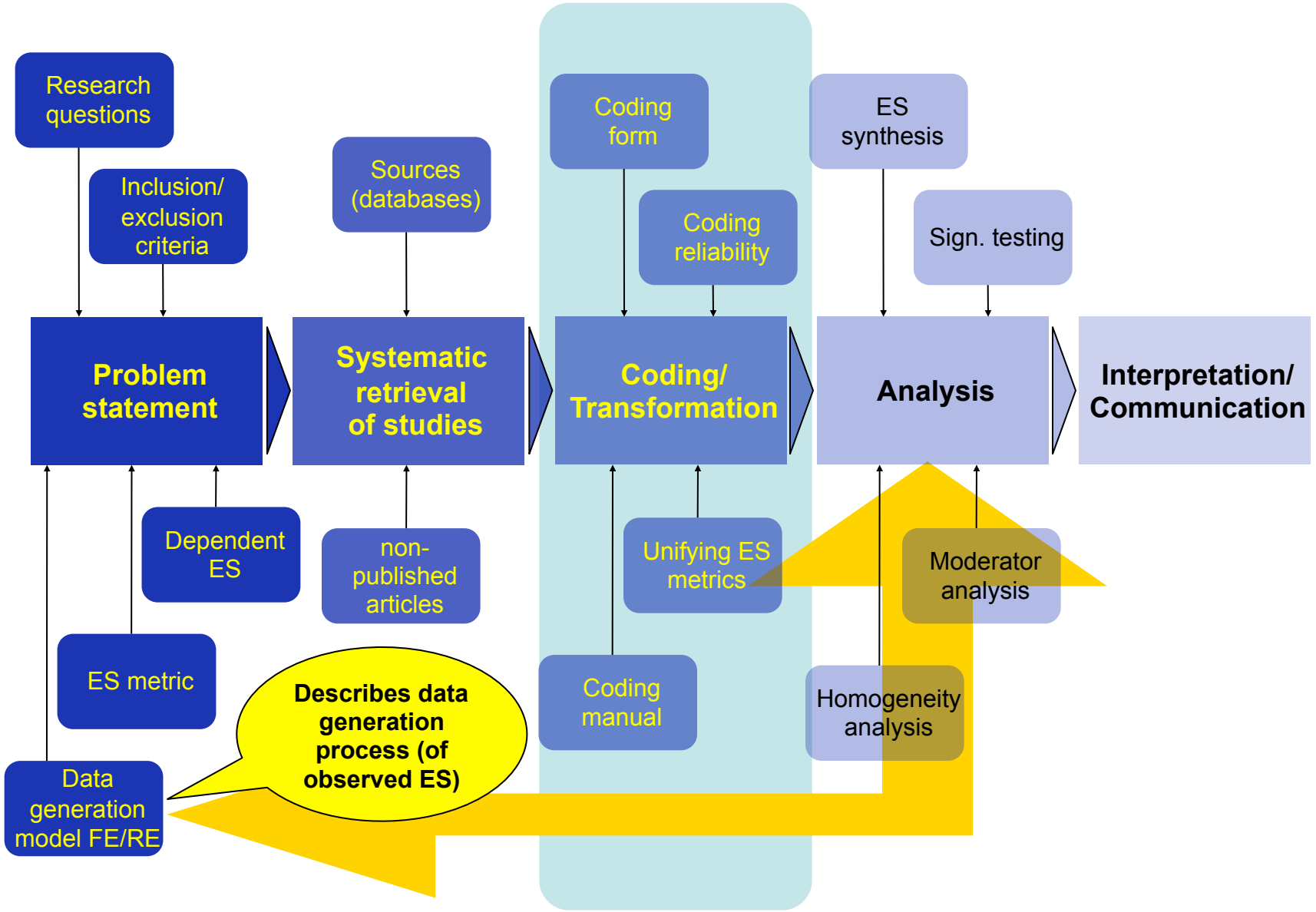
Meta-Analysis Workshop

Part 3: Data Extraction, Coding, and Unifying Effect Sizes

Michael Bošnjak
Free University of Bozen-Bolzano, School of Economics and Management

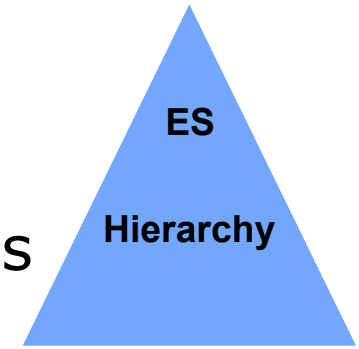
michael.bosnjak@unibz.it, www.bosnjak.eu

Generic Procedure



Overall Goal(s) and Tasks Involved

- Identify, extract, and *unify* effect sizes
  - Direct extraction (copy-and-paste)
    > Check plausibility / correctness
  - Computation / reconstruction of ES
  - Approximate ES estimate using assumptions
    > Vary assumptions
  - Caveats:
    - Simpson´s Paradox
    - ´Statistical fruit salad problem´
- Identify and extract information on/for moderators
- Determine *quality of coding process*
- Optional: Determine the *quality of primary studies*



Tools to Achieve the Overall Goal

- Tools for coding:
  - **Coding Form**
    - Effect sizes
    - Moderators
    - Study characteristics
    - Optional: Quality indicators
  - **Coding Manual**: Instructions for Coders

- Tools to compute interrater reliability
  - macros (e.g. for SPSS / PASW)
  - Web-based (e.g., Congruence Metrics Generator)

Agenda

- **Principles of data extraction**
- Development of Coding Form and Coding Manual
- Assessing Study Quality
- Determining Inter-Coder Reliability
- Unifying Effect Sizes

Principles of Data Extraction

The data extracted from each study forms the basis for all subsequent analyses.

What is ´Data Extraction´?

- The process by which one locates and transcribes information from a primary study.
- Enacted to systematically and reliably extract information from each study included in the meta-analysis.
- The process is ...
  - ... initiated after the appropriate pool of studies is identified.
  - ... accomplished by applying specific and detailed criteria to the information in a study report.
  - ... carried out by coders, e.g. review authors, graduate students, trained research assistants, etc.

Goals of Data Extraction

- Ensure reliable and orderly extraction of information from each study report.
- Minimize coder interpretation and bias.
- Create a balance between potential cost and expected benefit.
  - As items are added, both coding time and the likelihood of capitalizing on chance increase.
  - But additional information allows for more nuanced analyses.
- Overarching principles:

Preserve as much of the original information as possible!

Categories of Data to be Extracted

- Publication
- Design
- Participants
- Interventions
- Outcomes
- Measures
- Empirical findings

Error(s) in Coding

- Sources:
  - Inadequate reporting in study report
  - Ambiguities in judgment process
  - Coder bias
  - Coder mistakes
- Ways to decrease coding errors:
  - Contact original study authors for information
  - Train coders
  - Pilot test the coding instruments
  - Investigate discrepancies between coders (!)
  - Revise coding instrument

Agenda

- Principles of data extraction
- **Development of Coding Form and Coding Manual**
- Assessing Study Quality
- Determining Inter-Coder Reliability
- Unifying Effect Sizes

Coding Forms ...

- May be either paper-based or computer-based
- Are best completed by trained coders
- Should have items organized such that the order reflects the manner in which information is presented in the study report
- Address the same categories of information as the inclusion and exclusion criteria (but elicit more detail)

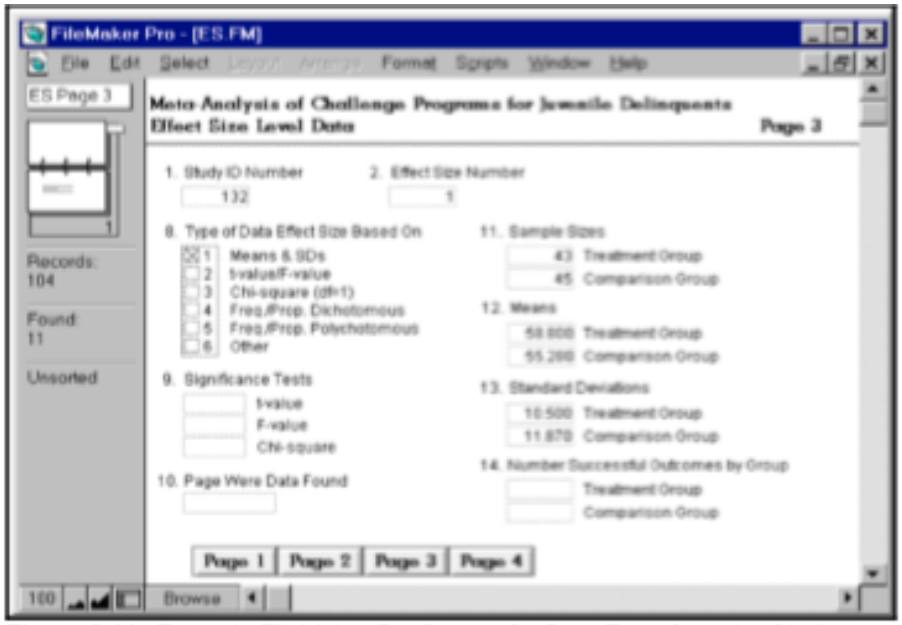


Figure 5.11: *Example FileMaker Pro Screen for Data Entry from the Challenge Meta-Analysis*

Coding Directly into a Computer Database

- Advantages
  - Avoids additional step of transferring data from paper to computer
  - Easy access to data for data cleanup
  - Data base can perform calculations during coding process (e.g., calculation of effect sizes)
  - Faster coding

- Disadvantages
  - Can be time consuming to set up
    - the bigger the meta-analysis the bigger the payoff
  - Requires a higher level of computer skill

Coding Manuals ...

- Provide detailed guidance to coders
- Serve as a historical record of the synthesis
- Incorporate elements of study reports, as well as elements of the process
- Helps avoiding low interrater agreements for ´high-inference codes´ by providing clear guidelines.
  - high-inference codes:
    require the coder to infer the information to be transferred into the coding sheet, e.g.
    - judgments of ´adequacy´ (e.g., quality)
    - interpretations by study subjects (e.g., involvement level)
  - low-inference codes:
    require the coder to locate the information and to transfer it into the coding sheet

Categories to be addressed

- Publication characteristics
  - Publication year and type (journal, dissertation etc.)
- Study design and methodological characteristics
  - Unit of assignment and unit of analysis (indiv., family, workplace, etc.)
  - Mechanism of allocation and its quality appraisal (random, matching, etc.)
- Participant characteristics
- Intervention characteristics
  - esp. ´dosis´ of treatment, method of delivery, specific characteristics of exp. and control group(s)
- Outcome and measure characteristics
  - type of outcome measure, reliability of measures (!), informants/ sources, personnel, time of assessment
- Empirical findings (extract information to compute ES)

General comments

- It is best to have closed-ended questions (!)
- Create mutually exclusive categories of variables
- Quantify whenever possible: Assign numerical values to categorical variables.
- Example of a coding manual and coding form:
  See supplement
  ´Lipsey_Wilson_2001_AppendixE.pdf´

Database Structures

- **Important:**
  Take the hierarchical nature of meta-analytic data into account!

- The familiar flat data file
- The relational data file
- Advantages and disadvantages of each

- What about the meta-analysis bibliography?

The Hierarchical Nature of Meta-Analytic Data

- Meta-analytic data is inherently hierarchical, e.g.:
  - Multiple studies per publication
  - Multiple outcomes/constructs per study
  - Multiple measurement(s) (points) per outcome/construct
  - Multiple sub-samples per study/outcome/measurement
  - Results in multiple effect sizes per study
- Any specific analysis can only include one effect size per study
  (or one effect size per sub-sample within a study)
- Analyses almost always are of a subset of coded effect sizes. Data structure needs to allow for the selection and creation of those subset

Example of a Flat Data File

Multiple ESs handled by having multiple variables, one for each potential ES.

ID	Paradigm	ES1	DV1	ES2	DV2	ES3	DV3	ES4	DV4
22	2	0.77	3						
23	2	0.77	3						
31	1	-0.1	5	-0.05	5			-0.2	11
36	2	0.94	3						
40	1	0.96	11						
82	1	0.29	11						
185	1	0.65	5	0.58	5	0.48	5	0.068	5
186	1			0.83	5				
204	2			0.88	3				
229	2	0.97	3						
246	2			0.91	3				
274	2	0.86	3	-0.31	3	0.79	3	1.17	3
295	2	7.03	3	6.46	3	.	3	0.57	.
626	1	0.87	3	-0.04	3	0.1	3	0.9	3
1366	2			0.5	3				

Note that there is only one record (row) per study

Advantages & Disadvantages of a Single Flat File Data Structure

- Advantages
  - All data is stored in a single location
  - Familiar and easy to work with
  - No manipulation of data files prior to analysis
- Disadvantages
  - Only a limited number of ESs can be calculated per study
  - Any adjustments applied to ESs must be done repeatedly
- When to use
  - Interested in a small predetermined set of ESs
  - Number of coded variables is modest
  - Comfort level with a multiple data file structure is low

Example of Relational Data Structure (Multiple Related Flat Files)

Study Level Data File

ID	PubYear	MeanAge	TxStyle
100	92	15.5	2
7049	82	14.5	1

Note that a single record in the file above is “related” to five records in the file to the right

Effect Size Level Data File

ID	ESNum	Outcome Type	TxN	CgN	ES
100	1	1	24	24	-0.39
100	2	1	24	24	0
100	3	1	24	24	0.09
100	4	1	24	24	-1.05
100	5	1	24	24	-0.44
7049	1	2	30	30	0.34
7049	2	4	30	30	0.78
7049	3	1	30	30	0

Example of a More Complex Multiple File Data Structure

Study Level Data File

ID	PubYear	MeanAge	TxStyle
100	92	15.5	2
7049	82	14.5	1

Outcome Level Data File

ID	OutNum	Constrct	Scale
100	1	2	1
100	2	6	1
100	3	4	2
7049	1	2	4
7049	2	6	3

Effect Size Level Data File

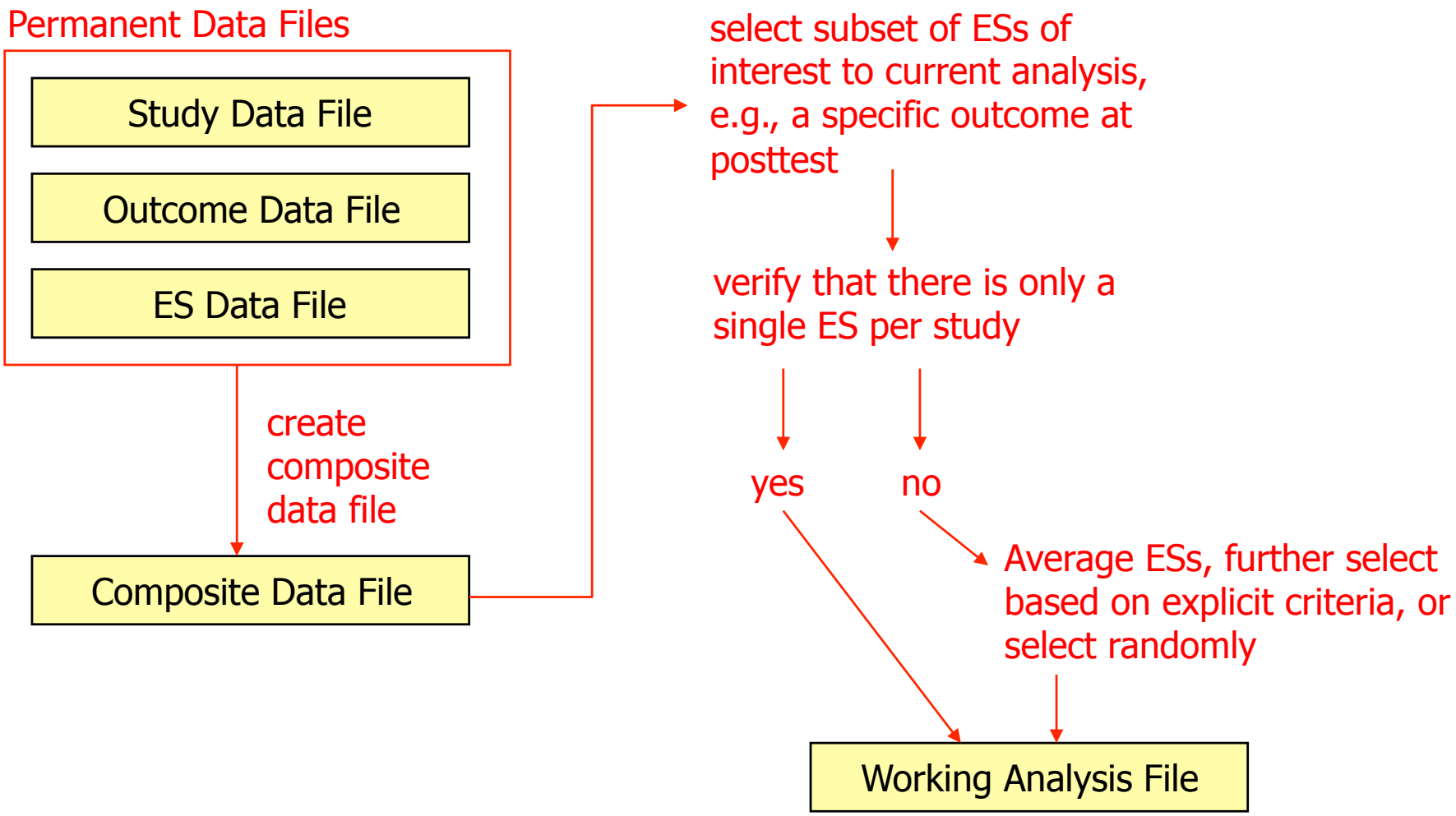
ID	OutNum	ESNum	Months	TxN	CgN	ES
100	1	1	0	24	24	-0.39
100	1	2	6	22	22	0
100	2	3	0	24	24	0.09
100	2	4	6	22	22	-1.05
100	3	5	0	24	24	-0.44
100	3	6	6	22	21	0.34
7049	1	2	0	30	30	0.78
7049	1	6	12	29	28	0.78
7049	2	2	0	30	30	0

Note that study 100 has 2 records in the outcomes data file and 6 outcomes in the effect size data file, 2 for each outcome measured at different points in time (Months)

Advantages & Disadvantages of Multiple Flat Files Data Structure

- Advantages
  - Can "grow" to any number of ESs
  - Reduces coding task (faster coding)
  - Simplifies data cleanup
  - Smaller data files to manipulate
- Disadvantages
  - Complex to implement
  - Data must be manipulated prior to analysis (creation of "working" analysis files)
  - Must be able to select a single ES per study for any given analysis
- When to use
  - Large number of ESs per study are possible

Concept of "Working" Analysis Files



What about Sub-Samples?

- So far I have assumed that the only ESs that have been coded were based on the full study sample
- What if you are interested in coding ESs separately for different sub-samples, such as, boys and girls, or high-risk and low-risk youth, etc?
  - Just say "no"!
    - Often not enough of such data for meaningful analysis
    - Complicates coding and data structure
  - Well, if you must, plan your data structure carefully
    - Include a full sample effect size for each dependent measure of interest
    - Place sub-sample in a separate data file

Agenda

- Principles of data extraction
- Development of Coding Form and Coding Manual
- **Assessing Study Quality**
- Determining Inter-Coder Reliability
- Unifying Effect Sizes

Quality of Primary Studies: Working Definitions

- „I use quality to refer to the fit between a study´s goals and the study´s design and implementation characteristics."
  (Valentine, 2009, p. 130)
- Quality of a primary study is „the degree to which the study´s design and implementation permit you to draw the inference that guide your work".
  (Cooper, 2010, p. 116)
- „high quality means high correspondence between methods and desired inferences"
  (Cooper, 2010, p. 117)
- (Sometimes) related, but **not** identical concepts:
  - study relevance
  - reporting quality

Sources of variance in quality **criteria**...

- ... inferential goal of meta-analysis
  -> needs to be stated clearly
  -> optimal research design(s) may differ.
- ... topic(s) addressed
  -> optimal research design(s) may differ.
- ... ethical/legal restrictions imposed on primary studies
  -> feasible research design(s) may differ.
- ... scientific culture
  -> (at least slightly) different quality criteria and weighting of quality aspects, e.g.
    - survey research: representativeness, external val.
    - experimental psychology: internal val.
    - personality and differential psych.: construct val.

Options for addressing study quality

- a-priori strategy:
  Exclusion of studies not meeting certain quality criteria, BUT
  - direction of bias(es) mostly unkown a-priori
  - Use only if (a) exclusion criteria are operationally defined and (b) empirical and/or theoretical evidence exists that their adoption do actually remove bias (or do otherwise improve the interpretability of the meta-analysis).

- a-posteriori strategies:
  - Accounting for the precision of study-generated evidence by weighting effect sizes for quality parameter(s)
    (in addition to inverse variance weighting)
  - Synthesis generated evidence:
    Moderator analysis over quality measure(s)

General Approaches to Assess Quality

- Threats-to-Validity Approach
  - Based on Cook & Campbell´s (1979) work
  - Four validity dimensions:
    internal, external, construct, statistical conclusion
  - Threats to validity identified as a coding guideline
    (see, e.g., Shadish, Cook & Campbell, 2002)
  - Example: Wittmann & Matt (1986)
- Methods Description Approach
  - objectives characteristics of the design are coded
  - low inference coding
  - quality of design and implementation as proxy
  - lengthy procedures, underreporting attenuates score
- **Mixed-Criteria Approach: The Study DIAD**
- Specific quality statements (CONSORT in the Medical Science; STROBE in Epidemiology)

Study DIAD (Valentine & Cooper, 2008)

- Study DIAD: Study Design and Implementation Device
- Hierarchical *framework* for *building* an evaluative scale for intervention studies
  - Combines threats to validity and methods description approach
  - ´Contextual questions´ serve as rating manual
  - Allows for describing (´profiling´) the correspondence between a study´s methods and desired inference on *different levels of abstractness*
    - *Global questions (4)*
      - *Composite questions (8)*
        - » *Design and implementation questions (32-34)*

Study DIAD: Structure and Sources I

Global Questions

Fit Between Concepts and Operations

Were the participants treated and the outcomes measured in a way that is consistent with the definition of the intervention and its proposed effects?

Clarity of Causal Inference

Did the research design permit an unambiguous conclusion about the intervention's effectiveness?

Generality of Findings

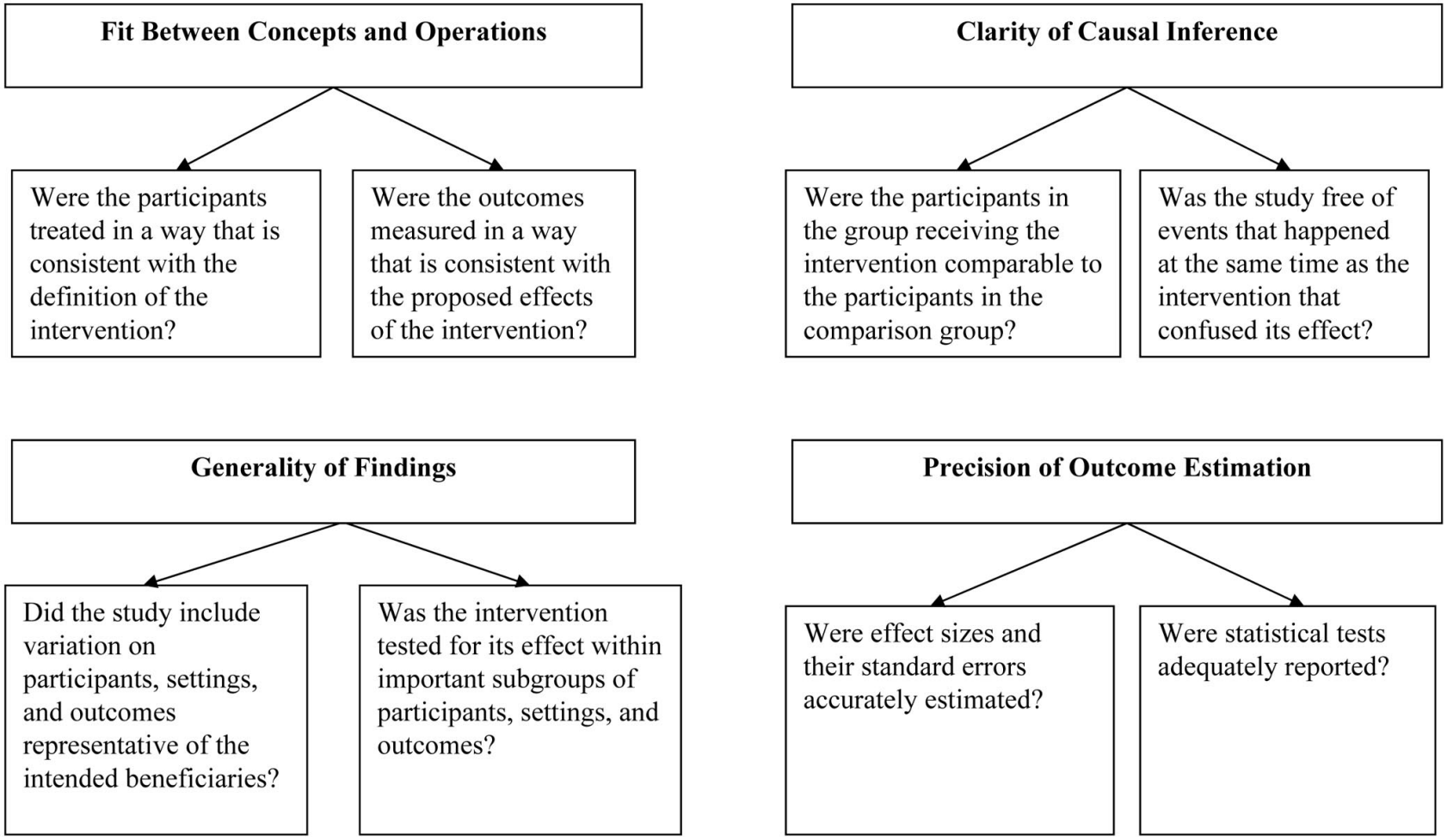
Was the intervention tested on participants, settings, outcomes, and occasions representative of its intended beneficiaries?

Precision of Outcome Estimation

Could accurate estimates of the intervention's impact be derived from the study report?

Study DIAD: Structure and Sources II

Composite Questions



Study DIAD: Details and Procedure

- Design and implementation questions:
  More fine-grained questions about study design and implementation from the Study DIAD:
  Valentine & Cooper, 2008, pp. 141-142

- Overview and example of conceptual questions (i.e., the coding manual of the Study DIAD used to rate study quality) which should be answered by the project PI:
  Valentine & Cooper, 2008, Table 1, pp. 136-138

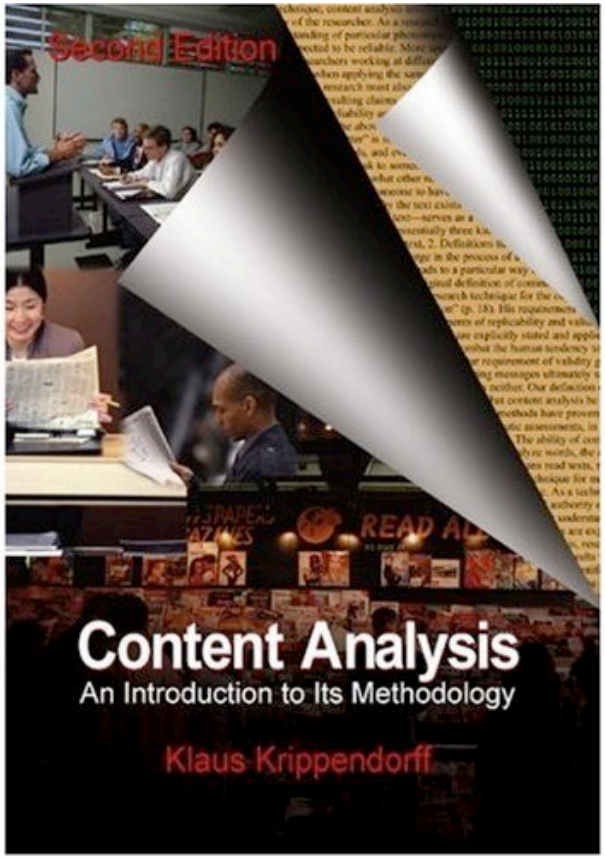
- Further details: see Literature and Supplements folders

Agenda

- Principles of data extraction
- Development of Coding Form and Coding Manual
- Assessing Study Quality
- **Determining Inter-Coder Reliability**
- Unifying Effect Sizes

Measures and Resources

- Most widely used metrices:
  - Percent agreement
  - Holsti's method
  - Scott's pi (p)
  - Cohen's kappa (k)
  - **Krippendorff's alpha (a)**
- Resources (in addition to the textbooks recommended):
  - Krippendorf: Content Analysis
  - Hayes & Krippendorf (2007)
  - Computing_Krippendorffs_ AlphaReliability.pdf
  - Section on www.meta-analysis.eu
  - Congruence metrics generator at cmg.bosnjak.eu



Agenda

- Principles of data extraction
- Development of Coding Form and Coding Manual
- Assessing Study Quality
- Determining Inter-Coder Reliability
- **Unifying Effect Sizes**

Why Unifying Effect Sizes?

- The dependent variable in a meta-analysis is a measures of effect, association, or central tendency that is comparable across studies.
- By unifying effect sizes, results from studies can be compared, synthesized, and analyzed (see part 1)
- Unified effect size measurements must be comparable across studies *on theoretical grounds*!

- Technical requirements for unification:
  - ES should be (approx.) normally distributed
    -> transformations are sometimes applied
    (e.g., r in the HO-Tradition, by using Fisher´s z-Transformation)
  - Sampling variance is known (or can be estimated)
    -> can be more easily estimated for some transformed ES (e.g., z)

Basic Types of Measures in Meta-Analyses

- Central tendency measures:
  - one dichotomous variable: proportion or logit
  - one continuous variable: mean
- Association measures:
  - both measures are continuous:
    ´r-family´: correlation coefficient(s) (raw and transformed)
  - both variables are dichotomous: odds ration, phi coefficient
- Effect measures:
  - all of them examine differences between groups
  - outcome variable is quantitative:
    ´d-family´: (standardized) mean difference
  - when outcome variable is dichotomous:
    - proportion/rate/risk difference
    - relative rate/risk
    - odds ration

Generic Measures in Meta-Analysis

Measure	Estimator *(y)*	Parameter *(θ)*	Sampling Variance *(v)*
mean difference	$\bar{x}_1 - \bar{x}_2$	$\mu_1 - \mu_2$	$s_p^2(1/n_1 + 1/n_2)$
standardized mean difference	$(\bar{x}_1 - \bar{x}_2)/s_p$	$(\mu_1 - \mu_2)/\sigma$	$\frac{n_1+n_2}{n_1 n_2} + \frac{y^2}{2(n_1+n_2)}$
risk difference	$p_1 - p_2$	$\pi_1 - \pi_2$	$p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$
risk ratio	$\ln[p_1/p_2]$	$\ln[\pi_1/\pi_2]$	$\frac{1}{p_1 n_1} - \frac{1}{n_1} + \frac{1}{p_2 n_2} - \frac{1}{n_2}$
odds ratio	$\ln\left[\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right]$	$\ln\left[\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}\right]$	$\frac{1}{p_1 n_1} + \frac{1}{(1-p_1)n_1} + \frac{1}{p_2 n_2} + \frac{1}{(1-p_1)n_2}$
correlation coefficient	r	ρ	$\frac{(1-r^2)^2}{n-1}$
transformed correlation	$\frac{1}{2}\ln\left[\frac{1+r}{1-r}\right]$	$\frac{1}{2}\ln\left[\frac{1+\rho}{1-\rho}\right]$	$1/(n-3)$
proportion	p	π	$p(1-p)/n$
log odds	$\ln[p/(1-p)]$	$\ln[\pi/(1-\pi)]$	$1/[np(1-p)]$
mean	$\bar{x}$	μ	s^2/n

Examples: d- versus r-family measures

d-family measures see also: Bernard_et_al_d_Family_calculations.pdf	r-family measures
$Glass\_\Delta = \frac{M_1 - M_2}{s_{control}}$	Pearson´s r
$Cohen's\_d = \frac{M_1 - M_2}{\sigma}$	Fisher´s Z_r
$Hedges's\_g = \frac{M_1 - M_2}{s_{pooled}}$	r_{pb} (dichotomous/continuos)
	ρ (ordinal/ordinal)

Example: Odds Ratio (OR)

Odds Ratio computation from a 2*2 frequency table

	with risk factor	**without risk factor**
symptoms	a	b
no symptoms	c	d

$$\text{Odds Ratio} = \frac{a/c}{b/d} = \frac{a \cdot d}{b \cdot c}$$

The Big Picture: Basic Steps

1. Choose a unified effect size measure that is natural, i.e. theoretically (and/or practically) justifiable for your meta-analytic research question(s)
2. Extract the information needed to estimate effect sizes and sampling variances/SEs from primary studies
   - take into account ´hierarchy of effect size approximations´ (see next)
   - use tools assisting in extracting the appropriate information (see next)
3. Estimate effect sizes and sampling variances SEs on the unified effect size measure
   - take into account ´hierarchy of effect size approximations´ (see next)
   - use tools assisting in extracting the appropriate information (see next)

Hierarchy of ES Approximations

Quality	Method of Calculating or Estimating Effect Size
High Quality ES	• Direct calculation based on means and standard deviations • Algebraically equivalent formulas (i.e., *t*-test, and two-group *ANOVA*) • Exact probability value for a *t*-test or *ANOVA* (two groups) • Approximations based on continuous data (correlation coefficient)
Medium Quality ES	• Estimates of the mean difference (e.g., adjusted means, regression β weight, gain score means) • Estimates of the pooled standard deviation (e.g., gain score standard deviation, one-way *ANOVA* with three or more groups, *ANCOVA*)
Low Quality ES	• Estimates based on a probability of a significant t using α (e.g., $p < .05$) • Approximations based on dichotomous data

Important: Always code calculation/approximation method!
Use this variable later for moderator analysis, i.e. to check if approximation method is systematically related to ES magnitude.

Tools for ES Selection, Computation, and Unification I

- Decision Trees to Select Effect Size Measures
  - ... for studies involving group contrasts on dependent variables:
    Lipsey & Wilson (2001), Chapter 3, Figure 3.1, p. 58
  - ... for studies involving correlation or association between variables:
    Lipsey & Wilson (2001), Chapter 3, Figure 3.2, p. 68
- Effect sizes, Standard Errors, and Inverse Variance Weights for each Effect Size Type:
  - Lipsey & Wilson (2001), Chapter 3, Table 3.2, p. 72
- Procedures to Compute and Convert ESs and SEs:
  - Lipsey & Wilson (2001), Appendix B, p. 172pp.
  - Cooper, Hedges & Valentine (2009), Chapter 12
  - Borenstein et. al (2009): Part II, esp. Chapter 7
  - Cooper (2010), Chapter 6
  - Bernard_et_al_d_Family_calculations.pdf (see supplements)

Tools for ES Selection, Computation, and Unification II

- MS Excel Effect Size Computation Program: ES_Calculator.xls
- Online Calculators listed at www.meta-analysis.eu
- Effect size calculators on meta-analytic software programs (see part 5), such as:
  - Meta-Win
  - Comprehensive Meta-Analysis
  - etc.

Exercises: *d*-Family

Overall: Please compute/estimate/approximate *d/g*

- High (to medium, for dependent samples) quality ES data provided:
  $t = 3.2$, $df = 10$, d=?
  (for dependent/independent samples)

- Low quality ES data provided:
  - „group differences for n(1)=n(2)=61 were significant at $p < .05$“
  - „50% improved in the treatment group, compared to only 10% in the control group; variances were not affected by the treatment“

Exercises: r-Family

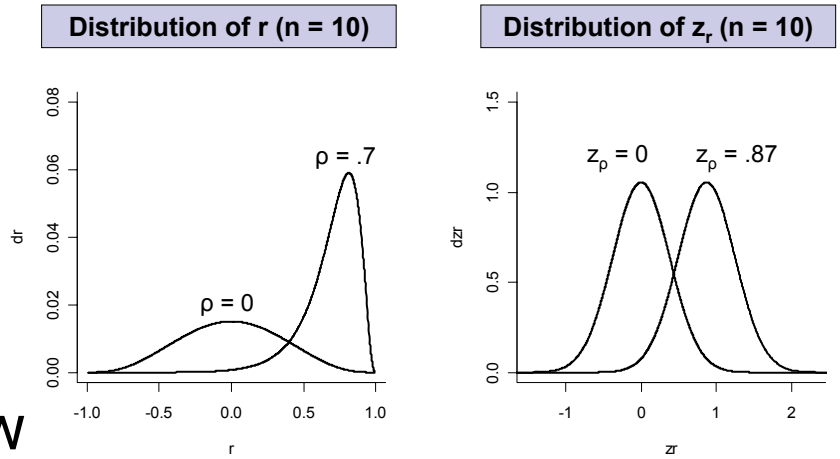
Overall: Please compute/estimate/approximate *r*

- High quality ES data provided: Variance-Covariance information available, e.g. *COV*(a,b)= 6.6, *VAR*(a)= 7.5, *VAR*(b)= 9.8, *r*= ?
- Medium quality ES data provided: Regression weight *beta*= .45 for relationship of interest in a two-predictor model
- Low quality ES data provided: 2*2 frequency table of *artificially dichotomized* continuous variables

	high V2	low V2
high V1	183	98
low V1	117	518

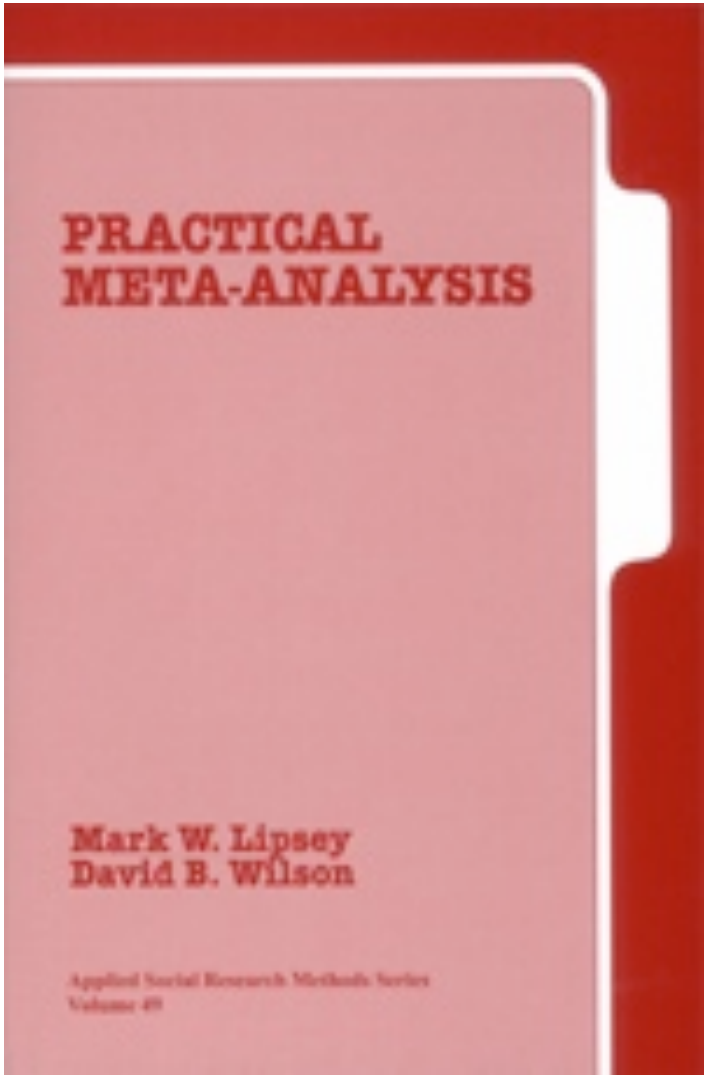
Special Topics

- Cohen´s *d* or Hedges *g* (/w small sample correction)?



- Fisher´s *z* or *r*?
  - The statisticians view:
    Transformation needed to achieve *N*(rho, thau); stat. theory well developed.
  - The Hunter & Schmidt view
    z-Transformation results in more serious ´upward bias´ than ´downward bias´ if *r* are used (corroborated with the aid of simulations by Field, 2001; Schulze, 2004)
- Standardized regression coefficients as *r*?
  - The ´statistical fruit salad´problem (Brüderl, 2004)
  - Imputation recommendation by Peterson & Brown (2005) based on simulations (esp. two predictor case)
  - Research literature still in its infancy

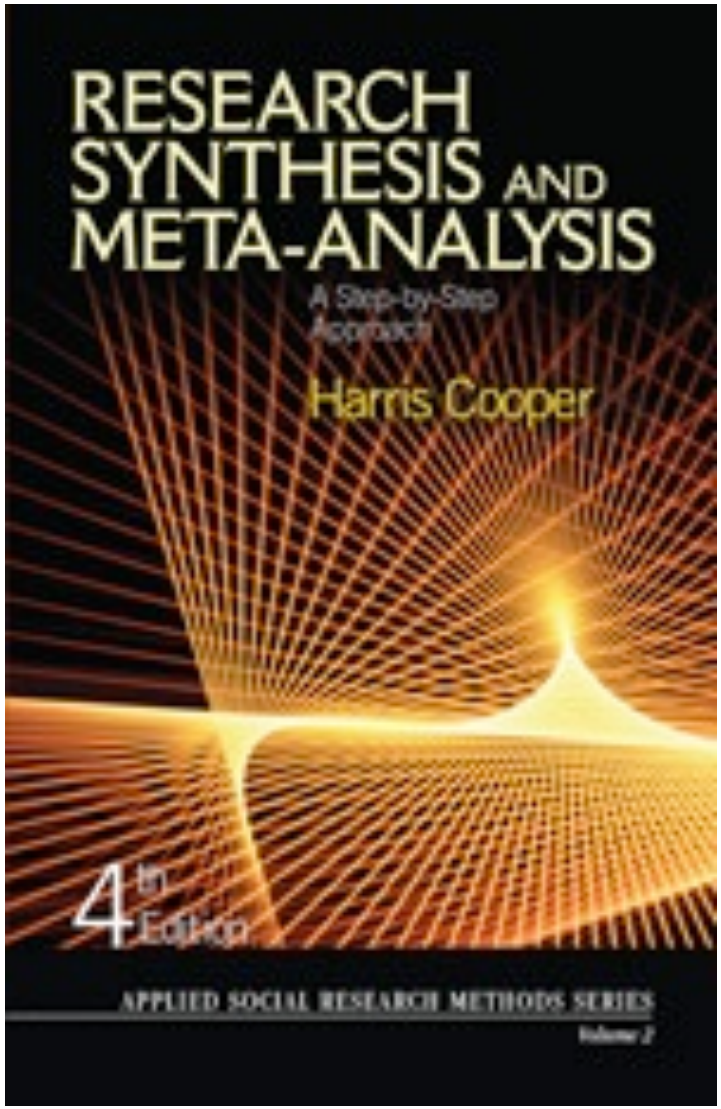
Lipsey & Wilson (2001)



Lipsey, M.W., & Wilson, D.B.(2001). *Practical Meta-analysis.* Thousand Oaks: Sage.

- Chapter 3: Selecting, Computing, and Coding the Effect Size Statistic
- Chapter 4: Developing a Coding Scheme and Coding Study Reports
- Chapter 5: Data management
- Appendix B: Procedures for computing effect size values from eligible study reports
- Appendix C: MS Excel effect size computation program
- Appendix E: Coding manual and coding forms .... (examples)

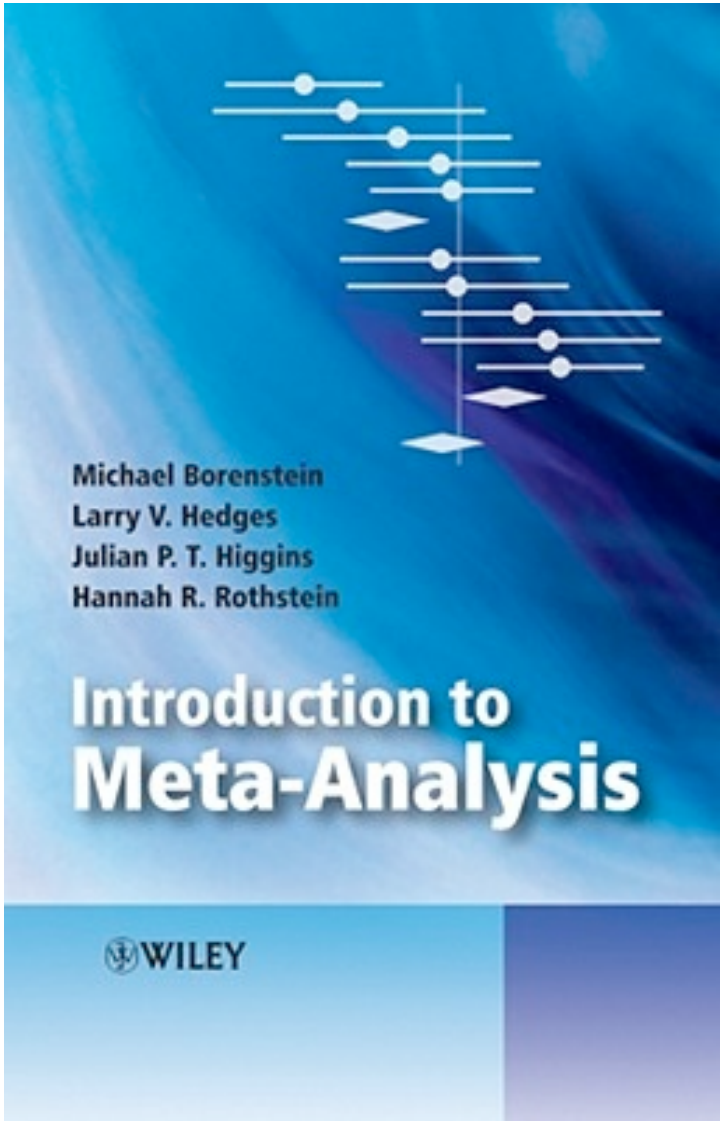
Cooper (2010)



Cooper, H. (2010). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach.* Thousand Oaks, CA: Sage.

- Chapter 4: Step 3: Gathering information from studies
- Chapter 5: Step 4: Evaluating the quality of studies
- Chapter 6: Step 5: Analyzing and integrating the outcomes of studies

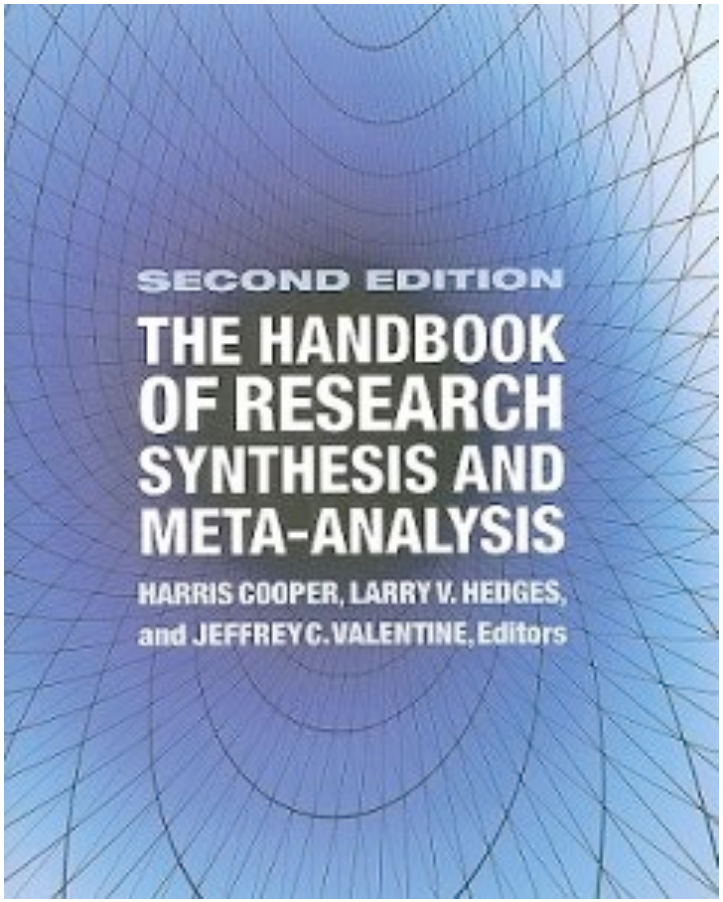
Borenstein et al. (2009)



Borenstein, M., Hedges, L.V., Higgins, J.P.T, & Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: Wiley.

- Part 2: Effect size and precision (encompassing chapters 3-9)

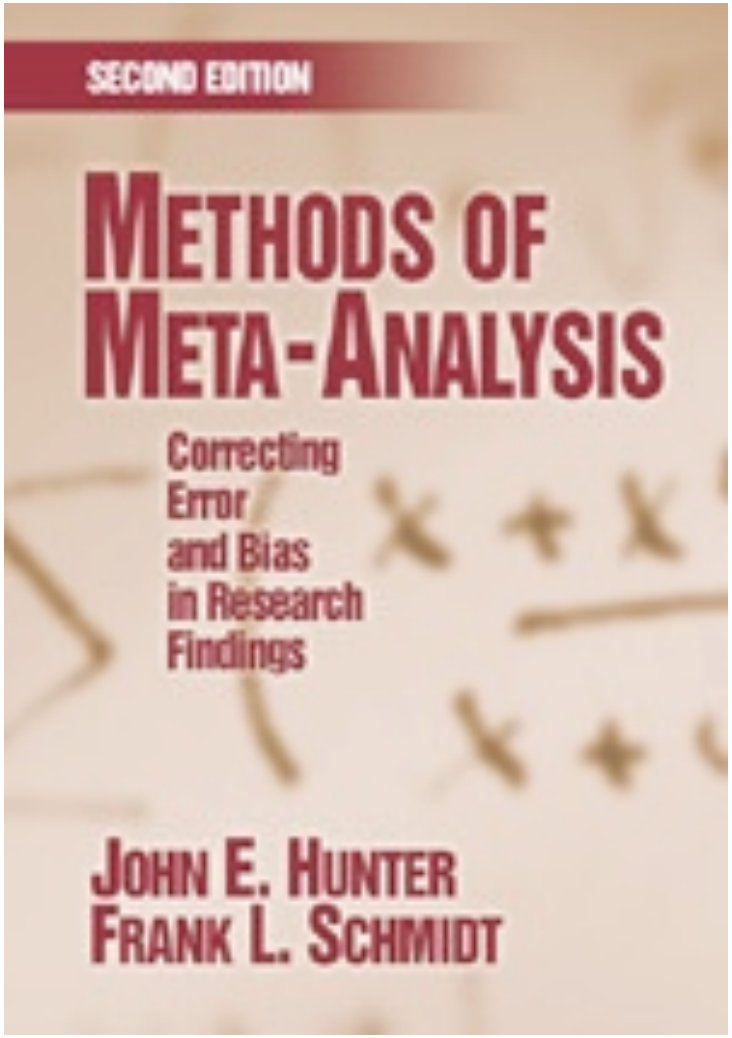
Cooper, Hedges & Valentine (2009)



Cooper, H., Hedges, L.V., & Valentine, J.C. (Eds.) (2009). *Handbook of Research Synthesis (2nd ed.)*. New York: Russell Sage Foundation.

- Chapter 7: Judging the quality of primary studies (by Jeff Valentine)
- Chapter 8: Identifying interesting variables and analysis opportunities (by Mark Lipsey)
- Chapter 9: Systematic coding (by David Wilson)
- Chapter 10: Evaluating coding decisions (by Robert Orwin and Jack Vevea)
- Chapter 12: Effect sizes for continuous data (by Michael Borenstein)

Hunter & Schmidt (2004)



Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.)*. Thousand Oaks, CA: Sage.

- Basic overview in part IV: General issues in meta-analysis