

An Introduction to Meta-analysis

Will G Hopkins

Sportscience 8, 20-24, 2004 (sportsci.org/jour/04/wghmeta.htm)

Sport and Recreation, Auckland University of Technology, Auckland 1020, New Zealand. [Email](#).

Reviewer: George A Kelley, Meta-Analytic Research Group, Community Medicine, West Virginia University, Morgantown, WV 26506-9190.

A meta-analysis is a systematic quantitative review of original research studies of some phenomenon, such as the effect of a specific treatment on some aspect of health or behavior. The meta-analyst expresses the magnitudes of effects from all relevant studies in the same units, then uses an appropriate weighting factor (the inverse of each effect's error variance) to combine the magnitudes into a mean value and its uncertainty (confidence limits). In a traditional meta-analysis, the true effects are assumed to be homogeneous (have the same value) in the analyzed studies, and some "outlier" studies may be eliminated to satisfy this assumption. In the more recent and realistic random-effect or mixed-model meta-analysis, true values of all effects are assumed to be heterogeneous (different), and the analysis provides an estimate of the heterogeneity as a standard deviation representing unexplained typical true variation in the effect between studies. Inclusion of study and mean subject characteristics in the analysis as covariates may reduce heterogeneity and provide further useful information about the magnitude of the effect in different locations and with different subjects. Published effects are usually larger than their true values, owing to the misuse of statistical significance as a criterion for publication. A funnel plot can detect such publication bias, but there is currently no satisfactory way to adjust for it in the meta-analysis, and the only long-term solution is to ban statistical significance.

KEYWORDS: Cochrane Collaboration, funnel plot, mixed model, quantitative analysis, random effect, research, systematic review.

[Reprint pdf](#) · [Reprint doc](#) · [Slideshow](#)

Update Aug 2007: minor improvements to slideshow. See also a more [succinct version](#) of the slideshow prepared for but not presented at the 2007 ACSM meeting, as explained in the [conference report](#).

The basis for this article is an updated version of a slideshow accompanying a talk on meta-analysis I presented this year locally and at the University of Bath. The article should meet a need for a straightforward and up-to-date account of meta-analysis suitable for research students and staff in the sport sciences and other biomedical disciplines.

My experience with meta-analysis is limited—one analysis published and three others completed recently—but most of my assertions appear to be consistent with those in the ultimate source of meta-analytic wisdom, the [handbook](#) of the [Cochrane Collaboration](#) (cochrane.org) I depart from the handbook with my emphasis or novel material on individual responses, standardized differences in means, log transformation, measures of physical performance, and correlations. You will need to refer to the Cochrane handbook for information on topics I don't cover, including survival analysis, intention-to-treat analysis, and meta-analysis of single-subject studies (cases or individual patient data).

The [reprint pdf](#) version of this article contains printer-friendly images of the PowerPoint [slideshow](#) and references to relevant publications.

An Introduction to Meta-analysis

Will G Hopkins
Faculty of Health Science
Auckland University of Technology, NZ

- What is a Meta-Analysis?
- Why is Meta-Analysis Important?
- What Happens in a Meta-Analysis?
 - Traditional (fixed-effects) vs random-effect meta-analysis
- Limitations to Meta-Analysis
- Generic Outcome Measures for Meta-Analysis
 - Difference in means, correlation coefficient, relative frequency
- How to Do a Meta-Analysis
- Main Points
- References

What is a Meta-Analysis?

- A **systematic review of literature** to address this question: **on the basis of the research to date, how big is a given effect**, such as...
 - the effect of endurance training on resting blood pressure;
 - the effect of bracing on ankle injury;
 - the effect of creatine supplementation on sprint performance;
 - the relationship between obesity and habitual physical activity.
- It is similar to a simple cross-sectional study, in which the **subjects are individual studies** rather than individual people.
 - But the stats are a lot harder.
- A review of literature is a meta-analytic review only if it includes **quantitative estimation** of the magnitude of the effect and its uncertainty (confidence limits).

Why is Meta-Analysis Important?

- Researchers used to think the aim of a single study was to decide if a given effect was **"real"** (statistically significant).
- But they put **little faith in a single study** of an effect, no matter how good the study and how statistically significant.
- When many studies were done, someone would write a **narrative (= qualitative) review** trying to explain why the effect was/wasn't real in the studies.
- Enlightened researchers now realize that all effects are real.
- The aim of research is therefore to get the **magnitude of an effect with adequate precision**.
- Each study produces a **different estimate** of the magnitude.
- Meta-analysis **combines the effects from all studies** to give an overall mean effect and other important statistics.

What Happens in a Meta-analysis?

- The main outcome is the **overall magnitude** of the effect...
 - ...and how it differs between subjects, protocols, researchers.
- It's **not a simple average** of the magnitude in all the studies.
- Meta-analysis gives **more weight** to studies with **more precise estimates**.
 - The weighting factor is almost always $1/(\text{standard error})^2$.
 - The standard error is the **expected variation in the effect** if the study was repeated again and again.
- Other things being equal, this weighting is equivalent to weighting the effect in each study by the study's **sample size**.
 - So, for example, a meta-analysis of 3 studies of 10, 20 and 30 subjects each amounts to a single study of 60 subjects.
 - But the weighting factor also takes into account differences in **error of measurement** between studies.

Traditional Meta-Analysis

- You can and should allow for **real differences** between studies: **heterogeneity** in the magnitude of the effect.
 - The I^2 statistic quantifies % of variation due to real differences.
- In traditional (fixed-effects) meta-analysis, you do so by testing for heterogeneity using the **Q statistic**.
 - The test has low power, so you use $p < 0.10$ rather than $p < 0.05$.
 - If $p < 0.10$, you exclude **"outlier"** studies and re-test, until $p > 0.10$.
 - When $p > 0.10$, you declare the effect **homogeneous**.
 - That is, you assume the differences in the effect between studies are due only to **sampling variation**.
 - Which makes it **easy to calculate** the weighted mean effect and its p value or confidence limits.
 - But the approach is **unrealistic, limited**, and suffers from all the **problems of statistical significance**.

Random-Effect (Mixed-Model) Meta-Analysis

- In random-effect meta-analysis, you assume there are **real differences between all studies** in the magnitude of the effect.
- The "random effect" is the **standard deviation** representing the **variation in the true magnitude** from study to study.
 - You get an **estimate of this SD** and its precision.
 - The mean effect \pm this SD is **what folks can expect typically** in another study or if they try to make use of the effect.
- A better term is **mixed-model** meta-analysis, because...
 - You can include study characteristics as **"fixed effects"**.
 - The study characteristics will partly **account for differences** in the magnitude of the effect between studies. Example: differences between studies of athletes and non-athletes.
- You need **more studies** than for traditional meta-analysis.
- The analysis is not yet available in a spreadsheet.

Limitations to Meta-Analysis

- It's focused on **mean effects** and **differences between studies**. But what really matters is effects on **individuals**.
 - So we need to know the magnitude of **individual responses**.
 - Solution: researchers should **quantify individual responses** as a **standard deviation**, which itself can be meta-analyzed.
 - And we need to know which **subject characteristics** (e.g. age, gender, genotype) **predict** individual responses well.
 - Use **mean** characteristics as **covariates** in the meta-analysis.
 - Better if researchers make available all data for all subjects, to allow **individual patient-data meta-analysis**.
 - **Confounding** by unmeasured characteristics can be a problem.
 - e.g., different effect in elites vs subelites could be due to different training phases (which weren't reported in enough studies to include).
- A meta-analysis reflects only what's **published**.
 - But statistically significant effects are more likely to get published.
 - Hence published effects are **biased high**.

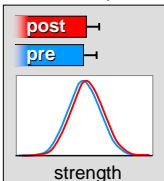
Generic Outcome Measures for Meta-Analysis

- You can combine effects from different studies only when they are expressed in the **same units**.
- In most meta-analyses, the effects are converted to a generic **dimensionless measure**. Main measures:
 - standardized difference or change in the mean (Cohen's *d*);
 - Other forms similar or less useful (Hedges' *g*, Glass's *d*)
 - percent or factor difference or change in the mean
 - correlation coefficient;
 - relative frequency (relative risk, odds ratio).

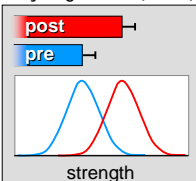
Standardized Difference or Change in the Mean (1)

- Express the difference or change in the mean as a fraction of the **between-subject standard deviation** ($\Delta\text{mean}/\text{SD}$).
 - Also known as the **Cohen effect size**.
 - This example of the effect of a treatment on strength shows why the SD is important:

Trivial effect (0.1x SD)



Very large effect (3x SD)


- The $\Delta\text{mean}/\text{SD}$ are **biased high for small sample sizes** and need correcting before including in the meta-analysis.

Standardized Difference or Change in the Mean (2)

- Problem:
 - Study samples are often drawn from **populations with different SDs**, so some differences in effect size between studies will be **due to the differences in SDs**.
 - Such differences are **irrelevant** and tend to **mask more interesting differences**.
- Solution:
 - Meta-analyze a better generic measure reflecting the **biological effect**, such as percent change.
 - **Combine the between-subject SDs** from the studies selectively and appropriately, to get one or more population SDs.
 - Express the overall effect from the meta-analysis as a standardized effect size using this/these SDs.
 - This approach also all but **eliminates the correction for sample-size bias**.

Percent or Factor Change in the Mean (1)

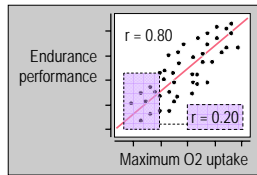
- The magnitude of many effects on humans can be expressed as a **percent or multiplicative factor** that tends to have the same value for every individual.
 - Example: effect of a treatment on performance is +2%, or a factor of 1.02.
- For such effects, **percent difference or change** can be the most appropriate generic measure in a meta-analysis.
- If all the studies have small percent effects (<10%), use percent effects **directly** in the meta-analysis.
- Otherwise express the effects as **factors** and **log-transform them** before meta-analysis.
 - **Back-transform** the outcomes into percents or factors.
 - Or calculate **standardized differences or changes** in the mean using the log transformed effects.

Percent or Factor Change in the Mean (2)

- Measures of **athletic performance** need special care.
- The best generic measure is **percent change**.
- But a given percent change in an athlete's ability to output power can result in **different percent changes** in performance in **different exercise modalities**.
 - Example: a 1% change in endurance power output produces the following changes...
 - 1% in running time-trial speed or time;
 - -0.4% in road-cycling time-trial time;
 - 0.3% in rowing-ergometer time-trial time;
 - -15% in time to exhaustion in a constant-power test.
- So convert all published effects to changes in **power output**.
- For **team-sport fitness tests**, convert percent changes back into standardized mean changes after meta-analysis.

Correlation Coefficient

- A good measure of association between **two numeric variables**.
 - If the correlation is, say, 0.80, then a 1 SD difference in the predictor variable is associated with a 0.80 SD difference in the dependent variable.
- Samples with **small between-subject SD** have **small correlations**, so correlation coefficient suffers from a similar problem as standardized effect size.
 - Solution: meta-analyze the **slope** then convert to a correlation using composite SD for predictor and dependent variables.
 - Divide each estimate of slope by the reliability correlation for the predictor to adjust for downward bias due to error of measurement.



Relative Frequencies

- When the dependent variable is a frequency of something, effects are usually expressed as **ratios**.
 - **Relative risk** or **risk ratio**: if 10% of active people and 25% of inactive people get heart disease, the relative risk of heart disease for inactive vs active is $25/10=2.5$.
 - **Hazard ratio** is similar, but is the **instantaneous** risk ratio.
 - **Odds ratio** for these data is $(25/75)/(10/90)=3.0$.
 - Risk and hazard ratios are mostly for **cohort studies**, to compare incidence of injury or disease between groups.
 - Odds ratio is mostly for **case-control studies**, to compare frequency of exposure to something in cases and controls (groups with and without injury or disease).
 - Most models with numeric covariates need odds ratio.
 - Odds ratio is hard to interpret, but it's about **the same as risk or hazard ratio** in value and meaning when **frequencies are <10%**.

How to Do a Meta-Analysis (1)

- Decide on an **interesting effect**.
- Do a thorough **search** of the literature.
 - If you find the effect has already been meta-analyzed...
 - The analysis was probably traditional fixed effect, so do a **mixed-model** meta-analysis.
 - Otherwise find another effect to meta-analyze.
 - As you assemble the published papers, **broaden or narrow the focus** of your review to make it manageable and relevant.
 - Design (e.g., only randomized controlled trials)
 - Population (e.g., only competitive athletes)
 - Treatment (e.g., only acute effects)
- Record **effect magnitudes** and convert into values on a **single scale** of magnitude.
 - In a randomized controlled trial, the effect is the **difference** (experimental-control) in the **change** (post-pre) in the **mean**.

How to Do a Meta-Analysis (2)

- Record **study characteristics** that might account for differences in the effect magnitude between studies.
- Include the study characteristics as **covariates** in the meta-analysis. Examples:
 - duration or dose of treatment;
 - method of measurement of dependent variable;
 - quality score;
 - gender and mean characteristics of subjects (age, status...).
 - Treat **separate outcomes** for females and males from the same study as if they came from **separate studies**.
 - If gender effects aren't shown separately in one or more studies, analyze gender as a **proportion** of one gender (e.g. for a study of 3 males and 7 females, "maleness" = 0.3).
 - Use this approach for all problematic **dichotomous characteristics** (sedentary vs active, non-athletes vs athletes, etc.).

How to Do a Meta-Analysis (3)

- Some meta-analysts score the **quality** of a study.
 - Examples (scored yes=1, no=0):
 - Published in a peer-reviewed journal?
 - Experienced researchers?
 - Research funded by impartial agency?
 - Study performed by impartial researchers?
 - Subjects selected randomly from a population?
 - Subjects assigned randomly to treatments?
 - High proportion of subjects entered and/or finished the study?
 - Subjects blind to treatment?
 - Data gatherers blind to treatment?
 - Analysis performed blind?
 - Use the score to **exclude some studies**, and/or...
 - **Include** as a **covariate** in the meta-analysis, but...
 - Some statisticians advise **caution** when using quality.

How to Do a Meta-Analysis (4)

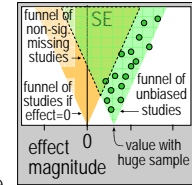
- Calculate the value of a **weighting factor** for each effect, using...
 - the **confidence interval** or limits
 - Editors, please **insist** on them for all outcome statistics.
 - the **test statistic** (t , χ^2 , F)
 - F ratios with numerator degrees of freedom >1 can't be used.
 - **p value**
 - If the **exact** p value is not given, try contacting the authors for it.
 - Otherwise, if "p<0.05"... analyze as p=0.05.
 - If "p>0.05" with no other info, deal with the study qualitatively.
 - For **controlled trials**, can also use...
 - **SDs of change scores**
 - **Post-test SDs** (but almost always gives much larger error variance).
 - **Incredibly**, many researchers report p-value inequalities for control and experimental groups separately, so **can't use any of the above**.
 - Use **sample size** as the weighting factor instead.

How to Do a Meta-Analysis (5)

- Perform a **mixed-model** meta-analysis.
 - Get **confidence limits** (preferably 90%) for everything.
 - Interpret the **clinical or practical magnitudes** of the effects and their confidence limits...
 - and/or calculate chances that the true mean effect is **clinically or practically** beneficial, trivial, and harmful.
 - Interpret the magnitude of the between-study random effect as the **typical variation** in the magnitude of the mean effect **between researchers** and therefore possibly between **practitioners**.
- For controlled trials, caution readers that there may also be **substantial individual responses** to the treatment.
 - Scrutinize the studies and report any **evidence** of such individual responses.
 - **Meta-analyze SDs** representing individual responses, if possible.
 - No-one has, yet. It's coming, perhaps by 2050.

How to Do a Meta-Analysis (6)

- Some meta-analysts present the effect magnitude of all the studies as a **funnel plot**, to address the issue of **publication bias**.
 - Published effects tend to be **larger** than true effects, because...
 - effects that are larger simply because of **sampling variation** have smaller p values,
 - and $p < 0.05$ is more likely to be **published**.
 - A plot of standard error vs effect magnitude has a triangular or **funnel** shape.
 - **Asymmetry** in the plot can indicate non-significant studies that weren't published.
 - But **heterogeneity** disrupts the funnel shape.
 - So a funnel plot of **residuals** is better & helps identify outlier studies.
 - It's still unclear how best to deal with publication bias.
 - Short-term wasteful solution: meta-analyze only the **larger studies**.
 - Long-term solution: **ban $p < 0.05$** as a publication criterion.



Main Points

- Meta-analysis is a **statistical literature review** of magnitude of an effect.
- Meta-analysis uses the magnitude of the effect and its precision from each study to produce a **weighted mean**.
- **Traditional** meta-analysis is based unrealistically on using a test for heterogeneity to **exclude outlier studies**.
- **Random-effect** (mixed-model) meta-analysis **estimates heterogeneity** and allows estimation of the **effect of study** and **subject characteristics** on the effect.
- For the analysis, the effects have to be converted into the same units, usually **percent** or other **dimensionless generic measure**.
- It's possible to visualize the impact of **publication bias** and identify outlier studies using a **funnel plot**.

References

- A good source of meta-analytic wisdom is the **Cochrane Collaboration**, an international non-profit academic group specializing in meta-analyses of healthcare interventions.
 - **Website:** <http://www.cochrane.org>
 - **Publication:** The Cochrane Reviewers' Handbook (2004). <http://www.cochrane.org/resources/handbook/index.htm>.
- **Simpler reference:** Bergman NG, Parker RA (2002). Meta-analysis: neither quick nor easy. BMC Medical Research Methodology 2, <http://www.biomedcentral.com/1471-2288/2/10>
- **Glossary:** Delgado-Rodríguez M (2001). Glossary on meta-analysis. Journal of Epidemiology and Community Health 55, 534-536.
- Recent reference for problems with **publication bias:** Terrin N, Schmid CH, Lau J, Olkin I (2003). Adjusting for publication bias in the presence of heterogeneity. Statistics in Medicine 22, 2113-2126.