15 Meta-Analysis

R. Heller · Date: 2026-05-11 · Last changed: 2026-10-30

Pairwise random- and fixed-effects meta-analysis, network meta-analysis, individual-patient-data meta-analysis, and PRISMA-aligned reporting. Heterogeneity is treated as the main result, not a nuisance.

This chapter contains **25 method pages** and **3 labs**. If you are not sure which method to read, return to Chapter 0 and follow the decision tree to the right node.

15.1 Method pages

Method	Source slug
Bivariate SROC Curves	`bivariate-sroc`
Converting Between Effect Sizes	`effect-size-conversion`
Cumulative Meta-Analysis	`cumulative-meta-analysis`
Egger's Test for Funnel Asymmetry	`egger-test`
Fixed-Effect vs. Random-Effects Meta-Analysis	`fixed-vs-random-effects`
Forest Plots in Meta-Analysis	`forest-plot-meta`
Funnel Plots	`funnel-plot-meta`
Individual Participant Data Meta-Analysis	`ipd-meta-analysis`
Leave-One-Out Meta-Analysis	`leave-one-out`
Meta-Analysis of Diagnostic Accuracy	`diagnostic-test-meta`
Meta-Regression	`meta-regression`
Network Meta-Analysis: Introduction	`network-meta-analysis-intro`
NMA Consistency and Inconsistency	`nma-consistency`
NMA League Tables	`nma-league-tables`
Pooling Correlation Coefficients	`correlation-pooling`
Pooling Log-Odds Ratios	`log-odds-ratio-pooling`
Pooling Risk Ratios	`risk-ratio-pooling`
Prediction Intervals in Meta-Analysis	`prediction-interval-meta`
Quantifying Heterogeneity: Q and I^2	`heterogeneity-q-i2`
Selection Models for Publication Bias	`selection-models-bias`
Standardised Mean Difference: Hedges' g	`smd-hedges-g`
Subgroup Meta-Analysis	`subgroup-meta-analysis`
SUCRA Rankings	`nma-sucra`
Tau-Squared Estimation	`tau-squared-estimation`
Trim-and-Fill	`trim-and-fill`

15.2 Labs

Lab
Systematic reviews and PRISMA
Meta-analysis basics
Network meta-analysis

15.3 Bivariate SROC Curves

15.3.1 Introduction

The summary ROC (SROC) curve visualises the relationship between pooled sensitivity and specificity across a meta-analysis of diagnostic studies. Modern SROCs are derived from bivariate or HSROC models that properly account for the correlation between sensitivity and specificity as thresholds vary.

15.3.2 Prerequisites

Sensitivity and specificity; bivariate meta-analysis.

15.3.3 Theory

The bivariate model fits

$$\begin{pmatrix} \operatorname{logit}(\mathrm{Sens}_i) \\ \operatorname{logit}(1-\mathrm{Spec}_i) \end{pmatrix} \sim N(\mu, \Sigma).$$

The SROC curve is derived by mapping a range of the logit(1 - Spec) axis back to expected logit(Sens), given the bivariate mean and covariance. The curve shows how sensitivity trades against specificity as threshold varies.

15.3.4 Assumptions

Logit sensitivity and specificity are bivariate normal; studies vary by threshold; between-study heterogeneity captured by the covariance matrix.

15.3.5 R Implementation

```r
library(mada)

set.seed(2026)
df ← data.frame(
  TP = c(45, 60, 32, 80, 28, 55, 70, 50),
  FN = c( 5, 10,  8, 20, 12,  5,  8, 15),
  FP = c(10, 15,  8, 20, 10, 12, 18,  9),
  TN = c(140, 165, 152, 180, 150, 138, 200, 175)
)

fit ← reitsma(df)

# Summary statistics
summary(fit)$coefficients[c("tsens", "tfpr"), ]

# SROC plot
plot(fit, sroclty = 1, sroclwd = 2, predict = TRUE, predlty = 2,
     main = "Bivariate SROC with 95% confidence and prediction regions")
points(fpr(df), sens(df), cex = 0.8, pch = 19, col = "grey40")
```

15.3.6 Output & Results

Pooled sensitivity and FPR estimates; SROC curve with confidence (mean) and prediction (future-study) regions; individual study points overlaid.

15.3.7 Interpretation

“The SROC curve had AUC = 0.94, with pooled sensitivity 0.84 and specificity 0.91 at the summary operating point. Individual studies clustered tightly along the curve, consistent with threshold variation rather than diagnostic accuracy variation.”

15.3.8 Practical Tips

- Always display both confidence (mean) and prediction (future-study) regions on the SROC.
- Individual study points should lie close to the curve; large off-curve outliers indicate effect modifiers.
- Do not average sensitivity and specificity independently – their correlation matters.
- For partial-threshold studies (e.g., only low-Sens/high-Spec), restrict the SROC to the covered threshold range.
- Report per PRISMA-DTA with a clear operating-point recommendation.

15.3.9 For Reviewers

Use the shared reviewer checklist, together with this method’s assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript’s application of the method.

15.3.10 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.4 Pooling Correlation Coefficients

15.4.1 Introduction

Pearson correlation coefficients are bounded to $[-1, 1]$, and their sampling distributions are skewed with heteroscedastic variance — the variance depends on the underlying ρ and on n. These properties make raw correlations unsuitable for inverse-variance meta-analysis. Fisher's z transformation converts each correlation to an approximately Normal scale with variance independent of ρ, enabling standard random-effects pooling. After pooling on the z scale, the result is back-transformed to a correlation for interpretation.

15.4.2 Prerequisites

A working understanding of Pearson correlation, variance-stabilising transformations, and inverse-variance meta-analysis under random effects.

15.4.3 Theory

The Fisher z transformation is

$$z = \tfrac{1}{2}\log\left(\frac{1+r}{1-r}\right) = \tanh^{-1}(r), \qquad \mathrm{Var}(z) \approx \frac{1}{n-3}.$$

The variance is constant (depends only on n, not ρ), so the inverse-variance weighting in meta-analysis is well-defined. Pool z values across studies via fixed-effect or random-effects meta-analysis, then back-transform $r = \tanh(z)$ for reporting on the natural correlation scale.

The back-transformation preserves coverage at the original scale provided the transformed pooling is correctly weighted; reporting both z (with SE) and r (with CI) is standard.

15.4.4 Assumptions

Bivariate-Normal data in each study (the basis of the variance formula); independent studies; sufficient sample size per study ($n > 10$ as a rough minimum) for the asymptotic distribution to be reasonable.

15.4.5 R Implementation

```r
library(metafor)

set.seed(2026)
k <- 10
# Simulated correlations and sample sizes
df <- data.frame(
  ri = c(0.25, 0.32, 0.18, 0.40, 0.28, 0.35, 0.22, 0.30, 0.45, 0.38),
  ni = sample(50:300, k, replace = TRUE)
)

es <- escalc(measure = "ZCOR", ri = ri, ni = ni, data = df)
re <- rma(yi, vi, data = es, method = "REML")

# Convert pooled z back to correlation
pooled_z  <- re$b
pooled_r  <- tanh(pooled_z)
ci_r      <- tanh(c(re$ci.lb, re$ci.ub))

c(pooled_r = round(pooled_r, 3),
  lower = round(ci_r[1], 3), upper = round(ci_r[2], 3))
re$I2
```

15.4.6 Output & Results

`escalc(measure = "ZCOR")` computes Fisher z and its variance per study; `rma()` fits the random-effects model on the z scale. The pooled correlation is approximately 0.31 with a tight 95 % CI; I^2 quantifies heterogeneity.

15.4.7 Interpretation

A reporting sentence: "Random-effects meta-analysis of 10 studies (REML, n total = 1{,}523) gave a pooled correlation of $r = 0.31$ (95 % CI 0.25 to 0.37) with modest heterogeneity ($I^2 = 28\%$, $\tau^2 = 0.012$); pooling on the Fisher z scale and back-transforming maintains correct coverage despite the $[-1, 1]$ bound on r." Always pool on the z scale and report both Fisher z and back-transformed r.

15.4.8 Practical Tips

- Always pool on the Fisher z scale; pooling raw r biases estimates for high correlations and produces incorrect coverage.
- `metafor::escalc(measure = "ZCOR")` handles the Fisher z transformation and variance computation in a single call.
- Report both z (with SE) and back-transformed r (with CI); r is more interpretable for substantive readers.
- For partial correlations, apply the same approach with the degrees-of-freedom-adjusted sample size $n - p$ where p is the number of partialled-out variables.
- Hartung-Knapp small-sample adjustment (`rma(..., test = "knha")`) is especially important when studies are few or unevenly sized; standard inverse-variance gives over-confident intervals in those settings.

- For meta-analysis of intraclass correlations or other correlation-like statistics, the same Fisher z framework applies with appropriate variance formulae.

15.4.9 R Packages Used

`metafor` for `escalc()`, `rma()`, and the canonical correlation-pooling workflow; `meta` for an alternative interface with cleaner forest-plot output; `psychmeta` for psychometric meta-analysis with Hunter-Schmidt artefact corrections; `clubSandwich` for robust variance estimation in correlated-effect meta-analyses.

15.4.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.4.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.5 Cumulative Meta-Analysis

15.5.1 Introduction

Cumulative meta-analysis plots the running pooled estimate as studies are added one by one in chronological (or any pre-specified) order. Each successive pool incorporates one more study, producing a sequence of estimates that traces how the body of evidence evolved over time. The resulting forest plot reveals when the cumulative effect first crossed a meaningful threshold — clinical relevance, statistical significance, or stability — and whether later studies overturned, attenuated, or simply added precision to the early consensus. Cumulative meta-analysis is widely used in retrospective evidence syntheses to assess research efficiency: if the evidence had been pooled in real time, when would the answer have been clear, and how many subsequent confirmatory trials were arguably redundant?

15.5.2 Prerequisites

A working understanding of inverse-variance meta-analytic pooling, the sequential-analysis idea, and an awareness of why repeated looks at accumulating evidence raise multiplicity concerns when the goal is decision-making rather than description.

15.5.3 Theory

For an ordering of studies $1, \ldots, k$, the cumulative estimate at step t is the pooled effect computed from studies $1, \ldots, t$ under the chosen meta-analysis model. Plotting these $\hat{\theta}_{(1:t)}$ with their confidence intervals along the ordering axis (typically year) reveals stability and convergence: a horizontal band of overlapping CIs indicates the answer settled early, while a shifting estimate signals that later evidence materially changed conclusions.

The same procedure can use precision-based ordering (largest studies first) to assess sensitivity to the small-study effect, or risk-of-bias ordering (low-risk studies first) to assess whether pooled conclusions hinge on lower-quality evidence.

15.5.4 Assumptions

Study ordering is correctly recorded; studies are independent; the meta-analysis model used at each step is the same as for the full pool. Cumulative meta-analysis is descriptive, not inferential — formal sequential-testing methods such as trial-sequential analysis adjust for repeated looks; ordinary cumulative meta-analysis does not.

15.5.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 12
year ← sort(sample(1990:2023, k))
yi ← rnorm(k, 0.35, 0.2)
vi ← 1 / sample(50:250, k, replace = TRUE)

df ← data.frame(year, yi, vi)

res ← rma(yi, vi, data = df, method = "REML")
cum ← cumul(res, order = df$year)

plot(cum, main = "Cumulative meta-analysis by year",
     xlab = "Year")
```

15.5.6 Output & Results

`cumul()` returns a running-pooled estimate, confidence interval, and heterogeneity statistics at each step in the specified ordering. The associated `plot()` method renders a forest-style display where each row corresponds to "evidence after the t-th study", making convergence visually obvious. Many pooled effects stabilise after four to six well-powered studies; remaining studies typically tighten the CI without shifting the centre.

15.5.7 Interpretation

A reporting sentence: "Cumulative meta-analysis ordered by publication year showed that the pooled effect became statistically significant after the third study (1994) and remained essentially unchanged through 2023; the 18 trials published thereafter added precision but

did not alter the direction or magnitude of the conclusion, suggesting an opportunity for earlier evidence-based decision-making had a living synthesis been in place." Always describe both when the evidence stabilised and what later studies contributed.

15.5.8 Practical Tips

- Order primarily by publication date for the temporal interpretation; precision-based and risk-of-bias orderings are useful complementary sensitivity analyses.
- Cumulative meta-analysis is descriptive — for inferential stopping rules under repeated looks, use trial-sequential analysis (TSA) with appropriate alpha-spending.
- Convergence to stability after the first few studies often reveals research waste: many subsequent trials confirmed what was already known.
- Combine with meta-regression on year to formally test whether effect sizes drifted with time, e.g. as practice patterns or populations changed.
- Plot both point estimate and confidence interval; a shrinking CI without estimate movement is a different signal from a moving estimate.
- Use cumulative meta-analysis to motivate living-systematic-review infrastructure: continuous evidence updating prevents the "many years late" problem that retrospective cumulative plots so often expose.

15.5.9 R Packages Used

`metafor::cumul()` for the canonical cumulative pooling and its `plot()` method; `meta::metacum()` for an alternative interface aligned with the `meta` package's forest-plot conventions; `RTSA` and `gsDesign` for trial-sequential analysis when inferential adjustment for multiple looks is required; `metaviz` for richer cumulative-forest visualisations.

15.5.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.5.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.6 Meta-Analysis of Diagnostic Accuracy

15.6.1 Introduction

Meta-analysis of diagnostic test accuracy pools per-study sensitivity and specificity to summarise how well a test discriminates disease from non-disease across multiple primary studies. Unlike meta-analysis of treatment effects, the two paired outcomes are intrinsically correlated: sensitivity and specificity are determined jointly by the test threshold, so a study choosing a stricter cut-off will trade higher specificity for lower sensitivity. Naively pooling each metric separately ignores this trade-off and produces estimates that misrepresent the test's underlying performance. The modern approach is bivariate random-effects modelling (Reitsma et al., 2005) or hierarchical summary ROC modelling (Rutter and Gatsonis, 2001), both of which respect the correlation structure and yield interpretable summary points and curves.

15.6.2 Prerequisites

A working understanding of sensitivity, specificity, and the receiver-operating characteristic (ROC) curve; familiarity with the logit transformation; and basic mixed-effects modelling concepts.

15.6.3 Theory

The bivariate random-effects model jointly fits logit-sensitivity and logit-specificity using a bivariate Normal random effect across studies, estimating pooled sensitivity, pooled specificity, between-study variances for each, and the correlation between them. The HSROC model parameterises the underlying ROC curve directly with random study-specific thresholds and accuracy parameters, producing a summary curve that explicitly accommodates threshold variation. Reitsma's bivariate model and the HSROC model are mathematically equivalent under standard parameterisation; the choice between them is mainly one of presentation — pooled point with confidence ellipse versus summary curve.

15.6.4 Assumptions

Each study contributes a 2×2 table of true positives, false negatives, false positives, and true negatives. Random effects on the logit scale are bivariate Normal; threshold variation across studies is captured implicitly by the random-effects structure; case ascertainment and reference-standard verification are comparable across studies (otherwise applicability concerns enter, captured by the QUADAS-2 framework).

15.6.5 R Implementation

```r
library(mada)

set.seed(2026)
df <- data.frame(
  TP = c(45, 60, 32, 80, 28, 55),
  FN = c(5, 10, 8, 20, 12, 5),
  FP = c(10, 15, 8, 20, 10, 12),
  TN = c(140, 165, 152, 180, 150, 138)
)

fit <- reitsma(df)
summary(fit)

plot(fit, sroclty = 1, sroclwd = 2,
     main = "Summary ROC curve")
```

15.6.6 Output & Results

`reitsma()` returns pooled logit-sensitivity and logit-specificity with their variances and the between-study correlation; back-transformation gives the pooled point estimates with confidence intervals and a confidence ellipse on the ROC plane. The summary ROC curve traces the implied trade-off across thresholds, and a summary AUC quantifies discrimination across the full operating range.

15.6.7 Interpretation

A reporting sentence: “Bivariate meta-analysis of six diagnostic-accuracy studies (Reitsma model, REML) pooled sensitivity 0.84 (95 % CI 0.78 to 0.89) and specificity 0.91 (95 % CI 0.87 to 0.94) on the logit scale; the summary ROC area under the curve was 0.94. The between-study correlation -0.3 indicates threshold variation contributes meaningfully to between-study heterogeneity, which the bivariate model accommodates explicitly. Reporting follows PRISMA-DTA.” Always report both summary points and the SROC curve, plus the between-study correlation.

15.6.8 Practical Tips

- Always use bivariate or HSROC models; separate univariate pooling of sensitivity and specificity is statistically incorrect because it ignores the threshold-induced correlation between the two.
- Report the pooled point estimate, the 95 % confidence ellipse on the ROC plane, and the summary curve — together they convey both central performance and uncertainty.
- Threshold variability is a major source of heterogeneity in DTA meta-analyses; subgroup analyses by pre-specified threshold or by clinical setting are often more interpretable than a single overall pool.
- For studies with zero counts in any cell, `mada` applies a 0.5 continuity correction by default; sparse-data settings may benefit from Bayesian DTA-NMA models in `bamdit` or `INLA`.

- PRISMA-DTA is the reporting standard for diagnostic-accuracy systematic reviews; QUADAS-2 is the standard risk-of-bias assessment tool for individual primary studies.
- For diagnostic-accuracy NMA — comparing multiple competing tests on the same disease — `BUGSnet`, `bamdit`, and recent extensions in `multinma` provide the appropriate framework.

15.6.9 R Packages Used

`mada::reitsma()` for the canonical bivariate random-effects model and `mada` for SROC plotting and AUC; `mada::HSROC()` for the equivalent HSROC parameterisation; `bamdit` for Bayesian DTA meta-analysis with rich diagnostics; `meta::metaprop()` for univariate proportion pooling when only one of sensitivity or specificity is of interest; `multinma` for Bayesian DTA-NMA across multiple competing tests.

15.6.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.6.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.7 Converting Between Effect Sizes

15.7.1 Introduction

A single meta-analysis often spans studies that report effect sizes in different metrics: some randomised trials report standardised mean differences (Cohen's d or Hedges's g), some observational analyses report correlations (r), and some binary-outcome studies report odds ratios (OR) or risk ratios (RR). Pooling such heterogeneous evidence requires converting every per-study effect to a common metric — typically d when most studies are continuous, OR when most are binary, or r when most are correlational. Analytical conversion formulas, derived under standard latent-distribution assumptions, map between metrics together with their variances. Doing the conversion correctly — and documenting the assumptions — is critical because naively recycling an effect size between scales without proper variance propagation produces falsely precise pooled estimates.

15.7.2 Prerequisites

A working understanding of the principal effect-size metrics (d, r, OR, RR), their variance formulas, and the delta method for propagating variance through non-linear transformations.

15.7.3 Theory

The standard conversion formulas, all derived under the assumption that the underlying latent variable is approximately Normal, are:

$$d \to r: \quad r = \frac{d}{\sqrt{d^2 + a}}, \qquad a = \frac{(n_1 + n_2)^2}{n_1 n_2};$$

$$\log \mathrm{OR} \to d: \quad d = \log(\mathrm{OR}) \cdot \frac{\sqrt{3}}{\pi};$$

$$r \to d: \quad d = \frac{2r}{\sqrt{1 - r^2}}.$$

Each conversion has a corresponding variance formula obtained by the delta method; using these (rather than the un-transformed variance) is essential, because the non-linear mapping changes the precision of the resulting estimate.

15.7.4 Assumptions

The underlying latent distributions are approximately Normal (for $d \leftrightarrow r$ conversions) or follow the standard logistic-Normal correspondence (for OR conversions), the dichotomisation underlying any reported OR was approximately at the median of the latent variable, and the variance is propagated through the conversion using the appropriate formula rather than the original metric's variance.

15.7.5 R Implementation

```r
library(esc)

set.seed(2026)
d ← 0.5; var_d ← 0.04; n1 ← 50; n2 ← 50
conv_dr ← esc::convert_d2r(d = d, v = var_d,
                           grp1n = n1, grp2n = n2)
conv_dr

log_or ← log(2.5); var_logor ← 0.15
conv_od ← esc::convert_or2d(or = exp(log_or), v = var_logor)
conv_od

r ← 0.3; var_r ← 0.004
conv_rd ← esc::convert_r2d(r = r, v = var_r)
conv_rd
```

15.7.6 Output & Results

The `esc` package returns a list with the converted point estimate, its delta-method variance, and the corresponding standard error and confidence interval. Each conversion preserves both the central effect and a properly propagated variance, allowing inverse-variance pooling on the harmonised metric without artificially inflating precision.

15.7.7 Interpretation

A reporting sentence: "Effect sizes were harmonised to standardised mean differences. Continuous-outcome studies contributed $d = 0.50$ directly; binary-outcome studies were converted from $\mathrm{OR} = 2.5$ to $d = 0.51$ using $d = \log(\mathrm{OR}) \cdot \sqrt{3}/\pi$; correlational studies were converted from $r = 0.30$ to $d = 0.63$ using $d = 2r/\sqrt{1 - r^2}$. Variances were propagated by the delta method. The common metric enabled inverse-variance pooling across study designs and is documented in PRISMA Supplementary Table S2." Always document every conversion explicitly.

15.7.8 Practical Tips

- Use validated conversion functions from `esc`, `compute.es`, or `metafor::escalc()`; manual arithmetic with these formulas is highly error-prone, especially for variance propagation.
- Always propagate variance through the conversion using the delta-method-derived formulas; using the original metric's variance produces falsely narrow confidence intervals on the converted scale.
- Converting between binary and continuous metrics assumes a logistic-Normal latent-variable structure; for outcomes that obviously violate this — categorical with more than two levels treated as binary, very skewed continuous variables — the conversion is approximate at best.
- For highly skewed continuous outcomes (e.g., counts, costs, biomarkers with long tails), d-based conversions may misrepresent the substantive effect; consider rank-based effect sizes or per-domain pooling instead.
- Document every conversion in a PRISMA supplementary table, including the formula used, the input quantities, the converted estimate, and its propagated variance — methodological reviewers routinely check this.
- When the meta-analysis is mostly one metric with a few studies in another, conversion is appropriate; when the metrics are evenly split, consider whether a combined-metric analysis (e.g., a generalised linear mixed model on the underlying outcome) might be more defensible.

15.7.9 R Packages Used

`esc` for conversion functions across d, r, OR, and other effect sizes with delta-method variance; `compute.es` for an alternative comprehensive conversion library; `metafor::escalc()` and `metafor::conv.delta()` for canonical conversions and general delta-method variance propagation; `effectsize` for tidyverse-style effect-size computation and conversion alongside primary analyses.

15.7.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.7.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.8 Egger's Test for Funnel Asymmetry

15.8.1 Introduction

Egger's regression test, introduced by Egger and colleagues in 1997, provides a formal hypothesis test for funnel-plot asymmetry — the most commonly cited indirect symptom of publication bias and small-study effects. Where the funnel plot offers a visual diagnostic that depends on the eye of the beholder, Egger's test reduces the question to a regression intercept whose distribution under symmetry is known. A statistically significant intercept indicates that small studies systematically report effects of different magnitude than larger studies, which may reflect selective publication, methodological heterogeneity confounded with sample size, or genuine small-study heterogeneity. Although interpretation requires nuance, Egger's test remains the de facto standard quantitative complement to a funnel plot and is recommended by the Cochrane Handbook for meta-analyses with at least ten studies.

15.8.2 Prerequisites

Familiarity with funnel plots, the small-study-effect concept, and ordinary least-squares regression with weighted variants.

15.8.3 Theory

The test regresses the standardised effect $y_i/\mathrm{SE}(y_i)$ on the precision $1/\mathrm{SE}(y_i)$. Under the null hypothesis of funnel symmetry the regression line passes through the origin; the intercept tests this directly. A non-zero intercept indicates that the standardised effect depends on precision — exactly the pattern produced when small, imprecise studies are over-represented at one tail of the funnel because of selective publication.

Statistical power is low with fewer than ten studies, and inflated when between-study heterogeneity is large; $k \geq 10$ is the conventional minimum for interpretable use, and the test should never be the sole basis for declaring publication bias.

15.8.4 Assumptions

Studies are independent; the effect estimate is approximately normally distributed on the analysis scale; the relationship between standardised effect and precision is linear under the null; and the meta-analysis contains enough studies that any asymmetry is detectable above the random-sampling noise. For binary outcomes the standard Egger test is biased — Harbord's and Peters's modifications are preferred.

15.8.5 R Implementation

```
library(metafor)

set.seed(2026)
k ← 20
yi ← rnorm(k, 0.4, 0.2)
vi ← 1 / sample(20:300, k, replace = TRUE)

keep ← !(sqrt(vi) > 0.1 & abs(yi) < 0.2)
yi_b ← yi[keep]; vi_b ← vi[keep]

res ← rma(yi_b, vi_b, method = "REML")

egger ← regtest(res, model = "lm")
print(egger)

fsn(yi_b, vi_b)
```

15.8.6 Output & Results

`regtest()` returns the regression intercept, its standard error, the test statistic, and the two-sided p-value for the null of funnel symmetry. The companion fail-safe-N statistic from `fsn()` reports how many additional null-effect studies would be needed to drive the pooled estimate to non-significance — a complementary, intuitive sensitivity measure that helps readers gauge robustness.

15.8.7 Interpretation

A reporting sentence: "Egger's regression suggested significant funnel asymmetry (intercept 1.28, 95 % CI 0.21 to 2.35, $p = 0.02$), consistent with small-study effects or publication bias. A fail-safe N of 42 indicates the pooled conclusion would be diluted to non-significance only by 42 unpublished null studies, suggesting moderate but not unlimited robustness; trim-and-fill and PET-PEESE sensitivity analyses are reported in the supplement." Always pair the test result with at least one bias-adjusted sensitivity analysis.

15.8.8 Practical Tips

- Egger's test has low power for $k < 10$; the Cochrane Handbook explicitly discourages its use below this threshold and recommends visual funnel inspection only.

- For binary outcomes, prefer Harbord's modified test (continuous data) or Peters's test (rare events); the standard Egger test inflates the type-I error rate when the effect measure is a log-OR or log-RR.
- A significant intercept is not synonymous with publication bias — true between-study heterogeneity correlated with study size (small studies systematically using different populations or methods) can produce the same signal.
- Sensitivity-analyse with trim-and-fill and PET-PEESE; both adjust the pooled estimate for apparent asymmetry and should be reported alongside the unadjusted estimate when Egger's test is significant.
- For low-prevalence binary outcomes, asymmetry can arise from continuity-correction artefacts rather than genuine bias; Mantel–Haenszel without correction is a useful diagnostic alternative.
- Report the intercept estimate with its CI, not only its p-value; the magnitude of asymmetry conveys more than dichotomous significance.

15.8.9 R Packages Used

`metafor::regtest()` for Egger's, Harbord's, Peters's, and the more general regression-based asymmetry tests; `metafor::fsn()` for fail-safe N across multiple definitions (Rosenthal, Orwin, Rosenberg); `meta::metabias()` as the alternative interface with built-in plotting; `puniform` and `weightr` for selection-model-based bias adjustment; `metasens` for trim-and-fill and PET-PEESE workflows.

15.8.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.8.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.9 Fixed-Effect vs. Random-Effects Meta-Analysis

15.9.1 Introduction

A meta-analysis pools effect-size estimates from multiple studies into a single summary. The two classical approaches – the fixed-effect model and the random-effects model – rest on different assumptions about what those studies represent. Choosing between them is not a technical afterthought; it is a substantive decision about the scope of inference, and it can change both the point estimate and its uncertainty.

15.9.2 Prerequisites

The reader should know what an effect size is (mean difference, standardised mean difference, log-odds ratio, etc.), how its standard error is computed, and how a weighted average works. Familiarity with the `meta` or `metafor` packages is helpful.

15.9.3 Theory

Let $\hat{\theta}_i$ be the effect-size estimate from study i with variance v_i, for $i = 1, \ldots, k$.

15.9.3.1 Fixed-effect model

Assumes all studies estimate the *same* true effect θ:

$$\hat{\theta}_i = \theta + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, v_i).$$

The optimal pooled estimate is the inverse-variance-weighted mean

$$\hat{\theta}_{\mathrm{FE}} = \frac{\sum w_i \hat{\theta}_i}{\sum w_i}, \qquad w_i = 1/v_i,$$

with standard error $1/\sqrt{\sum w_i}$. Scope of inference: *the set of studies included.*

15.9.3.2 Random-effects model

Assumes each study estimates its own study-specific true effect, drawn from a distribution of true effects across a population of plausible studies:

$$\hat{\theta}_i = \theta_i + \varepsilon_i, \qquad \theta_i \sim \mathcal{N}(\mu, \tau^2), \qquad \varepsilon_i \sim \mathcal{N}(0, v_i).$$

The weights become $w_i^* = 1/(v_i + \tau^2)$, where τ^2 is the between-study variance, estimated by DerSimonian-Laird, REML, Paule-Mandel, or Sidik-Jonkman. The pooled estimate is $\hat{\mu} = \sum w_i^* \hat{\theta}_i / \sum w_i^*$. Scope of inference: *the population of studies from which the included ones are a sample.*

When $\tau^2 = 0$ (no heterogeneity), the random-effects model reduces to the fixed-effect model. When τ^2 is large, random-effects weights are more balanced (small studies get relatively more weight), and the random-effects CI is wider than the fixed-effect CI – properly so, since it reflects between-study variability.

15.9.4 Assumptions

Both models assume:

- Study-level effect estimates are approximately normal with known variance.
- Studies are independent of each other.
- No publication bias or other selection distorting the set of included studies.

The fixed-effect model additionally assumes a single true effect. The random-effects model assumes a normal distribution of true effects; the Hartung-Knapp adjustment corrects CIs under small k or imprecise τ^2 estimates.

15.9.5 R Implementation

```r
library(meta)

dat ← data.frame(
  study = paste("Study", 1:8),
  mean_t = c(11.2, 9.8, 12.5, 10.1, 13.0, 8.7, 11.6, 10.9),
  sd_t   = c(2.1, 2.4, 2.0, 2.6, 1.9, 2.8, 2.2, 2.3),
  n_t    = c(40, 35, 50, 30, 45, 25, 38, 42),
  mean_c = c(9.8, 9.5, 10.1, 9.9, 10.5, 9.0, 10.0, 10.2),
  sd_c   = c(2.0, 2.3, 2.1, 2.5, 2.0, 2.7, 2.1, 2.2),
  n_c    = c(40, 35, 50, 30, 45, 25, 38, 42)
)

m ← metacont(n.e = n_t, mean.e = mean_t, sd.e = sd_t,
             n.c = n_c, mean.c = mean_c, sd.c = sd_c,
             studlab = study, data = dat,
             sm = "MD",
             method.tau = "REML",
             method.random.ci = "HK")

summary(m)

forest(m, leftcols = c("studlab", "n.e", "mean.e", "sd.e",
                       "n.c", "mean.c", "sd.c"))
```

`metacont()` handles continuous outcomes with mean differences. Setting `method.tau = "REML"` uses REML for τ^2; `method.random.ci = "HK"` applies the Hartung-Knapp correction to the random-effects confidence interval.

15.9.6 Output & Results

Typical output:

```
                        MD            95%-CI %W(fixed) %W(random)
Fixed effect model   1.210 [ 0.802; 1.618]    100.0        --
Random effects model 1.195 [ 0.567; 1.822]       --    100.0

Quantifying heterogeneity:
 tau^2 = 0.3521; tau = 0.5934;
 I^2 = 52.8% [0.0%; 78.4%]; H = 1.46 [1.00; 2.15]

Test of heterogeneity:
     Q d.f.  p-value
 14.83    7   0.0381
```

The fixed-effect and random-effects point estimates are similar here (1.21 vs. 1.20), but the random-effects confidence interval is more than 50% wider, reflecting the moderate heterogeneity ($I^2 = 53\%$).

15.9.7 Interpretation

Report both models when I^2 is non-negligible, and emphasise the random-effects result with its Hartung-Knapp CI as the primary inference when heterogeneity is plausible on substantive grounds. Include the τ^2 estimate and the prediction interval – the latter answers "what true effect should I expect in a future, similar study?" and is often the most informative summary for a clinician.

15.9.8 Practical Tips

- Choose the model based on subject-matter plausibility, not on whether heterogeneity is statistically significant – Q-tests are underpowered with few studies.
- Report the prediction interval alongside the pooled estimate; it is the honest summary of what the pooled result implies for a new study.
- Use REML or Paule-Mandel for τ^2; DerSimonian-Laird is the historical default but understates τ^2 with few studies.
- Never pool fewer than 3 studies without explicit caveat; with 5 or fewer, consider narrative synthesis rather than meta-analysis.
- When heterogeneity is high, investigate it with subgroup analysis and meta-regression before pooling.

15.9.9 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.9.10 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.10 Forest Plots in Meta-Analysis

15.10.1 Introduction

The forest plot is the canonical meta-analysis figure and the single most important visual artefact a systematic review delivers. Each included study appears as a row with its point estimate, confidence interval, and pooling weight; the overall summary is rendered at the bottom as a diamond whose width represents the pooled confidence interval. This compact layout simultaneously conveys per-study effects, study-level precision, between-study scatter, and the overall conclusion — all in a glance, without needing to consult a

separate table. Editorial standards in major clinical journals routinely require a forest plot for every meta-analytic synthesis, and reviewers rely on it to judge whether the pooled estimate fairly represents the constituent studies.

15.10.2 Prerequisites

Inverse-variance pooling, the distinction between fixed-effect and random-effects models, and an understanding of how meta-analytic weights translate study precision into influence on the summary estimate.

15.10.3 Theory

Each study's square is plotted at its point estimate with size proportional to its meta-analytic weight (inverse variance under fixed-effect, or DerSimonian–Laird / REML weights under random-effects). Horizontal whiskers extend to the 95 % confidence interval. The pooled estimate is a diamond centred at the summary, with horizontal extent equal to the pooled CI; under random-effects models, an additional prediction interval is often overlaid as a dashed bar to convey where the next study's true effect is plausibly expected.

Standard columns include: study label (often author + year), sample sizes per arm, point estimate with CI, and the percent weight contribution. Heterogeneity statistics (Q, I^2, τ^2) typically appear in the figure footer.

15.10.4 Assumptions

Study-level estimates and their variances are on the same scale, expressed in the same direction (so that "to the right of zero" or "to the right of one" has a consistent clinical meaning), and pooled under a stated model. Forest plots themselves do not impose extra assumptions — they visualise whatever model produced the inputs.

15.10.5 R Implementation

```
library(metafor)

set.seed(2026)
k ← 8
yi ← c(0.32, 0.51, 0.28, 0.45, 0.38, 0.22, 0.60, 0.35)
vi ← c(0.02, 0.03, 0.015, 0.025, 0.04, 0.018, 0.05, 0.022)
authors ← paste("Study", LETTERS[1:k])

res ← rma(yi, vi, method = "REML")

forest(res, slab = authors,
       header = c("Study", "SMD [95% CI]"),
       xlab = "Standardised mean difference (Hedges' g)",
       mlab = "Random-effects (REML)",
       cex = 0.9)
```

15.10.6 Output & Results

The `forest()` method renders study-wise rows with squares sized by random-effects weight, whiskers showing 95 % CIs, and a summary diamond at the bottom. The header and label arguments customise the axis label and column heads, and `addpred = TRUE` overlays the random-effects prediction interval as a dashed bar around the diamond.

15.10.7 Interpretation

A reporting sentence: "The forest plot summarises eight randomised trials; seven favour the intervention with confidence intervals excluding the null, and the random-effects pooled SMD is 0.39 (95 % CI 0.28 to 0.50, $I^2 = 27\%$, $\tau^2 = 0.012$). The 95 % prediction interval (0.10 to 0.68) suggests the true effect in a future trial is likely positive but variable in magnitude." Always pair the forest plot with prose that quantifies both the central effect and the heterogeneity context.

15.10.8 Practical Tips

- Sort studies by publication year (for temporal reading), by effect size (for visual scanning of dispersion), or by subgroup (for pre-specified comparisons); the choice should match the scientific question.
- Always include heterogeneity statistics (I^2, τ^2, Q) in the figure footer; readers cannot judge "is this pool reasonable?" without them.
- Overlay the prediction interval (`addpred = TRUE` in `metafor`); a CI describes uncertainty about the mean, but the prediction interval is what readers actually want to know about future studies.
- For complex layouts — multiple subgroups, side-by-side outcomes, risk-of-bias columns — `forestplot`, `forester`, or `metafor::forest()` with custom `ilab` arguments give finer control than the default `meta::forest()`.
- Always include a clear directional legend ("favours intervention ← → favours control"); ambiguity here is the most common cause of reader misinterpretation.
- Round to a sensible number of digits (typically 2 for SMDs, 2 for log-ratios back-transformed); excessive precision suggests false confidence and clutters the plot.

15.10.9 R Packages Used

`metafor::forest()` for the canonical method, with `addpred`, `ilab`, and slab customisation; `meta::forest()` for the alternative `meta`-package interface with simpler defaults; `forestplot` and `forester` for highly customised layouts including categorical variables, multiple effects, and risk-of-bias columns; `metaviz` for ggplot-based forest plots with publication-ready theming.

15.10.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.10.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.11 Funnel Plots

15.11.1 Introduction

The funnel plot, introduced by Light and Pillemer in 1984, is the classical visual diagnostic for publication bias and small-study effects in meta-analysis. Each study contributes a single point whose horizontal coordinate is the effect estimate and whose vertical coordinate is precision (typically standard error or its inverse). Under the null of no bias and homogeneous true effect, the plot resembles a symmetric inverted funnel: large, precise studies cluster near the pooled estimate at the top and smaller, less precise studies fan out symmetrically below. Asymmetry — a visible empty zone in one of the two lower corners — is the canonical visual red flag that smaller studies with effects in one direction may be missing from the synthesis, often because they were never published.

15.11.2 Prerequisites

A working understanding of pooled meta-analysis, the relationship between sample size and standard error, and the small-study-effect concept that links study size to the magnitude of reported effects.

15.11.3 Theory

The horizontal axis carries the effect estimate on its analysis scale (log-OR, log-RR, SMD); the vertical axis carries precision, typically standard error inverted so that large, precise studies sit at the top. Pseudo-confidence-interval lines fan out symmetrically from the pooled estimate, giving the funnel its name. Under symmetry, small studies scatter on both sides; an empty lower corner suggests small studies with null or opposite-direction effects may be missing from the synthesis, the visual signature of publication bias.

A central conceptual point: asymmetry indicates a small-study effect, of which publication bias is one possible cause among several. Treating asymmetry as a synonym for bias is a common interpretive error.

15.11.4 Assumptions

Studies are independent; under the null of no bias the effect estimates scatter symmetrically around the true pooled effect; precision is well measured. The interpretation requires enough studies that asymmetry is detectable above sampling noise — typically at least ten.

15.11.5 R Implementation

```
library(metafor)

set.seed(2026)
k ← 20
yi ← rnorm(k, 0.4, 0.2)
vi ← 1 / sample(20:300, k, replace = TRUE)

keep ← !(sqrt(vi) > 0.1 & abs(yi) < 0.2)
yi_biased ← yi[keep]; vi_biased ← vi[keep]

res_biased ← rma(yi_biased, vi_biased, method = "REML")
funnel(res_biased, main = "Funnel plot (potentially biased)")

regtest(res_biased, model = "lm")
```

15.11.6 Output & Results

`funnel()` renders the canonical scatter with pseudo-CI lines. Visual asymmetry — typically an empty lower corner — combined with a significant Egger or Harbord regression test on the same data quantifies the impression. Reporting both views together (visual plus formal test) is the standard approach because the two are complementary diagnostics rather than redundant.

15.11.7 Interpretation

A reporting sentence: “The funnel plot was visibly asymmetric with an empty region in the small-study, low-effect quadrant; Egger’s regression test confirmed asymmetry (intercept 1.28, $p = 0.02$). Publication bias is a plausible explanation, but true heterogeneity correlated with study size cannot be ruled out; trim-and-fill and PET-PEESE sensitivity analyses are reported in the supplement and showed an attenuated but still positive pooled effect.” Always describe both the visual pattern and the formal-test result, and acknowledge alternative explanations.

15.11.8 Practical Tips

- Funnel plots require at least ten studies for meaningful interpretation; with fewer studies, the visual is dominated by sampling noise and asymmetry impressions are unreliable.
- Asymmetry has many possible causes besides publication bias: true between-study heterogeneity correlated with size, selective outcome reporting, methodological quality differences, or chance with small k. Discuss each explicitly when reporting an asymmetric funnel.

- Combine visual inspection with at least one formal test: Egger's regression for continuous outcomes, Harbord's modification for binary outcomes (log-OR), and Peters's test for rare events.
- For binary outcomes, plot log(OR) versus standard error on the log scale; raw OR funnels are intrinsically asymmetric for mathematical reasons unrelated to bias.
- Include sensitivity analyses — trim-and-fill, PET-PEESE, or Vevea–Hedges weight-function selection models — when asymmetry is detected; the difference between adjusted and unadjusted estimates is what readers actually need.
- Contour-enhanced funnel plots (Peters et al., 2008) overlay significance contours and help distinguish missingness near the null (suggestive of publication bias) from missingness elsewhere (suggestive of other explanations).

15.11.9 R Packages Used

`metafor::funnel()` for the canonical and contour-enhanced funnel plot, with `regtest()` for Egger/Harbord/Peters tests; `meta::funnel()` as the alternative interface; `metafor::trimfill()` for trim-and-fill adjustment; `metasens` for limit-meta-analysis and Copas-model bias adjustment alongside the funnel; `metaviz` for ggplot-based publication-ready funnel-plot output.

15.11.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.11.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.12 Quantifying Heterogeneity: Q and I^2

15.12.1 Introduction

Heterogeneity in meta-analysis refers to the variation in true effects across studies that exceeds what would be expected from sampling variability alone. Cochran's Q test asks whether any heterogeneity is present at all; I^2 quantifies how much of the total observed variability is attributable to genuine between-study differences rather than within-study sampling noise; and τ^2 provides the actual variance of true effects on the analysis scale. Together these three statistics inform the choice between fixed-effect and random-effects

models, the width of prediction intervals, and the appropriateness of subgroup or meta-regression follow-up analyses. Reporting all three is now the universal standard in systematic-review publications.

15.12.2 Prerequisites

A working understanding of fixed-effect and random-effects meta-analytic pooling, the chi-squared distribution, and inverse-variance weighting.

15.12.3 Theory

Cochran's Q is the weighted sum of squared deviations of study effects from the fixed-effect pooled estimate:

$$Q = \sum_{i} w_i (\theta_i - \hat{\theta}_{\mathrm{FE}})^2.$$

Under the null of homogeneity, Q follows a χ^2_{k-1} distribution. The I^2 statistic rescales Q to a proportion:

$$I^2 = \max\left\{0,\ \frac{Q - (k-1)}{Q}\right\} \cdot 100\%,$$

interpretable as the proportion of total variation attributable to between-study heterogeneity rather than within-study error. Higgins's benchmarks are: 0–40 % low, 30–60 % moderate, 50–90 % substantial, 75–100 % considerable. The variance component τ^2 — typically estimated by REML, DerSimonian–Laird, or Paule–Mandel — gives heterogeneity on the effect-size scale and feeds directly into prediction-interval computation.

15.12.4 Assumptions

Studies are independent and within-study variances are known (or precisely estimated). The Q test is famously underpowered for small numbers of studies and famously over-powered for large meta-analyses with very precise studies, so its p-value should never be the sole criterion for a heterogeneity decision.

15.12.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 12
theta ← rnorm(k, 0.4, 0.3)
n_per ← sample(50:400, k, replace = TRUE)
var_i ← 4 / n_per

yi ← theta + rnorm(k, 0, sqrt(var_i))
vi ← var_i

res ← rma(yi, vi, method = "REML")
cat("Q =", round(res$QE, 2),
    "(df=", res$k - 1, ", p =", round(res$QEp, 3), ")\n")
cat("I^2 =", round(res$I2, 1), "%;  tau^2 =", round(res$tau2, 3), "\n")
```

15.12.6 Output & Results

`rma()` returns the Q statistic with its p-value, I^2 as a percentage, and τ^2 on the effect-size scale. Confidence intervals on τ^2 and I^2 are obtained from `confint(res)`, and the prediction interval comes from `predict(res)`. Reporting all four — Q (with its df and p), I^2, τ^2, and the prediction interval — is now standard practice.

15.12.7 Interpretation

A reporting sentence: "Between-study heterogeneity was substantial: $Q = 31.4$ (df $= 11$, $p = 0.001$), $I^2 = 65\%$, $\tau^2 = 0.07$ ($\tau = 0.26$). The 95 % prediction interval for the true effect in a future study was $[-0.10, 0.85]$, far wider than the confidence interval on the pooled mean ($[0.21, 0.55]$). The random-effects model is therefore appropriate, and a pre-specified meta-regression on study year and population age was conducted to explain the heterogeneity." Always combine the three statistics with the prediction interval.

15.12.8 Practical Tips

- Do not rely on Q's p-value alone; it is underpowered when k is small and over-powered when individual studies are very precise. Always interpret it alongside I^2 and τ^2.
- Report Q, I^2, τ^2, and the 95 % prediction interval together; each tells a different and complementary story about heterogeneity.
- Substantial heterogeneity should prompt follow-up analyses: pre-specified meta-regression for continuous moderators, subgroup analysis for categorical ones, and explicit reporting of prediction intervals to communicate the practical range of true effects.
- I^2 depends on the precision of individual studies as well as on τ^2: large, precise studies inflate I^2 for the same true heterogeneity. The variance component τ^2 is therefore more directly comparable across meta-analyses than I^2.
- Confidence intervals on I^2 and τ^2 via `confint()` in `metafor` are routinely reported; a wide CI on I^2 should temper interpretation of any single point estimate.

- Use REML for τ^2 as the default; DerSimonian–Laird is convenient and traditional but can be biased downward when heterogeneity is large, and the Paule–Mandel estimator is also widely accepted.

15.12.9 R Packages Used

`metafor::rma()` for the canonical Q, I^2, and τ^2 output, with `confint()` for variance-component intervals and `predict()` for prediction intervals; `meta::metagen()` and related `meta` functions for the alternative interface; `dmetar` for Knapp–Hartung-corrected heterogeneity reporting; `metafor` `permutest()` for permutation-based heterogeneity inference under small k.

15.12.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.12.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.13 Individual Participant Data Meta-Analysis

15.13.1 Introduction

Individual participant data (IPD) meta-analysis pools raw participant-level data from multiple studies, enabling analyses that aggregate-data meta-analyses cannot – time-to-event across studies, individual-level effect modification, and harmonised outcome definitions. It is the gold standard when feasible but requires data-sharing agreements.

15.13.2 Prerequisites

Meta-analysis pooling; mixed-effects models.

15.13.3 Theory

Two-stage: analyse each study separately to obtain effect and variance; pool with standard meta-analysis methods.

One-stage: fit a single hierarchical model on the combined data with study as a random effect. Allows individual-participant-level interactions.

Both approaches typically yield similar results for main effects; one-stage is preferred for individual-level moderator analyses.

15.13.4 Assumptions

Data-sharing agreements in place; common outcome definitions across studies; correct handling of study-level heterogeneity.

15.13.5 R Implementation

```
library(lme4)

set.seed(2026)
n_studies ← 5
participants_per ← 100

# Simulate IPD
df ← do.call(rbind, lapply(1:n_studies, function(s) {
  n ← participants_per
  data.frame(
    study = s,
    arm = factor(sample(c("trt", "ctrl"), n, replace = TRUE)),
    age = rnorm(n, 50, 10)
  )
}))
df$arm_num ← as.integer(df$arm == "trt")
df$y ← 0.4 * df$arm_num + 0.05 * df$age +
        rnorm(nrow(df), 0, 1) +
        rep(rnorm(n_studies, 0, 0.3), each = participants_per)

# One-stage mixed-effects model
one_stage ← lmer(y ~ arm + age + (1 | study), data = df)
summary(one_stage)$coefficients

# Two-stage: per-study effects, then pool
per_study ← split(df, df$study) ▷ sapply(function(s) {
  m ← lm(y ~ arm + age, data = s)
  c(est = coef(m)["armtrt"], var = vcov(m)["armtrt", "armtrt"])
})
library(metafor)
ma ← rma(yi = per_study["est.armtrt", ], vi = per_study["var", ],
         method = "REML")
ma$b
```

15.13.6 Output & Results

One-stage arm coefficient and the equivalent two-stage pooled estimate; typically close agreement for simple models.

15.13.7 Interpretation

“IPD meta-analysis estimated the intervention effect at 0.42 (95 % CI 0.28-0.56, p < 0.001); two-stage and one-stage approaches agreed closely, supporting the finding’s robustness to analytic framing.”

15.13.8 Practical Tips

- Prospective IPD consortium studies avoid selection bias from data-availability issues.
- Harmonise outcomes before pooling; different outcome definitions across studies cause bias.
- One-stage model allows individual-level interactions (effect modifiers); more valuable than ecological moderators.
- For survival endpoints, IPD is required (aggregate hazard ratios are difficult to pool correctly).
- Report PRISMA-IPD checklist for transparency.

15.13.9 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.13.10 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.14 Leave-One-Out Meta-Analysis

15.14.1 Introduction

Leave-one-out (LOO) meta-analysis systematically removes each study and re-computes the pooled estimate using the remaining $k - 1$ studies. The resulting per-removal estimates show how much each study drives the overall conclusion. Studies whose removal substantially changes the pooled effect — magnitude, significance, or both — are influential and warrant careful interpretation. LOO is a routine sensitivity analysis in modern meta-analysis reporting.

15.14.2 Prerequisites

A working understanding of pooled meta-analysis (fixed or random effects) and the concept of influence diagnostics from regression analysis.

15.14.3 Theory

For each study $i = 1, \ldots, k$, compute $\hat{\theta}_{(-i)}$ by pooling the other $k - 1$ studies under the same meta-analysis model. Plot $\hat{\theta}_{(-i)}$ with its confidence interval across i; large deviations from the full-model pooled estimate indicate influential studies.

LOO complements other influence measures: Cook's distance and DFBETAS quantify how much each study perturbs the overall fit, and Baujat plots show each study's contribution to heterogeneity (Q) versus its contribution to the pooled estimate. Used together, these diagnostics provide a comprehensive picture of which studies the conclusion depends on.

15.14.4 Assumptions

Same as the underlying random-effects meta-analysis; each leave-one-out fit uses the same model and method as the full pooling.

15.14.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 10
yi ← c(0.35, 0.42, 0.25, 0.55, 0.28, 0.39, 0.95, 0.31, 0.44, 0.36)
vi ← c(0.02, 0.03, 0.015, 0.02, 0.025, 0.018, 0.01, 0.04, 0.02, 0.022)

res ← rma(yi, vi, method = "REML")
loo ← leave1out(res)

print(loo[, c("estimate", "ci.lb", "ci.ub", "I2")])

# Influence diagnostics: Cook's distance, DFBETAS
inf ← influence(res)
print(inf$inf)
```

15.14.6 Output & Results

`leave1out()` returns a per-study table of the recomputed pooled estimate, confidence interval, and I^2 when each study is omitted. `influence()` provides Cook's distance, DFBETAS, hat values, and other diagnostics computed jointly.

15.14.7 Interpretation

A reporting sentence: "Leave-one-out analysis revealed study 7 (SMD = 0.95, the largest single estimate) as highly influential; removing it reduced the pooled SMD from 0.41 to 0.35 and shrunk I^2 from 38 % to 18 %. Conclusions are robust to removal of any other study; the sensitivity analysis without study 7 is reported as supplementary." Always describe what an influential study removal does to both effect size and heterogeneity.

15.14.8 Practical Tips

- Pair LOO with Cook's distance and DFBETAS from `metafor::influence()`; the combination flags both effect-size influence and overall fit influence.
- If removing any single study crosses the significance line, the conclusion is fragile and should be reported as such.
- LOO is descriptive — it does not correct any bias. Use it to inform readers about robustness, not to "clean up" the meta-analysis post-hoc.

- Report LOO results only if at least one study is markedly influential; otherwise the table conveys no useful information beyond the headline pooled estimate.
- Combine with Baujat plots (`metafor::baujat()`) for a richer visualisation: studies in the upper-right corner contribute disproportionately to both heterogeneity and effect.
- For meta-analyses with fewer than 5 studies, LOO is unstable and largely uninformative; report it but interpret cautiously.

15.14.9 R Packages Used

`metafor::leave1out()` for the canonical LOO output; `metafor::influence()` for Cook's distance, DFBETAS, and hat values; `metafor::baujat()` for Baujat plots; `meta` for analogous sensitivity tools in an alternative interface; `dmetar::infA()` for combined influence and outlier diagnostics in an integrated workflow.

15.14.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.14.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.15 Pooling Log-Odds Ratios

15.15.1 Introduction

For binary outcomes, the odds ratio (OR) is the most widely reported effect measure across case-control studies, randomised trials with binary endpoints, and observational comparative studies. Because the OR is bounded below at zero with a long right tail, its sampling distribution is markedly skewed and inverse-variance pooling on the natural scale would be poorly behaved. The standard approach is therefore to pool on the log scale, where the distribution is approximately Normal, and back-transform the pooled estimate and its confidence interval to the OR scale for reporting. Pooling on the log scale stabilises variance, allows symmetric confidence intervals on the analysis scale, and keeps the multiplicative-on-back-transform interpretation that clinicians and epidemiologists expect.

15.15.2 Prerequisites

A working understanding of odds ratios as a measure of association in 2×2 tables, inverse-variance meta-analytic weighting, and the distinction between fixed-effect and random-effects models.

15.15.3 Theory

For a study with a_i, b_i, c_i, d_i in the four cells of the exposure-by-outcome 2×2 table, the per-study log-OR and its variance are

$$\log \mathrm{OR}_i = \log \frac{a_i d_i}{b_i c_i}, \qquad \mathrm{Var}(\log \mathrm{OR}_i) = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}.$$

A fixed-effect pool is $\hat{\theta} = \sum w_i \theta_i / \sum w_i$ with $w_i = 1/\mathrm{Var}(\log \mathrm{OR}_i)$. A random-effects pool adds the between-study variance τ^2 to each study's weight, widening the confidence interval to reflect between-study uncertainty about the underlying effect. Both pools are fit on the log scale and back-transformed via $\exp(\cdot)$ for reporting.

15.15.4 Assumptions

Studies are independent; the log-OR is approximately Normal (typically holds when each cell has ≥5 events); the within-study variance is accurately estimated; the pooled population is reasonably homogeneous (fixed-effect) or, more commonly, is heterogeneous with effects drawn from a Normal distribution around a common mean (random-effects).

15.15.5 R Implementation

```
library(metafor)

set.seed(2026)
k ← 10
df ← data.frame(
  ai = rpois(k, 30), bi = rpois(k, 170),
  ci = rpois(k, 20), di = rpois(k, 180)
)

es ← escalc(measure = "OR", ai = ai, bi = bi,
             ci = ci, di = di, data = df)

fe ← rma(yi, vi, data = es, method = "FE")
re ← rma(yi, vi, data = es, method = "REML")

cat("FE pooled OR:", round(exp(fe$b), 2),
    " 95% CI:", paste(round(exp(c(fe$ci.lb, fe$ci.ub)), 2), collapse = "-"), "\n")
cat("RE pooled OR:", round(exp(re$b), 2),
    " 95% CI:", paste(round(exp(c(re$ci.lb, re$ci.ub)), 2), collapse = "-"), "\n")
```

15.15.6 Output & Results

`escalc(measure = "OR")` computes the per-study log-OR and its variance, applying continuity correction to zero cells when needed. `rma()` fits the meta-analysis on the log scale; back-transformation via `exp()` yields the pooled OR and its confidence interval on the natural scale. The random-effects model additionally returns τ^2 and I^2 describing between-study heterogeneity.

15.15.7 Interpretation

A reporting sentence: "Random-effects meta-analysis of 10 studies (REML, total $n = 4{,}000$) gave a pooled odds ratio of 1.56 (95 % CI 1.31 to 1.87) with modest between-study heterogeneity ($I^2 = 22\%$, $\tau^2 = 0.01$); the fixed-effect estimate was almost identical (1.54), confirming that heterogeneity did not materially shift the central effect." Always report both the pooled estimate and the heterogeneity metrics; readers cannot judge the pool without them.

15.15.8 Practical Tips

- Always pool on the log scale; ORs are multiplicative and pooling on the natural scale produces biased and asymmetrically distributed estimates.
- Apply a continuity correction (typically add 0.5 to every cell of a study with any zero cell) for studies with zero events; `metafor::escalc()` handles this automatically by default.
- Choose between fixed-effect and random-effects models based on the substantive research question and expected sources of heterogeneity, not on the Q-test p-value alone — I^2 and τ^2 are more informative about the magnitude of heterogeneity.
- Always report heterogeneity statistics (τ^2, I^2, and the prediction interval) alongside the pooled point estimate.
- Apply the Hartung–Knapp–Sidik–Jonkman adjustment for random-effects pooling with small k (usually fewer than 20 studies); it produces wider, more conservative CIs that better match coverage in simulations.
- Mantel–Haenszel pooling (`metafor::rma.mh()`) is preferable for sparse-data settings with many zero cells, where continuity correction biases the standard inverse-variance pool.

15.15.9 R Packages Used

`metafor::escalc()` and `rma()` for the canonical log-OR pooling workflow; `metafor::rma.mh()` for Mantel–Haenszel pooling and `rma.peto()` for the Peto OR method on rare events; `meta::metabin()` as the alternative interface with cleaner forest-plot defaults; `metafor` Hartung–Knapp via `test = "knha"`.

15.15.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.15.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.16 Meta-Regression

15.16.1 Introduction

Meta-regression extends the random-effects meta-analysis by allowing study-level covariates — moderators such as publication year, mean participant age, average dose, or risk-of-bias score — to enter the model as predictors of the underlying effect. The fitted slope quantifies how much the true study effect changes per unit of moderator, and a residual between-study variance τ^2 describes whatever heterogeneity the moderator did not explain. Meta-regression is the appropriate tool whenever the candidate moderator is continuous, and is generally more powerful than dichotomising into subgroups, which discards information and can produce a categorisation-dependent answer. It is now standard practice for any meta-analysis with $k \geq 10$ studies and a pre-specified hypothesis about effect modification.

15.16.2 Prerequisites

A working understanding of random-effects meta-analysis (notably the role of τ^2), ordinary linear regression, and the difference between fixed-effect, random-effects, and mixed-effects models in the meta-analytic context.

15.16.3 Theory

The mixed-effects meta-regression model is

$$y_i = \beta_0 + \beta_1 x_i + u_i + \varepsilon_i, \qquad u_i \sim N(0, \tau^2), \quad \varepsilon_i \sim N(0, s_i^2),$$

where u_i is between-study residual heterogeneity and ε_i is within-study sampling error with known variance s_i^2. REML estimates τ^2 jointly with the regression coefficients; Wald or likelihood-ratio tests assess moderator effects. The pseudo-$R^2 = 1 - \tau^2_{\text{residual}} / \tau^2_{\text{null}}$ summarises the proportion of between-study variance explained by the moderator and is the analogue of the regression R^2.

15.16.4 Assumptions

The moderator is pre-specified rather than data-mined; the functional form (linear unless explicitly modelled otherwise) is correct; the residual heterogeneity is normal; and there are sufficient studies — a frequently-cited rule of thumb is at least ten studies per moderator coefficient — to support reliable estimation.

15.16.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 15
year ← sample(1995:2023, k, replace = TRUE)
mean_age ← runif(k, 40, 70)

yi ← 0.5 - 0.02 * (year - 1995) + rnorm(k, 0, 0.15)
vi ← 1 / sample(50:250, k, replace = TRUE)

df ← data.frame(yi, vi, year, mean_age)

res1 ← rma(yi, vi, mods = ~ year, data = df, method = "REML")
summary(res1)

res2 ← rma(yi, vi, mods = ~ year + mean_age, data = df, method = "REML")
summary(res2)

cat("R^2:", round(res1$R2, 3), "\n")
```

15.16.6 Output & Results

`rma()` with a `mods` formula returns moderator coefficients with confidence intervals and p-values, the residual heterogeneity statistics (τ^2, I^2, residual Q_E), an omnibus moderator test (Q_M), and the pseudo-R^2 summarising explained heterogeneity. Reporting all four elements — slope, residual τ^2, Q_M, and R^2 — is standard practice.

15.16.7 Interpretation

A reporting sentence: “Year of publication significantly moderated the pooled effect (slope -0.017, 95 % CI -0.029 to -0.005, $p = 0.01$): each additional year reduced the standardised mean difference by 0.017 units. Meta-regression explained 52 % of the between-study variance (R^2); residual heterogeneity remained moderate ($I^2_{\text{residual}} = 21\%$). The temporal trend was pre-specified in the PROSPERO protocol.” Always report both the slope and the residual heterogeneity to convey what the moderator did and did not explain.

15.16.8 Practical Tips

- Restrict to roughly one moderator per ten studies; meta-regression overfits aggressively because between-study sample sizes are tiny by regression standards.
- Pre-specify all moderators in the PROSPERO protocol; post-hoc moderators are at best exploratory and should be flagged as such in reporting.

- Centre continuous moderators (e.g., year minus mean year) so the intercept has a meaningful interpretation as the effect at the mean moderator value.
- For small k, permutation tests via `permutest()` provide more conservative inference than the Wald-based default and are widely recommended.
- Beware ecological fallacy: a study-level moderator effect does not imply the same effect at the individual-participant level; that requires IPD meta-analysis.
- Plot bubble plots (`metafor::regplot()`) showing study effects, moderator value, and weighted regression line; they make the moderator relationship visible and reveal influential outliers.

15.16.9 R Packages Used

`metafor::rma()` with `mods = ~ ...` for the canonical mixed-effects meta-regression; `metafor::regplot()` for bubble plots; `metafor::permutest()` for permutation-based inference under small k; `meta::metareg()` for the alternative `meta`-package interface; `robumeta` for robust-variance meta-regression with dependent effect sizes.

15.16.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.16.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.17 Network Meta-Analysis: Introduction

15.17.1 Introduction

Network meta-analysis (NMA), or mixed-treatment comparison (MTC), pools evidence across studies comparing different pairs of treatments to estimate a coherent set of relative effects. It enables comparisons for which no direct head-to-head trials exist, provided indirect paths in the evidence network connect the treatments.

15.17.2 Prerequisites

Pairwise meta-analysis; graph structure; consistency concept.

15.17.3 Theory

A network has nodes (treatments) and edges (pairwise comparisons with at least one study). Direct evidence = studies comparing treatments directly; indirect evidence = inferred via a common comparator. The pooled estimate combines both under an assumption of **consistency** (direct = indirect estimates).

Implementation: frequentist (`netmeta`) or Bayesian (`gemtc` , `BUGSnet`).

15.17.4 Assumptions

Transitivity: studies in the network are sufficiently similar that indirect comparisons are valid. **Consistency**: direct and indirect estimates of the same contrast agree within sampling error.

15.17.5 R Implementation

```r
library(netmeta)

# Example dataset
set.seed(2026)
net_df ← data.frame(
  study = c("S1", "S1", "S2", "S2", "S3", "S3", "S4", "S4", "S5", "S5"),
  treat = c("A", "B", "A", "C", "B", "C", "A", "D", "C", "D"),
  TE    = c(0, -0.3, 0, -0.5, 0.4, 0.2, 0, -0.2, 0, 0.3),
  seTE  = c(0.2, 0.2, 0.25, 0.25, 0.3, 0.3, 0.2, 0.2, 0.22, 0.22)
)
# Rearrange to pairs
p_df ← pairwise(treat = treat, TE = TE, seTE = seTE,
                studlab = study, data = net_df)

net ← netmeta(p_df, common = FALSE, reference = "A")
summary(net)
netgraph(net)
```

15.17.6 Output & Results

A network graph, pooled treatment effects relative to the reference, and overall I^2.

15.17.7 Interpretation

“Network meta-analysis of 5 studies covering 4 treatments (A reference, B, C, D) yielded pooled SMDs: B vs A -0.22 (-0.41, -0.03); C vs A -0.35 (-0.58, -0.12); D vs A 0.05 (-0.15, 0.25). Transitivity was plausible; consistency was not violated.”

15.17.8 Practical Tips

- Always produce a network graph first; disconnected components cannot be compared.
- Check transitivity via study characteristics (populations, doses, outcomes).
- Use SUCRA to rank treatments by their probability of being best.

- Report PRISMA-NMA checklist; network meta-analyses are complex and demand transparency.
- Bayesian NMA via `gemtc` is more flexible for hierarchical models and missing-arm data.

15.17.9 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.17.10 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.18 NMA Consistency and Inconsistency

15.18.1 Introduction

Network meta-analysis combines direct evidence from head-to-head trials with indirect evidence inferred through common comparators. The validity of this synthesis hinges on the consistency assumption: across closed loops in the network, the direct and indirect estimates of any pairwise contrast must agree up to sampling error. When they disagree systematically — when, for example, the direct A-vs-C estimate from head-to-head trials differs materially from the indirect estimate inferred via B — the network is inconsistent, and the pooled NMA estimates cannot be interpreted at face value. Detecting and characterising inconsistency is therefore a mandatory step in any rigorous NMA workflow, and modern guidance (NICE DSU, ISPOR, PRISMA-NMA) requires that consistency be reported alongside the league table and SUCRA results.

15.18.2 Prerequisites

A working understanding of network meta-analysis structure (treatments, designs, contrasts), the difference between direct and indirect evidence, and the transitivity assumption that underpins indirect comparisons.

15.18.3 Theory

Three complementary tests address different scales of inconsistency. The loop-specific Bucher test, for a triangle A-B-C, compares the direct A-vs-C estimate against the indirect estimate $\hat{\theta}_{AB} + \hat{\theta}_{BC}$ via a Wald-type difference. Node-splitting (Dias et al., 2010) generalises this for any contrast: it splits the evidence for one contrast into direct-only

and indirect-only components and tests whether they agree. The design-by-treatment interaction model (Higgins et al., 2012) provides a global test via the generalised-inconsistency framework, comparing the consistent NMA model against an unconstrained model that allows different contrasts to take different values across study designs.

15.18.4 Assumptions

The network contains at least one closed loop (otherwise consistency cannot be tested), transitivity holds in principle, and the underlying meta-analytic model is correctly specified. Inconsistency tests have low power when loops are short or sparsely populated; absence of detected inconsistency does not prove consistency.

15.18.5 R Implementation

```r
library(netmeta)

set.seed(2026)
net_df <- data.frame(
  study = c("S1", "S1", "S2", "S2", "S3", "S3", "S4", "S4",
            "S5", "S5", "S6", "S6"),
  treat = c("A", "B", "A", "C", "B", "C", "A", "D",
            "B", "D", "C", "D"),
  TE    = c(0, -0.3, 0, -0.5, 0.4, 0.2, 0, -0.2,
            0.1, 0.3, 0, 0.4),
  seTE  = rep(0.25, 12)
)

p_df <- pairwise(treat = treat, TE = TE, seTE = seTE,
                 studlab = study, data = net_df)
net  <- netmeta(p_df, common = FALSE, reference = "A")

netsplit(net)

decomp.design(net)
```

15.18.6 Output & Results

`netsplit()` returns per-comparison direct vs indirect estimates, their difference, and a Wald test on the difference. `decomp.design()` decomposes the heterogeneity statistic Q into within-design and between-design components and reports the global design-by-treatment interaction test. Non-significant results in both diagnostics support the consistency assumption.

15.18.7 Interpretation

A reporting sentence: “Node-splitting (`netsplit`) showed agreement between direct and indirect estimates for all six pairwise comparisons (smallest two-sided p-value $= 0.15$). The global design-by-treatment interaction test (`decomp.design`) was non-significant ($p = 0.36$), and the between-design heterogeneity component was negligible ($Q_{\text{inc}} = 1.2$, df $= 3$). The consistency assumption is therefore supported, and the random-effects league table can be interpreted as reflecting the joint direct and indirect evidence.” Always report both tests.

15.18.8 Practical Tips

- Always test consistency; reporting an NMA without any consistency check is no longer acceptable in major journals or HTA submissions.
- If inconsistency is detected, investigate study-level effect modifiers (population age, baseline severity, dose, year, risk of bias) that differ across designs; meta-regression within the NMA framework can then partially resolve the inconsistency.
- `netsplit` provides per-comparison tests (high resolution, low power per test); `decomp.design` provides the global test (low resolution, higher power overall) — they are complementary, not redundant.
- For simple triangles with sufficient direct evidence, the Bucher test is a quick first-pass diagnostic; it is what node-splitting reduces to in the three-treatment case.
- Inconsistency can be local (a single loop) or global (the whole network); the diagnostic patterns differ, and the discussion section should reflect which kind was found.
- Bayesian implementations in `gemtc` and `multinma` provide UME (unrelated mean effects) consistency models that compare against the consistent NMA via DIC; this is the Bayesian analogue of the design-by-treatment test.

15.18.9 R Packages Used

`netmeta::netsplit()` and `netmeta::decomp.design()` for the canonical frequentist consistency diagnostics, alongside `netmeta()` itself; `gemtc` and `multinma` for Bayesian NMA with built-in UME and node-splitting; `pcnetmeta` for component-NMA inconsistency analysis; `BUGSnet` for combined Bayesian NMA workflow including consistency reporting.

15.18.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.18.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.19 NMA League Tables

15.19.1 Introduction

The league table is the standard summary display for a network meta-analysis (NMA). It packs every pairwise treatment contrast — direct, indirect, and mixed — into a single triangular or square grid, making the relative ranking of all interventions visible at a glance. Each diagonal cell labels a treatment; each off-diagonal cell reports the pooled contrast between row and column with its 95 % confidence (or credible) interval. League tables are now a standard reporting requirement in the PRISMA-NMA extension because they communicate the entire network's information in a way that no single forest plot can.

15.19.2 Prerequisites

A working understanding of network meta-analysis (transitivity, consistency), how mixed treatment-comparison estimates combine direct and indirect evidence, and the conventions of pairwise effect sizes (OR, RR, SMD).

15.19.3 Theory

For T treatments the league table is a $T \times T$ matrix. The convention used by `netmeta` and most software places the random- or common-effect mixed estimates in the upper triangle and direct-only pooled estimates in the lower triangle, with the same treatment ordering on both axes. Each off-diagonal cell reports the pooled contrast (row-vs-column or column-vs-row, with the convention noted in the caption) and its uncertainty.

Ordering treatments by their rank (e.g., SUCRA, P-score, or median ranking from a Bayesian NMA) groups the best-performing treatments in one corner, making the table substantially easier to interpret than alphabetical ordering.

15.19.4 Assumptions

The underlying NMA is valid: transitivity holds across studies that compare different treatment subsets, consistency holds between direct and indirect evidence, and all pairwise contrasts are estimable from the connected network. Reporting league tables for a disconnected network is meaningless.

15.19.5 R Implementation

```
library(netmeta)

set.seed(2026)
net_df ← data.frame(
  study = c("S1", "S1", "S2", "S2", "S3", "S3", "S4", "S4",
            "S5", "S5", "S6", "S6"),
  treat = c("A", "B", "A", "C", "B", "C", "A", "D",
            "B", "D", "C", "D"),
  TE    = c(0, -0.3, 0, -0.5, 0.4, 0.2, 0, -0.2,
            0.1, 0.3, 0, 0.4),
  seTE  = rep(0.25, 12)
)

p_df ← pairwise(treat = treat, TE = TE, seTE = seTE,
                studlab = study, data = net_df)
net  ← netmeta(p_df, common = FALSE, reference = "A")

league ← netleague(net, bracket = "(",
                   separator = " to ", digits = 2)
league$random
```

15.19.6 Output & Results

`netleague()` returns a structured matrix of pairwise contrast estimates with their confidence intervals, separately for the common-effect and random-effects models. The diagonal carries treatment labels; off-diagonal cells contain the pooled contrast and its CI in a formatted string. Exporting the table to LaTeX or HTML for publication uses standard formatting tools.

15.19.7 Interpretation

A reporting sentence: “The league table summarises all six pairwise contrasts in the network of four treatments. Treatment C was consistently the most effective: C vs A had a pooled effect of -0.55 (95 % CI -0.88 to -0.22) and C vs B was -0.34 (95 % CI -0.67 to -0.01); both contrasts excluded the null. The lower-triangle direct-only estimates were qualitatively concordant, suggesting consistency of direct and indirect evidence.” Always pair league-table reading with a node-splitting consistency check.

15.19.8 Practical Tips

- Order treatments by their NMA-derived ranking (e.g., SUCRA from highest to lowest); the best-vs-worst contrast then sits in one corner and visual scanning becomes intuitive.
- Report both the point estimate and the CI; a significant contrast is one whose CI excludes the null on the analysis scale, but readers benefit from seeing the magnitude as well.
- For binary outcomes, exponentiate to OR or RR for clinical interpretation; on the multiplicative scale, a CI excluding 1 is the significance criterion.
- Complement the league table with a forest plot of contrasts versus a clinically meaningful reference treatment; the two views together communicate more than

either alone.
- For networks with more than ten treatments, a full league table becomes unwieldy; consider a heatmap visualisation or restrict the table to clinically relevant pairs.
- Always cite the consistency assumption in the caption and report a node-splitting or design-by-treatment interaction check elsewhere in the manuscript.

15.19.9 R Packages Used

`netmeta::netleague()` for the canonical league-table output and `netmeta()` for the underlying frequentist NMA; `gemtc` and `multinma` for Bayesian NMA league-table equivalents using rank tables; `pcnetmeta` for component network meta-analysis with extended league-table summaries; `dmetar` for combined NMA workflow and reporting helpers.

15.19.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.19.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.20 SUCRA Rankings

15.20.1 Introduction

SUCRA — the Surface Under the Cumulative Ranking curve, introduced by Salanti and colleagues in 2011 — summarises a network meta-analysis's treatment ranking into a single interpretable number between 0 and 1. A SUCRA of 1 means the treatment is always ranked best across the posterior or simulation distribution; a SUCRA of 0 means it is always ranked worst. Because clinicians, guideline panels, and patients want to know "which treatment should I use?", and a triangular league table cannot answer that directly, SUCRA quickly became the most-cited summary in NMA papers. Its widespread use also brings risks: SUCRA is sensitive to the number of treatments included, can exaggerate small effect-size differences, and must always be interpreted alongside the absolute effect estimates and their uncertainty.

15.20.2 Prerequisites

A working understanding of network meta-analysis, posterior or simulated rank distributions, and the conceptual difference between a treatment ranking and a treatment effect.

15.20.3 Theory

For each treatment j in a network of T treatments, the rank distribution gives $P(\text{rank}_j = k)$ for $k = 1, \ldots, T$ — typically derived from a Bayesian posterior with MCMC or from frequentist parametric simulation. SUCRA is

$$\text{SUCRA}_j = \frac{1}{T-1} \sum_{k=1}^{T-1} P(\text{rank}_j \leq k),$$

the area under the cumulative rank curve scaled to $[0, 1]$. Rücker and Schwarzer's P-score (2015) is the frequentist analogue computed directly from NMA point estimates and standard errors; it does not require MCMC and reproduces SUCRA values closely in most networks.

15.20.4 Assumptions

The underlying NMA is valid (transitivity across studies, consistency between direct and indirect evidence), and the network is connected so all rankings are estimable. The clinical interpretability of "best treatment" requires that the treatments are exchangeable in indication and that no contraindications make a high-ranked option unusable for some patient groups.

15.20.5 R Implementation

```
library(netmeta)

set.seed(2026)
net_df ← data.frame(
  study = c("S1", "S1", "S2", "S2", "S3", "S3", "S4", "S4",
            "S5", "S5", "S6", "S6"),
  treat = c("A", "B", "A", "C", "B", "C", "A", "D",
            "B", "D", "C", "D"),
  TE    = c(0, -0.3, 0, -0.5, 0.4, 0.2, 0, -0.2,
            0.1, 0.3, 0, 0.4),
  seTE  = rep(0.25, 12)
)

p_df ← pairwise(treat = treat, TE = TE, seTE = seTE,
                studlab = study, data = net_df)
net  ← netmeta(p_df, common = FALSE, reference = "A",
               small.values = "good")

rankings ← netrank(net, method = "P-score", small.values = "good")
rankings$ranking.random
```

15.20.6 Output & Results

`netrank()` returns a P-score per treatment (between 0 and 1) under both common-effect and random-effects models. Higher scores indicate better ranking on the chosen analysis scale. For Bayesian NMA, packages such as `gemtc` and `multinma` produce SUCRA values directly, along with rank probabilities $P(\text{rank} = k)$ for each k.

15.20.7 Interpretation

A reporting sentence: "Treatment C had the highest P-score (0.89) under the random-effects NMA, identifying it as most likely the best of the four candidates. However, the pairwise contrast C vs B (P-score 0.61) had a 95 % CI that crossed the null, so the ranking is uncertain and absolute differences between top-ranked options are small. Rankograms reporting the full distribution over ranks accompany the SUCRA summary in the supplement." Always report rankings together with absolute effect estimates and the surrounding uncertainty.

15.20.8 Practical Tips

- Always report SUCRA or P-score alongside the pairwise effect estimates and their CIs; rankings alone can exaggerate small differences and obscure clinically irrelevant gaps.
- Treatments with similar point estimates can show very different SUCRA values when uncertainty differs; the more precise treatment will rank higher even if its mean effect is no better. This is a feature, not a bug — but it must be interpreted carefully.
- Use rankograms (bar plots of $P(\text{rank} = k)$ for each k) to visualise the entire rank distribution; they convey the uncertainty that the single SUCRA number compresses.
- Frequentist P-scores and Bayesian SUCRA values typically agree to within rounding error on connected networks; the choice between them is mainly methodological preference.
- For regulatory and reimbursement decisions, the absolute pairwise effect and its CI matter more than the ranking; SUCRA is a clinical-communication tool, not a decision rule.
- Check `small.values` carefully: misspecifying whether small effect sizes are "good" (e.g., for a harm outcome) or "bad" inverts the entire ranking and is one of the most common analytic errors in published NMAs.

15.20.9 R Packages Used

`netmeta::netrank()` for P-scores in frequentist NMA; `gemtc::rank.probability()` and `multinma::posterior_rank_probs()` for Bayesian SUCRA and rank tables; `BUGSnet` for rankograms with built-in convergence diagnostics; `pcnetmeta` for component-NMA ranking when treatments are decomposable into pieces.

15.20.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.20.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.21 Prediction Intervals in Meta-Analysis

15.21.1 Introduction

A 95 % confidence interval on a pooled meta-analytic estimate describes uncertainty about the **mean** true effect across the included studies. The prediction interval (PI), in contrast, describes the plausible range of the **true effect in a future study** drawn from the same population, propagating both the uncertainty in the mean and the between-study heterogeneity τ^2. Under non-trivial heterogeneity the PI is substantially wider than the CI, and it is far more clinically meaningful: a clinician treating the next patient cares less about uncertainty in the average across past studies than about the range of effects they might plausibly see in their own clinical context. Modern meta-analysis reporting standards (PRISMA, GRADE) increasingly require the PI alongside the CI for any random-effects synthesis.

15.21.2 Prerequisites

A working understanding of random-effects meta-analysis, the role of between-study variance τ^2, and the conceptual difference between uncertainty about a mean and uncertainty about a future realisation.

15.21.3 Theory

The 95 % prediction interval under a random-effects meta-analysis with k studies is

$$\mathrm{PI}_{95\%} = \hat{\theta} \pm t_{0.975,k-2}\sqrt{\hat{\tau}^2 + \mathrm{SE}(\hat{\theta})^2}.$$

Two features distinguish it from the CI: the use of Student's t with $k-2$ degrees of freedom rather than the normal quantile, and the addition of $\hat{\tau}^2$ inside the square root. As heterogeneity grows the PI widens considerably while the CI is comparatively unchanged, so the gap between them is itself a useful diagnostic for how much pooling has — or has not — clarified the underlying effect.

15.21.4 Assumptions

The random-effects model is valid; τ^2 is reasonably estimated (typically by REML, which is the recommended default); future studies are exchangeable with the included ones, drawing their true effects from the same Normal distribution centred on the pooled mean.

15.21.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 10
yi ← rnorm(k, 0.4, 0.25)
vi ← 1 / sample(50:200, k, replace = TRUE)

res ← rma(yi, vi, method = "REML")
summary(res)

pi ← predict(res, addx = NULL)
pi[, c("pred", "pi.lb", "pi.ub")]
```

15.21.6 Output & Results

`predict()` on an `rma` object returns the pooled estimate, its confidence interval (for the mean), and the 95 % prediction interval (for a future study). Whenever $\tau^2 > 0$ the PI is strictly wider than the CI, and the gap is informative: a small gap means heterogeneity is negligible and the pool is informative about future studies, while a large gap means the pool tells you the average but not what to expect next.

15.21.7 Interpretation

A reporting sentence: “The pooled standardised mean difference was 0.38 (95 % CI 0.20 to 0.56) under random-effects REML pooling. The 95 % prediction interval was wider, $[-0.10, 0.86]$, indicating that in a new study the true effect could plausibly range from a small negative value to a large positive value, consistent with the moderate heterogeneity observed ($I^2 = 41\%$, $\tau^2 = 0.04$). Clinical interpretation should account for this dispersion rather than rely solely on the pooled mean.” Always report the CI and PI together for random-effects pools.

15.21.8 Practical Tips

- Always report both the CI (uncertainty about the mean) and the PI (uncertainty about a future study) in a random-effects meta-analysis; they answer different questions and the difference is itself informative.
- A wide PI alongside a narrow CI is a distinctive pattern indicating heterogeneity that does not diminish with pooling; it should be acknowledged explicitly in the discussion as a limitation of the synthesis.
- For meta-analyses with very few studies (typically $k < 5$), the PI is unstable and very wide because τ^2 is poorly estimated; report it but interpret with caution and consider Hartung–Knapp adjustment.

- A PI that crosses the null while the CI does not is a common and important pattern: it means the pooled mean is "significant" but a meaningful proportion of future studies might find no effect at all. This finding typically reshapes interpretation and should be highlighted, not buried.
- The Hartung–Knapp–Sidik–Jonkman adjustment widens the CI in small samples and modestly affects the PI; for $k \leq 20$, it is the recommended default.
- Display the PI graphically on the forest plot as a dashed bar around the pooled diamond (`addpred = TRUE` in `metafor::forest()`); seeing it visually conveys more than reporting it numerically.

15.21.9 R Packages Used

`metafor::predict.rma()` for the canonical PI computation alongside the CI; `metafor::forest()` with `addpred = TRUE` for graphical display; `meta::print.meta()` reports PI in summary output for `meta` objects; `dmetar::pi()` for an explicit prediction-interval helper used in teaching contexts.

15.21.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.21.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.22 Pooling Risk Ratios

15.22.1 Introduction

Risk ratios (RR) — also called relative risks — are more interpretable than odds ratios for common outcomes; they directly express the ratio of event probabilities between two arms. RR approximates OR only when events are rare (probability below ~10 %); for common outcomes, the two diverge substantially. Meta-analytic pooling of binary outcomes from cohort studies and randomised trials therefore typically proceeds on the log-RR scale, analogously to log-OR pooling.

15.22.2 Prerequisites

A working understanding of risk ratios, log transformations for ratio measures, and inverse-variance meta-analysis under random effects.

15.22.3 Theory

For a study with a_i events in $a_i + b_i = n_{1i}$ exposed and c_i events in $c_i + d_i = n_{2i}$ unexposed:

$$\log \mathrm{RR}_i = \log \frac{a_i / n_{1i}}{c_i / n_{2i}}, \qquad \mathrm{Var}(\log \mathrm{RR}_i) = \frac{1}{a_i} - \frac{1}{n_{1i}} + \frac{1}{c_i} - \frac{1}{n_{2i}}.$$

Inverse-variance pooling on the log-RR scale follows the same fixed-effect or random-effects distinction as any other meta-analysis. Back-transformation to the natural RR scale via $\exp$ produces interpretable summaries with confidence intervals that respect the multiplicative scale.

15.22.4 Assumptions

Studies are independent; log-RR is approximately Normal (typically a good approximation when each study has at least 5 events per arm); the population RR is the same across studies (fixed-effect) or distributed around a mean with between-study variance τ^2 (random-effects).

15.22.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 10
df ← data.frame(
  ai = rpois(k, 30), n1i = rep(100, k),
  ci = rpois(k, 20), n2i = rep(100, k)
)

es ← escalc(measure = "RR",
             ai = ai, n1i = n1i, ci = ci, n2i = n2i,
             data = df)

re ← rma(yi, vi, data = es, method = "REML")
cat("Pooled RR:", round(exp(re$b), 2),
    " 95% CI:", paste(round(exp(c(re$ci.lb, re$ci.ub)), 2), collapse = "-"), "\n")
cat("tau^2:", round(re$tau2, 3), "  I^2:", round(re$I2, 1), "%\n")
```

15.22.6 Output & Results

`escalc(measure = "RR")` computes per-study log-RR and its variance. `rma()` fits the random-effects model on the log scale; back-transformation gives the pooled RR with confidence intervals on the natural scale, plus τ^2 and I^2 for heterogeneity.

15.22.7 Interpretation

A reporting sentence: “Random-effects meta-analysis of 10 studies (REML, n total = 2{,}000) gave a pooled risk ratio of 1.48 (95 % CI 1.24 to 1.77) with moderate heterogeneity ($I^2 = 35\%$, $\tau^2 = 0.02$); the intervention is associated with a 48 % higher event rate, with

the 95 % CI excluding the null." Always report the heterogeneity statistics alongside the pooled estimate.

15.22.8 Practical Tips

- Use RR for cohort studies and randomised controlled trials, where baseline risk is meaningful; use OR for case-control studies, where RR is not estimable from the design.
- For rare events (event rate below 10 %), RR ≈ OR; for common events, the two diverge and the choice matters substantively.
- Zero-event studies require a continuity correction (`escalc(..., add = 0.5)`) or a rare-events method (Peto OR, Mantel-Haenszel without correction); standard log-RR is undefined when any cell is zero.
- Present forest plots and reports on the natural scale (RR), not the log scale; readers reason multiplicatively about ratios.
- Mantel-Haenszel pooling (`metafor::rma.mh()`) is a common alternative for sparse-data settings; it does not require log-transformation and handles zero events more gracefully.
- For individual-participant-data meta-analysis, fit a logistic regression with study as a random effect; this is more powerful than aggregate-data pooling when the IPD are available.

15.22.9 R Packages Used

`metafor::escalc()` with `measure = "RR"` and `rma()` for the canonical workflow; `meta::metabin()` as an alternative interface with cleaner forest-plot output; `metafor::rma.mh()` for Mantel-Haenszel pooling; `metafor::rma.peto()` for the Peto method on rare events.

15.22.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.22.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.23 Selection Models for Publication Bias

15.23.1 Introduction

Selection models address publication bias more rigorously than trim-and-fill or simple regression-based asymmetry tests. They posit an explicit weight function describing how the probability of publication depends on a study's p-value or effect magnitude, then re-estimate the pooled effect under that assumed selection mechanism to recover a bias-adjusted estimate. Where trim-and-fill is descriptive and triangular in its assumption, selection models are formally model-based and yield likelihood-ratio tests of the selection hypothesis itself. The cost is stronger parametric assumptions about the form of the selection process, which means selection models are most useful when the meta-analysis is large enough that the selection function can be reasonably identified from the data.

15.23.2 Prerequisites

A working understanding of publication bias as a selection process, maximum-likelihood estimation, and the logic of inverse-probability weighting on a sampled population.

15.23.3 Theory

Given observed effects y_i with standard errors SE_i from published studies, the observed density is

$$p_{\text{obs}}(y) = \frac{w(y, \mathrm{SE})\, p(y)}{\int w(y, \mathrm{SE})\, p(y)\, dy},$$

where $p(y)$ is the population density of true effects and $w(\cdot)$ is a selection weight function — typically a step function in the one-sided p-value, with selection probability dropping at conventional thresholds (0.025, 0.05, 0.10). Maximum-likelihood estimation of the joint (μ, τ^2, w) parameters then inverts the selection mechanism to recover the bias-adjusted pooled effect, and a likelihood-ratio test against the null of $w(\cdot) \equiv 1$ formally tests for selection.

15.23.4 Assumptions

The functional form of the selection weight is correctly specified (or at least a reasonable approximation), the meta-analysis contains enough studies to identify the selection parameters (typically 15–20+), and the studies are exchangeable conditional on their effect size and standard error.

15.23.5 R Implementation

```r
library(metafor)

set.seed(2026)
k <- 20
yi <- rnorm(k, 0.3, 0.15)
vi <- 1 / sample(40:250, k, replace = TRUE)

res <- rma(yi, vi, method = "REML")

sel <- selmodel(res, type = "stepfun",
                steps = c(0.025, 0.10, 0.50, 1.0),
                alternative = "greater")
sel
```

15.23.6 Output & Results

`selmodel()` returns the bias-adjusted pooled estimate, its standard error and confidence interval, the estimated selection weights at each step, and a likelihood-ratio test against the null of no selection. Comparing the adjusted and unadjusted estimates side by side quantifies how much the conclusion shifts under the assumed selection mechanism.

15.23.7 Interpretation

A reporting sentence: "A step-function selection model with thresholds at $p = 0.025, 0.10, 0.50$ provided strong evidence for selective publication (likelihood-ratio test vs no-selection, $\chi^2 = 13.1$, df $= 3$, $p = 0.004$). The bias-adjusted standardised mean difference was 0.19 (95 % CI 0.05 to 0.33), substantially attenuated from the unadjusted estimate of 0.32 (0.21 to 0.43). The result suggests publication bias inflates the unadjusted pooled effect by roughly 40 %." Always report both estimates and the LR test.

15.23.8 Practical Tips

- Selection models require at least 15–20 studies for stable estimates; below this threshold the selection function is poorly identified and confidence intervals become uninformatively wide.
- Try several weight-function structures (different step thresholds, smooth-monotone forms) as sensitivity analyses; the bias-adjusted estimate can be quite sensitive to the assumed selection mechanism, and convergent results across forms are reassuring.
- Always report both the unadjusted and the selection-adjusted estimates, plus the LR test for selection; the unadjusted estimate remains the primary, and the selection-adjusted estimate is sensitivity.
- Modern alternatives — PET-PEESE regression-based correction, p-uniform and p-curve, and Vevea–Hedges weight-function models — are useful when standard step-function selection models fit poorly or fail to converge.
- Bayesian selection models with informative priors on the weight parameters (e.g., in `RoBMA` and related packages) are a stable alternative when frequentist MLE is unstable.

- Selection models adjust only for the selection mechanism they are told to assume; other forms of bias (selective outcome reporting, dependent effect-size selection) require separate handling.

15.23.9 R Packages Used

`metafor::selmodel()` for the canonical step-function and smooth selection-model implementations; `weightr` for Vevea–Hedges weight-function models; `puniform` and `puniforms` for p-uniform corrections; `RoBMA` for Bayesian model-averaged publication-bias adjustment; `metasens` for limit-meta-analysis and Copas-model alternatives.

15.23.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.23.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.24 Standardised Mean Difference: Hedges' g

15.24.1 Introduction

Hedges' g is the small-sample-corrected standardised mean difference (SMD), the preferred effect-size metric for meta-analysis of continuous outcomes. Cohen's d — the unstandardised SMD — over-estimates the population effect in small samples; Hedges' correction factor J de-biases it. The two are nearly identical for studies with combined sample size above 20 or so, but g should be used routinely whenever any included study is small.

15.24.2 Prerequisites

A working understanding of standardised effect sizes (Cohen's d), pooled standard deviation, and the bias of d in small samples.

15.24.3 Theory

Cohen's d is

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{\text{pool}}}, \qquad s_{\text{pool}} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}.$$

Hedges' correction is $J = 1 - 3/(4\,\mathrm{df} - 1)$ (with $\mathrm{df} = n_1 + n_2 - 2$), and $g = Jd$. The correction is close to 1 for $\mathrm{df} > 20$ but can be substantially below 1 for very small studies. The variance of g is

$$\mathrm{Var}(g) = J^2\left[\frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)}\right].$$

15.24.4 Assumptions

Outcomes are approximately Normal within groups; group variances are approximately equal (use `SMDH` for unequal-variance variants); independent observations.

15.24.5 R Implementation

```r
library(metafor)

set.seed(2026)
n ← 10                    # small study
x1 ← rnorm(n, mean = 1, sd = 1)
x2 ← rnorm(n, mean = 0, sd = 1)

# Compute g and its variance via metafor
es ← escalc(measure = "SMD",
             m1i = mean(x1), sd1i = sd(x1), n1i = n,
             m2i = mean(x2), sd2i = sd(x2), n2i = n)
es

# Manual computation
s_pool ← sqrt(((n - 1) * var(x1) + (n - 1) * var(x2)) / (2 * n - 2))
d ← (mean(x1) - mean(x2)) / s_pool
J ← 1 - 3 / (4 * (2 * n - 2) - 1)
c(d = d, g = d * J)
```

15.24.6 Output & Results

`escalc(measure = "SMD")` returns Hedges' g (in the `yi` column) and its variance (`vi`) per study, ready for inverse-variance meta-analysis. The manual calculation confirms the J-correction; for $n_1 = n_2 = 10$, $J \approx 0.92$, reducing $d \approx 1.04$ to $g \approx 0.96$.

15.24.7 Interpretation

A reporting sentence: "Hedges' g for the simulated small study was 0.92 (95 % CI 0.01 to 1.83), corresponding to a large effect by Cohen's benchmarks ($d = 0.5$ medium, $d = 0.8$ large); the small-sample correction reduced the raw SMD from 1.04 to 0.92, reflecting the upward bias inherent in d for small studies." Always state the use of g vs d, especially when sample sizes are small.

15.24.8 Practical Tips

- Always use Hedges' g rather than Cohen's d for meta-analysis when any included study has $n_1 + n_2 < 20$; the bias correction is meaningful in that range.
- `metafor::escalc(measure = "SMD")` computes g and its variance in a single call; this is the standard workflow.
- For unequal variances (Levene's test rejects), use `measure = "SMDH"` which computes g using a pooled SD based on the separate group SDs.
- Report 95 % CIs alongside g; the effect size alone does not convey precision, and meta-analysis weighting depends on the precision.
- For pre-post or within-subject designs, use SMCR or SMCC variants in `metafor::escalc()`; they account for the within-subject correlation that ordinary SMD ignores.
- For binary outcomes, switch to log-odds ratio or risk ratio; SMD is for continuous outcomes only.

15.24.9 R Packages Used

`metafor::escalc()` and `rma()` for the canonical effect-size and meta-analysis workflow; `meta::metacont()` for an alternative interface with continuous-outcome conventions; `metan` for a tidymodels-flavoured wrapper; `dmetar` for additional meta-analysis tools and visualisations.

15.24.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.24.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.25 Subgroup Meta-Analysis

15.25.1 Introduction

Subgroup meta-analysis partitions the included studies according to a pre-specified categorical moderator — intervention intensity (low vs high dose), population (paediatric vs adult), risk of bias (low vs unclear vs high), or any clinically meaningful split — and pools within each level. The within-subgroup pooled effects describe how the intervention behaves in each stratum, and a formal between-subgroup test assesses whether the

moderator explains a measurable share of between-study heterogeneity. Subgroup analysis is the standard tool for categorical effect modification in evidence synthesis and is required by the Cochrane Handbook whenever the protocol pre-specifies a moderator hypothesis.

15.25.2 Prerequisites

A working understanding of pooled meta-analysis (preferably random-effects), the heterogeneity statistics τ^2, I^2, and Q, and the difference between within-group and between-group sources of variation in stratified analyses.

15.25.3 Theory

Within each subgroup $g = 1, \dots, G$ the studies are pooled to obtain $\hat{\theta}_g$ and its standard error. The between-subgroup heterogeneity statistic is

$$Q_B = \sum_{g} w_g (\hat{\theta}_g - \hat{\theta})^2,$$

with $G - 1$ degrees of freedom; a significant Q_B indicates the moderator accounts for some of the between-study variance. Equivalently, fitting a mixed-effects meta-regression with subgroup as a categorical moderator produces a Q_M test that recovers exactly this comparison and is the modern preferred implementation, because it can also share a common τ^2 across subgroups for more stable inference.

15.25.4 Assumptions

The moderator is pre-specified rather than chosen post hoc, within-subgroup studies are reasonably homogeneous (small within-group τ^2), and each subgroup contains enough studies — typically at least four to five — to support a stable within-group pool. Subgroup analyses with only one or two studies per stratum are largely uninformative and should be reported only descriptively.

15.25.5 R Implementation

```r
library(metafor)

set.seed(2026)
k ← 16
subgroup ← factor(rep(c("low_dose", "high_dose"), each = k/2))
yi ← c(rnorm(k/2, 0.2, 0.15),
       rnorm(k/2, 0.5, 0.15))
vi ← 1 / sample(50:200, k, replace = TRUE)

df ← data.frame(yi, vi, subgroup)

lapply(split(df, df$subgroup), function(d) {
  rma(yi, vi, data = d, method = "REML")$b
})

res_mod ← rma(yi, vi, mods = ~ subgroup, data = df, method = "REML")
summary(res_mod)
```

15.25.6 Output & Results

The script returns per-subgroup pooled estimates from independent random-effects pools and the joint mixed-effects meta-regression with subgroup as moderator. The latter reports the omnibus Q_M test for between-subgroup heterogeneity, the per-level coefficients, and the residual heterogeneity remaining after the moderator is included — the canonical reporting set for subgroup analyses.

15.25.7 Interpretation

A reporting sentence: "Pre-specified subgroup analysis showed significant between-subgroup heterogeneity ($Q_M = 18.4$, df $= 1$, $p < 0.001$): high-dose studies had a pooled standardised mean difference of 0.52 (95 % CI 0.39 to 0.65) compared with 0.18 (0.08 to 0.28) in low-dose studies, supporting the hypothesis that dose moderates the treatment effect. The moderator explained 47 % of the between-study heterogeneity (R^2); residual I^2 was 18 %." Always report both the within-subgroup pools and the between-subgroup test.

15.25.8 Practical Tips

- Pre-specify all subgroup analyses in the PROSPERO protocol; post-hoc subgrouping is exploratory and should be flagged as such — never as primary inference.
- Small per-subgroup study counts produce unstable pooled estimates; the Cochrane Handbook recommends at least 10 studies per subgroup before drawing conclusions, and warns explicitly against single-stratum subgroup pools.
- Prefer meta-regression over subgroup analysis whenever the moderator is continuous (age, dose, year); dichotomising loses information and depends on an arbitrary cutpoint.
- Report the between-subgroup Q_M or Q_B p-value as the primary inference; comparing within-subgroup p-values across strata ("significant in one, not the

other") is a well-known statistical fallacy because tests in small subgroups are underpowered.
- Use a shared τ^2 across subgroups (the default in mixed-effects meta-regression) rather than separate estimates per subgroup unless heterogeneity is clearly different in each; the shared specification is more efficient and more conservative.
- When subgroup analysis is exploratory, apply a multiplicity adjustment or interpret with explicit caution; the false-positive rate inflates rapidly with the number of subgroup tests.

15.25.9 R Packages Used

`metafor::rma()` with `mods = ~ factor(subgroup)` for the canonical mixed-effects implementation; `meta::metabin()` and `meta::metacont()` with `subgroup =` for built-in subgroup pooling and forest-plot grouping; `dmetar` for Knapp–Hartung-adjusted subgroup analyses; `robumeta` for robust-variance subgroup pooling under correlated effects.

15.25.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.25.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.26 Tau-Squared Estimation

15.26.1 Introduction

The random-effects meta-analysis requires an estimate of τ^2, the between-study variance. Several estimators exist with different bias-variance trade-offs; the choice affects pooled estimates, CIs, and downstream inference. REML is the modern default; DerSimonian-Laird (DL) the historical default.

15.26.2 Prerequisites

Random-effects meta-analysis; method-of-moments vs maximum likelihood.

15.26.3 Theory

Common estimators: - **DerSimonian-Laird (DL)**: method-of-moments; fastest, historically default. - **REML (Restricted Maximum Likelihood)**: likelihood-based, generally less biased than DL; modern default. - **Paule-Mandel (PM)**: iterative method-of-moments; often similar to REML. - **Sidik-Jonkman (SJ)**: analytic, robust to model misspecification. - **Empirical Bayes (EB)**: iterative Bayes-style estimator.

DL systematically underestimates τ^2 under moderate heterogeneity; REML is close to unbiased.

15.26.4 Assumptions

True effects are drawn from a distribution (typically normal) with common variance τ^2; within-study variances are known.

15.26.5 R Implementation

```
library(metafor)

set.seed(2026)
k ← 10
n_per ← sample(30:200, k, replace = TRUE)
true_theta ← rnorm(k, 0.3, 0.25)
yi ← true_theta + rnorm(k, 0, sqrt(4 / n_per))
vi ← 4 / n_per

methods ← c("DL", "REML", "PM", "SJ", "EB")
tau2_est ← sapply(methods, function(m) {
  rma(yi, vi, method = m)$tau2
})
round(tau2_est, 3)
```

15.26.6 Output & Results

τ^2 estimates under the five methods. DL typically gives the smallest value; REML, PM, SJ, EB are often close.

15.26.7 Interpretation

"τ^2 estimates ranged from 0.041 (DL) to 0.061 (SJ); we report REML ($\tau^2 = 0.055$, 95 % CI 0.012-0.142) as the primary estimator following modern meta-analysis recommendations."

15.26.8 Practical Tips

- Use REML by default; it is less biased than DL under most conditions.
- Hartung-Knapp adjustment (`test = "knha"` in metafor) improves CI coverage, especially with few studies.
- Compare multiple estimators as sensitivity; if DL and REML disagree substantially, heterogeneity is likely real.

- Report 95 % CI on τ^2 (profile likelihood or Q-profile methods) alongside the point estimate.
- For very few studies ($k < 5$), any τ^2 estimate is unreliable; use Bayesian methods with informative priors.

15.26.9 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.26.10 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.27 Trim-and-Fill

15.27.1 Introduction

Trim-and-fill, introduced by Duval and Tweedie in 2000, is the most widely used algorithmic approach for adjusting a pooled meta-analytic estimate when the funnel plot looks asymmetric. The idea is intuitive: if some smaller studies on the side opposite the dominant tail of the funnel appear to be missing — possibly because they were never published, never indexed, or never reached the systematic-review screening stage — the procedure imputes mirror-image counterparts and re-pools the augmented set. The difference between the unadjusted and adjusted estimates quantifies how sensitive the conclusion is to the presence of these hypothetical missing studies. Crucially, trim-and-fill is a sensitivity analysis, not a method of bias correction in the strict statistical sense; it cannot recover the true unbiased effect.

15.27.2 Prerequisites

Familiarity with funnel plots, the small-study-effect concept, the inverse-variance random-effects pool, and the distinction between sensitivity analysis and bias correction.

15.27.3 Theory

The algorithm proceeds iteratively:

1. Estimate the number of asymmetric studies in the dominant tail using one of three rank-based estimators (L_0, R_0, Q_0).

2. Trim those studies from the dominant side and re-estimate the pooled effect; this estimate becomes the new symmetry centre.
3. Fill the analysis by reflecting the trimmed studies across the new centre to the opposite side.
4. Re-pool with the augmented dataset to obtain the adjusted estimate.

Iterating to convergence yields the final imputed studies and adjusted pooled estimate. The number of imputed studies is itself an interpretable summary: zero imputations indicate no detectable asymmetry, and a large number indicates substantial sensitivity.

15.27.4 Assumptions

The observed funnel asymmetry is attributable to selection or publication bias rather than to true between-study heterogeneity correlated with study size; the underlying effect distribution is mirror-symmetric around the true effect; and the pooled model is otherwise correctly specified.

15.27.5 R Implementation

```
library(metafor)

set.seed(2026)
k ← 18
yi ← rnorm(k, 0.35, 0.25)
vi ← 1 / sample(15:300, k, replace = TRUE)

keep ← !(sqrt(vi) > 0.1 & abs(yi) < 0.15)
yi_b ← yi[keep]; vi_b ← vi[keep]

res ← rma(yi_b, vi_b, method = "REML")
tf  ← trimfill(res)

data.frame(method = c("Original", "Trim-and-fill"),
           estimate = round(c(res$b, tf$b), 3),
           ci = c(sprintf("[%.2f, %.2f]", res$ci.lb, res$ci.ub),
                  sprintf("[%.2f, %.2f]", tf$ci.lb, tf$ci.ub)))

funnel(tf, main = "Trim-and-fill adjusted funnel")
```

15.27.6 Output & Results

`trimfill()` returns the number of imputed studies, the adjusted pooled estimate with its confidence interval, and an augmented dataset suitable for plotting on a funnel diagram. Reporting both the original pool and the adjusted pool side by side — together with the count of imputations — lets readers see exactly how much the conclusion shifts when the missing studies are introduced.

15.27.7 Interpretation

A reporting sentence: “Trim-and-fill (Duval and Tweedie, L_0 estimator) imputed four potentially missing studies on the left side of the funnel; the adjusted standardised mean difference was 0.18 (95 % CI 0.03 to 0.33), notably attenuated from the unadjusted

estimate of 0.32 (0.21 to 0.43). The result suggests the pooled conclusion is sensitive to plausible publication bias, and Egger's regression and PET-PEESE were reported as triangulating sensitivity analyses." Always report the unadjusted estimate as primary and the adjusted estimate as sensitivity.

15.27.8 Practical Tips

- Trim-and-fill is a sensitivity analysis, not a proof of publication bias; report it as such, and never use the adjusted estimate as the primary effect.
- The method performs poorly when funnel asymmetry is driven by genuine between-study heterogeneity correlated with study size; in such settings it can spuriously imply bias.
- Apply alongside Egger's regression test and PET-PEESE for triangulation; consistent signals strengthen the bias interpretation, divergent signals warn against over-interpretation.
- Report both original and adjusted estimates with their CIs and the count of imputed studies; an imputation count of zero indicates the funnel was already symmetric and no adjustment was needed.
- The L_0 estimator (default in `metafor`) is the most widely used; R_0 is more conservative; Q_0 is rarely used in practice. Sensitivity to estimator choice should be noted when results are borderline.
- Selection-model approaches (such as Vevea–Hedges weight-function models) and PET-PEESE are increasingly preferred over trim-and-fill in methodological literature; consider them when the small-study effect is consequential to your conclusion.

15.27.9 R Packages Used

`metafor::trimfill()` for the canonical implementation with all three estimators and `funnel()` overlay; `meta::trimfill()` for the alternative interface within the `meta` ecosystem; `metasens` for trim-and-fill alongside copas and limit-meta-analysis bias-adjustment methods; `weightr` and `puniform` for selection-model alternatives that handle complex bias mechanisms.

15.27.10 For Reviewers

Use the shared reviewer checklist, together with this method's assumptions and practical tips.

Method-specific reviewer notes have not yet been developed for this entry. The links above support a structured review; they do not certify the example or a manuscript's application of the method.

15.27.11 See also — labs in this chapter

- Systematic reviews and PRISMA
- Meta-analysis basics
- Network meta-analysis

15.28 Week 4, Session 2 — Meta-analysis basics

Inference lab: Hypothesis → Visualise → Assumptions → Conduct → Conclude.

15.28.1 Learning objectives

- Fit a random-effects meta-analysis with `metafor::rma`.
- Produce forest and funnel plots, and interpret between-study heterogeneity (τ^2, I^2).
- Assess small-study effects and publication bias informally.

15.28.2 Prerequisites

Comfort with effect sizes on the log scale.

15.28.3 Background

Meta-analysis pools effect estimates across studies to sharpen an overall answer. Fixed-effect models assume all studies estimate one true effect; random-effects models assume true effects vary around a common mean. In biomedical research, heterogeneity is the rule rather than the exception, so random effects are the default — and between-study variance (τ^2) and the proportion of total variance due to heterogeneity (I^2) become part of the primary reporting, not an afterthought.

Forest plots display each study's estimate with its confidence interval next to the pooled estimate. Funnel plots display study precision against effect size; asymmetry hints at small-study effects and potential publication bias. Neither plot settles the question — Egger's test, trim-and-fill, and careful inspection of which small studies are missing are all part of the routine.

15.28.4 Setup

```
library(tidyverse)
library(metafor)
set.seed(42)
theme_set(theme_minimal(base_size = 12))
```

15.28.5 1. Hypothesis

H_0: pooled log risk ratio = 0 (no overall effect). H_1: pooled log RR ≠ 0. $\alpha = 0.05$.

15.28.6 2. Visualise — use a built-in dataset

```
dat ← escalc(measure = "RR",
             ai = tpos, bi = tneg, ci = cpos, di = cneg,
             data = dat.bcg, slab = paste(author, year))
```

```
forest(rma(yi, vi, data = dat), header = TRUE, cex = 0.8)
```

15.28.7 3. Assumptions

- Between-study heterogeneity exists ($\tau^2 > 0$) in virtually every real meta-analysis — a random-effects model absorbs it, a fixed-effect model does not.
- Funnel-plot asymmetry does not prove publication bias; it is consistent with several mechanisms.

15.28.8 4. Conduct

```
fit ← rma(yi, vi, data = dat, method = "REML")
fit
```

```
funnel(fit, main = "Funnel plot — BCG vaccine dataset")
regtest(fit, model = "lm")
```

15.28.9 5. Concluding statement

```
rr   ← exp(fit$b[1])
lo   ← exp(fit$ci.lb)
hi   ← exp(fit$ci.ub)
```

> Across `nrow(dat)` trials of the BCG vaccine, the pooled risk ratio for tuberculosis was `round(rr, 2)` (95% CI `round(lo, 2)` to `round(hi, 2)`; τ^2 = `round(fit$tau2, 2)`, I^2 = `round(fit$I2, 1)` %). Between-trial heterogeneity was substantial; the pooled estimate should be interpreted with the heterogeneity statistics in view.

15.28.10 Common pitfalls

- Fixed-effect pooling when between-study heterogeneity is real.
- Quoting a pooled effect without its heterogeneity measures.
- Treating funnel-plot symmetry as a formal test rather than a visual prompt.

15.28.11 Further reading

- Viechtbauer W. (2010). *Conducting meta-analyses in R with the metafor package.* J Stat Softw.
- Higgins JPT et al. *Cochrane Handbook for Systematic Reviews of Interventions.*

15.28.12 Session info

```
sessionInfo()
```

15.28.13 See also — chapter index

- Methods in this chapter
- Find your method (Chapter 0)

15.29 Week 4, Session 3 — Network meta-analysis

Inference lab: Hypothesis → Visualise → Assumptions → Conduct → Conclude.

15.29.1 Learning objectives

- Fit a network meta-analysis (NMA) using `netmeta`.
- Draw and read the network graph.
- Derive SUCRA rankings and interpret them cautiously.

15.29.2 Prerequisites

Pairwise meta-analysis (W4 S2).

15.29.3 Background

Network meta-analysis extends pairwise meta-analysis to more than two treatments by combining direct comparisons (from head-to-head trials) with indirect comparisons (inferred through a shared comparator). The central assumption is *transitivity*: the studies compared indirectly are exchangeable — no moderator distinguishes them in a way that systematically shifts the effect. The statistical analogue, *consistency*, checks whether direct and indirect estimates agree.

SUCRA (surface under the cumulative ranking curve) summarises each treatment's rank distribution on a 0-1 scale, where 1 is best. SUCRA values compress a lot of uncertainty into a single number; they are useful for signalling but should be reported with interval estimates for every pairwise comparison.

15.29.4 Setup

```
library(tidyverse)
library(netmeta)
set.seed(42)
theme_set(theme_minimal(base_size = 12))
```

15.29.5 1. Hypothesis

H_0: all treatments in the network have equal efficacy. H_1: at least two differ.

15.29.6 2. Visualise

```
data(Senn2013)
head(Senn2013)

net ← netmeta(TE, seTE, treat1, treat2, studlab,
              data = Senn2013, sm = "MD",
              common = FALSE, random = TRUE,
              reference.group = "plac")
```

```
netgraph(net, plastic = FALSE, thickness = "number.of.studies",
         points = TRUE, cex.points = 2, cex = 0.8)
```

15.29.7 3. Assumptions

```
decomp.design(net)
```

The within-design heterogeneity and between-design inconsistency variances are the key diagnostics.

15.29.8 4. Conduct

```
summary(net)
rank_df ← netrank(net, small.values = "good")
rank_df
```

15.29.9 5. Concluding statement

```
ranked ← as_tibble(rank_df$ranking.random, rownames = "treatment") ▷
  rename(sucra = value) ▷
  arrange(desc(sucra))
top ← head(ranked, 1)
```

> In a network meta-analysis of `length(unique(c(Senn2013$treat1, Senn2013$treat2)))` treatments across `length(unique(Senn2013$studlab))` studies (`nrow(Senn2013)` pairwise comparisons), the treatment with the highest SUCRA for HbA1c lowering was **`top$treatment`** (SUCRA = `round(top$sucra, 2)`). SUCRA rankings should be interpreted alongside the pairwise effect estimates and their intervals, not as a standalone conclusion.

15.29.10 Common pitfalls

- Treating SUCRA as if it were the effect size.
- Ignoring the transitivity assumption because the output compiled.
- Mixing direct and indirect evidence without quantifying inconsistency.

15.29.11 Further reading

- Rücker G, Schwarzer G. (2015). *Ranking treatments in frequentist network meta-analysis.*
- Salanti G. (2012). *Indirect and mixed-treatment comparison…*

15.29.12 Session info

```
sessionInfo()
```

15.29.13 See also — chapter index

- Methods in this chapter
- Find your method (Chapter 0)

15.30 Week 4, Session 1 — Systematic reviews and PRISMA

Workflow lab: Goal → Approach → Execution → Check → Report.

15.30.1 Learning objectives

- Frame a systematic-review question using PICO.
- Draft a reproducible search strategy across at least two databases.
- Produce a PRISMA flow diagram from counts at each screening stage.

15.30.2 Prerequisites

None beyond reading a paper.

15.30.3 Background

A systematic review is itself a study: it has a protocol, a pre-specified search strategy, explicit inclusion and exclusion criteria, and a plan for synthesis. The PRISMA 2020 statement is the reporting standard, which means the diagram showing what you searched, what you excluded, and why is not optional. PROSPERO (https://www.crd.york.ac.uk/prospero/) is the standard registry for review protocols — registering early protects you from post-hoc drift.

Good search strategies combine controlled-vocabulary terms (MeSH on PubMed, Emtree on Embase) with free-text terms, connect concepts with Boolean AND, and expand synonyms with OR. A librarian-reviewed strategy typically outperforms a DIY one by a wide margin in recall; co-authoring with a health-sciences librarian is the single best investment in review quality.

15.30.4 Setup

```
library(tidyverse)
set.seed(42)
theme_set(theme_minimal(base_size = 12))
```

15.30.5 1. Goal

Illustrate a PICO-framed question, sketch a search strategy, and simulate screening counts to generate a PRISMA flow diagram.

15.30.6 2. Approach

A fictional question in PICO form:

> In adults with type 2 diabetes (**P**), does metformin (**I**) compared with placebo (**C**) reduce all-cause mortality (**O**) at 24 months?

A sketch of a search strategy (PubMed syntax):

```
("Diabetes Mellitus, Type 2"[MeSH] OR "type 2 diabetes"[tiab])
AND (metformin[MeSH] OR metformin[tiab])
AND (randomised[tiab] OR randomized[tiab] OR trial[tiab] OR RCT[tiab])
```

15.30.7 3. Execution — simulated screening counts

```
prisma ← tibble(
  stage = c("Identified (PubMed)", "Identified (Embase)",
            "After deduplication",
            "Title/abstract screened", "Full-text assessed",
            "Included in qualitative synthesis",
            "Included in meta-analysis"),
  n = c(812, 640, 1098, 1098, 84, 22, 18)
)
prisma
```

```
prisma ▷
  mutate(stage = factor(stage, levels = rev(stage))) ▷
  ggplot(aes(n, stage)) +
  geom_col(fill = "#1a73e8", alpha = 0.8) +
  geom_text(aes(label = n), hjust = -0.1, size = 3.5) +
  scale_x_continuous(expand = expansion(mult = c(0, 0.15))) +
  labs(x = "n", y = NULL)
```

15.30.8 4. Check

PRISMA 2020 checklist pointer (items most often missed):

- Item 5: eligibility criteria pre-specified.
- Item 6: information sources with date of last search.
- Item 7: complete search strategy for every database, including limits.
- Item 16a: reasons for excluding full-text articles.

15.30.9 5. Report

> A systematic review of the effect of metformin on all-cause mortality in adults with type 2 diabetes was conducted following PRISMA 2020. Database searches identified 812 records from PubMed and 640 from Embase. After deduplication, 1098 records were screened by title and abstract; 84 full texts were assessed, and 22 studies were included in the qualitative synthesis, of which 18 were pooled by meta-analysis.

15.30.10 Common pitfalls

- Ad-hoc searches that cannot be re-run.

- Failing to double-screen at each stage (reviewer drift).
- Reporting included-study counts without the flow that produced them.

15.30.11 Further reading

- Page MJ et al. (2021). *The PRISMA 2020 statement.* BMJ.
- Cochrane Handbook for Systematic Reviews of Interventions.

15.30.12 Session info

```
sessionInfo()
```

15.30.13 See also — chapter index

- Methods in this chapter
- Find your method (Chapter 0)