.85                 Prob > Q = 0.0068
```

The 80% credibility interval, reported at the bottom of the table, is [0.086, 0.567]. You may specify the `credinterval(#)` option to request credibility levels other than the default 80%. The credibility interval can be interpreted as a plausible range for the middle 80% of values in the distribution of population true-score correlations, $\rho_{TP,u}$.

We can also present the results of the psychometric meta-analysis and display the credibility interval graphically by using a forest plot. This can be done as follows:

```
. meta forestplot, credinterval tdistribution esrefline
  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
        Study label: studylbl
     Correcting for: Measurement errors in X and Y.
                     Univariate indirect range restriction in X.
```



The overall effect size corresponds to the green diamond centered at the estimate of the mean corrected correlation. The width of the diamond corresponds to the width of the overall CI, $[0.18, 0.47]$. The green whiskers extending from the overall diamond span the width of the credibility interval displayed in `meta summarize` $[0.086, 0.567]$. We also specified the `esrefline` option to draw a vertical red line at the mean corrected correlation value. For more information about the forest plot, see [META] **meta forestplot**.

◁

▷ Example 5: Reliability estimates from the unrestricted sample

In example 3, the reliability estimates specified within the `xreliability()` and `yreliability()` options are, by default, assumed to be computed from the restricted samples. If we only have access to reliability estimates from the unrestricted samples, we should specify the `unrestricted` suboption within these options. Continuing with example 3, we illustrate this by typing `xreliability(rxxu, unrestricted)`, where the `rxxu` variable stores the reliability estimates for X in the unrestricted sample.

```
. meta psycorr rho n, xreliability(rxxu, unrestricted) yreliability(ryyr)
> xuratios(ux) studylabel(studylbl)

Psychometric meta-analysis setting information

 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Reliability for X
            Values: rxxu
              Type: unrestricted

 Reliability for Y
            Values: ryyr
              Type: restricted

 Range restriction
       u_X values: ux
         u_X type: observed
             Type: indirect
```

The output is similar to that in example 3 except now `meta psycorr` reports under `Reliability for X` that the reliability estimates for X are from the unrestricted sample (`Type: unrestricted`). The expression for the corrected correlation, $\rho^{\text{uvirr}_x}_{TP,u}$, defined in (8) depends on the restricted reliability estimates $r_{XX,r}$ and $r_{YY,r}$ as well as the true-score u-ratios u_T; therefore, `meta psycorr` uses the values of unrestricted reliability estimates, $r_{XX,u}$ (stored in the `rxxu` variable), to compute $r_{XX,r}$ as follows:

$$r_{XX,r} = 1 - \frac{1 - r_{XX,u}}{u_X^2}$$

See *Estimating the values of artifacts for univariate range restriction* for more details.

Alternatively, suppose some studies report reliability estimates for X based on restricted samples, whereas others use unrestricted samples. You can specify the `xreliability(rxx, restricted(idx_r))` option, where the `rxx` variable stores the reliability estimates and the `idx_r` variable identifies whether the sample is restricted (`idx_r` $= 1$) or unrestricted (`idx_r` $= 0$). Compared with the previous specification of `meta psycorr`, we are only modifying how the reliability estimates for X are specified; therefore, it is syntactically convenient to use `meta update` to avoid respecifying other settings (for example, range restriction) that remain unchanged.

```
. meta update, xreliability(rxx, restricted(idx_r))
-> meta psycorr rho n, xreliability(rxx, restricted(idx_r)) yreliability(ryyr)
>  xuratios(ux) studylabel(studylbl)
Psychometric meta-analysis setting information from meta psycorr
 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
          Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Reliability for X
            Values: rxx
              Type: restricted, identified by idx_r

 Reliability for Y
            Values: ryyr
              Type: restricted

 Range restriction
        u_X values: ux
          u_X type: observed
              Type: indirect
```

The output now reports under `Reliability for X` that the restricted reliability estimates for X are identified by the indicator variable `idx_r` (`Type: restricted, identified by idx_r`). If we were to run `meta summarize` after both `meta psycorr` specifications in this example to compute the mean corrected correlation, the output would be identical to that in example 3.

◁

▷ Example 6: True-score u-ratios

In example 3, the u-ratios specified in the `xuratios()` option are, by default, assumed to be observed-score ratios of standard deviations of X in the restricted sample to those in the unrestricted sample (that is, $u_X = S_{X,r}/S_{X,u}$). If we have access to true-score u-ratios, $u_T = S_{T,r}/S_{T,u}$, we can use the `true` suboption of `xuratios()`. $S_{T,r}$ and $S_{T,u}$ are the standard deviations of T in the restricted and unrestricted samples, respectively. Continuing with example 3, we specify `xuratios(ut, true)`, where the `ut` variable stores the true-score u-ratios.

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr)
> xuratios(ut, true) studylabel(studylbl)
Psychometric meta-analysis setting information
 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
          Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Reliability for X
            Values: rxxr
              Type: restricted

 Reliability for Y
            Values: ryyr
              Type: restricted

 Range restriction
       u_X values: ut
         u_X type: true
              Type: indirect
```

The output is similar to that in example 3 except that `meta psycorr` now reports under `Range restriction` that the u-ratio values are specified in the `ut` variable (`u_X values: ut`) and that these are true-score u-ratios (`u_X type: true`) instead of the default observed-score u-ratios.

Alternatively, suppose some studies report observed-score u-ratios, whereas others report true-score u-ratios. You can specify the `xuratios(u, observed(idx_o))` option, where the `u` variable stores the u-ratio values and the `idx_o` variable identifies whether they are observed-score u-ratios (`idx_o` $= 1$) or true (`idx_o` $= 0$). Like we did in example 5, we use `meta update` for syntactical convenience to avoid respecifying other settings (for example, measurement error) that remain unchanged.

```
. meta update, xuratios(u, observed(idx_o))
-> meta psycorr rho n , xreliability(rxxr) yreliability(ryyr)
> xuratios(u, observed(idx_o)) studylabel(studylbl)

Psychometric meta-analysis setting information from meta psycorr

 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Reliability for X
            Values: rxxr
              Type: restricted

 Reliability for Y
            Values: ryyr
              Type: restricted

 Range restriction
        u_X values: u
          u_X type: observed, identified by idx_o
              Type: indirect
```

The output now reports under `Range restriction` that the observed-score u-ratios are identified by the indicator variable `idx_o` (`Type: observed, identified by idx_o`). If we were to run `meta summarize` after both `meta psycorr` specifications in this example to compute the mean corrected correlation, the output would be identical to that in example 3.

◁

▷ Example 7: Univariate direct range restriction and other estimation methods

As mentioned in *Range restriction*, direct range restriction rarely occurs in practice; that is, almost all range restrictions encountered in practice are indirect. But for illustration purposes, we show how you can specify direct range restriction with `meta psycorr`. We repeat the `meta psycorr` command used in example 5, but we now specify the `direct` option.

```
. meta psycorr rho n, xreliability(rxxu, unrestricted) yreliability(ryyr)
> xuratios(ux) direct studylabel(studylbl)

Psychometric meta-analysis setting information

 Study information
    No. of studies: 12
       Study label: studylbl
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Reliability for X
            Values: rxxu
              Type: unrestricted

 Reliability for Y
            Values: ryyr
              Type: restricted

 Range restriction
        u_X values: ux
          u_X type: observed
              Type: direct
```

We can compute the mean corrected correlation by using `meta summarize`:

```
. meta summarize, nostudies

  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
        Study label: studylbl
     Correcting for: Measurement errors in X and Y.
                     Univariate direct range restriction in X.

Meta-analysis summary                                    Number of studies =     12
Random-effects model                                     Heterogeneity:
Method: Individual-correction MA                                      tau2 =   0.0284
                                                                    I2 (%) =    62.08
                                                                        H2 =     2.64

        theta: Overall corrected correlation

             |   Estimate   Std. err.       z    P>|z|     [95% conf. interval]
-------------+------------------------------------------------------------------
       theta |   .2690783    .0617478    4.36    0.000     .1480548    .3901018
--------------------------------------------------------------------------------
Test of homogeneity: Q = chi2(11) = 29.01                          Prob > Q = 0.0023
```

After computing the corrected correlations (_meta_es) and their standard errors (_meta_se), we use the default individual-correction meta-analysis random-effects estimation method, which also corresponds to the random(icma) option, to obtain the mean corrected correlation. This method is the preferred choice in psychometric meta-analysis and is the only one described in the literature. However, there is no conceptual barrier preventing the use of other traditional meta-analysis estimation methods. For example, after calculating _meta_es and _meta_se with meta psycorr, we could request the DerSimonian–Laird method by specifying the random(dlaird) option with the desired command. Below, we demonstrate this by using meta summarize.

```
. meta summarize, nostudies random(dlaird)
  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
        Study label: studylbl
     Correcting for: Measurement errors in X and Y.
                     Univariate direct range restriction in X.

Meta-analysis summary                                Number of studies =     12
Random-effects model                                 Heterogeneity:
Method: DerSimonian-Laird                                        tau2 =   0.0449
                                                               I2 (%) =    71.75
                                                                   H2 =     3.54

        theta: Overall corrected correlation
```

	Estimate	Std. err.	z	P>|z|	[95% conf. interval]	
theta	.336066	.073458	4.57	0.000	.1920909	.4800411

```
Test of homogeneity: Q = chi2(11) = 38.93                       Prob > Q = 0.0001
```

The estimate for the DerSimonian–Laird between-study heterogeneity is $\hat{\tau}^2_{\rm DL} = 0.0449$. This is larger than the one we obtained from the default individual-correction meta-analysis method ($\hat{\tau}^2_{\rm ICMA} = 0.0284$). Because the weights in the traditional meta-analysis depend on $\hat{\tau}^2$, the mean corrected correlation $\hat{\theta}_{\rm DL} = 0.336$ is also different from $\hat{\theta}_{\rm ICMA} = 0.269$. Heterogeneity statistics such as I^2, H^2, and Q are computed as described in *Heterogeneity measures* in *Methods and formulas* in [META] **meta summarize**, whereas the statistics reported in the first table of this example are defined in *I^2 and H^2 statistics* in *Methods and formulas* of [META] **meta summarize**.

◁

So far, we have assumed a univariate indirect range restriction in most of the previous examples and used the Hunter, Schmidt, and Le (2006) formula to correct the correlations for range restriction. A fundamental assumption in deriving the correction formula is that the effect of selection on a third (suitability) variable Z on the dependent variable Y is fully caused (mediated) by the independent variable true-score T. In practice, this full-mediation assumption can be difficult to verify (Le et al. 2016). Beatty et al. (2014) and Fife, Mendoza, and Terry (2013) provided scenarios where that assumption is unlikely to hold. In the next example, we show how to correct for indirect range restriction while avoiding the full-mediation assumption. However, this requires knowledge of both u-ratio variables, u_X and u_Y.

▷ Example 8: Bivariate indirect range restriction

Consider a fictional dataset, inspired by the example in Le et al. (2016), with 45 studies that explores the relation between two latent constructs T (conscientiousness) and P [Organizational Citizenship Behavior (OCB)] in the population of all working adults. These constructs are measured (with error) using instruments X and Y, respectively (for example, self-report surveys, peer evaluations, or supervisor ratings). The studies include only long-term employees (restricted sample), excluding those who left because of low organizational commitment (construct Z). This self-selection process leads to indirect range restriction, reducing the observed correlation between conscientiousness and OCB if both are linked to organizational commitment.

Suppose we wish to check for small-study effects by constructing a funnel plot (or by using Egger's test; see [META] **meta bias**). It is better to assess small-study effects using the corrected correlations for artifacts and their corresponding standard errors because they are better representations of the true correlation distribution. Let's first describe our dataset:

```
. use https://www.stata-press.com/data/r19/mapsycorr_bvirr, clear
(Fictional data for bivariate range restriction)

. describe

Contains data from dta/mapsycorr_bvirr.dta
 Observations:            45                  Fictional data for bivariate
                                                range restriction
    Variables:             8                  4 Sep 2025 12:50
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
n               int     %9.0g                 Study sample size
rho             double  %9.0g                 X-Y correlation in restricted
                                                sample
rxxr            double  %9.0g                 Reliability estimates for X in
                                                restricted sample
ux              double  %9.0g                 Std. dev. ratio
                                                (restricted/unrestricted) of X
ryyr            double  %9.0g                 Reliability estimates for Y in
                                                restricted sample
uy              double  %9.0g                 Std. dev. ratio
                                                (restricted/unrestricted) of Y
nu              int     %9.0g                 Unrestricted sample size
moderator       float   %9.0g                 Moderator variable
-------------------------------------------------------------------------------
Sorted by:
```

Variables `rho`, `n`, `rxxr`, and `ryyr` are observed correlations in the restricted sample, the within-study sample size, and the reliability estimate for X and Y in the restricted samples, respectively. Variables `ux` and `uy` are the observed-score u-ratios of X and Y, respectively.

Because we have access to u-ratios for Y, we can bypass the full-mediation assumption and compute corrected correlations, $\rho^{\text{bvirr}}_{TP,u}$, as defined in (9). We use the `xuratios(ux)` and `yuratios(uy)` options to specify both u-ratio variables and instruct `meta psycorr` to correct for bivariate indirect range restriction (and measurement error).

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr)
> xuratios(ux) yuratios(uy)

Psychometric meta-analysis setting information

 Study information
    No. of studies: 45
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Reliability for X
            Values: rxxr
              Type: restricted

 Reliability for Y
            Values: ryyr
              Type: restricted

 Range restriction
       u_X values: ux
         u_X type: observed
       u_Y values: uy
         u_Y type: observed
             Type: indirect
```

Compared with the output of example 3, we have additional information under `Range restriction` for the u-ratio of Y (`u_Y values: uy` and `u_Y type: observed`). The corrected correlations, stored in `_meta_es` (and `_meta_rtpu`) are computed according to (9). The sampling standard errors, stored in `_meta_se`, are computed as $\sqrt{\widehat{\text{Var}}\left(\rho_{TP,u}\right)}$, where the sampling variances, $\widehat{\text{Var}}\left(\rho_{TP,u}\right)$, are defined in (12).

The computation of the corrected correlations, $\rho^{\text{bvirr}}_{TP,u}$, assumes that the correlations between the latent selection variable, Z, and variables X and Y, ρ_{XZ} and ρ_{YZ}, are positive across all studies. You may use the `signrxz()` and `signryz()` options to specify study-specific signs for these correlations; see λ in (9) for details.

Additionally, the computation of the sampling variances also assumes that the standard deviations of X and Y in the unrestricted sample are estimated without error (that is, based on an infinite sample size n_u). If values of n_u are available, you should specify them via the `nu()` option to obtain more accurate estimates. Here, the `nu` variable stores the sample sizes n_u of the unrestricted sample in each study. The `nu()` option is required when reliability estimates from the unrestricted samples are specified (for example, via `xreliability(rxxu, unrestricted)`), which is not the case in our example, but see *Sampling variances of the corrected correlations* $\rho_{TP,u}$.

```
. meta update, nu(nu)
-> meta psycorr rho n, xreliability(rxxr) yreliability(ryyr) xuratios(ux)
> yuratios(uy) nu(nu)
Psychometric meta-analysis setting information from meta psycorr
 Study information
    No. of studies: 45
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n
       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es
         Precision
         Std. err.: _meta_se
 Unrestricted size: nu
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%
  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis
 Reliability for X
            Values: rxxr
              Type: restricted
 Reliability for Y
            Values: ryyr
              Type: restricted
 Range restriction
       u_X values: ux
         u_X type: observed
       u_Y values: uy
         u_Y type: observed
             Type: indirect
```

The output now additionally lists `Unrestricted size: nu` under `Precision` to highlight that the unrestricted sample sizes, n_u, were used to obtain more accurate estimates of the sampling variances of $\rho^{\text{bvirr}}_{TP,u}$.

When n_u is not known, a sensitivity analysis can be conducted by specifying different n_u values corresponding to low, medium, or high selection ratios (n/n_u) and assessing their impact on the corrected correlations and their mean.

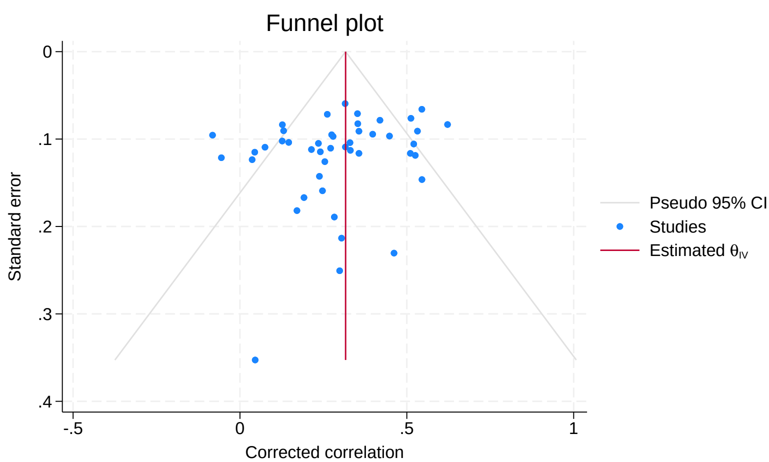
Let's now compute the mean corrected correlation by using meta summarize:

```
. meta summarize, nostudies tdistribution

  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
     Correcting for: Measurement errors in X and Y.
                     Bivariate indirect range restriction.

Meta-analysis summary                                Number of studies =     45
Random-effects model                                 Heterogeneity:
Method: Individual-correction MA                                 tau2 =  0.0177
                                                               I2 (%) =   62.88
                                                                   H2 =    2.69

        theta: Overall corrected correlation

             Estimate    Std. err.      t     P>|t|     [95% conf. interval]

       theta  .3166293    .0250018   12.66    0.000     .2662416    .3670171

Test of homogeneity: Q = chi2(44) = 118.54                       Prob > Q = 0.0000
```

The mean corrected correlation is 0.3166 with the 95% CI of [0.266, 0.367].

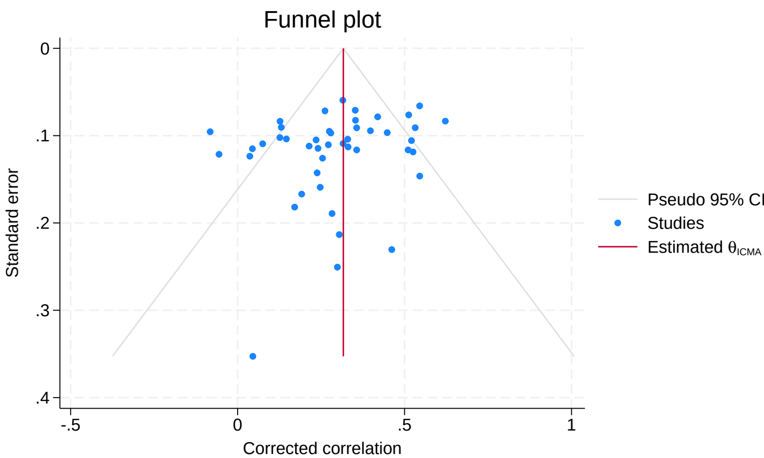
We are now ready to construct our funnel plot by using meta funnelplot.

```
. meta funnel

  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
     Correcting for: Measurement errors in X and Y.
                     Bivariate indirect range restriction.
              Model: Common effect
             Method: Inverse-variance
```



By default, historically, the assumed model to compute the overall correlation is a common-effect model with the inverse-variance method. If you prefer to compute the mean corrected correlation by using the individual-correction meta-analysis (ICMA) method, you can specify the `random(icma)` option.

```
. meta funnel, random(icma)

  Effect-size label: Corrected correlation
        Effect size: _meta_es
          Std. err.: _meta_se
     Correcting for: Measurement errors in X and Y.
                     Bivariate indirect range restriction.
              Model: Random effects
             Method: Individual-correction MA
```



In this example, $\widehat{\theta}_{\rm IV} = 0.31658$ (stored as `r(theta)` after running `meta funnelplot`) is very close to $\widehat{\theta}_{\rm ICMA} = 0.31663$, but this is not always the case. The majority of the studies are large (they have small standard errors) and are more or less evenly distributed around the mean of the corrected correlation, indicating that the small-study effect may not be a major issue for this dataset.

◁

▷ Example 9: Bivariate direct range restriction

Continuing with example 8, for illustration purposes, we will show how to correct correlations for bivariate direct range restriction on both X and Y. We only need to specify the `direct` option.

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr)
> xuratios(ux) yuratios(uy) direct

Psychometric meta-analysis setting information

 Study information
    No. of studies: 45
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
             Type: correlation
            Label: Corrected correlation
         Variable: _meta_es

         Precision
         Std. err.: _meta_se
               CI: [_meta_cil, _meta_ciu]
         CI level: 95%

  Model and method
            Model: Random effects
           Method: Individual-correction meta-analysis

 Reliability for X
           Values: rxxr
             Type: restricted

 Reliability for Y
           Values: ryyr
             Type: restricted

 Range restriction
       u_X values: ux
         u_X type: observed
       u_Y values: uy
         u_Y type: observed
             Type: direct
```

Below, we illustrate how to conduct a meta-regression by using meta regress to investigate whether a moderator variable (moderator) can account for some of the variability among the corrected correlations:

```
. meta regress moderator
note: method icma is not allowed with meta regress; using reml method.

  Effect-size label: Corrected correlation
        Effect size: _meta_es
           Std. err.: _meta_se
      Correcting for: Measurement errors in X and Y.
                      Bivariate direct range restriction.

Random-effects meta-regression                     Number of obs  =        45
Method: REML                                       Residual heterogeneity:
                                                               tau2 = .002961
                                                             I2 (%) =   28.82
                                                                 H2 =    1.40
                                                      R-squared (%) =   91.42
                                                   Wald chi2(1)   =    139.91
                                                   Prob > chi2    =    0.0000

------------------------------------------------------------------------------
    _meta_es | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   moderator |   .4556272   .0385198    11.83   0.000     .3801298    .5311246
       _cons |  -.3775785   .0510615    -7.39   0.000    -.4776573   -.2774997
------------------------------------------------------------------------------
Test of residual homogeneity: Q_res = chi2(43) = 54.44   Prob > Q_res = 0.1132
```

The note after the command specification is produced because the ICMA method (random(icma)) is only available to compute overall effect sizes (mean correlations) and cannot be used in meta-regression. The default estimation method in this case is the REML method. Unlike with many Stata regression commands, we do not specify the dependent variable with meta regress. The command assumes automatically that it is _meta_es (corrected correlations) from the declared meta settings.

The output header includes information about the artifacts that were used for correcting the study correlations. The reported I^2_{res} statistic is 29%, which suggests low heterogeneity, using the categorization of Higgins et al. (2003), after including moderator as the moderator. In other words, 29% of the variability in the residuals is attributed to the between-study variation, whereas 71% is attributed to the sampling error, measurement error, and range restriction. The adjusted R^2 statistic can be used to assess the proportion of between-study variance explained by the covariates; see (6) in *Residual heterogeneity measures* in *Methods and formulas* of [META] **meta regress** for its definition used in the meta-analysis literature. Here roughly 91% of the between-study variance is explained by the covariate moderator. See example 1 of [META] **meta regress** for details about the interpretation of the meta regress output.

◁

▷ Example 10: Missing artifact information and imputation methods

Consider a fictional psychometric meta-analysis of 12 studies where a few artifact values are missing. In particular, the rxxr variable has two missing reliability estimates for X in the restricted sample, and the ryyr variable has one missing reliability estimate for Y in the restricted sample. Additionally, the observed-score u-ratio ux has two missing values. Below, we list our dataset:

```
. use https://www.stata-press.com/data/r19/mapsycorr_miss, clear
(Fictional data for psychometric MA with missing artifact information)
. list
```

	studylbl	rho	n	rxxr	ryyr	ux
1.	Samwell et al. (2012)	.24	64	.61	.8	.81
2.	Cressen et al. (2007)	.16	228	.83	.57	.77
3.	Pycelle et al. (2018)	.21	111	.	.52	.79
4.	Qyburn et al. (2005)	.39	44	.74	.55	.76
5.	Wolkan et al. (2016)	.12	214	.71	.64	.78
6.	Volarik et al. (2019)	.52	61	.81	.	.
7.	Luwin et al. (2011)	.03	202	.61	.69	.74
8.	Assaad et al. (2023)	.32	92	.76	.63	.75
9.	Aemon et al. (2008)	.05	224	.61	.75	.
10.	Creylen et al. (2013)	.11	283	.	.58	.7
11.	Colemon et al. (2017)	.03	221	.66	.66	.75
12.	Vyman et al. (2015)	.27	186	.79	.64	.67

Variables rho and n represent the attenuated observed correlations and study sample sizes, respectively.

We use `meta psycorr` to correct the correlations for measurement error in both X and Y and to correct the correlations for range restriction.

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr)
> xuratios(ux) impute(bootstrap, rseed(19))
Psychometric meta-analysis setting information
 Study information
    No. of studies: 12
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n
       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es
         Precision
          Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%
  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis
 Imputation method: bootstrap
 Reliability for X
            Values: rxxr
       Imputations: 2
              Type: restricted
 Reliability for Y
            Values: ryyr
       Imputations: 1
              Type: restricted
 Range restriction
        u_X values: ux
   u_X imputations: 2
          u_X type: observed
              Type: indirect
```

In the presence of missing artifact information, `meta psycorr` will, by default, use bootstrap (sampling with replacement) to replace the missing reliability estimates and u-ratios. We also specified the `rseed()` suboption for reproducibility. This imputation method is labeled `bootstrap` in the above output (`Imputation method: bootstrap`). The number of imputed values is reported in the section corresponding to each artifact. For example, under `Range restriction`, we see that two values of the observed-score u-ratio, u_X, were imputed (`u_X imputations: 2`).

The imputed values are stored in the system variable corresponding to the artifact of interest. For example, to see the imputed values via bootstrap for study 6, we type

```
. list ryyr ux _meta_ryyr _meta_ux in 6
```

	ryyr	ux	_met~yyr	_meta_ux
6.	.	.	.57	.81

Alternatively, you can specify an imputation method that is different from the default one (`bootstrap`) by using the `impute()` option. Below, we use the `impute(perfect)` option to impute a value of 1 for the missing artifact information. This is equivalent to assuming a perfectly measured variable (no measurement error for the variable with a missing reliability estimate) and no range restriction.

```
. meta psycorr rho n, xreliability(rxxr) yreliability(ryyr)
> xuratios(ux) impute(perfect)
Psychometric meta-analysis setting information
 Study information
    No. of studies: 12
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n
       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es
         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%
  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis
 Imputation method: perfect
 Reliability for X
            Values: rxxr
       Imputations: 2
              Type: restricted
 Reliability for Y
            Values: ryyr
       Imputations: 1
              Type: restricted
 Range restriction
        u_X values: ux
   u_X imputations: 2
          u_X type: observed
              Type: indirect
```

The output now shows that `perfect` was used as the imputation method (`Imputation method: perfect`) instead of the default `bootstrap` method. For example, for study 6, we can confirm that missing values for `ryyr` and `ux` were imputed with a value of 1.

```
. list ryyr ux _meta_ryyr _meta_ux in 6
```

	ryyr	ux	_met~yyr	_meta_ux
6.	.	.	1	1

Other imputation methods are available, where missing values are replaced with the mean, `impute(mean)`, or the sample-size weighted mean, `impute(wmean)`, of the available artifact values. You may also specify `impute(none)` if you want to ignore studies with missing artifact information and perform no imputation.

◁

❑ Technical note

Direct imputation methods, such as bootstrapping or mean-value imputation, are not recommended when a substantive proportion of artifact information is missing, because they do not fully account for the uncertainty introduced by missing artifacts.

❑

▷ Example 11: Dichotomization and small-study bias

Consider a meta-analysis where both X and Y variables have been dichotomized. The proportions of successes (proportions of 1s) after dichotomization of X and Y are given in variables px and py, respectively. Variables rxy and n represent the attenuated observed correlation and study sample sizes, respectively. Additionally, the rxx variable contains the reliability estimates for X. We do not have range restriction in this meta-analysis; therefore, the distinction between restricted and unrestricted samples is not applicable here.

There are many studies with a small sample size (less than 20). In such cases, corrections for both dichotomization and small-study bias are necessary. Below, we describe our dataset.

```
. use https://www.stata-press.com/data/r19/mapsycorr_dich, clear
(Fictional data for psychometric MA with dichotomization and small-study bias)

. describe

Contains data from dta/mapsycorr_dich.dta
 Observations:            12                  Fictional data for psychometric
                                                MA with dichotomization and
                                                small-study bias
    Variables:             6                  4 Sep 2025 12:51
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
studylbl        str21   %21s                  Study label
rho             double  %9.0g                 X-Y correlation
n               byte    %9.0g                 Study sample size
rxx             double  %9.0g                 Reliability estimates for X
px              float   %9.0g                 Success proportion after
                                                dichotomization of X
py              float   %9.0g                 Success proportion after
                                                dichotomization of Y
-------------------------------------------------------------------------------
Sorted by:
```

We use the `xdich()` and `ydich()` options to correct our observed correlations for the distorting effects of dichotomization. The proportion of successes in the `px` variable is constant across studies and is equal to 0.5. Therefore, we can specify either `xdich(.5)` or `xdich(px)` with `meta psycorr`. And we also specify `ydich(py)`. The small-study bias is corrected by specifying the `small` option.

```
. meta psycorr rho n, small xdich(.5) ydich(py)
Psychometric meta-analysis setting information

 Study information
    No. of studies: 12
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

 Initial artifacts: Dichotomization and small-sample bias
```

As mentioned in the *Technical note* in *Artificial dichotomization and small-sample bias*, correcting for dichotomization and small-study bias occurs before correcting for measurement error and range restriction (when present in the meta-analysis). These artifacts are labeled as `Initial artifacts: Dichotomization and small-sample bias` in the above output. The corrected correlations for dichotomization and small-study bias, ρ^*_{XY} defined in (3), are stored in the system variable `_meta_rxy`. Also, the adjusted sample sizes after these corrections, n^* defined in (4), are stored in the system variable `_meta_studysize`.

```
. list rho _meta_rxy n _meta_studysize in 1/5
```

	rho	_meta_rxy	n	_meta~ze
1.	.24	.43744055	29	8
2.	.16	.28992161	43	13
3.	.21	.3448679	16	6
4.	.37	.62477508	9	3
5.	.12	.21872028	29	9

In the presence of these "initial" artifacts, variables _meta_rxy and _meta_studysize (instead of variables rho and n) serve as the input for psychometric meta-analysis. For example, we will additionally correct for measurement error in X (via the xreliability(rxx) option).

```
. meta psycorr rho n, small xdich(.5) ydich(py) xreliability(rxx)
Psychometric meta-analysis setting information

 Study information
    No. of studies: 12
       Study label: Generic
        Study size: _meta_studysize
      Summary data: rho n

       Effect size
              Type: correlation
             Label: Corrected correlation
          Variable: _meta_es

         Precision
         Std. err.: _meta_se
                CI: [_meta_cil, _meta_ciu]
          CI level: 95%

  Model and method
             Model: Random effects
            Method: Individual-correction meta-analysis

Initial artifacts: Dichotomization and small-sample bias

 Reliability for X
            Values: rxx
```

Given that we are only correcting for measurement error in X (in measuring the true-score variable T), the corrected correlations are given by $\rho_{TY} = \rho^*_{XY}/\sqrt{r_{XX}}$. These values are stored in _meta_rty (and _meta_es, which is equal to _meta_rty in this case).

```
. list rho _meta_rxy rxx _meta_rty in 1/5
```

	rho	_meta_rxy	rxx	_meta_rty
1.	.24	.43744055	.61	.56008524
2.	.16	.28992161	.83	.31823031
3.	.21	.3448679	.65	.42775598
4.	.37	.62477508	.74	.72628628
5.	.12	.21872028	.71	.25957321

We can easily verify that _meta_rty = _meta_rxy/$\sqrt{\texttt{rxx}}$ as opposed to _meta_rty = rho/$\sqrt{\texttt{rxx}}$, which would have been the case had we not corrected for these initial artifacts (dichotomization and small-study bias).

◁

Stored results

meta psycorr stores the following characteristics and system variables:

Characteristics

_dta[_meta_marker]	"_metapsych_ds_1"
_dta[_meta_K]	number of studies in the meta-analysis
_dta[_meta_studylabel]	name of string variable containing study labels or Generic
_dta[_meta_estype]	type of effect size; correlation
_dta[_meta_eslabelopt]	eslabel(*eslab*), if specified

_dta[_meta_eslabel]	effect-size label from eslabel(); default is Corrected correlation
_dta[_meta_eslabeldb]	effect-size label for dialog box
_dta[_meta_esvardb]	_meta_es
_dta[_meta_level]	default confidence level for meta-analysis
_dta[_meta_modellabel]	Random effects
_dta[_meta_model]	random
_dta[_meta_methodlabel]	meta-analysis method label; varies by meta-analysis model
_dta[_meta_method]	meta-analysis method; varies by meta-analysis model
_dta[_meta_show]	nometashow, if specified
_dta[_meta_nvar]	name of sample-size variable
_dta[_meta_rvar]	variable containing observed correlations
_dta[_meta_datatype]	data type; correlation
_dta[_meta_datavars]	variables specified with meta psycorr
_dta[_meta_setcmdline]	meta psycorr command line
_dta[_meta_ifexp]	*if* specification
_dta[_meta_inexp]	*in* specification
_dta[_meta_relx]	reliability estimates for X
_dta[_meta_rely]	reliability estimates for Y
_dta[_meta_relx_type]	type of reliability estimates for X: restricted or unrestricted
_dta[_meta_rely_type]	type of reliability estimates for Y: restricted or unrestricted
_dta[_meta_relx_restvar]	indicator variable identifying restricted reliability estimates for X, if specified
_dta[_meta_rely_restvar]	indicator variable identifying restricted reliability estimates for Y, if specified
_dta[_meta_relx_unrestvar]	indicator variable identifying unrestricted reliability estimates for X, if specified
_dta[_meta_rely_unrestvar]	indicator variable identifying unrestricted reliability estimates for Y, if specified
_dta[_meta_uratiox]	u-ratio for X
_dta[_meta_uratioy]	u-ratio for Y
_dta[_meta_uratiox_type]	type of u-ratio for X: observed or true
_dta[_meta_uratioy_type]	type of u-ratio for Y: observed or true
_dta[_meta_uratiox_obsvar]	indicator variable identifying observed u-ratio values for X, if specified
_dta[_meta_uratioy_obsvar]	indicator variable identifying observed u-ratio values for Y, if specified
_dta[_meta_uratiox_truevar]	indicator variable identifying true u-ratio values for X (u_T), if specified
_dta[_meta_uratioy_truevar]	indicator variable identifying true u-ratio values for Y (u_P), if specified
_dta[_meta_rrtype]	range restriction type: indirect or direct
_dta[_meta_rrtype_desc]	range restriction description: one of uvirr_x, uvirr_y, uvdrr_x, uvdrr_y, bvdrr, or bvirr
_dta[_meta_impute]	imputation method for missing artifact information: bootstrap, perfect, wmean, mean, or none
_dta[_meta_rseed]	random-number seed for bootstrap imputation method
_dta[_meta_relx_nmiss]	number of missing reliability estimates for X
_dta[_meta_rely_nmiss]	number of missing reliability estimates for Y
_dta[_meta_ux_nmiss]	number of missing u-ratio values for X
_dta[_meta_uy_nmiss]	number of missing u-ratio values for Y
_dta[_meta_signrxzvar]	variable containing signs of ρ_{XZ_u} used in indirect bivariate range restriction
_dta[_meta_signryzvar]	variable containing signs of ρ_{YZ_u} used in indirect bivariate range restriction
_dta[_meta_xdich]	specified proportions of successes (or failures) after dichotomization of X
_dta[_meta_ydich]	specified proportions of successes (or failures) after dichotomization of Y
_dta[_meta_small]	1, if small is specified

System variables

_meta_id	study ID variable
_meta_es	variable containing corrected correlations
_meta_se	variable containing standard errors for corrected correlations
_meta_cil	variable containing lower bounds of CIs for corrected correlations
_meta_ciu	variable containing upper bounds of CIs for corrected correlations
_meta_studylabel	string variable containing study labels
_meta_studysize	variable containing total sample size per study
_meta_var	sampling variances for observed correlations
_meta_A	compound attenuation (due to measurement error and range restriction) factor
_meta_a	refining factor used in computing _meta_se

Models with range restriction	
_meta_rtpu	T–P correlation, unrestricted, corrected for measurement error (ME) and range restriction (RR); $\rho_{TP,u}$; same as _meta_es
_meta_rtpr	T–P correlation, restricted, corrected for ME; $\rho_{TP,r}$
_meta_rtyr	T–Y correlation, restricted, corrected for ME in X; $\rho_{TY,r}$
_meta_rtyu	T–Y correlation, unrestricted, corrected for ME in X and RR; $\rho_{TY,u}$
_meta_rxpr	X–P correlation, restricted, corrected for ME in Y; $\rho_{XP,r}$
_meta_rxpu	X–P correlation, unrestricted, corrected for ME in Y and RR; $\rho_{XP,u}$
_meta_rxyu	X–Y correlation, unrestricted, corrected for RR; $\rho_{XY,u}$
_meta_rxyr	X–Y correlation, restricted, uncorrected; $\rho_{XY,r}$
_meta_rxxr	reliability of X in the restricted sample; $r_{XX,r}$
_meta_rxxu	reliability of X in the unrestricted sample; $r_{XX,u}$
_meta_ryyr	reliability of Y in the restricted sample; $r_{YY,r}$
_meta_ryyu	reliability of Y in the unrestricted sample; $r_{YY,u}$
_meta_ux	standard deviations (SD) ratio of X (restricted/unrestricted); $S_{X,r}/S_{X,u}$
_meta_ut	SD ratio of T (restricted/unrestricted); $S_{T,r}/S_{T,u}$
_meta_uy	SD ratio of Y (restricted/unrestricted); $S_{Y,r}/S_{Y,u}$
_meta_up	SD ratio of P (restricted/unrestricted); $S_{P,r}/S_{P,u}$
Models without range restriction	
_meta_rxx	reliability of X; r_{XX}
_meta_ryy	reliability of Y; r_{YY}
_meta_rtp	T–P correlation, corrected for ME; ρ_{TP}; same as _meta_es
_meta_rty	T–Y correlation, corrected for ME in X; ρ_{TY}
_meta_rxp	X–P correlation, corrected for ME in Y; ρ_{XP}
_meta_rxy	X–Y correlation, uncorrected; ρ_{XY}

Methods and formulas

Methods and formulas are presented under the following headings:

See *Remarks and examples* for the model setup and definitions of the relevant statistical concepts.

Estimating the correlation between T and P

In the following discussion, the subscripts r and u represent the restricted and unrestricted samples, respectively. In personnel selection research, these samples are often referred to as the incumbent and applicant samples. For notational simplicity, we omit the hat symbol typically used for estimated quantities, writing $r_{XX,r}$, $r_{YY,r}$, etc., instead of $\widehat{r_{XX,r}}$, $\widehat{r_{YY,r}}$, etc., because all quantities in the formulas below are estimates. Additionally, we omit the subscript j indicating that the quantities in the formulas below are specific to the jth study.

The terms u_X and u_Y represent the observed-score u-ratios of X and Y (the ratio of standard deviation in the restricted sample to that in the unrestricted sample), whereas u_T and u_P denote the true-score u-ratios of T and P. Using the observed correlation between X and Y in the restricted sample, $\rho_{XY,r}$, and study sample sizes, n, (specified immediately after `meta psycorr`), along with other artifact information, we aim to estimate the correlation between the perfectly measured variables T and P (the true scores) in the unrestricted population, $\rho_{TP,u}$.

Notice that $\rho_{XY,r}$ and n are replaced by $\rho^*_{XY,r}$ and n^* defined in (3) and (4) if correction for dichotomization or small-sample bias was specified in the model via the `xdich()`, `ydich()`, or `small` option.

Correcting correlations for measurement errors only

When no range restriction is present, there is no distinction between restricted and unrestricted samples, so the subscripts r and u are not needed. The correction for measurement error is given as follows (Schmidt and Hunter 2015, 112):

$$\rho_{TP} = \frac{\rho_{XY}}{\sqrt{r_{XX} r_{YY}}} \tag{5}$$

where r_{XX} and r_{YY} are the reliability estimates of X and Y, respectively. These values are specified in the `xreliability()` and `yreliability()` options, respectively. The values of ρ_{TP} are stored in the system variables `_meta_rtp` and `_meta_es`. You may correct for measurement error in the independent or the dependent variable alone as follows:

$$\rho_{TY} = \frac{\rho_{XY}}{\sqrt{r_{XX}}}, \qquad \rho_{XP} = \frac{\rho_{XY}}{\sqrt{r_{YY}}}$$

These values are stored in the system variables `_meta_rty` and `_meta_rxp`, respectively. Equation (5) can be expressed in terms of ρ_{TY} and ρ_{XP} as follows:

$$\rho_{TP} = \frac{\rho_{TY}}{\sqrt{r_{YY}}}, \qquad \rho_{TP} = \frac{\rho_{XP}}{\sqrt{r_{XX}}}$$

When range restriction is present, the above quantities will be defined for both the restricted and the unrestricted samples. Subscripts r and u will be used to distinguish between the two samples. For example, for the restricted sample, we have $\rho_{TP,r} = \rho_{XY,r}/\sqrt{r_{XX,r} r_{YY,r}}$, and for the unrestricted sample, we have $\rho_{TP,u} = \rho_{XP,u}/\sqrt{r_{XX,u}}$.

Correcting correlations for univariate range restriction only

As mentioned earlier in *Measurement error* of *Remarks and examples*, measurement errors are always present in psychometric applications. However, isolating the effect of correcting for range restriction only is useful as a foundation for the subsequent section that addresses both measurement error and range restriction. The correlation between variables X and Y in the unrestricted sample, $\rho_{XY,u}$, can be calculated using Thorndike's Case II formula (Pearson 1903, eq. 51; Thorndike 1949):

$$\rho^{\text{direct}}_{XY,u} = h(u_X, \rho_{XY,r}) = \frac{\rho_{XY,r}}{u_X \sqrt{\left(\frac{1}{u_X^2} - 1\right) \rho^2_{XY,r} + 1}} \tag{6}$$

The above formula assumes that range restriction is direct on the independent variable, X. When range restriction is indirect (that is, when selection is based on a third variable that correlates with X), we use the true-score u-ratio, u_T, instead of the observed-score u-ratio, u_X, to make range correction; otherwise, the two formulas are functionally identical (Schmidt and Hunter 2015, 126):

$$\rho_{XY,u}^{\text{indirect}} = h(u_T, \rho_{XY,r}) = \frac{\rho_{XY,r}}{u_T\sqrt{\left(\frac{1}{u_T^2}-1\right)\rho_{XY,r}^2+1}} \tag{7}$$

Here, u_T is specified using the `xuratios(`*varname*`, true)` option. Recall that the above formulas are appropriate under the full-mediation assumption discussed in item 3 in *Range restriction*.

The above formulas assume that the values of X are restricted either by direct selection on X or by indirectly selecting values based on a third variable that correlates with X. When the range restriction is assumed direct in Y or indirect through a third variable that correlates with Y, similar formulas can be derived with u_X replaced by u_Y and u_T replaced by u_P.

Correcting correlations for measurement errors and univariate range restriction

Below, we present formulas for correlations corrected for univariate range restriction in either X or Y under various scenarios, accounting for measurement error in both variables. Range restriction is univariate (as opposed to bivariate) when either u_X (or u_T) or u_Y (or u_P) is known, but not both. Additionally, the full-mediation assumption must hold for univariate indirect range restriction. When both u_X and u_Y are known, we can correct for bivariate range restriction without requiring the full-mediation assumption; see *Correcting correlations for measurement errors and bivariate range restriction*. The values of $\rho_{TP,u}$ are stored in the system variables `_meta_rtpu` and `_meta_es`.

In what follows, it is useful to think of the function $h(\cdot)$, defined in (6) and (7), as a transformation that takes a correlation (for example, $\rho_{XY,r}$, $\rho_{XP,r}$, $\rho_{TP,r}$) from the restricted sample as input and uses u_X (for direct range restriction) or u_T (for indirect range restriction) to compute the corresponding correlation (for example, $\rho_{XY,u}$, $\rho_{XP,u}$, $\rho_{TP,u}$) in the unrestricted sample.

The correction for univariate direct range restriction in X is provided by Hunter, Schmidt, and Le (2006, eq. 17):

$$\begin{aligned}\rho_{TP,u}^{\text{uvdrr}_x} &= \frac{\rho_{XP,u}}{\sqrt{r_{XX,u}}} = h\left(u_X, \rho_{XP,r}\right)/\sqrt{r_{XX,u}}\\ &= \left[\frac{\rho_{XY,r}}{u_X\sqrt{r_{YY,r}}\left\{\left(\frac{1}{u_X^2}-1\right)\frac{\rho_{XY,r}^2}{r_{YY,r}}+1\right\}^{1/2}}\right]/\sqrt{r_{XX,u}}\end{aligned}$$

The above equation reduces to (5) when no range restriction is present. In that case, we have $u_X = 1$, $r_{XX,r} = r_{XX,u} = r_{XX}$, and $r_{YY,r} = r_{YY,u} = r_{YY}$ because there is no distinction between restricted and unrestricted samples. It also reduces to (6) when there is no measurement error in the data ($r_{XX,r} = r_{XX,u} = r_{YY,r} = r_{YY,u} = 1$).

The correction for univariate indirect range restriction through the effect on X of selecting on a third latent variable Z is as follows (Hunter, Schmidt, and Le 2006, eq. 29):

$$\begin{aligned}\rho_{TP,u}^{\text{uvirr}_x} &= h\left(u_T, \rho_{TP,r}\right) \\ &= \frac{\rho_{TP,r}}{u_T\left\{\left(\frac{1}{u_T^2}-1\right)\rho_{TP,r}^2+1\right\}^{1/2}} \\ &= \frac{\rho_{XY,r}}{u_T\sqrt{r_{XX,r}r_{YY,r}}\left\{\left(\frac{1}{u_T^2}-1\right)\frac{\rho_{XY,r}^2}{r_{XX,r}r_{YY,r}}+1\right\}^{1/2}}\end{aligned} \tag{8}$$

The authors referred to the above formula as Case IV [Thorndike (1949) had already presented Cases I, II, and III, of which only II was used in practice for univariate direct range restriction; see (6) for details]. The above equation reduces to (7) when there is no measurement error in the data ($r_{XX,r} = r_{XX,u} = r_{YY,r} = r_{YY,u} = 1$).

The correction for univariate indirect range restriction through the effect on Y of selecting on a third latent variable Z is as follows:

$$\begin{aligned}\rho_{TP,u}^{\text{uvirr}_y} &= h\left(u_P, \rho_{TP,r}\right) \\ &= \frac{\rho_{TP,r}}{u_P\left\{\left(\frac{1}{u_P^2}-1\right)\rho_{TP,r}^2+1\right\}^{1/2}} \\ &= \frac{\rho_{XY,r}}{u_P\sqrt{r_{XX,r}r_{YY,r}}\left\{\left(\frac{1}{u_P^2}-1\right)\frac{\rho_{XY,r}^2}{r_{XX,r}r_{YY,r}}+1\right\}^{1/2}}\end{aligned}$$

The correction for univariate direct range restriction in Y is as follows:

$$\begin{aligned}\rho_{TP,u}^{\text{uvdrr}_y} &= \frac{\rho_{TY,u}}{\sqrt{r_{YY,u}}} = h\left(u_Y, \rho_{TY,r}\right)/\sqrt{r_{YY,u}} \\ &= \left[\frac{\rho_{XY,r}}{u_Y\sqrt{r_{XX,r}}\left\{\left(\frac{1}{u_Y^2}-1\right)\frac{\rho_{XY,r}^2}{r_{XX,r}}+1\right\}^{1/2}}\right]/\sqrt{r_{YY,u}}\end{aligned}$$

Correcting correlations for measurement errors and bivariate range restriction

When the mediation assumption in univariate range restriction is violated, correcting for indirect range restriction is still possible without requiring that assumption if values of both u_X and u_Y are available. Bryant and Gokhale (1972) and Alexander et al. (1987) derived methods to correct for bivariate indirect and direct range restriction, respectively. These methods require that both u-ratio variables for X and Y, u_X and u_Y, be available. The correction for bivariate direct range restriction is as follows (Alexander et al. 1987):

$$\rho_{TP,u}^{\text{bvdrr}} = \left\{\frac{\rho_{XY,r}^2-1}{2\rho_{XY,r}}u_X u_Y + \text{sgn}(\rho_{XY,r})\sqrt{\frac{\left(1-\rho_{XY,r}^2\right)^2}{4\rho_{XY,r}^2}u_X^2 u_Y^2+1}\right\}/\sqrt{r_{XX,u}r_{YY,u}}$$

where $\text{sgn}(\cdot)$ is the sign operator.

The correction for bivariate indirect range restriction is as follows (Bryant and Gokhale 1972; Alexander 1990; Le et al. 2016 ; Dahlke and Wiernik 2020):

$$\rho^{\text{bvirr}}_{TP,u} = \frac{\rho_{XY,r} u_X u_Y + \lambda \sqrt{\left|1 - u_X^2\right| \left|1 - u_Y^2\right|}}{\sqrt{r_{XX,u} r_{YY,u}}} \tag{9}$$

where the λ value is computed as

$$\lambda = \text{sgn}\left\{\rho_{XZ_u} \rho_{YZ_u} (1 - u_X)(1 - u_Y)\right\} \times \frac{\text{sgn}(1 - u_X) \min\left(u_X, \frac{1}{u_X}\right) + \text{sgn}(1 - u_Y) \min\left(u_Y, \frac{1}{u_Y}\right)}{\min\left(u_X, \frac{1}{u_X}\right) \min\left(u_Y, \frac{1}{u_Y}\right)}$$

ρ_{XZ_u} and ρ_{YZ_u} are the correlations of variables X and Y with Z, respectively, where Z is the third variable on which selection occurred. The signs of ρ_{XZ_u} and ρ_{YZ_u} may be specified using the `signrxz()` and `signryz()` options. Le et al. (2016) referred to the above formula as Case V.

Dahlke and Wiernik (2020) derived the formula for λ to allow for mixed patterns of range restriction ($u_X = S_{X,r}/S_{X,u} < 1$) and range enhancement ($u_X = S_{X,r}/S_{X,u} > 1$). However, because we are exclusively dealing with range restriction, which is the most common scenario in psychometric applications, λ simplifies to 1 when ρ_{XZ_u} and ρ_{YZ_u} have the same sign and to -1 when they have opposite signs. Also, the absolute values in the formula for $\rho^{\text{bvirr}}_{TP,u}$ are unnecessary in the case of range restriction because both u_X and u_Y are less than 1.

Estimating the values of artifacts for univariate range restriction

The formulas below are used to estimate artifact values from other related artifacts. These formulas enable the adjustment of reliability estimates for range restriction (that is, obtaining reliability estimates in both restricted and unrestricted samples) and facilitate the conversion of u-ratios between observed-score and true-score frameworks. All formulas are presented for the case of univariate range restriction (when only u_X is known), where selection occurs either directly on the independent variable X (uvdrr$_x$ case) or indirectly through a third variable that is correlated with X (uvirr$_x$ case, where the full-mediation assumption holds). Formulas for the case of univariate range restriction, where selection occurs on Y (directly or indirectly), can be obtained by replacing X with Y and T with P in all expressions below.

The formulas to estimate the observed-score u-ratio u_X from other artifacts, such as true-score u-ratio u_T and one of $r_{XX,u}$ or $r_{XX,r}$, are given by Hunter, Schmidt, and Le (2006, eqs. 19, 19a):

$$u_X = \sqrt{r_{XX,u} u_T^2 - r_{XX,u} + 1}$$

$$u_X = \sqrt{\frac{u_T^2}{r_{XX,r} + u_T^2 - r_{XX,r} u_T^2}}$$

These values are stored in the system variable `_meta_ux`.

You can use the above equations to express u_T in terms of u_X and one of $r_{XX,u}$ or $r_{XX,r}$ as follows:

$$u_T = \sqrt{\frac{u_X^2 - 1 + r_{XX,u}}{r_{XX,u}}}$$

$$u_T = \sqrt{\frac{r_{XX,r} u_X^2}{1 + r_{XX,r} u_X^2 - u_X^2}}$$

These values are stored in the system variable `_meta_ut`.

Below, we present formulas for estimating $r_{XX,r}$, $r_{XX,u}$, $r_{YY,r}$, and $r_{YY,u}$ when range restriction is either direct or indirect in X. The case where range restriction is in Y can be obtained by swapping X and Y in the formulas below and replacing T with P.

When range restriction is indirect, we can estimate the reliability of X in the unrestricted sample, $r_{XX,u}$, from the reliability of X in the restricted sample, $r_{XX,r}$, and one of either the true-score u-ratio, u_T, or the observed-score u-ratio, u_X. This is possible by expressing $r_{XX,u}$ as a correlation according to (1) and then by using the $h(\cdot)$ function as follows:

$$r_{XX,u} = \rho^2_{XT,u} = h^2\left(u_T, \rho_{XT,r}\right) = \frac{r_{XX,r}}{r_{XX,r} + u_T^2 - r_{XX,r}u_T^2}$$

$$r_{XX,u} = 1 - u_X^2\left(1 - r_{XX,r}\right)$$

For direct range restriction, we have

$$r_{XX,u} = \rho^2_{XT,u} = h^2\left(u_X, \rho_{XT,r}\right) = \frac{r_{XX,r}}{r_{XX,r} + u_X^2 - r_{XX,r}u_X^2}$$

These values are stored in the system variable _meta_rxxu.

We can solve the above equations for $r_{XX,r}$ and obtain expressions to compute $r_{XX,r}$ from $r_{XX,u}$ and one of u_X or u_T. For indirect range restriction, we have the following (Hunter, Schmidt, and Le 2006, eqs. 26, 27):

$$r_{XX,r} = \frac{r_{XX,u}u_T^2}{r_{XX,u}u_T^2 - r_{XX,u} + 1}$$

$$r_{XX,r} = 1 - \frac{1 - r_{XX,u}}{u_X^2}$$

For direct range restriction, we have

$$r_{XX,r} = \frac{r_{XX,u}u_X^2}{r_{XX,u}u_X^2 - r_{XX,u} + 1}$$

These values are stored in the system variable _meta_rxxr.

For direct range restriction (based on X), we can estimate the reliability of the dependent variable Y in the unrestricted sample, $r_{YY,u}$, without the usual requirement of knowing the values of the u-ratio variable u_Y. This can be done as follows (Hunter, Schmidt, and Le 2006, eq. 13):

$$r_{YY,u} = 1 - \frac{1 - r_{YY,r}}{1 - \rho^2_{XY,r}\left(1 - \frac{1}{u_X^2}\right)} \tag{10}$$

Similarly, when range restriction is indirect (that is, when selection is based on a third variable that is correlated with X), we estimate

$$r_{YY,u} = 1 - \frac{1 - r_{YY,r}}{1 - \rho^2_{XY,r}\left(1 - \frac{1}{u_T^2}\right)}$$

These values are stored in the system variable _meta_ryyu.

The following formulas are obtained by solving the above two equations for $r_{YY,r}$. When the range restriction is direct in X, we have

$$r_{YY,r} = 1 - (1 - r_{YY,u}) \left\{ 1 - \rho^2_{XY,r} \left(1 - \frac{1}{u^2_X} \right) \right\}$$

When the range restriction is indirect (that is, when selection is based on a third variable that is correlated with X):

$$r_{YY,r} = 1 - (1 - r_{YY,u}) \left\{ 1 - \rho^2_{XY,r} \left(1 - \frac{1}{u^2_T} \right) \right\}$$

These values are stored in the system variable _meta_ryyr.

Estimating the values of artifacts for bivariate range restriction

For bivariate range restriction on both X and Y, the artifacts of interest that determine $\rho^{\text{bvdrr}}_{TP,u}$ and $\rho^{\text{bvirr}}_{TP,u}$ are u_X, u_Y, $r_{XX,u}$, and $r_{YY,u}$. If any of these quantities is not directly available in your dataset, you may use the following formulas to compute them.

$$u_X = \sqrt{r_{XX,u} u^2_T - r_{XX,u} + 1}$$

$$u_X = \sqrt{\frac{u^2_T}{r_{XX,r} + u^2_T - r_{XX,r} u^2_T}}$$

$$u_Y = \sqrt{r_{YY,u} u^2_P - r_{YY,u} + 1}$$

$$u_Y = \sqrt{\frac{u^2_P}{r_{YY,r} + u^2_P - r_{YY,r} u^2_P}}$$

For bivariate indirect range restriction, we have

$$r_{XX,u} = \rho^2_{XT,u} = h^2\left(u_T, \rho_{XT,r}\right) = \frac{r_{XX,r}}{r_{XX,r} + u^2_T - r_{XX,r} u^2_T}$$

$$r_{XX,u} = 1 - u^2_X \left(1 - r_{XX,r}\right)$$

Here, because the values of u_Y are known, we do not need to estimate $r_{YY,u}$ as in (10), but we express it as a correlation, $\rho_{YP,u}$, according to (2) and by using the $h(\cdot)$ function as follows:

$$r_{YY,u} = \rho^2_{YP,u} = h^2\left(u_P, \rho_{YP,r}\right) = \frac{r_{YY,r}}{r_{YY,r} + u^2_P - r_{YY,r} u^2_P}$$

$$r_{YY,u} = 1 - u^2_Y \left(1 - r_{YY,r}\right)$$

For bivariate direct range restriction, we have

$$r_{XX,u} = \rho^2_{XT,u} = h^2\left(u_X, \rho_{XT,r}\right) = \frac{r_{XX,r}}{r_{XX,r} + u_X^2 - r_{XX,r}u_X^2}$$

$$r_{YY,u} = \rho^2_{YP,u} = h^2\left(u_Y, \rho_{YP,r}\right) = \frac{r_{YY,r}}{r_{YY,r} + u_Y^2 - r_{YY,r}u_Y^2}$$

Sampling variances of the corrected correlations $\rho_{TP,u}$

The sampling variances for the observed correlations, $\rho_{XY,r}$, are given by Schmidt and Hunter (2015, 144):

$$\widehat{\text{Var}}\left(\rho_{XY,r}\right) = v = \frac{\left(1 - \overline{\rho}^2_{XY_r}\right)^2}{n-1}$$

where $\overline{\rho}_{XY_r}$ is the weighted average of the study-specific $\rho_{XY,r}$ values with weights equal to the studies' sample sizes. These values are stored in the system variable `_meta_var`. Schmidt and Hunter (2015, 237) discuss the usage of $\overline{\rho}_{XY_r}$ instead of the $\rho_{XY,r}$ in the above formula, because the latter tend to underestimate the sampling error, which leads to overestimating the between-study heterogeneity parameter, τ^2.

Except for bivariate indirect range restriction, sampling variances of the corrected correlations are computed as

$$\widehat{\text{Var}}\left(\rho_{TP,u}\right) = a^2 \times \frac{v}{A^2} \tag{11}$$

where A is the compound artifact (measurement error and range restriction) attenuation factor and a is a refining factor defined in Schmidt and Hunter (2015, 144–145) as

$$A = \rho_{XY,r}/\rho_{TP,u} \qquad \text{and} \qquad a = \sqrt{\left(\frac{1}{u_\dagger^2} - 1\right)\rho^2_{XY,r} + 1}$$

where $u_\dagger$ is one of u_X, u_T, u_Y, or u_P depending on which type of univariate range restriction is specified. For bivariate range restriction, no refining factor is used to correct the sampling variances; therefore, $a = 1$ in this case. The values of A and a for each study are stored in the system variables `_meta_A` and `_meta_a`.

Additionally, when correlations are corrected only for measurement error, the sampling variance is computed using (11), but with $a = 1$ and $A = \sqrt{r_{XX}} \times \sqrt{r_{YY}}$.

For the case of bivariate indirect range restriction, Dahlke and Wiernik (2020) derived a linear approximation of the sampling variances of $\rho_{TP,u}$ that is more accurate than (11). Recall the bivariate indirect range restriction formula introduced in *Correcting correlations for measurement errors and bivariate range restriction*:

$$\rho_{TP,u} = \frac{\rho_{XY,r}u_X u_Y + \lambda\sqrt{\left|1 - u_X^2\right|\left|1 - u_Y^2\right|}}{q_{X,u}q_{Y,u}}$$

where $q_{X,u} = \sqrt{r_{XX,u}}$ and $q_{Y,u} = \sqrt{r_{YY,u}}$. The Delta method with a working independence assumption of $q_{X,u}$, $q_{Y,u}$, u_X, u_Y, and $\rho_{XY,r}$ implies the following linear approximation of the sampling variance of $\rho_{TP,u}$:

$$\widehat{\text{Var}}\left(\rho_{TP,u}\right) = b_1^2\widehat{\text{Var}}\left(q_{X,u}\right) + b_2^2\widehat{\text{Var}}\left(q_{Y,u}\right) + b_3^2\widehat{\text{Var}}\left(u_X\right) + b_4^2\widehat{\text{Var}}\left(u_Y\right) + b_5^2\widehat{\text{Var}}\left(\rho_{XY,r}\right) \tag{12}$$

where b_1, b_2, b_3, b_4, and b_5 are the first-order partial derivatives of the corrected (disattenuated) correlation, $\rho_{TP,u}$, with respect to $q_{X,u}$, $q_{Y,u}$, u_X, u_Y, and $\rho_{XY,r}$, respectively. These partial derivatives are given as follows:

$$b_1 = \frac{\partial \rho_{TP,u}}{\partial q_{X,u}} = -\frac{\rho_{TP,u}}{q_{X,u}}$$

$$b_2 = \frac{\partial \rho_{TP,u}}{\partial q_{Y,u}} = -\frac{\rho_{TP,u}}{q_{Y,u}}$$

$$b_3 = \frac{\partial \rho_{TP,u}}{\partial u_X} = \left\{ \rho_{XY,r} u_Y - \frac{\lambda u_X (1 - u_X^2) \sqrt{|1 - u_Y^2|}}{|1 - u_X^2|^{1.5}} \right\} \Big/ (q_{X,u} q_{Y,u})$$

$$b_4 = \frac{\partial \rho_{TP,u}}{\partial u_Y} = \left\{ \rho_{XY,r} u_X - \frac{\lambda u_Y (1 - u_Y^2) \sqrt{|1 - u_X^2|}}{|1 - u_Y^2|^{1.5}} \right\} \Big/ (q_{X,u} q_{Y,u})$$

$$b_5 = \frac{\partial \rho_{TP,u}}{\partial \rho_{XY,r}} = \frac{u_X u_Y}{q_{X,u} q_{Y,u}}$$

Expressions for $\widehat{\text{Var}}\left(q_{X,u}\right)$, $\widehat{\text{Var}}\left(q_{Y,u}\right)$, $\widehat{\text{Var}}\left(u_X\right)$, $\widehat{\text{Var}}\left(u_Y\right)$, and $\widehat{\text{Var}}\left(\rho_{XY,r}\right)$ are provided in Appendix C of the supplemental material in Dahlke and Wiernik (2020). When reliability estimates from the unrestricted sample are specified (for example, via `xreliability(rxxu, unrestricted)`), the sample size n_u of the unrestricted sample is required to compute $\widehat{\text{Var}}\left(q_{X,u}\right)$ and $\widehat{\text{Var}}\left(q_{Y,u}\right)$. The sample size n_u can be specified via the `nu()` option.

If reliability estimates from the restricted sample are specified within `xreliability()` and `yreliability()`, the `nu()` option is not required. However, specifying n_u if it is available in your dataset will yield slightly more accurate estimates of the sampling variances in (12). This improvement occurs because omitting the `nu()` option forces the computation of (12) to assume that the standard deviations of X and Y in the unrestricted sample are estimated without error (that is, based on an infinite sample size, n_u). For instance, $\widehat{\text{Var}}\left(u_X\right)$ is given by

$$\widehat{\text{Var}}\left(u_X\right) \approx \frac{1}{2} u_X^2 \left(\frac{1}{n+1} + \frac{1}{n_u + 1} \right)$$

and an implication of the simplifying assumption of an infinite sample size n_u would be to approximate $\widehat{\text{Var}}\left(u_X\right)$ by

$$\widehat{\text{Var}}\left(u_X\right) \approx \frac{1}{2} u_X^2 \left(\frac{1}{n+1} \right)$$

The sampling standard errors of the corrected correlations, $\sqrt{\widehat{\text{Var}}\left(\rho_{TP,u}\right)}$, defined in (11) and (12), are stored in the system variable `_meta_se`.

Confidence intervals for effect sizes

For the jth study in a given meta-analysis, let $\widehat{\theta}_j$ be one of the $\rho_{TP,u}$'s described above corresponding to the jth study (or ρ_{TP} when no range restriction is present). Then the asymptotic $100(1-\alpha)\%$ confidence interval computed by `meta psycorr` is

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

where $z_{1-\alpha/2}$ is a critical value from the standard normal distribution. The lower and upper bounds of the study CIs are stored in system variables `_meta_cil` and `_meta_ciu`, respectively.

References

Alexander, R. A. 1990. Correction formulas for correlations restricted by selection on an unmeasured variable. *Journal of Educational Measurement* 27: 187–189. https://doi.org/10.1111/j.1745-3984.1990.tb00742.x.

Alexander, R. A., K. P. Carson, G. M. Alliger, and L. Carr. 1987. Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. *Journal of Educational Measurement* 47: 309–315. https://doi.org/10.1177/0013164487472002.

Beatty, A. S., C. L. Barratt, C. M. Berry, and P. R. Sackett. 2014. Testing the generalizability of indirect range restriction corrections. *Journal of Applied Psychology* 99: 587–598. https://doi.org/10.1037/a0036361.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2021. *Introduction to Meta-Analysis.* 2nd ed. Chichester, UK: Wiley.

Bryant, N. D., and S. Gokhale. 1972. Correcting correlations for restrictions in range due to selection on an unmeasured variable. *Educational and Psychological Measurement* 32: 305–310. https://doi.org/10.1177/001316447203200207.

Dahlke, J. A., and B. M. Wiernik. 2020. Not restricted to selection research: Accounting for indirect range restriction in organizational research. *Organizational Research Methods* 23: 717–749. https://doi.org/10.1177/1094428119859398.

Fife, D. A., J. L. Mendoza, and R. Terry. 2013. Revisiting Case IV: A reassessment of bias and standard errors of Case IV under range restriction. *British Journal of Mathematical and Statistical Psychology* 66: 521–542. https://doi.org/10.1111/j.2044-8317.2012.02060.x.

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

Hunter, J. E., F. L. Schmidt, and H. Le. 2006. Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology* 91: 594–612. https://doi.org/10.1037/0021-9010.91.3.594.

Le, H., I.-S. Oh, F. L. Schmidt, and C. D. Wooldridge. 2016. Correction for range restriction in meta-analysis revisited: Improvements and implications for organizational research. *Personnel Psychology* 69: 975–1008. https://doi.org/10.1111/peps.12122.

Lord, F. M., and M. R. Novick, eds. 1968. *Statistical Theories of Mental Test Scores.* Reading, MA: Addison–Wesley.

Pearson, K. 1903. Mathematical contributions to the theory of evolution—XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society,* A ser., 200: 321–330. https://doi.org/10.1098/rsta.1903.0001.

Sackett, P. R., and H. Yang. 2000. Correction for range restriction: An expanded typology. *Journal of Applied Psychology* 85: 112–118. https://doi.org/10.1037/0021-9010.85.1.112.

Schmidt, F. L., and J. E. Hunter. 2015. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* 3rd ed. Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781483398105.

Schmidt, F. L., H. Le, and I.-S. Oh. 2019. "Correcting for the distorting effects of study artifacts in meta-analysis and second order meta-analysis". In *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. Cooper, L. V. Hedges, and J. C. Valentine, 315–338. 3rd ed. New York: Russell Sage Foundation. https://doi.org/10.7758/9781610448864.18.

Thorndike, R. L. 1949. *Personnel Selection: Test and Measurement Techniques.* New York: Wiley.

Also see

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

[META] **meta set** — Declare meta-analysis data using generic effect sizes

[META] **meta esize** — Compute effect sizes and declare meta-analysis data

[META] **meta update** — Update, describe, and clear meta-analysis settings

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

[META] **Glossary**

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