Meta-analysis in biological and environmental sciences

yDiv/HIGRADE course 23-26 October 2017

Dr. Dylan Craven, Dr. Katharina Gerstner

iDiv is a research centre of the

DFG Deutsche Forschungsgemeinschaft

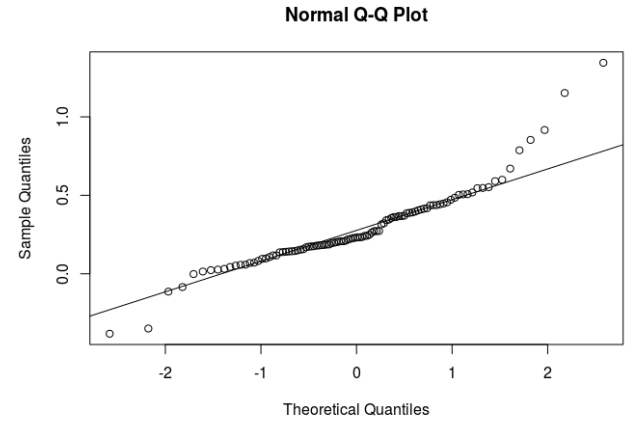
Day 1	**Introduction to meta-analysis** • What is a meta-analysis? • Examples of meta-analyses • Why performing a meta-analysis? • Procedure of meta-analysis in a nutshell **Searching the literature** **Effect sizes and moderators** **Data extraction/Coding**
Day 2	**Meta-analytic models** • Fixed effects model • Random effects model • Mixed effects/hierarchical model **Quantifying and explaining heterogeneity**
Day 3	**Assumptions, biases and confounding effects** • Variance homogeneity and normality of residuals • Publication bias • Sensitivity analysis **Interpretation and presentation of results** • Format for meta-analysis report • PRISMA flow diagram • Forest plots
Day 4	**Methodological issues, advances, and common mistakes** • Non-independence among effect sizes • Non-independence of moderators • Missing data **Criticism of meta-analysis**

Meta-analysis in biological and environmental sciences

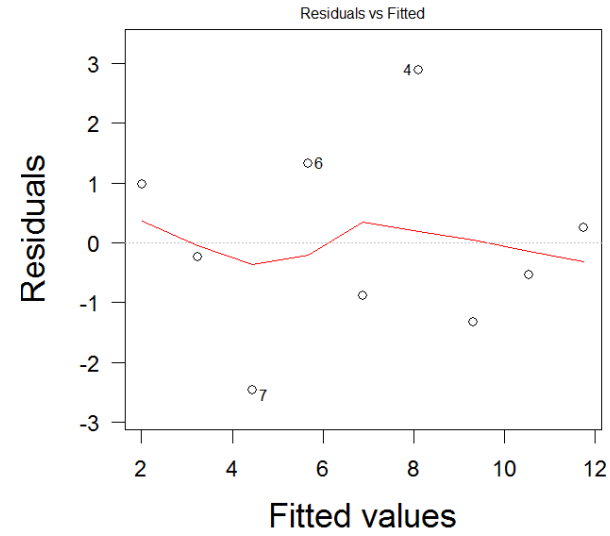
Assumptions, biases and confounding effects

Checking assumptions

1. Normality of residuals
   Quantile-Quantile plots



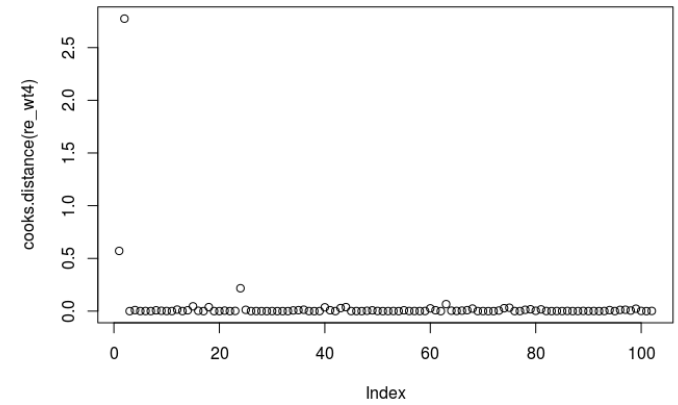
2. Homogeneity of variances
   Diagnostic plot Residuals vs. Fitted values
   Relationship between mean and variance?
   Is there a pattern or trend?
   NO = “The sky at night”
   YES = Linear relationship?,
   Homogeneity of variance?



Sensitivity analysis

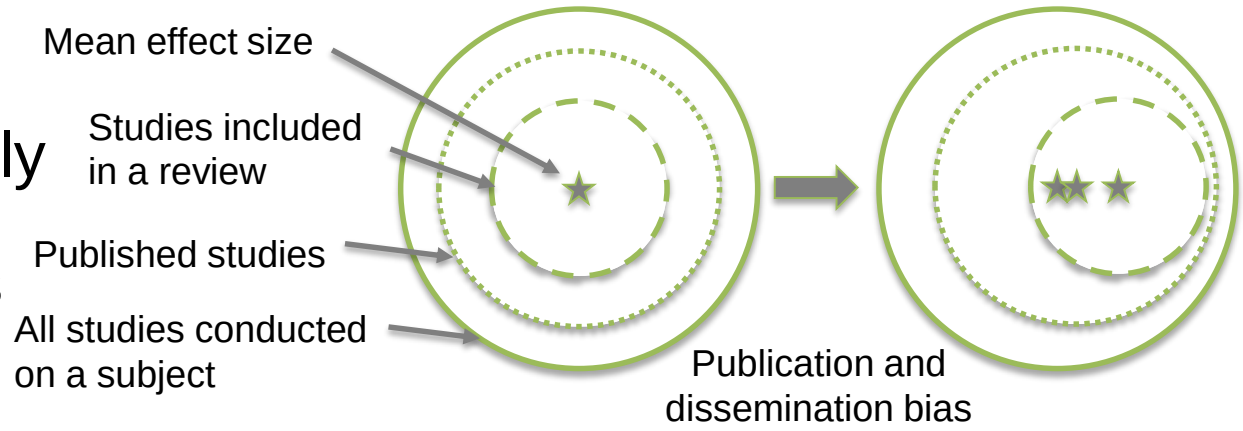
Case Diagnostics

- A single outlying trial could be the source of substantial heterogeneity.
- To identify suspicious cases, a leave-one-out method can be used whereby we rerun the meta-analysis, iteratively removing studies.
- In the metafor package this is accomplished with the `leave1out` function.
- Data points with strong influence on effect size estimates?
  - Outliers are no property of the data but of the model
  - Is this data point erroneous?
  - Is there a covariable missing?
  - Transformation of covariables needed?
  - How do parameter estimates vary if outlier is excluded from the fit?
  - Function `cooks.distance` in metafor



Publication bias

… findings of the set of included studies inaccurately represent the findings of all correctly conducted studies



Reasons for detecting publication bias:

- Narrow sense publication bias (file-drawer problem)
  - Influence of research findings on the probability of a study being published
- Selective reporting
- Dissemination bias
  - Dependence of accessibility of research findings on the direction or strength of these findings
- *But also:* Real heterogeneity between effect sizes, which can be accounted for by covariables

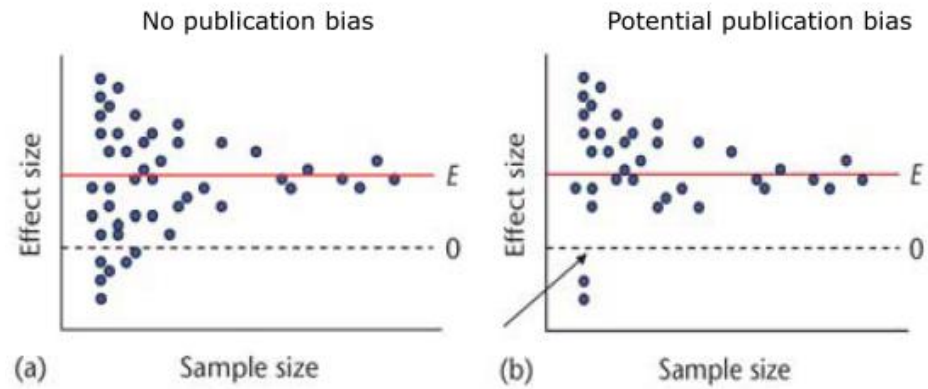
Dealing with publication bias

- **Avoiding it**
  - including unpublished studies and grey literature
  - using studies in which testing the hypothesis of interest is not the main aim
- **Detecting and estimating it**
  - direct methods (comparison of results of published vs unpublished studies)
  - indirect methods
- **Adjusting for it**
  - ’trim and fill’ method (Duval & Tweedie 2000)

Detecting publication bias

The funnel plot

- a scatter plot of effect size vs. sample size or variance
- Asymmetry (gaps) in the funnel may be indicative of publication bias.
- `funnel` is the function in metafor for generating funnel plots from rma objects.
- In the absence of bias:
  - the variance of effect sizes should increase as sample sizes decrease
  - the mean effect should be independent of sample size
  - at a given sample size, individual studies should be normally distributed around the cumulative effect size



Hillebrand (2008) Meta-analysis in Ecology. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0003272

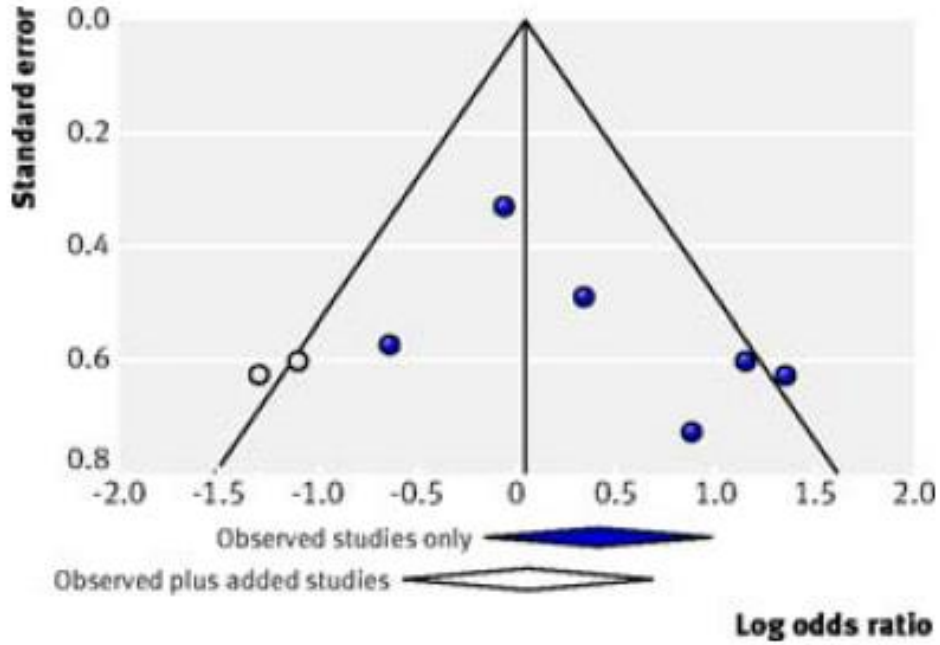
Detecting publication bias

- Judging asymmetry in the funnel plot can be difficult due to:
  - Low efficiency for small number of studies
  - Rely on visual detection of asymmetry
  - Assume that all studies share the same effect size
  - Plot asymmetry may be caused by other factors
- Additional ways of assessing the threat of publication bias:
  - Rank-correlation tests between effect sizes and sample size/variance (Kendall‘s tau, Spearman‘s rank order)
  - Linear regression between effect sizes and sample size/variance (Egger‘s regression)
  - Trim-and-Fill
  - Fail Safe N

Detecting publication bias

Trim-and-fill method

- The trim and fill method estimates the number of missing NULL studies from the meta-analysis.
- The function `trimfill` of the metafor package augments the observed data and returns the fitted rma object with the missing studies included.
- These points can be added to the funnel plot.
- Note: `trimfill` is only applicable on fixed- and random-effects models



Detecting publication bias

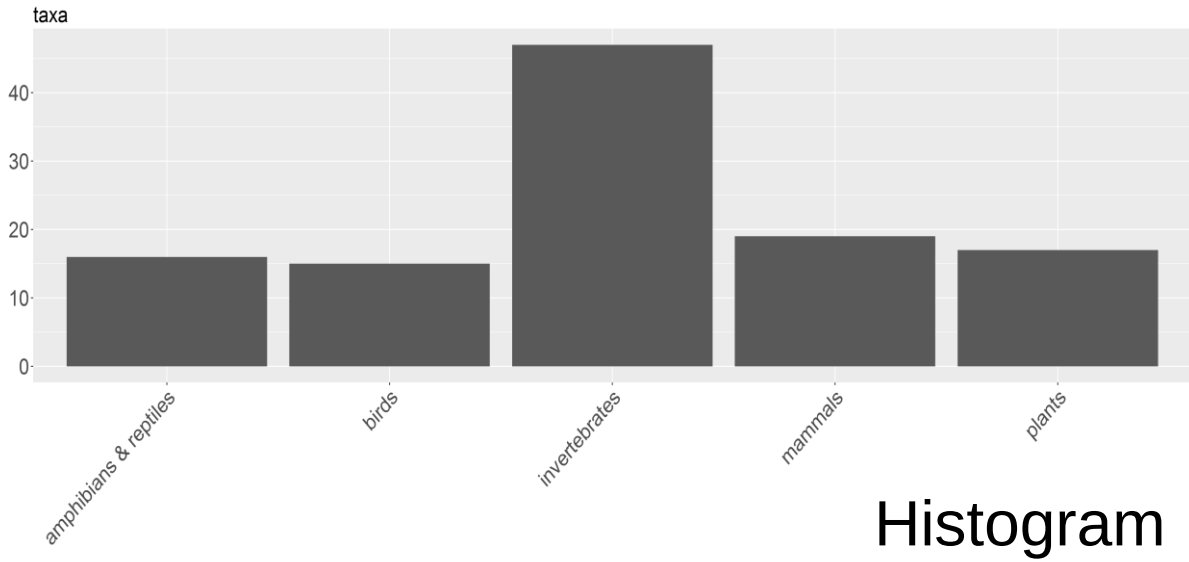
Fail-safe number

- Rosenthal method (sometimes called a ‘file-drawer analysis’)
- Is the number of NULL studies that have to be added to reduce the significance of the meta-analysis to α (usually 0.05)
- Function `fsn` in metafor calculates the fail-safe number using the Rosenthal approach

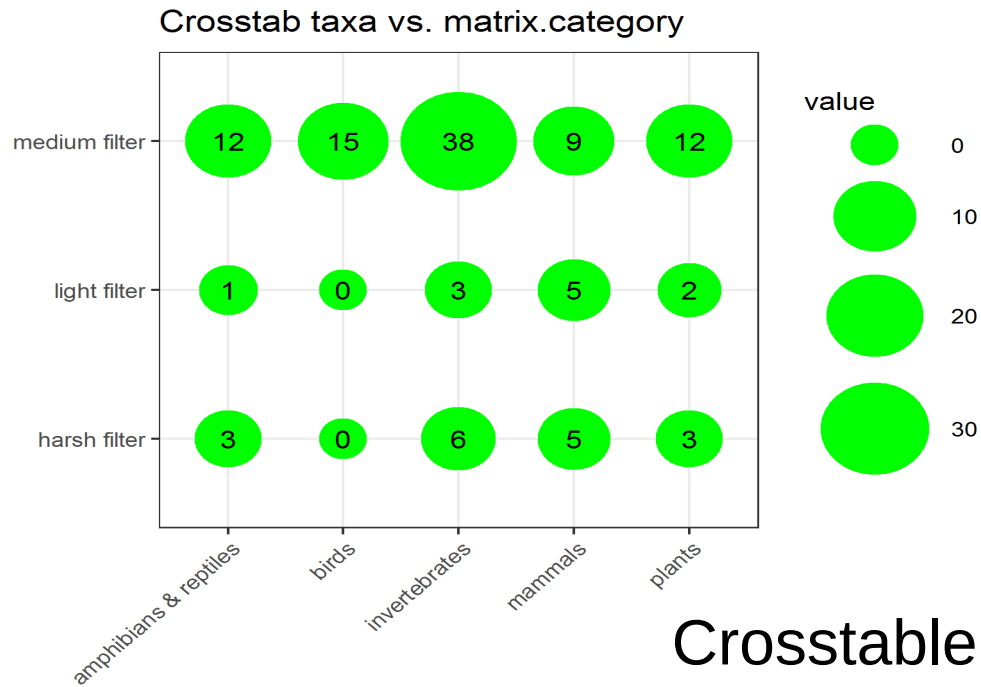
Biased sampling of the natural world

There is a disproportionate amount of information on certain ecosystems, taxa, and species

Even if there is no formal publication bias, the fact that some systems are better studied than others means that we should exercise caution when making statements about effect sizes in the natural world.



Histogram



Crosstable

Biased sampling of the natural world

Non-independence of moderators

- Example “unbalanced design”:
  - Moderator 1: plant type (woody vs herbaceous)
  - Moderator 2: study type (field vs lab)

	Woody	Herbs
Field	35	15
Lab	1	40

- Effect of study type can be properly tested only within herbaceous plants
- Comparison of plant types might be confounded by differences between study types (most studies on woody plants are field studies whereas for herbs it is a mixture of field and lab studies)

Dealing with non-independence of moderators

- Always consider and test for potential dependencies between moderators
- If dependencies are extreme (as in the example above), use hierarchical approach (e.g. test for study type for herbs only) or use 3 categories in 1-way analysis (woody-field, herbs-field, herbs-lab)
- Always aim to include both of moderators into the model instead of testing them one by one, but also check whether there are enough studies to test for interactions between moderators.

EXAMPLE 3.1

1. Residuals and outlier diagnostics
2. Publication bias

Exercise 3.1

Data

Stewart, G.B. A database on windfarm impacts on birds.

1. Perform residual and outlier diagnostics.
2. Study publication bias.
3. Check for sampling biases.

Meta-analysis in biological and environmental sciences

Interpretation and presentation of results

Format for meta-analysis report

- Same sections as in a primary study (Intro, Methods, Results, Discussion)
- Main requirements:
  - Clarity: how the data were used and why
  - Transparency: search strategy, inclusion criteria
  - Technical accuracy: assumptions of statistical tools match the structure of the data
  - Availability: all datasets available along with published results

Introduction

- An overview of the research question (theoretical and methodological history)
- Debates surrounding the research question
- Previous syntheses of the research topic
- Aims and scope of the present review

Materials and methods

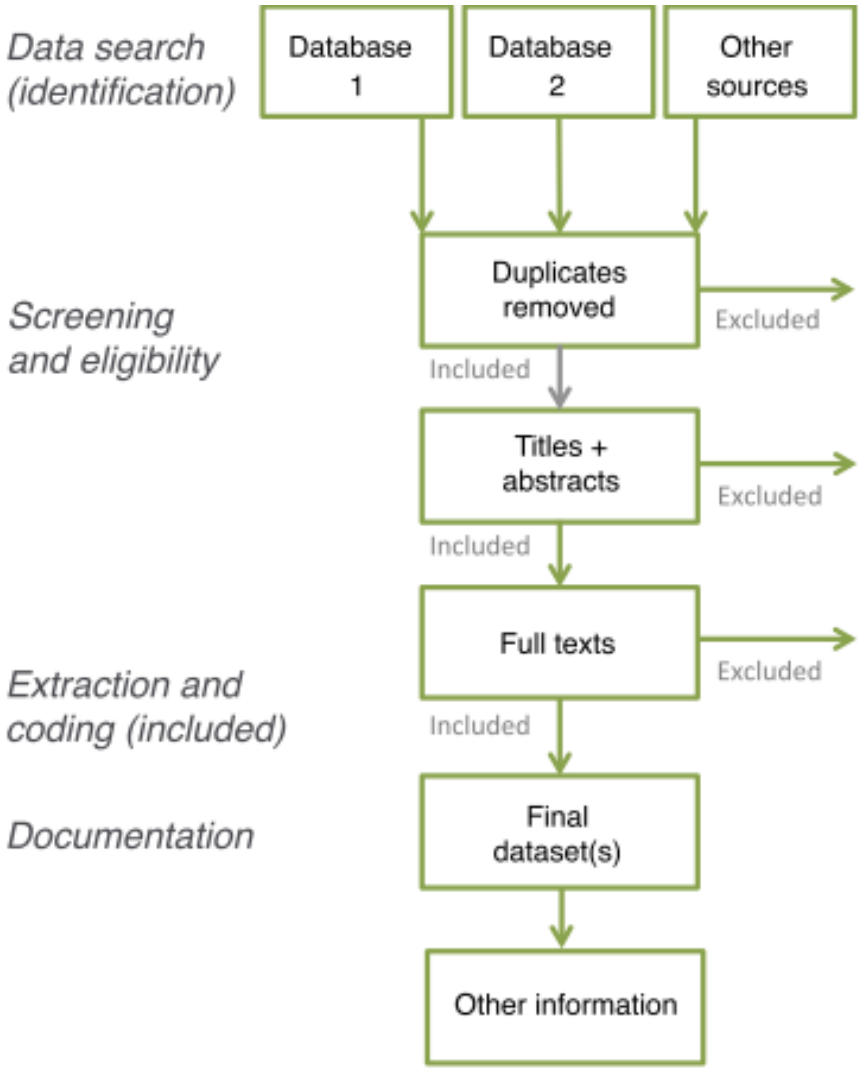
"Based on the ''Methods'' section in a meta-analysis, a researcher should be able to redo the entire analysis starting with the literature search and database build-up, proceeding with the statistical analysis, and finally coming to the same conclusions. Of course, these criteria are no different than the standards that should be imposed by reviewers on any publishable research."

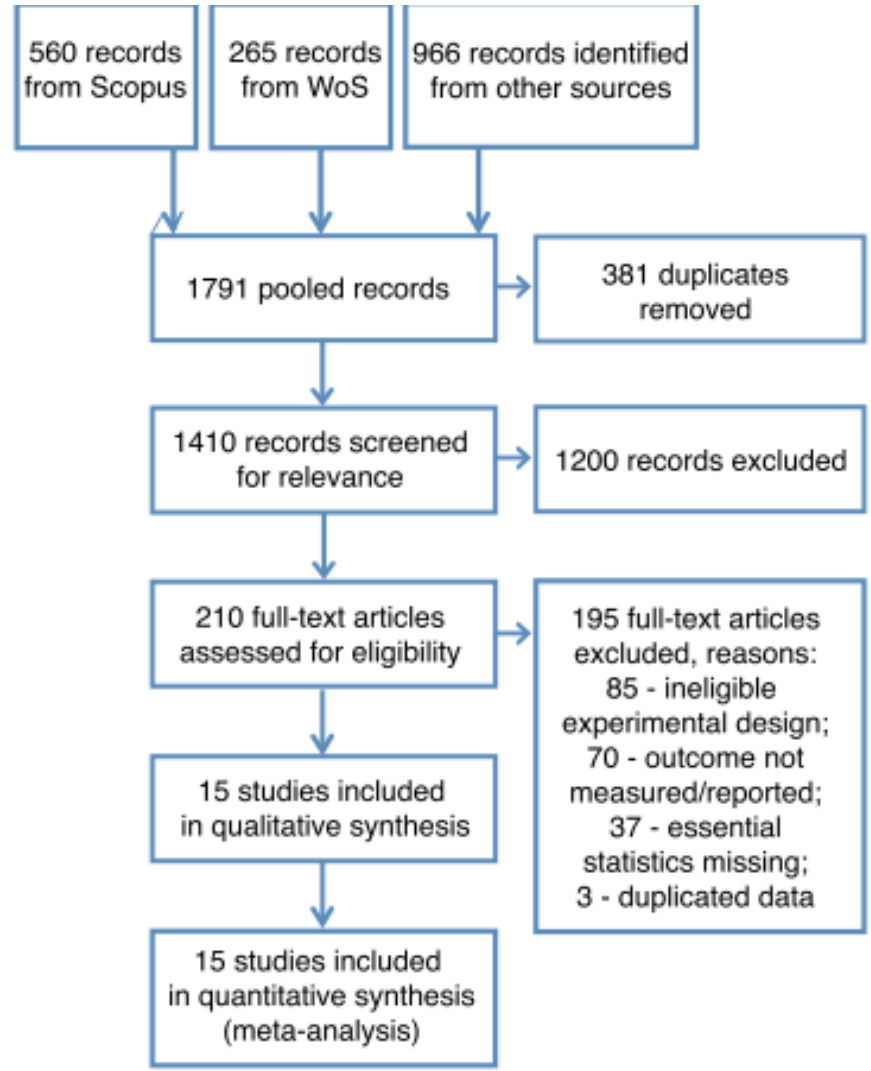
Hillebrand & Cardinale 2010 Ecology 91: 2545-2549

- Details of literature search (reference databases, keyword combinations, years)
- Inclusion criteria
- Determination of independent observations
- Details of study coding (response variables and moderators)
- Effect size metric
- Statistical models (fixed-, random-, mixed-effects)
- Testing for publication bias
- Sensitivity analysis
- Software used

Materials and methods

PRISMA diagram





Nakagawa et al. (2017) BMC Biology

Materials and methods

Sensitivity analyses

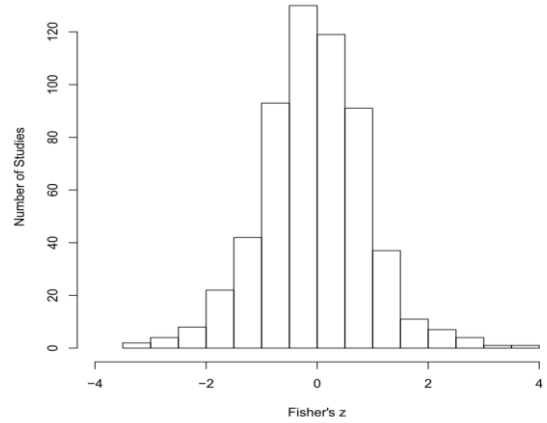
- Testing the robustness of results of meta-analysis to decisions made during the data compilation and the analysis
- Examples:
  - Testing effects of outliers (± individual studies)
  - Comparing results based on single and multiple estimates of effects per study
  - Results using different metrics of effect size (e.g. d vs lnR)
  - Results using fixed- vs random-effects models
  - Fail safe numbers (robustness to publication bias)
  - Traditional vs phylogenetically controlled analysis
  - Cumulative meta-analysis (temporal stability)

Results

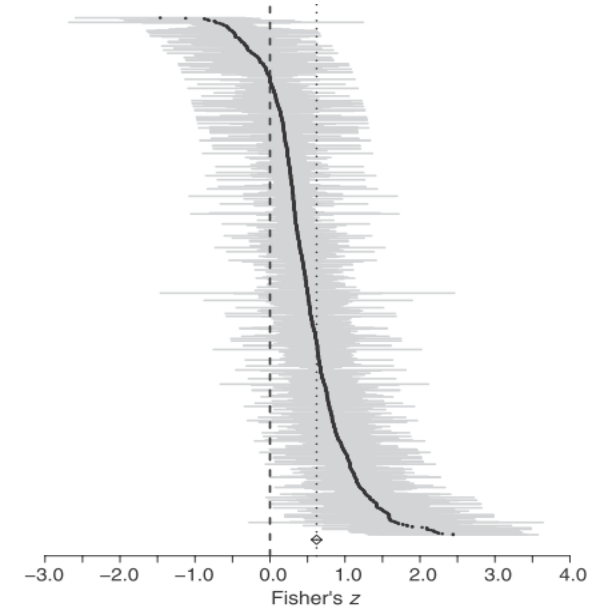
- Descriptive statistics:
  - number of studies and observations
  - range of study organisms
  - range, average and median for sample sizes, duration of experiments etc.
- Overall effect size, 95% CI and total heterogeneity
- Analysis of sources of heterogeneity (moderators)

Results

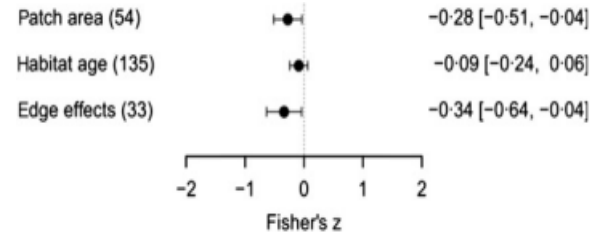
- Frequency distribution of effect sizes (histogram)



- Forest plots of effect sizes per study



- Plots of predicted effect sizes using categorical moderators (forest plots) or continuous moderators (scatter plot)



Discussion

- Summary of the major results
- Magnitude of the effects sizes (comparison with other meta-analyses on similar topics)
- Consistency with existing theory
- Interpretation and generality of findings
- Theory development
- Study limitations
- Directions for future research (research gaps)

Criteria for technically complete meta-analysis

- Generating an effect size metric based on continuous data, binary data or correlations
- Weighting effect sizes by sample size or precision
- Pooling of effect sizes into a summary effect or reasoning against pooling (e.g., due to high variation between effect sizes)
- Calculating confidence intervals for each effect size and the summary effect
- Quantifying total heterogeneity/variability
- Exploring existent heterogeneity/variability in effect sizes by considering explanatory variables (e.g., in subgroup analyses or meta-regressions)
- Presenting results in form of a forest plot or providing respective data (effect sizes, corresponding confidence intervals and weights for all included studies) elsewhere (e.g., in a table)

Vetter, Rücker & Storch (2013) Ecosphere

Other forms of research synthesis

"Researchers connect the term meta-analysis with the idea of some quantitative statistical calculations combining independent studies from the literature; an idea that is not wrong in itself, but too vague and not sufficient to be qualified as a meta-analysis" (Vetter, Rücker & Storch (2013) Ecosphere)

TABLE 1.1. Comparison of methods of research synthesis.

Characteristics of the review type	Narrative review	Vote counting	Combining probabilities	Meta-analysis
Imposes restrictions on the type of studies that can be used in review	No	No	No	Yes
Interprets study outcome based on its statistical significance	Yes	Yes	Yes	No
Takes into account sample size and statistical power of the individual studies being combined	No	No	Yes	Yes
Assesses statistical significance of the mean (overall) effect (i.e., whether it is significantly different than zero)	No	No	Yes	Yes
Assesses the magnitude of the mean effect	No	No	No	Yes
Allows analysis of sources of variation among studies	No	No	No	Yes

Koricheva et al. (2013)

Other forms of research synthesis

Vote-counting

- Studies are classified into 3 groups:
  - significant results in the predicted direction
  - significant results in the opposite direction
  - non-significant results
- Advantages:
  - Simplicity
  - broad applicability
- Problems:
  - sample size is not incorporated into a vote
  - does not provide a measure of the effect size
  - low power for medium to small effect sizes
  - power decreases as the number of studies increases

Other forms of research synthesis

Vote-counting

- Vote-counting ≠ meta-analysis!
- Vote-counting has very poor statistical properties and will often lead to wrong conclusions
- Vote-counting should never be used instead of meta-analysis
- If the studies to be reviewed are too heterogeneous in terms of outcomes or do not report results in the way which allows formal meta-analysis, using flawed method to summarize their outcomes will not help!

Checklist of quality criteria for meta-analysis for use by research synthesists, peer reviewers and journal editors

Journal of Ecology 2014, **102**, 828–844 doi: 10.1111/1365-2745.12224

SPECIAL FEATURE – ESSAY REVIEW

META-ANALYSIS IN PLANT ECOLOGY

Uses and misuses of meta-analysis in plant ecology

Julia Koricheva[1]* and Jessica Gurevitch[2]

[1]*School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK; and*
[2]*Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA*

- Are details of bibliographic search (electronic data bases used, keyword combinations, years) reported in sufficient detail to allow replication? Have full bibliographic details of primary studies included in a meta-analysis been provided?
- Are criteria for study inclusion/exclusion explicitly listed?

Checklist of quality criteria for meta-analysis for use by research synthesists, peer reviewers and journal editors

- Is the distribution of effect size metrics known and have the authors explained how they calculated variances for such metrics?
- Have statistical model for meta-analysis and the software used been described?
- Has the data set used for meta-analysis, including effect sizes and variances/sample sizes from individual primary studies and moderator variables, been provided as electronic appendix?
- Has potential non-independence been taken into account?
  - Of effect size estimates
  - Of and interactions between moderators
  - Due to phylogenetic relatedness of species been taken into account
- Have tests for publication bias been conducted?
- If meta-analysis combines studies published over considerable time span, have possible temporal changes in effect size been tested?
- Have sensitivity analysis been performed to test the robustness of results?

Exercise 3.2

Find a meta-analysis of interest and go through the study. Check the quality of the report using the quality criteria of Vetter, Rücker & Storch (2013) and Koricheva & Gurevitch (2014).