SPORTSCIENCE sportsci.org
Perspectives / Research Resources

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: percent units are best for most effect representing differences or changes in means; risk, odds and hazard ratios are appropriate for proportions; and ratios are appropriate for counts. The meta-analyst 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 or plot of standardized residuals can reveal such publication bias, and deletion of studies with larger standard errors reduces the bias.

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

Reprint pdf · Reprint doc · Slideshow

Update Oct 2013. Substantial make-over to take into account the unified approach to ratios of risks or proportions, odds, hazards and counts in the article on linear models and effect magnitudes. A novel approach of including separate effects for each group from controlled trials or other studies with control, reference or other comparison groups is also described.

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.

Published Nov 2004
©2004

An Introduction to Meta-analysis

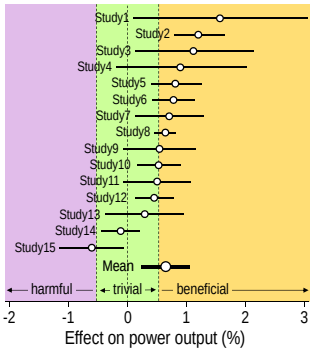
Will G Hopkins
Victoria University, Melbourne, Australia

- What is a Meta-Analysis?
  - Definition, weighted average, heterogeneity, mixed-model meta-regression
- Limitations to Meta-Analysis
  - Individual differences or responses, publication bias
- How to Do a Meta-Analysis
  - Generic measures, finding effects, study characteristics, study quality, weighting factor, model, publication bias
- Summary and 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 bit 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).

- The main outcome is the overall magnitude of the effect.
- The effect in each study (the study estimates) and the meta-analyzed mean outcome are often shown in a forest plot:



Data are means and 95% confidence intervals.

- The main outcome is not a simple average of the study estimates.
  - Meta-analysis uses the standard error of each study estimate to give more weight to studies with more precise estimates.
  - The standard error is the expected variation in the study estimate if the study was repeated again and again.
- The weighting factor is 1/(standard error)^2.
  - Other things being equal, use of this factor 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.
  - For controlled trials, this factor also takes into account differences in standard error of measurement between studies.

- You can and should allow for real differences or heterogeneity in the magnitude of the effect between studies.
  - In early (fixed-effects only) meta-analysis, you did so by testing for heterogeneity using the Q or chi-squared statistic.
  - The test has low power, so you used p<0.10 rather than p<0.05.
  - If p<0.10, you excluded "outlier" studies and re-tested, until p>0.10.
  - When p>0.10, you declared the effect homogeneous.
    - That is, you assumed the differences in the effect between studies were due only to sampling variation.
    - Which made it easy to analyze the weighted mean effect.
  - But the approach was unrealistic, limited, and suffered from the problem of whether statistical non-significance means negligible.
- In random-effect meta-analysis, you accept there are always 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.
    - It is sometimes known as tau and is provided as tau^2.
  - The mean effect ± this SD is what folks can expect typically in another study or if they try to make use of the effect.
  - Instead of this SD, most researchers provide the Q and the related I^2 statistic, representing percent of variation due to real differences.
    - Ignore any conclusions based on this uninterpretable statistic.
- Mixed-model meta-analysis or meta-regression is best of all.
  - You 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.
  - The random effect now represents residual variation in the effect between studies (i.e., not explained by the study characteristics).
  - Include an extra random effect to account for repeated measurement (multiple estimates) within studies.
  - Custom software or an advanced package (e.g., SAS) is required.

Limitations to Meta-Analysis

- It's focused on mean effects and differences between studies.
  - But what really matters is effects on individuals.
  - So we should also quantify individual differences or responses.
  - These can be expressed as standard deviations, but researchers usually don't provide enough info to allow their meta-analysis.
  - Inclusion of mean subject characteristics (e.g., age, gender, genotype) as predictors in the meta-analytic model only partly addresses this problem.
    - It would be better if researchers made available all data for all subjects, to allow individual patient-data meta-analysis.
- A meta-analysis reflects only published effects.
  - But statistically significant effects are more likely to get published.
  - Hence published effects are biased high.
  - Funnel or related plots can be used to assess and reduce publication bias.

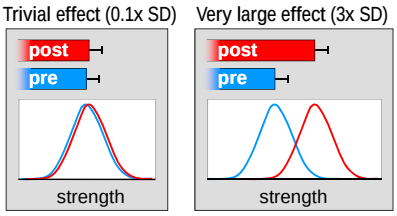
How to Do a Meta-Analysis: Opt for a Generic Measure

- 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 are similar or less useful (Hedges' *g*, Glass's δ)
  - percent or factor difference or change in the mean;
  - correlation coefficient and slope;
  - risk, odds, hazard and count ratios.

Standardized Difference or Change in the Mean

- Express the difference or change in the mean as a fraction of the between-subject standard deviation (Δmean/SD).
  - Also known as Cohen's *d* (*d* stands for *difference*).
  - This example of the effect of a treatment on strength shows why the SD is important:



- The Δmean/SD are biased high for small sample sizes and need correcting before including in the meta-analysis.

- A problem with standardization:
  - 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.
- The solution:
  - Meta-analyze a better generic measure reflecting the biological effect, usually percent or factor differences or changes.
    - Rarely, the raw measure is best; for example, joint angles representing flexibility.
  - 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 using this/these SDs.
  - This approach also effectively eliminates the correction for sample-size bias with standardized effects.

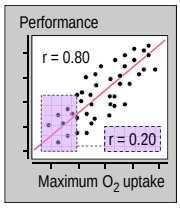
Percent or Factor Difference or Change in the Mean

- The magnitude of many effects 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, regardless of the raw value of the performance.
- 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 (and their standard errors) 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 and logs of factor SD.

- 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.
    - A difficult and time-consuming task; you have been warned!
    - See recent meta-analyses by my students and colleagues.
- For team-sport fitness tests, convert percent changes back into standardized mean changes after meta-analysis.

Correlation Coefficient and Slope

- These measures of association between two numeric variables are seldom meta-analyzed.
- Studies with small between-subject SD have small correlations, so correlation suffers from a similar SD problem as standardized effects.
- Solution: meta-analyze the slope.
  - The slope is biased low (degraded) only by random error in the predictor.
  - Adjust for this bias by dividing the slope by the short-term reliability intraclass correlation coefficient.
  - Express the meta-analyzed slope as either…
    - a correlation using SD for an appropriate population, or
    - the effect of two SD of the predictor in that population.



Risk, Odds, Hazard and Count Ratios

- When the dependent variable is a proportion or count of something, effects should be expressed as ratios.
- Risk ratio, relative risk, proportion ratio…
  - Example: if proportions of inactive and active adults who get heart disease after 20 years are 25% and 10%, risk ratio = 25/10 = 2.5.
- If proportions are time-independent classifications, convert all effects to odds ratios for meta-analysis.
  - Convert meta-analyzed odds ratios back into proportions and proportion ratios by choosing a sensible proportion for the reference group.
- Hazard ratio is the risk ratio for new occurrences in the next brief instant of time (the "right-now" risk ratio).
  - If proportions change with time, the proportion and odds ratios also change, but the hazard ratio usually doesn't.

- So, to meta-analyze studies with different time periods, convert any proportion and odds ratios to hazard ratios.
- Odds ratios from time-dependent case-control studies are already hazard ratios, if controls were sampled as the cases came in (incidence-density sampling).
- If proportions in the two groups in all studies are low (<10%), all proportion, odds and hazard ratios are effectively equal and need not be interconverted.
- Express effects on counts as count ratios.
- Express standard errors of ratio effects as ×/÷ factor errors, then log transform the ratios and errors for meta-analysis.
  - Some software does the log transformation and works out the standard errors for you.

How to Do a Meta-Analysis: Find and Record Effects

- Do a search of the literature for studies of a specific effect.
  - If the effect has been meta-analyzed already…
    - You can do another, if the analysis was done badly or if there have been many new studies since the previous meta.
    - 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 via…
    - design (e.g., only randomized controlled trials), population (e.g., only competitive athletes), treatment (e.g., only acute effects)…
  - Document your searches, inclusions and exclusions.
- Record each effect magnitude and inferential information (sample size, p value, confidence limits, SD of change scores).
  - Convert effects into values on a single scale of magnitude.
  - In studies with a control or other reference group, record the effect and inferential information in each group to enhance the analysis.

How to Do a Meta-Analysis: Get Study Characteristics

- 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…).
    - Record separate outcomes for females and males from the same study, if possible.
    - Otherwise analyze gender as a proportion of one gender; for example, in 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: Assess Study Quality?

- Most 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.

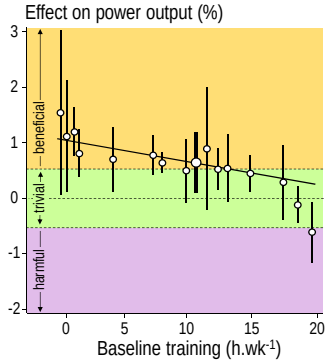
How to Do a Meta-Analysis: Get the Weighting Factor

- Calculate the standard error for each effect via one or more of…
  - the confidence interval or limits
  - the test statistic (t, χ^2, F)
    - F ratios with numerator degrees of freedom >1 can't be used.
  - the p value
    - If the exact p value is not given and you can't calculate the standard error from the data, 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.
  - SD of change scores (for controlled trials)
    - For studies lacking sufficient information to calculate standard errors, calculate the typical error (standard error of measurement) in every other study and impute typical errors (and standard errors via SD of change scores) from these. The spreadsheet for sample-size estimation at Sportscience calculates the typical errors.
  - the data: done for you by the software, depending on the effect.

How to Do a Meta-Analysis: Develop the Model

- Do a mixed-model meta-regression.
  - Estimate and interpret the effect for interesting types of study or mean subject.
  - Show a moderator plot rather than a forest plot:
  - In a mediator plot, the subject characteristic is a difference score or change score.

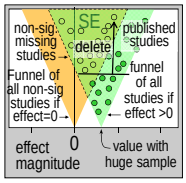
Data are means and 90% confidence intervals



- For any linear covariate, estimate and interpret the effect of 2x the average of between-subject SD from appropriate studies.
- Double the SD representing the between-study random effect to interpret its magnitude as the unexplained typical differences in the magnitude of the effect between settings.
  - The random effect provided as tau^2 may be in log-transformed units and may need back-transformation to a factor SD.
    - Apply the p value for tau^2 directly to the SD to get confidence limits.
- For effects where there are control or reference groups…
  - include each group effect separately, if possible;
  - include a within-study random effect to account for the resulting repeated measurement;
  - include fixed effects to estimate uncontrolled effects and effects relative to control, best-practice or other reference groups.
- Inspect between-subject SD between and within studies for evidence of individual differences or responses.

How to Do a Meta-Analysis: Deal with Publication Bias

- 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 should have a triangular or funnel shape.
  - If some non-significant studies weren't published, the plot will be asymmetrical.
    - The missing studies are generally smaller (therefore larger SE).
  - Effect heterogeneity also disrupts the funnel shape.
  - So plot standardized residuals (random-effect solution) vs standard error (not shown) to spot publication bias and also outlier studies.
- Delete studies with larger SE to give a symmetrical plot.



Summary

- 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 account for publication bias and identify outlier studies using a funnel plot or residuals plot.

References

- Best: read recent meta-analyses co-authored by me.
- 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.
  - But the (free) Cochrane meta-analysis software is somewhat limited.

These references are getting out of date:

- 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.
- 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.

This presentation is available from:

SPORTSCIENCE sportsci.org
A Peer-Reviewed Site for Sport Research

See Sportscience 8, 2004