<https://en.wikipedia.org/wiki/Meta-analysis>

# Meta-analysis

From Wikipedia, the free encyclopedia

[Jump to navigation](https://en.wikipedia.org/wiki/Meta-analysis#mw-head) [Jump to search](https://en.wikipedia.org/wiki/Meta-analysis#p-search)

For the process in historical linguistics known as metanalysis, see [Rebracketing](https://en.wikipedia.org/wiki/Rebracketing).

Graphical summary of a meta analysis of over 1,000 cases of [diffuse intrinsic pontine glioma](https://en.wikipedia.org/wiki/Diffuse_intrinsic_pontine_glioma) and other pediatric gliomas, in which information about the [mutations](https://en.wikipedia.org/wiki/Mutation) involved as well as generic outcomes were distilled from the underlying [primary literature](https://en.wikipedia.org/wiki/Primary_literature).

A **meta-analysis** is a statistical analysis that combines the results of multiple [scientific studies](https://en.wikipedia.org/wiki/Randomized_controlled_trial).

The basic tenet behind meta-analyses is that there is a common truth behind all conceptually similar scientific studies, but which has been measured with a certain error within individual studies. The aim then is to use approaches from [statistics](https://en.wikipedia.org/wiki/Statistics) to derive a pooled estimate closest to the unknown common truth based on how this error is perceived. In essence, all existing methods yield a [weighted average](https://en.wikipedia.org/wiki/Weighted_average) from the results of the individual studies and what differs is the manner in which these weights are allocated and also the manner in which the uncertainty is computed around the point estimate thus generated. In addition to providing an estimate of the unknown common truth, meta-analysis has the capacity to contrast results from different studies and identify patterns among study results, sources of disagreement among those results, or other interesting relationships that may come to light in the context of multiple studies.[[1]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-1)

A key benefit of this approach is the aggregation of information leading to a higher [statistical power](https://en.wikipedia.org/wiki/Statistical_power) and more robust point estimate than is possible from the measure derived from any individual study. However, in performing a meta-analysis, an investigator must make choices which can affect the results, including deciding how to search for studies, selecting studies based on a set of objective criteria, dealing with incomplete data, analyzing the data, and accounting for or choosing not to account for [publication bias](https://en.wikipedia.org/wiki/Publication_bias).[[2]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-2)

Meta-analyses are often, but not always, important components of a [systematic review](https://en.wikipedia.org/wiki/Systematic_review) procedure. For instance, a meta-analysis may be conducted on several clinical trials of a medical treatment, in an effort to obtain a better understanding of how well the treatment works. Here it is convenient to follow the terminology used by the [Cochrane Collaboration](https://en.wikipedia.org/wiki/Cochrane_Collaboration),[[3]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-3) and use "meta-analysis" to refer to statistical methods of combining evidence, leaving other aspects of 'research synthesis' or 'evidence synthesis', such as combining information from qualitative studies, for the more general context of systematic reviews. A meta-analysis is a [secondary source](https://en.wikipedia.org/wiki/Secondary_source).[[4]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-4)[[5]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-5)

## History

The historical roots of meta-analysis can be traced back to 17th century studies of astronomy,[[6]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Plackett1958-6) while a paper published in 1904 by the statistician [Karl Pearson](https://en.wikipedia.org/wiki/Karl_Pearson) in the [*British Medical Journal*](https://en.wikipedia.org/wiki/British_Medical_Journal)[[7]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-7) which collated data from several studies of typhoid inoculation is seen as the first time a meta-analytic approach was used to aggregate the outcomes of multiple clinical studies.[[8]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Nordmann2012-8)[[9]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-9) The first meta-analysis of all conceptually identical experiments concerning a particular research issue, and conducted by independent researchers, has been identified as the 1940 book-length publication [*Extrasensory Perception After Sixty Years*](https://en.wikipedia.org/wiki/Extrasensory_Perception_After_Sixty_Years), authored by Duke University psychologists [J. G. Pratt](https://en.wikipedia.org/wiki/Joseph_Gaither_Pratt), [J. B. Rhine](https://en.wikipedia.org/wiki/Joseph_Banks_Rhine), and associates.[[10]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-10) This encompassed a review of 145 reports on [ESP](https://en.wikipedia.org/wiki/Extrasensory_perception) experiments published from 1882 to 1939, and included an estimate of the influence of unpublished papers on the overall effect (the [*file-drawer problem*](https://en.wikipedia.org/wiki/Meta-analysis#Publication_bias:_the_file_drawer_problem)). Although meta-analysis is widely used in [epidemiology](https://en.wikipedia.org/wiki/Epidemiology) and [evidence-based medicine](https://en.wikipedia.org/wiki/Evidence-based_medicine) today, a meta-analysis of a medical treatment was not published until 1955. In the 1970s, more sophisticated analytical techniques were introduced in [educational research](https://en.wikipedia.org/wiki/Educational_research), starting with the work of [Gene V. Glass](https://en.wikipedia.org/wiki/Gene_V._Glass), [Frank L. Schmidt](https://en.wikipedia.org/wiki/Frank_L._Schmidt) and [John E. Hunter](https://en.wikipedia.org/wiki/John_E._Hunter).

The term "meta-analysis" was coined in 1976 by the statistician [Gene V. Glass](https://en.wikipedia.org/wiki/Gene_V._Glass),[[11]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Glass-11) who stated *"my major interest currently is in what we have come to call ...the meta-analysis of research. The term is a bit grand, but it is precise and apt ... Meta-analysis refers to the analysis of analyses"*. Although this led to him being widely recognized as the modern founder of the method, the methodology behind what he termed "meta-analysis" predates his work by several decades.[[12]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-12)[[13]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-13) The statistical theory surrounding meta-analysis was greatly advanced by the work of [Nambury S. Raju](https://en.wikipedia.org/wiki/Nambury_S._Raju), [Larry V. Hedges](https://en.wikipedia.org/wiki/Larry_V._Hedges), Harris Cooper, [Ingram Olkin](https://en.wikipedia.org/wiki/Ingram_Olkin), [John E. Hunter](https://en.wikipedia.org/wiki/John_E._Hunter), [Jacob Cohen](https://en.wikipedia.org/wiki/Jacob_Cohen_%28born_1923%29), [Thomas C. Chalmers](https://en.wikipedia.org/wiki/Thomas_C._Chalmers), [Robert Rosenthal](https://en.wikipedia.org/wiki/Robert_Rosenthal_%28psychologist%29), [Frank L. Schmidt](https://en.wikipedia.org/wiki/Frank_L._Schmidt), and Douglas G. Bonett.

## Advantages

Conceptually, a meta-analysis uses a statistical approach to combine the results from multiple studies in an effort to increase power (over individual studies), improve estimates of the size of the effect and/or to resolve uncertainty when reports disagree. A meta-analysis is a statistical overview of the results from one or more systematic reviews. Basically, it produces a weighted average of the included study results and this approach has several advantages:

* Results can be generalized to a larger population
* The precision and accuracy of estimates can be improved as more data is used. This, in turn, may increase the statistical power to detect an effect
* Inconsistency of results across studies can be quantified and analyzed. For instance, inconsistency may arise from [sampling error](https://en.wikipedia.org/wiki/Sampling_error), or study results (partially) influenced by differences between study protocols
* Hypothesis testing can be applied on summary estimates
* Moderators can be included to explain variation between studies
* The presence of [publication bias](https://en.wikipedia.org/wiki/Publication_bias) can be investigated

## Problems

A meta-analysis of several small studies does not predict the results of a single large study.[[14]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-14) Some have argued that a weakness of the method is that sources of bias are not controlled by the method: a good meta-analysis cannot correct for poor design or bias in the original studies.[[15]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Slavin-15) This would mean that only methodologically sound studies should be included in a meta-analysis, a practice called 'best evidence synthesis'.[[15]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Slavin-15) Other meta-analysts would include weaker studies, and add a study-level predictor variable that reflects the methodological quality of the studies to examine the effect of study quality on the effect size.[[16]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-16) However, others have argued that a better approach is to preserve information about the variance in the study sample, casting as wide a net as possible, and that methodological selection criteria introduce unwanted subjectivity, defeating the purpose of the approach.[[17]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-17)

## Steps in a meta-analysis

A meta-analysis is usually preceded by a systematic review, as this allows identification and critical appraisal of all the relevant evidence (thereby limiting the risk of bias in summary estimates). The general steps are then as follows:

1. Formulation of the research question, e.g. using the PICO model (Population, Intervention, Comparison, Outcome).
2. Search of literature
3. Selection of studies ('incorporation criteria')
	1. Based on quality criteria, e.g. the requirement of randomization and blinding in a clinical trial
	2. Selection of specific studies on a well-specified subject, e.g. the treatment of breast cancer.
	3. Decide whether unpublished studies are included to avoid publication bias ([file drawer problem](https://en.wikipedia.org/wiki/Meta-analysis#Publication_bias:_the_file_drawer_problem))
4. Decide which dependent variables or summary measures are allowed. For instance, when considering a meta-analysis of published (aggregate) data:
	1. Differences (discrete data)
	2. Means (continuous data)
	3. [Hedges' *g*](https://en.wikipedia.org/wiki/Effect_size#Hedges'_g) is a popular summary measure for continuous data that is standardized in order to eliminate scale differences, but it incorporates an index of variation between groups:
		1. δ = μ t − μ c σ , {\displaystyle \delta ={\frac {\mu \_{t}-\mu \_{c}}{\sigma }},} in which μ t {\displaystyle \mu \_{t}} is the treatment mean, μ c {\displaystyle \mu \_{c}} is the control mean, σ 2 {\displaystyle \sigma ^{2}} the pooled variance.
5. Selection of a meta-analysis model, e.g. fixed effect or random effects meta-analysis.
6. Examine sources of between-study heterogeneity, e.g. using subgroup analysis or [meta-regression](https://en.wikipedia.org/wiki/Meta-regression).

Formal guidance for the conduct and reporting of meta-analyses is provided by the [Cochrane Handbook](http://training.cochrane.org/handbook).

For reporting guidelines, see the [Preferred Reporting Items for Systematic Reviews and Meta-Analyses](https://en.wikipedia.org/wiki/Preferred_Reporting_Items_for_Systematic_Reviews_and_Meta-Analyses) (PRISMA) statement.[[32]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-32)

## Methods and assumptions

### Approaches

In general, two types of evidence can be distinguished when performing a meta-analysis: [individual participant data](https://en.wikipedia.org/wiki/Individual_participant_data) (IPD), and aggregate data (AD). The aggregate data can be direct or indirect.

AD is more commonly available (e.g. from the literature) and typically represents summary estimates such as odds ratios or relative risks. This can be directly synthesized across conceptually similar studies using several approaches (see below). On the other hand, indirect aggregate data measures the effect of two treatments that were each compared against a similar control group in a meta-analysis. For example, if treatment A and treatment B were directly compared vs placebo in separate meta-analyses, we can use these two pooled results to get an estimate of the effects of A vs B in an indirect comparison as effect A vs Placebo minus effect B vs Placebo.

IPD evidence represents raw data as collected by the study centers. This distinction has raised the need for different meta-analytic methods when evidence synthesis is desired, and has led to the development of one-stage and two-stage methods.[[33]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-33) In one-stage methods the IPD from all studies are modeled simultaneously whilst accounting for the clustering of participants within studies. Two-stage methods first compute summary statistics for AD from each study and then calculate overall statistics as a weighted average of the study statistics. By reducing IPD to AD, two-stage methods can also be applied when IPD is available; this makes them an appealing choice when performing a meta-analysis. Although it is conventionally believed that one-stage and two-stage methods yield similar results, recent studies have shown that they may occasionally lead to different conclusions.[[34]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-pmid23585842-34)[[35]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-35)

### Statistical models for aggregate data

#### Direct evidence: Models incorporating study effects only

##### Fixed effects model

The fixed effect model provides a weighted average of a series of study estimates. The inverse of the estimates' variance is commonly used as study weight, so that larger studies tend to contribute more than smaller studies to the weighted average. Consequently, when studies within a meta-analysis are dominated by a very large study, the findings from smaller studies are practically ignored.[[36]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-pmid11884693-36) Most importantly, the fixed effects model assumes that all included studies investigate the same population, use the same variable and outcome definitions, etc. This assumption is typically unrealistic as research is often prone to several sources of heterogeneity; e.g. treatment effects may differ according to locale, dosage levels, study conditions, ...

##### Random effects model

A common model used to synthesize heterogeneous research is the random effects model of meta-analysis. This is simply the weighted average of the effect sizes of a group of studies. The weight that is applied in this process of weighted averaging with a random effects meta-analysis is achieved in two steps:[[37]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-37)

1. Step 1: Inverse variance weighting
2. Step 2: Un-weighting of this inverse variance weighting by applying a random effects variance component (REVC) that is simply derived from the extent of variability of the effect sizes of the underlying studies.

This means that the greater this variability in effect sizes (otherwise known as heterogeneity), the greater the un-weighting and this can reach a point when the random effects meta-analysis result becomes simply the un-weighted average effect size across the studies. At the other extreme, when all effect sizes are similar (or variability does not exceed sampling error), no REVC is applied and the random effects meta-analysis defaults to simply a fixed effect meta-analysis (only inverse variance weighting).

The extent of this reversal is solely dependent on two factors:[[38]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-ReferenceA-38)

1. Heterogeneity of precision
2. Heterogeneity of effect size

Since neither of these factors automatically indicates a faulty larger study or more reliable smaller studies, the re-distribution of weights under this model will not bear a relationship to what these studies actually might offer. Indeed, it has been demonstrated that redistribution of weights is simply in one direction from larger to smaller studies as heterogeneity increases until eventually all studies have equal weight and no more redistribution is possible.[[38]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-ReferenceA-38) Another issue with the random effects model is that the most commonly used confidence intervals generally do not retain their coverage probability above the specified nominal level and thus substantially underestimate the statistical error and are potentially overconfident in their conclusions.[[39]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Brockwell2001-39)[[40]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Noma2011-40) Several fixes have been suggested[[41]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-41)[[42]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-42) but the debate continues on.[[40]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Noma2011-40)[[43]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-pmid19016302-43) A further concern is that the average treatment effect can sometimes be even less conservative compared to the fixed effect model[[44]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-44) and therefore misleading in practice. One interpretational fix that has been suggested is to create a prediction interval around the random effects estimate to portray the range of possible effects in practice.[[45]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-45) However, an assumption behind the calculation of such a prediction interval is that trials are considered more or less homogeneous entities and that included patient populations and comparator treatments should be considered exchangeable[[46]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-pmid23494781-46) and this is usually unattainable in practice.

The most widely used method to estimate between studies variance (REVC) is the DerSimonian-Laird (DL) approach.[[47]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-pmid3802833-47) Several advanced iterative (and computationally expensive) techniques for computing the between studies variance exist (such as maximum likelihood, profile likelihood and restricted maximum likelihood methods) and random effects models using these methods can be run in Stata with the metaan command.[[48]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-metaan-48) The metaan command must be distinguished from the classic metan (single "a") command in Stata that uses the DL estimator. These advanced methods have also been implemented in a free and easy to use Microsoft Excel add-on, MetaEasy.[[49]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-MetaEasy1-49)[[50]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-MetaEasy2-50) However, a comparison between these advanced methods and the DL method of computing the between studies variance demonstrated that there is little to gain and DL is quite adequate in most scenarios.[[51]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-KontopantelisSMMR10-51)[[52]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-KontopantelisSMMR11-52)

However, most meta-analyses include between 2 and 4 studies and such a sample is more often than not inadequate to accurately estimate heterogeneity. Thus it appears that in small meta-analyses, an incorrect zero between study variance estimate is obtained, leading to a false homogeneity assumption. Overall, it appears that heterogeneity is being consistently underestimated in meta-analyses and sensitivity analyses in which high heterogeneity levels are assumed could be informative.[[53]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-KontopantelisP1-53) These random effects models and software packages mentioned above relate to study-aggregate meta-analyses and researchers wishing to conduct individual patient data (IPD) meta-analyses need to consider mixed-effects modelling approaches.[[54]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-ipdforest-54)

##### IVhet model

Doi & Barendregt working in collaboration with Khan, Thalib and Williams (from the University of Queensland, University of Southern Queensland and Kuwait University), have created an inverse variance quasi likelihood based alternative (IVhet) to the random effects (RE) model for which details are available online.[[55]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Manual-55) This was incorporated into MetaXL version 2.0,[[56]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Epigear-56) a free Microsoft excel add-in for meta-analysis produced by Epigear International Pty Ltd, and made available on 5 April 2014. The authors state that a clear advantage of this model is that it resolves the two main problems of the random effects model. The first advantage of the IVhet model is that coverage remains at the nominal (usually 95%) level for the confidence interval unlike the random effects model which drops in coverage with increasing heterogeneity.[[39]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Brockwell2001-39)[[40]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Noma2011-40) The second advantage is that the IVhet model maintains the inverse variance weights of individual studies, unlike the RE model which gives small studies more weight (and therefore larger studies less) with increasing heterogeneity. When heterogeneity becomes large, the individual study weights under the RE model become equal and thus the RE model returns an arithmetic mean rather than a weighted average. This side-effect of the RE model does not occur with the IVhet model which thus differs from the RE model estimate in two perspectives:[[55]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Manual-55) Pooled estimates will favor larger trials (as opposed to penalizing larger trials in the RE model) and will have a confidence interval that remains within the nominal coverage under uncertainty (heterogeneity). Doi & Barendregt suggest that while the RE model provides an alternative method of pooling the study data, their simulation results[[57]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-57) demonstrate that using a more specified probability model with untenable assumptions, as with the RE model, does not necessarily provide better results. The latter study also reports that the IVhet model resolves the problems related to underestimation of the statistical error, poor coverage of the confidence interval and increased MSE seen with the random effects model and the authors conclude that researchers should henceforth abandon use of the random effects model in meta-analysis. While their data is compelling, the ramifications (in terms of the magnitude of spuriously positive results within the Cochrane database) are huge and thus accepting this conclusion requires careful independent confirmation. The availability of a free software (MetaXL)[[56]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Epigear-56) that runs the IVhet model (and all other models for comparison) facilitates this for the research community.

#### Direct evidence: Models incorporating additional information

##### Quality effects model

Doi and Thalib originally introduced the quality effects model.[[58]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Doi_2008-58) They[[59]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-59) introduced a new approach to adjustment for inter-study variability by incorporating the contribution of variance due to a relevant component (quality) in addition to the contribution of variance due to random error that is used in any fixed effects meta-analysis model to generate weights for each study. The strength of the quality effects meta-analysis is that it allows available methodological evidence to be used over subjective random effects, and thereby helps to close the damaging gap which has opened up between methodology and statistics in clinical research. To do this a synthetic bias variance is computed based on quality information to adjust inverse variance weights and the quality adjusted weight of the *i*th study is introduced.[[58]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Doi_2008-58) These adjusted weights are then used in meta-analysis. In other words, if study *i* is of good quality and other studies are of poor quality, a proportion of their quality adjusted weights is mathematically redistributed to study *i* giving it more weight towards the overall effect size. As studies become increasingly similar in terms of quality, re-distribution becomes progressively less and ceases when all studies are of equal quality (in the case of equal quality, the quality effects model defaults to the IVhet model – see previous section). A recent evaluation of the quality effects model (with some updates) demonstrates that despite the subjectivity of quality assessment, the performance (MSE and true variance under simulation) is superior to that achievable with the random effects model.[[60]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-60)[[61]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-61) This model thus replaces the untenable interpretations that abound in the literature and a software is available to explore this method further.[[56]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Epigear-56)

#### Indirect evidence: Network meta-analysis methods



A network meta-analysis looks at indirect comparisons. In the image, A has been analyzed in relation to C and C has been analyzed in relation to b. However the relation between A and B is only known indirectly, and a network meta-analysis looks at such indirect evidence of differences between methods and interventions using statistical method.

Indirect comparison meta-analysis methods (also called network meta-analyses, in particular when multiple treatments are assessed simultaneously) generally use two main methodologies. First, is the Bucher method[[62]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-62) which is a single or repeated comparison of a closed loop of three-treatments such that one of them is common to the two studies and forms the node where the loop begins and ends. Therefore, multiple two-by-two comparisons (3-treatment loops) are needed to compare multiple treatments. This methodology requires that trials with more than two arms have two arms only selected as independent pair-wise comparisons are required. The alternative methodology uses complex [statistical modelling](https://en.wikipedia.org/wiki/Statistical_model) to include the multiple arm trials and comparisons simultaneously between all competing treatments. These have been executed using Bayesian methods, mixed linear models and meta-regression approaches.[[*citation needed*](https://en.wikipedia.org/wiki/Wikipedia%3ACitation_needed)]

##### Bayesian framework

Specifying a Bayesian network meta-analysis model involves writing a directed acyclic graph (DAG) model for general-purpose [Markov chain Monte Carlo](https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo) (MCMC) software such as WinBUGS.[[63]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Valkenhoef,_G._2012-63) In addition, prior distributions have to be specified for a number of the parameters, and the data have to be supplied in a specific format.[[63]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Valkenhoef,_G._2012-63) Together, the DAG, priors, and data form a Bayesian hierarchical model. To complicate matters further, because of the nature of MCMC estimation, overdispersed starting values have to be chosen for a number of independent chains so that convergence can be assessed.[[64]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-64) Currently, there is no software that automatically generates such models, although there are some tools to aid in the process. The complexity of the Bayesian approach has limited usage of this methodology. Methodology for automation of this method has been suggested[[65]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-65) but requires that arm-level outcome data are available, and this is usually unavailable. Great claims are sometimes made for the inherent ability of the Bayesian framework to handle network meta-analysis and its greater flexibility. However, this choice of implementation of framework for inference, Bayesian or frequentist, may be less important than other choices regarding the modeling of effects[[66]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-ReferenceC-66) (see discussion on models above).

##### Frequentist multivariate framework

On the other hand, the frequentist multivariate methods involve approximations and assumptions that are not stated explicitly or verified when the methods are applied (see discussion on meta-analysis models above). For example, the mvmeta package for Stata enables network meta-analysis in a frequentist framework.[[67]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-67) However, if there is no common comparator in the network, then this has to be handled by augmenting the dataset with fictional arms with high variance, which is not very objective and requires a decision as to what constitutes a sufficiently high variance.[[68]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-68) The other issue is use of the random effects model in both this frequentist framework and the Bayesian framework. Senn advises analysts to be cautious about interpreting the 'random effects' analysis since only one random effect is allowed for but one could envisage many.[[66]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-ReferenceC-66) Senn goes on to say that it is rather naıve, even in the case where only two treatments are being compared to assume that random-effects analysis accounts for all uncertainty about the way effects can vary from trial to trial. Newer models of meta-analysis such as those discussed above would certainly help alleviate this situation and have been implemented in the next framework.

##### Generalized pairwise modelling framework

An approach that has been tried since the late 1990s is the implementation of the multiple three-treatment closed-loop analysis. This has not been popular because the process rapidly becomes overwhelming as network complexity increases. Development in this area was then abandoned in favor of the Bayesian and multivariate frequentist methods which emerged as alternatives. Very recently, automation of the three-treatment closed loop method has been developed for complex networks by some researchers[[55]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Manual-55) as a way to make this methodology available to the mainstream research community. This proposal does restrict each trial to two interventions, but also introduces a workaround for multiple arm trials: a different fixed control node can be selected in different runs. It also utilizes robust meta-analysis methods so that many of the problems highlighted above are avoided. Further research around this framework is required to determine if this is indeed superior to the Bayesian or multivariate frequentist frameworks. Researchers willing to try this out have access to this framework through a free software.[[56]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-Epigear-56)

##### Tailored meta-analysis

Another form of additional information comes from the intended setting. If the target setting for applying the meta-analysis results is known then it may be possible to use data from the setting to tailor the results thus producing a ‘tailored meta-analysis’.,[[69]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-69)[[70]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-70) This has been used in test accuracy meta-analyses, where empirical knowledge of the test positive rate and the prevalence have been used to derive a region in [Receiver Operating Characteristic](https://en.wikipedia.org/wiki/Receiver_Operating_Characteristic) (ROC) space known as an ‘applicable region’. Studies are then selected for the target setting based on comparison with this region and aggregated to produce a summary estimate which is tailored to the target setting.

### Validation of meta-analysis results

The meta-analysis estimate represents a weighted average across studies and when there is [heterogeneity](https://en.wikipedia.org/wiki/Study_heterogeneity) this may result in the summary estimate not being representative of individual studies. Qualitative appraisal of the primary studies using established tools can uncover potential biases,[[71]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-71)[[72]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-72) but does not quantify the aggregate effect of these biases on the summary estimate. Although the meta-analysis result could be compared with an independent prospective primary study, such external validation is often impractical. This has led to the development of methods that exploit a form of [leave-one-out cross validation](https://en.wikipedia.org/wiki/Cross-validation_%28statistics%29), sometimes referred to as internal-external cross validation (IOCV).[[73]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-73) Here each of the k included studies in turn is omitted and compared with the summary estimate derived from aggregating the remaining k- 1 studies. A general **validation statistic, Vn** based on IOCV has been developed to measure the statistical validity of meta-analysis results.[[74]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-74) For test accuracy and prediction, particularly when there are multivariate effects, other approaches which seek to estimate the prediction error have also been proposed.[[75]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-75)

…

## Applications in modern science

Modern statistical meta-analysis does more than just combine the effect sizes of a set of studies using a weighted average. It can test if the outcomes of studies show more variation than the variation that is expected because of the sampling of different numbers of research participants. Additionally, study characteristics such as measurement instrument used, population sampled, or aspects of the studies' design can be coded and used to reduce variance of the estimator (see statistical models above). Thus some methodological weaknesses in studies can be corrected statistically. Other uses of meta-analytic methods include the development and validation of clinical prediction models, where meta-analysis may be used to combine individual participant data from different research centers and to assess the model's generalisability,[[76]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-76)[[77]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-77) or even to aggregate existing prediction models.[[78]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-78)

Meta-analysis can be done with [single-subject design](https://en.wikipedia.org/wiki/Single-subject_design) as well as group research designs. This is important because much research has been done with [single-subject research](https://en.wikipedia.org/wiki/Single-subject_research) designs. Considerable dispute exists for the most appropriate meta-analytic technique for single subject research.[[79]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-79)

Meta-analysis leads to a shift of emphasis from single studies to multiple studies. It emphasizes the practical importance of the effect size instead of the statistical significance of individual studies. This shift in thinking has been termed "meta-analytic thinking". The results of a meta-analysis are often shown in a [forest plot](https://en.wikipedia.org/wiki/Forest_plot).

Results from studies are combined using different approaches. One approach frequently used in meta-analysis in health care research is termed '[inverse variance method](https://en.wikipedia.org/wiki/Inverse-variance_weighting)'. The average [effect size](https://en.wikipedia.org/wiki/Effect_size) across all studies is computed as a *weighted mean*, whereby the weights are equal to the inverse variance of each study's effect estimator. Larger studies and studies with less random variation are given greater weight than smaller studies. Other common approaches include the [Mantel–Haenszel method](https://en.wikipedia.org/wiki/Cochran%E2%80%93Mantel%E2%80%93Haenszel_statistics)[[80]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-80) and the [Peto method](https://en.wikipedia.org/wiki/Richard_Peto).[[81]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-81)

[Seed-based d mapping](https://en.wikipedia.org/wiki/Seed-based_d_mapping) (formerly signed differential mapping, SDM) is a statistical technique for meta-analyzing studies on differences in brain activity or structure which used neuroimaging techniques such as fMRI, VBM or PET.

Different high throughput techniques such as [microarrays](https://en.wikipedia.org/wiki/DNA_microarray) have been used to understand [Gene expression](https://en.wikipedia.org/wiki/Gene_expression). [MicroRNA](https://en.wikipedia.org/wiki/MicroRNA) expression profiles have been used to identify differentially expressed microRNAs in particular cell or tissue type or disease conditions or to check the effect of a treatment. A meta-analysis of such expression profiles was performed to derive novel conclusions and to validate the known findings.[[82]](https://en.wikipedia.org/wiki/Meta-analysis#cite_note-82)