Meta-Analysis Workshop

Part 7: Special Topics

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

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

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Agenda

- **Criticism of meta-analytic techniques**

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Criticism: Overview

- One number cannot summarize a research field
- Selective publication (bias): The file drawer problem invalidates meta-analyses
- Incommensurability
  - ... of aggregated constructs (´mixing apples and oranges problem´)
  - ... of corrected/adjusted versus uncorrected/ unadjusted effect sizes (´statistical fruit salad problem´)
- Garbage-in, garbage out (GiGo)
- Meta-analyses are performed poorly

„One number cannot summarize a research field“

- Criticism: Summary effect reporting ignores (´substantial´)variability among studies

- Response(s):
  - critique concerns bad application and interpretation/ reporting, not meta-analysis methodology per se
  - no dispersion: mean adequately representing studies included
  - modest dispersion: placing the mean in context needed
  - large dispersion: focus should shift to explain dispersion itself!

Publication Bias Invalidates Meta-Analyses

- Criticism: Mean ES estimates are biased towards over-reporting of significant/larger effects

- Response(s):
  - legitimate concern, but file drawer problem general problem in primary research, all summaries are facing it
  - meta-analysis comes with formal procedures to check/test and quantify (!) the amount of publication bias
  - specific methods to quantitatively correct for, or account for publication bias

Incommensurability: Apples-and-Oranges

- Criticism: Summary effects does ignore important differences between studies

- Response(s):
  - meta-analysis do almost always address broad research questions (i.e., fruit, not only apples/ oranges), generalizability of primary studies is therefore informed by theory
  - potential differences can be detected and investigated formally, e.g., with the aid of
    - homogeneity/heterogeneity analysis
    - subgroup/moderator analyses etc.

Incommensurability: Stat. Fruit Salad

- Criticism: Corrected/adjusted versus uncorrected/ unadjusted effect sizes must not be pooled, because they estimate a different population parameter

- Response(s):
  - true based on statistical theory
  - meta-analytic theory does not permit this, therefore errors on the level of implementation by researchers
  - influence/magnitude can be formally detected (sensitivity analysis)
  - correction formula, modeling and reversing the correction(s) used, might be applied

Garbage-in, Garbage-out

- Criticism: Many low-quality studies are carried over and will (a) invalidate the results and (b) will be hard to identify.

- Response:
  - meta-analysis can be regarded as a ´waste management´process, i.e. does provide formal procedures to account for varying quality in primary studies
    - via inclusion/exclusion criteria
    - via coding study quality and using this information to investigate whether variation in characteristics os studies is related to the size of the effect(s)

Poorly performed meta-analyses

- Criticism: Mistake are common in meta-analyses because of them being inherently complicated (Bailar, 1997)
- Response:
  - correct statement (that many meta-analyses contain errors)
  - implementation problem, not due to meta-analytic theory
  - needs to be addressed with the aid of improving standards (education, peer-reviewing, protocols required ahead of time, etc.)
  - impact of subjective and apparently ´erroneous´ decisions can be modeled (see Bosnjak & Viechtbauer, 2009, and below)

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - **Simpson´s paradox**
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Simpson´s Paradox

- Instead of synthesizing effect sizes, we could simply pool the raw data to estimate the overall effect (assuming we do have the raw data)!
- What do you think about this claim?
- Problem(s):
  - we may get wrong answers due to confounding factors and disproportionate Ns between treatment/control groups (for case-control studies)
  - meta-analyses ensures that each study serves as its own control, canceling confounders (partly) out
  - meta-analysis allows us to determine the consistency of studies (raw data pooling would assume homogeneity ex ante)
  - ´Wrong´ answers from raw data pooling with possible large scale public policy implications (Example: Van Howe, 1999, erroneously concluding that circumcision increases HIV risk; see Bornstein et. al, 2009, Chapter 33)

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - **(Stochastically) dependent effect sizes**
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Dependent Effect Sizes

- Types of dependencies among effect sizes:
  - stochastic dependence:
    ES estimates are based on the same subjects (and therefore correlated)
    - reported within one primary publication
    - repeatedly reported in more than one primary publication
  - other types of dependences (i.e., systematic between ES correlations):
    - same author(s) as proxy for systematic biases
    - same scientific ´circles´ as proxy for systematic biases
    - etc.

Types of Stochastically Dependent ES

- *multiple treatment studies*:
  More than one treatment group is compared against the same control group

- *multiple-endpoint studies*:
  - more than one dependent variable is used, e.g. role of an intervention on both math and reading scores (representing basic cognitive skills) in the same subjects.
  - more than one data point in time is used, e.g. repeated assessment of a parameter using the same subjects

Handling Stochastically Dependent ES

- Select one per subject sample
  - randomly
  - on the basis of some criteria
- Average (see, e.g. Marín-Martínez & Sánchez-Meca, 1999, for an overview)
  - Rosenthal/Rubin procedure (preferable)
  - Hedges/Olkin procedure
  - Simple averaging (not recommended)
- Modeling dependencies (complex data structures)and adjusting the variances according to the within-study correlations
  - described for all usual ES types and cases in Bornstein et. al (2009), Part 5
  - Some programs allow for modeling complex data structures (e.g., CMA)

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - **Publication bias (File drawer problem)**
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Publication Bias (File-Drawer Problem)

- Probability of publication/dissemination of results may depend on the ´significance´ or size of the effect(s) reported, and could therefore bias the meta-analytically computed estimates *systematically*
- Past studies suggest that ´significant´ results are more likely to be published, and that ´gray´ literature (including dissertations) are underrepresented in meta-analyses
- Publication bias may influence all types of research, including primary studies (partly) based on past research and other reviews
- Meta-Analysis should try to assess its impact

Publication Bias

General (implicit) assumption: Publication bias *increases* as the sample size goes down

- Large studies are likely to be published regardless of stat. significance (large N > ´credibility´)
- Moderately sized studies more likely not to be published if results insignificant
- Small studies most likely not to be published of insignificant results (´underpowered´, least ´credible´)

- ***Important:*** Assumption cannot be (dis)confirmed in each an every case:
  - Larger small study effect might be ´real´ (a small, extreme group might have benefitted more from a treatment, e.g. severely ill)
  - Small studies could have been more carefully conducted and therefore more ´valid´ than larger (less carefully implemented) ones

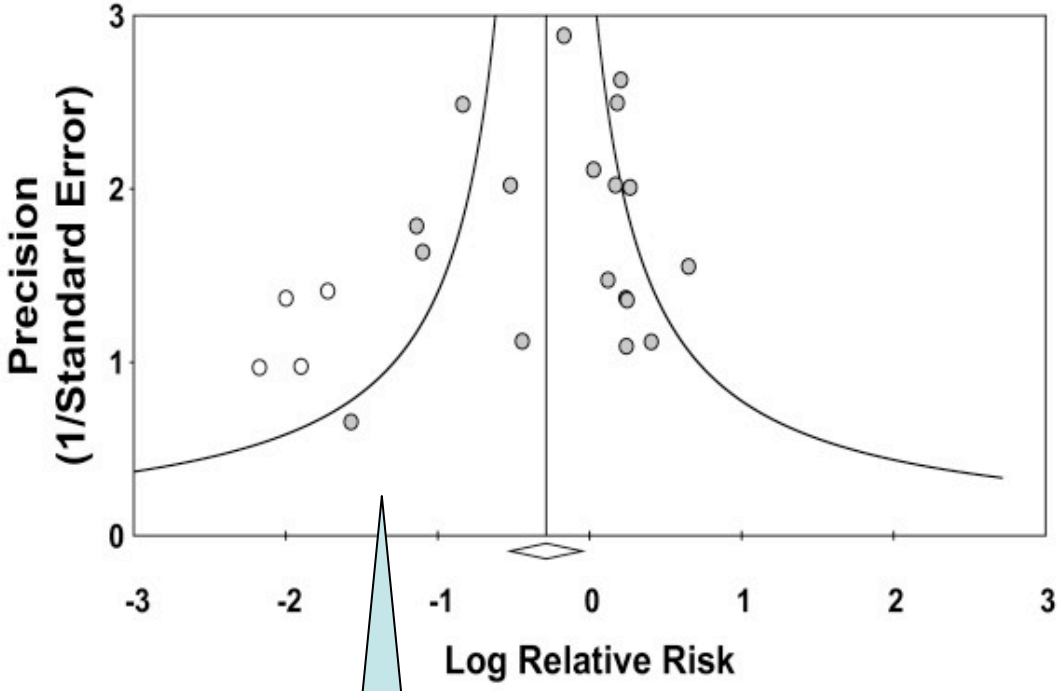
Publication Bias: Related Concepts

- Language bias (selective inclusion of English literature)
- Availability bias (selective inclusion of studies that are easily to access)
- Cost bias (selective inclusion of free / low-cost material)
- Familiarity bias (selective inclusion of studies only from own discipline)
- Duplication bias (stat. significant results are are more likely to be published more than once)
- Citation bias (significant studies are more cited, and have therefore a higher probability to be found)
- ....

Addressing Publication Bias

- **Ex ante: Prevention is the best solution!**
  - Broad inclusion criteria
  - Importance of comprehensive search process
  - Prospective registries / protocols
- Ex post: Employ methods aimed at
  - identifying bias („Is there a bias at all?“)
    - forest plots ordered according to relative weights
    - funnel plot
    - rank correlation test
    - Egger´s regression test
  - determining if the entire effect is an artifact of bias
    - Fail-Safe-N approaches (Rosenthal-method, Orwin-method, Rosenberg-method)
  - estimating the impact of publication bias
    - Trim-and-Fill Analysis
    - Selection Modeling

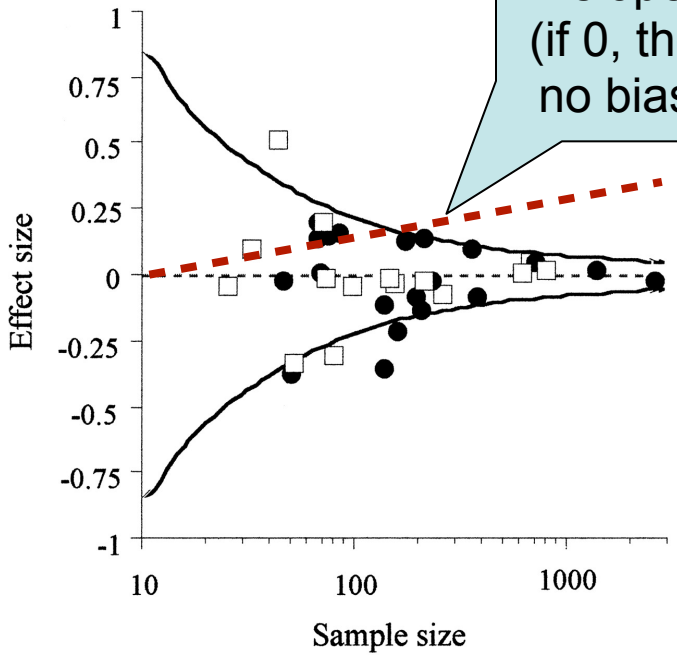
Funnel Plot & Related Tests



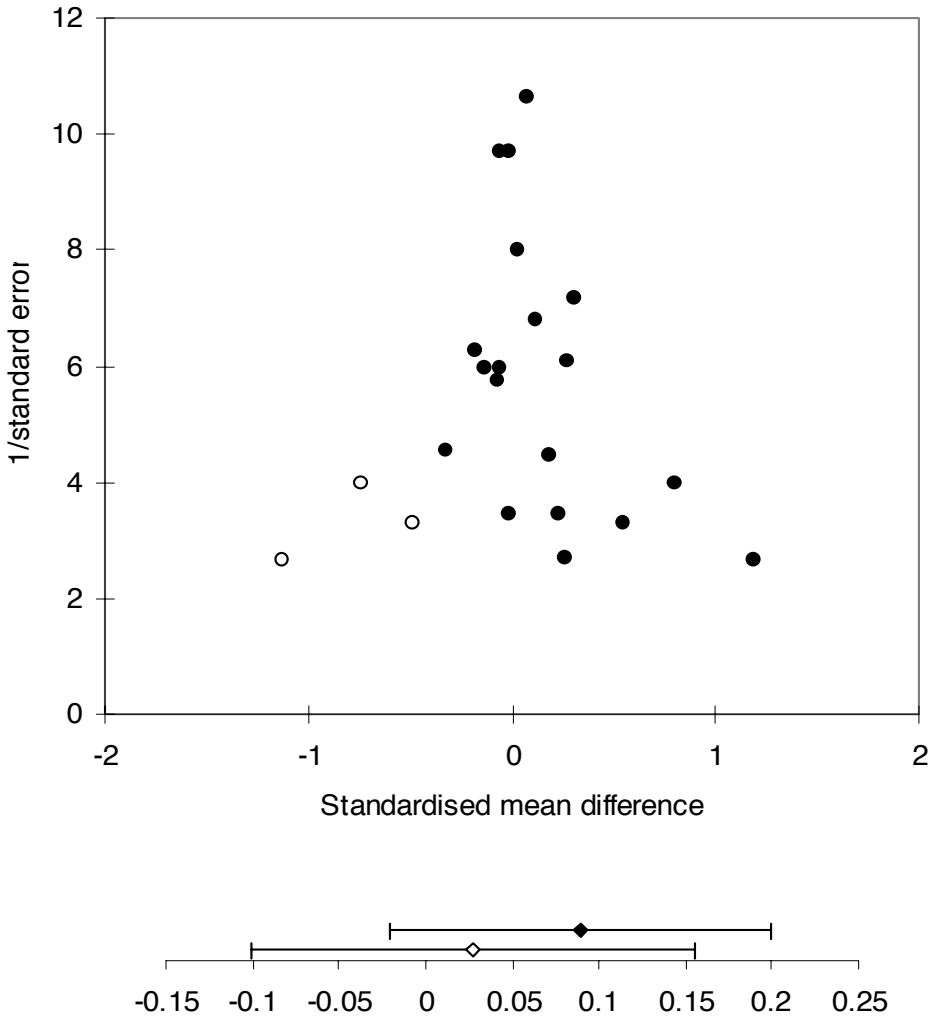
Computing a rank correlation (> detects also non-linear dependences) between ES and SE

Regression test: Rotation

Estimating slope (if 0, then no bias)



Funnel Plot & Trim-and-Fill



Iterative procedure:

Step 1: ´Trimming´ the funnel plot by elimination the X most extreme ESs on one side (direction of bias and X may be defined by the researcher)

Step 2: Computing the new summary ES (pooled estimate)

Step 3: Fill funnel plot by mirroring (around the new mean ES) all, including trimmed and mirrored, data points

Step 4: Calculate new variance of pooled estimate

Step 5: Determining symmetry of funnel plot. If symmetry, STOP. If asymmetry, re-start at step 1.

Fail-Safe-N Approaches: Rosenthal I

- Rosenthal´s method (1979)
  - Based on the Stouffer-method (aggregation of exact p-values)

$$Z_s = \frac{\sum_{i=1}^{k} z_{p(i)}}{\sqrt{k}}$$

z-transformed p-values of research finding i

k No of included research findings

  - determines the No of missing studies with *insignificant* effects needed (in addition to those included) to render the summary effect non-significant
  - suggested tolerance level to judge whether or not it is likely that missing zero-effect studies are of a concern: $5k+10$

Fail-Safe-N Approaches: Rosenthal II

Major shortcomings have been brought up, such as:

- p-Value aggregation, focusing stat. (in)significance, does therefore not follow contemporary meta-analytic logic (substantive significance: effect size magnitudes)
- precision(s) / sample size(s) are not account for
- shape of funnel plot / ES distributions not accounted for
- Underlying assumption about missing studies: Non-significant, zero-effect: How about negative effect sizes? Fewer undetected studies could nullify the effect.

Fail-Safe-N Approaches: Orwin & Rosenberg

- Orwin (1983):
  - fail-safe-N variant for effect size magnitudes (not p-levels)
  - No of studies needed to bring the overall effect to a *pre-defined (substantively important)* overall ES level
  - allows the researcher to specific the mean effect of missing studies other than zero (modeling of other distributions possible)
- Rosenberg (2005):
  - proposes a weighting method for the Orwin approach, accounting for study precision(s) in fail-safe-N estimation
  - weighting formula for both fixed and random effects

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - **Model-based meta-analysis (MASEM)**
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Model-based Approaches (MASEM)

- How to test multivariate models with the ad of meta-analyses?

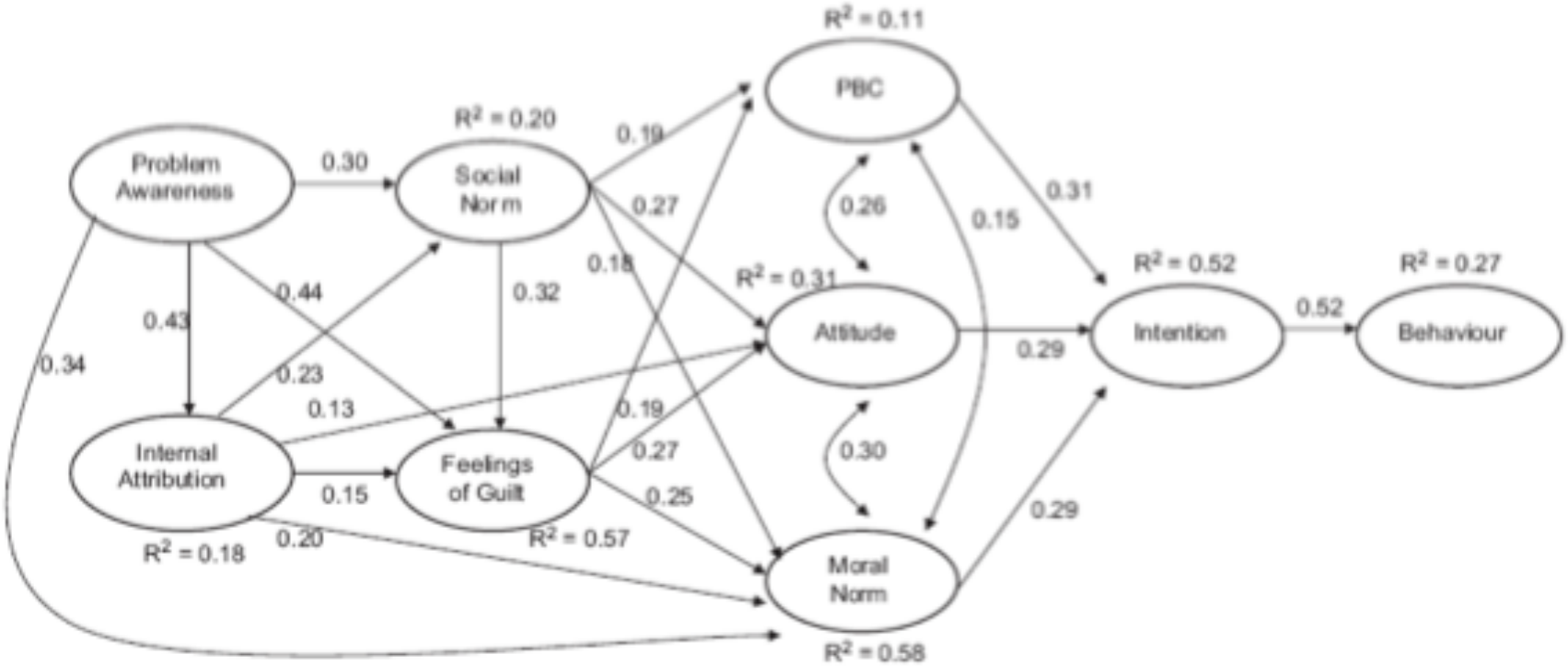


Fig. 1. Results of the MASEM based on pooled random-effects correlations, PBC = perceived behavioural control, single-headed arrows = standardised path-coefficients; double-headed arrows = correlations, R^2 = explained variance.

MASEM: Univariate Synthesis

Table 1
Total sample size (upper row) and number of independent correlation matrices (lower row) obtained for each construct

Construct	1	2	3	4	5	6	7	8	9
1. Problem	11,740								
	(23)								
2. Attribution	2471	3565							
	(7)	(9)							
3. Social norm	3358	175	12,255						
	(6)	(1)	(33)						
4. Guilt	2760	443	4154	4597					
	(4)	(1)	(6)	(7)					
5. PBC	3789	175	10,467	4154	13,378				
	(8)	(1)	(25)	(6)	(34)				
6. Attitude	1727	175	9053	2627	9588	12,345			
	(4)	(1)	(26)	(4)	(24)	(33)			
7. Moral norm	8957	1540	8126	4154	8836	6646	14,022		
	(15)	(4)	(16)	(6)	(18)	(14)	(26)		
8. Intention	6571	3068	7900	3070	8356	8551	8907	14,365	
	(13)	(7)	(22)	(5)	(24)	(23)	(19)	(36)	
9. Behaviour	8276	1866	7325	3203	8029	6751	6840	5654	14,394
	(18)	(6)	(18)	(5)	(18)	(17)	(11)	(15)	(36)

Table 2
Fisher's Z-back-transformed pooled "true" correlation matrix under the fix-effects (upper triangular matrix) and random-effects assumption (lower triangular matrix)

Construct	1	2	3	4	5	6	7	8	9
1. Problem	—	.43	.42	.61	.10	.30	.61	.38	.22
2. Attribution	.43	—	.36*	.45*	.18*	.36*	.53	.33	.25
3. Social norm	.40	.36*	—	.57	.31	.49	.61	.46	.31
4. Guilt	.63	.45*	.55	—	.31	.52	.66	.55	.31
5. PBC	.11	.18*	.29	.29	—	.49	.41	.61	.30
6. Attitude	.27	.36*	.47	.48	.44	—	.81	.66	.54
7. Moral norm	.63	.53	.53	.66	.35	.67	—	.63	.58
8. Intention	.40	.33	.42	.50	.54	.62	.59	—	.52
9. Behaviour	.19	.24	.31	.30	.30	.42	.39	.52	—

*No pooled correlation.

MASEM: Two-Stage Approach (TSSEM)

- TSSEM: *One* MASEM approach performing multivariate meta-analyses under the SEM umbrella
- TSSEM essential stages:
  1. Estimating a pooled correlation matrix using multiple group SEM (groups: indiv. corr. matrices)
  2. SEM with ADF estimation method on the pooled correlation matrix (total sample size as N for SEM)
- Overall: MASEM in an early infancy stage, despite almost 20 years of multivariate MA research questions
- Mike Cheung´s Website on recent papers and a LISREL syntax to perform TSSEM
  http://courses.nus.edu.sg/course/psycwlm/Internet/publications.html

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - **Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)**
  - Errors in published meta-analyses

- Overview of threats to validity of meta-analytic conclusions

Recap: 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

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

Relative Risk (*RR*) versus *OR*

	with risk factor	without risk factor
symptoms	130	70
no symptoms	1870	7930

$$RR = \frac{130/(130+1870)}{70/(70+7930)} \approx 7,4$$

$$\text{Odds Ratio} = \frac{130 \cdot 7930}{70 \cdot 1870} \approx 7,88$$

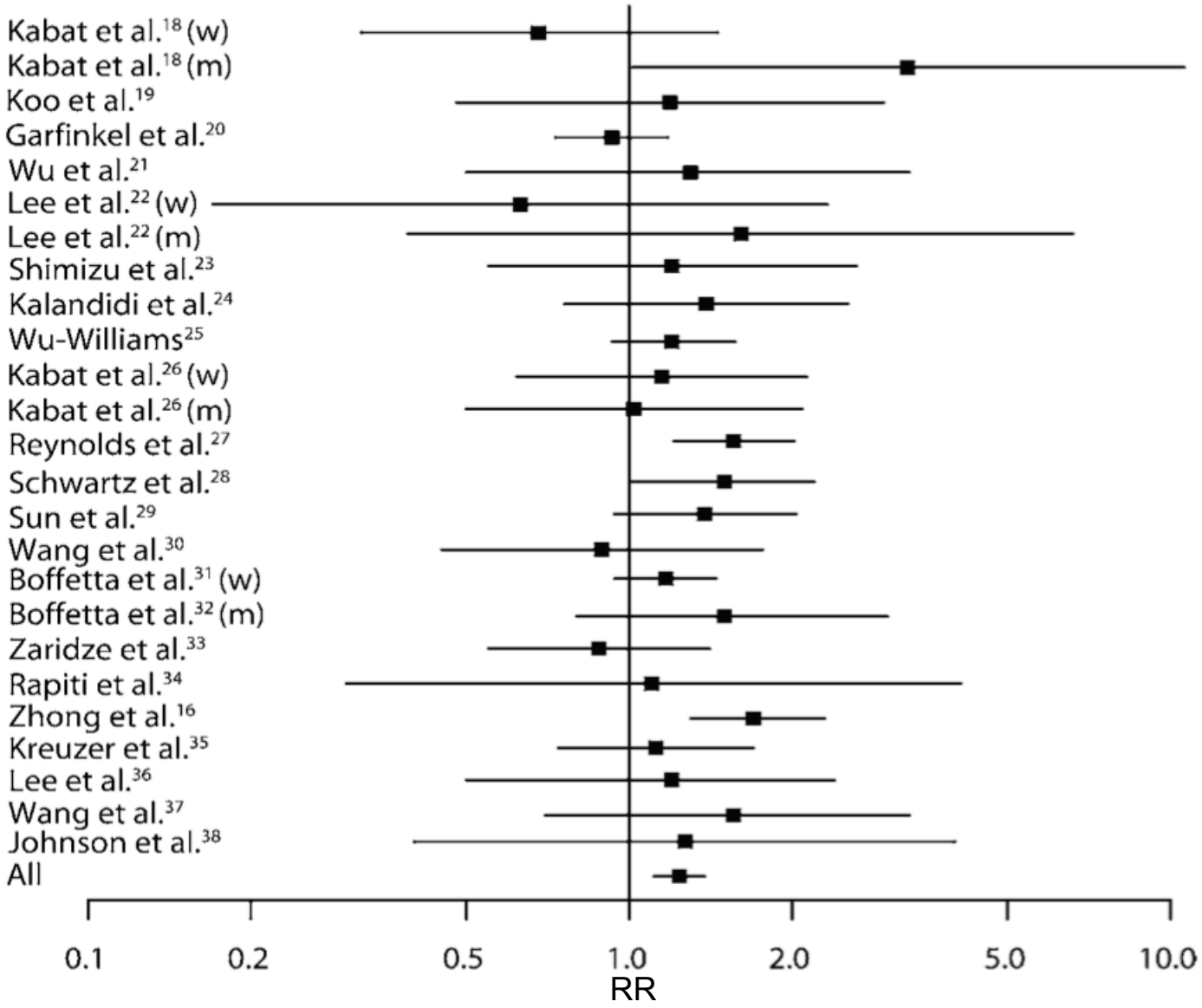
Important:
RR depends on table design (see Bosnjak & Viechtbauer, 2009), therefore *NOT* recommended

Meta-Analysis Stayner et al. (2007)

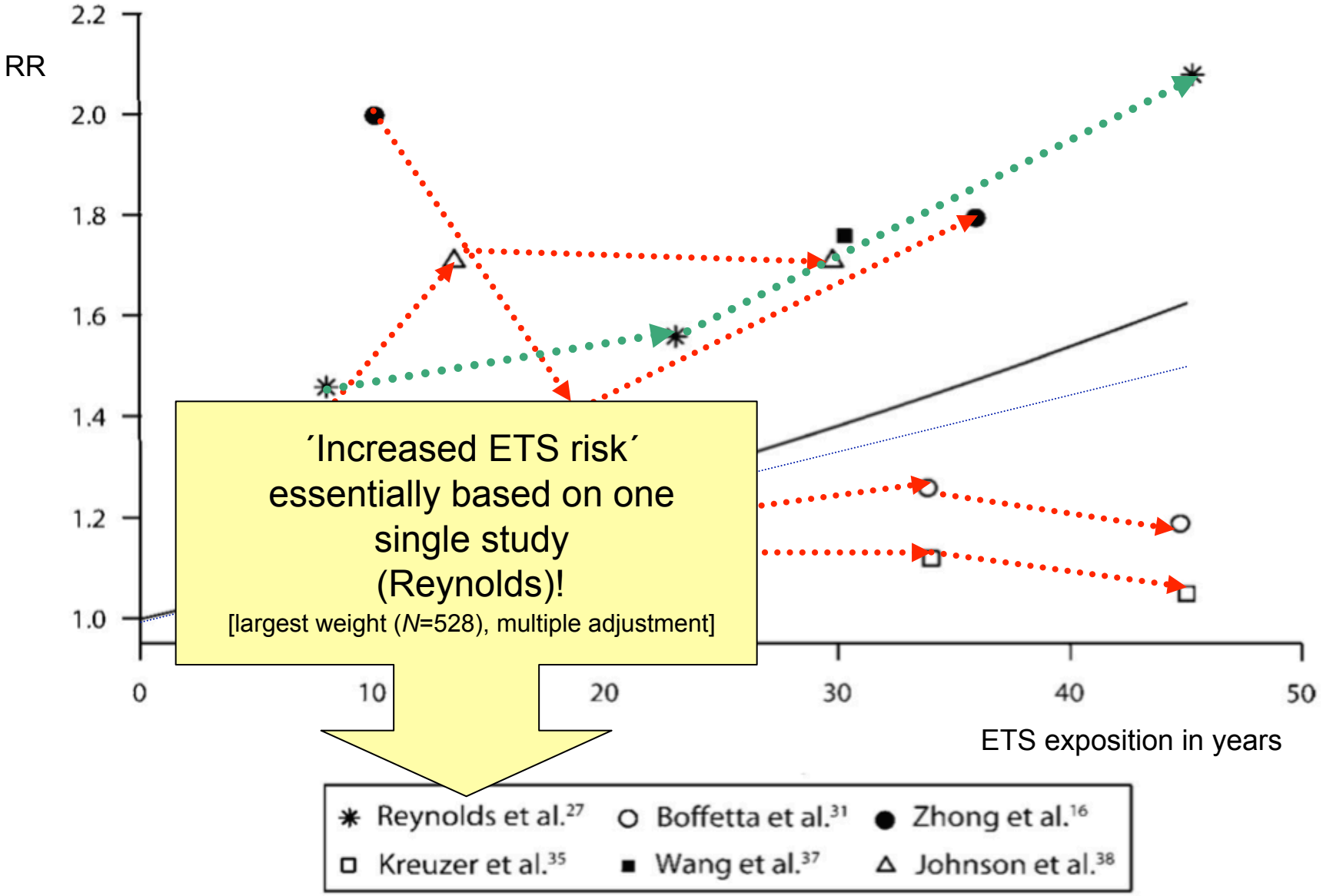
- Increased lung cancer risk in nonsmokers due to ETS exposition at workplace?

- 25 effect sizes (RR) based on 22 papers
- 18 effect sizes were (multiply) adjusted (statistically corrected for the influence of potential confounding variables)
- Mean RR was estimated, homogeneity analysis
- Meta regression to detect moderators
- Dosage-response analysis for time based on 6 primary studies and 17 RR estimates
- Technical procedure: Normand (1999), SAS Proc Mixed, HO meta-analysis

Stayner et al. (2007) : Results I



Stayner et al. (2007) : Results II



Critical Issues : Stayner et al. (2007) I

Quality of documentation (selection):

- Criteria for inclusion/exclusion? Search strategy?
- Information of coding quality?
- Problem of dependent effect sizes?
- Procedure followed for adjusting ES?
- „Statistical fruit salad“: Adjusted and unadjusted estimates mixed.
- Sensitivity and outlier analysis only sketched.

> *Replicabililty low, logical flow of analyses partly unclear*

Documentation in (Medical) Meta-Analyses

Cochrane Collaboration**:

- QUORUM Statement
  (Standards of Reporting of Meta-Analyses)
- STARLITE Statement
  (Standards for Reporting Literature searches)

Coding quality**:

- CONSORT Statement (for parallel-group randomized trials)
- TREND Statement (for non-randomized trials)
- STROBE Statement (Strengthening the Reporting of Observational studies in Epidemiology)

**http://www.cochrane.de/de/statements.htm

Stayner et al. (2007): Conclusions

Robustness of results in this specific case?

General principle(s) influencing robustness of meta-analyses in general? (*Crossroads model*)

Subjectivity & Meta-Analyses: Past Studies

Mengersen et al. (1995):

- Method choice influences meta-analytic results in passive smoking research syntheses.
- Choices investigated:
  (a) approximate versus exact statistical techniques, (b) fixed versus random effects models, (c) publication bias / choice of studies included (> *data generation phase*).

Barnes & Bero (1998):

Affiliation of ETS research synthesis authors (tobacco industry / non tobacco industry) single best predictor of results and final conclusion drawn.

Overall Research Question(s)

Replicability of the Stayner et. al. (2007) meta-analysis based on …

1. … the data reported in the paper?
2. … newly coded data?
3. … data taking <u>subjective decisions within a meta-analytic process</u> into account?

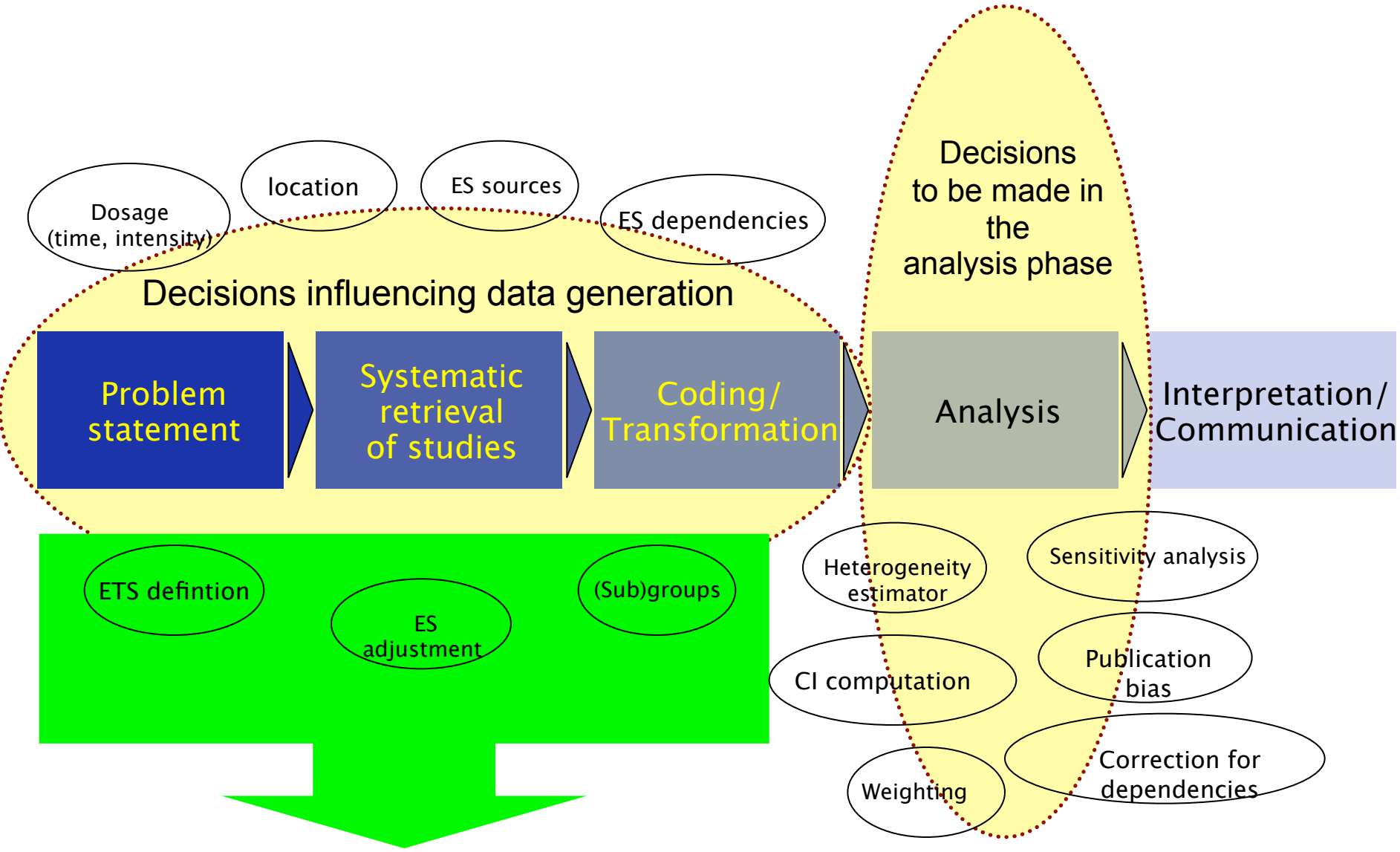
A ´Crossroads Model´ : Basic Idea

- Modeling all possible and reasonable (but still subjective) decisions within a meta-analytic process on the level of (a) data generation and (b) data analysis.

- Current focus of project: Data generation phase (excluding literature retrieval)

- Next step(s), not reported here, simulating the effect of different data analysis approaches on the results (or ranges of "reasonable" results).

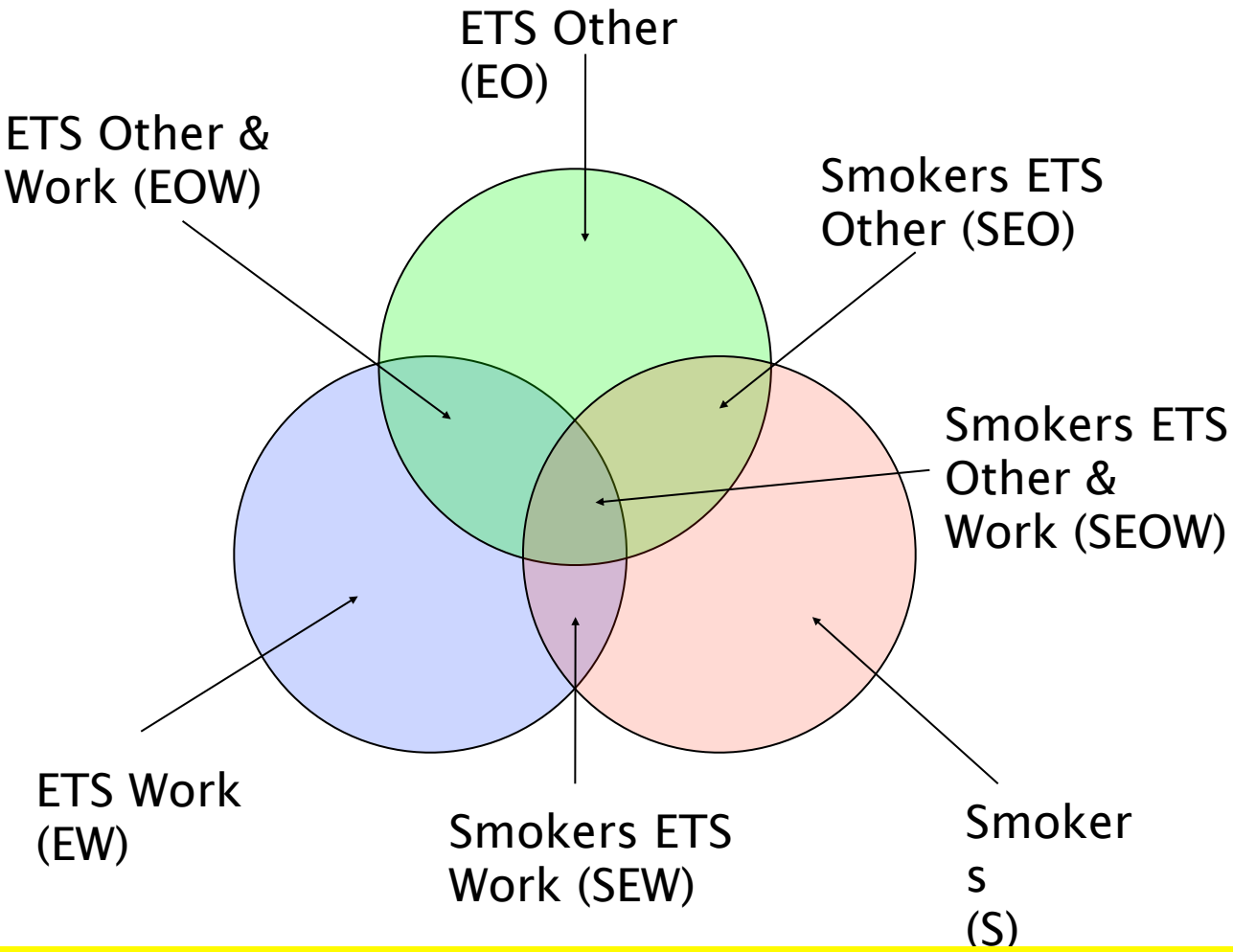
A ´Crossroads Model´ : Examples



Overview of Methods

- Retrieval of all primary studies used in the Stayner et al. (2007) meta-analysis
- Comprehensive re-coding of data by employing:
  - Different ETS definitions implicitly used in the Stayner paper
  - ES computed for different groups (male/female/ both)
  - Adjusted and unadjusted ES computed
  - Quality criteria, intercoder reliability, etc.
- Simulating the effect of all subjective decisions under different restrictions (work in progress)
- Today: Tails of the simulation results (max./min. RR) if only three decisions on the level of data generation are addressed.

ETS (at Work) Definitions



Definitions (implicitly) used in the Stayner et al (2007) paper:
(a) EW, (b) EW & EOW (c) EW & EOW & EO

BGN_Stellschrauben über alle ES bei a...

	A	B	C	D	E	G	H	I	J
1	Study	Erstautor (Jahr)	Definition	Un(corrected)?	Gender	Odds Ratio1	N OR1	OR1 Erklärung	OR2
2	Studiennur		1=Eng, 2= mittel, 3=weit	Adjustment der ES allgemein		keine LnORs!	Hier alle N, nicht nur Cases; NR= Not Reported	Erst über alle Kategorien	
147	14	Boffetta (1998)	1	uncorrected	male				
148	14	Boffetta (1998)	1	uncorrected	female				
149	14	Boffetta (1998)	1	uncorrected	both				
150	14	Boffetta (1998)	1	corrected	male				
151	14	Boffetta (1998)	1	corrected	female				
152	14	Boffetta (1998)	1	corrected	both				
153	14	Boffetta (1998)	2	uncorrected	male	1.16974494283201	672	Ever exposed yes/no	1.2037825059101
154	14	Boffetta (1998)	2	uncorrected	female	1.25976015406162	1520	Ever exposed yes/no	1.1788324979114
155	14	Boffetta (1998)	2	uncorrected	both	1.08881261123824	2192	Ever exposed yes/no	1.091448360993
156	14	Boffetta (1998)	2	corrected	male	1.13	484	Angaben aus dem Text S. 1444 links	
157	14	Boffetta (1998)	2	corrected	female	1.19	1520	Ever exposed yes/no	1.14
158	14	Boffetta (1998)	2	corrected	both	1.17	2192	Ever exposed yes/no	1.15
159	14	Boffetta (1998)	3	uncorrected	male	1.13778280542986	672	Ever exposed yes/no	1.1633663366336
160	14	Boffetta (1998)	3	uncorrected	female	1.1652281134402	1517	Ever exposed yes/no	1.1266006434113
161	14	Boffetta (1998)	3	uncorrected	both	1.21929130096504	2189	Ever exposed yes/no	1.1894183045463
	14	Boffetta (1998)	3	corrected	male				

Overview of Analyses

Block 0: Recap.: Published Results (fixed/random)

Block 1: Replication based on the data reported in the published paper (fixed, random)

Block 2: Replication based on newly coded data (fixed, random)

Block 3: Exploring the tails of the first simulation

- „good guy" analysis (unfavorable towards ETS; fixed, random)
- „bad guy" analysis (favorable towards ETS; fixed random)

Results : Block 0

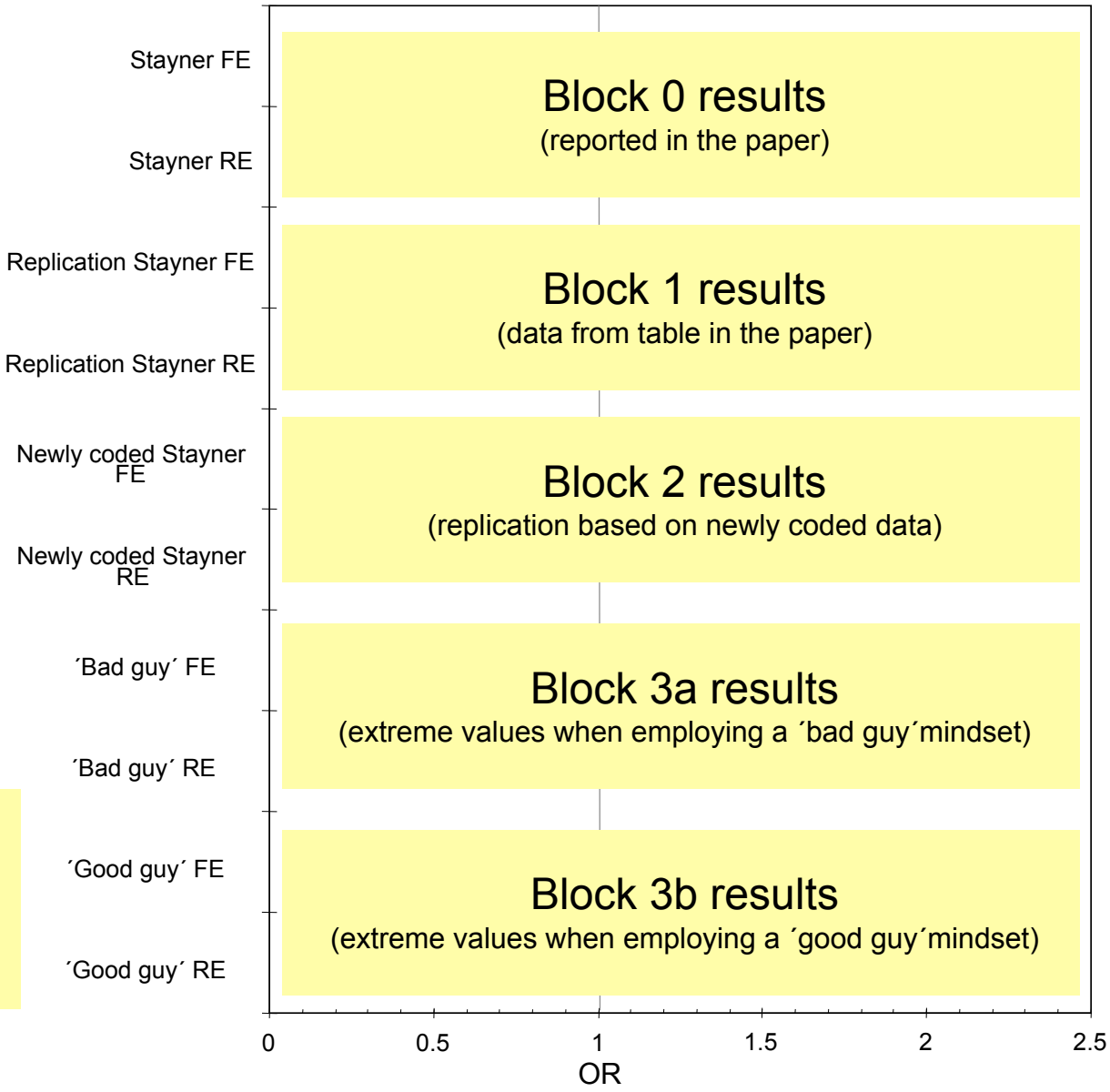
Block 0 results:

FE model:
OR= 1.24
95% CI= 1.18, 1.29

RE model:
OR= 1.24
95% CI= 1.17, 1.31

… based on 25 ES.

Please note:
We are using OR here, not RR (as done in the paper).



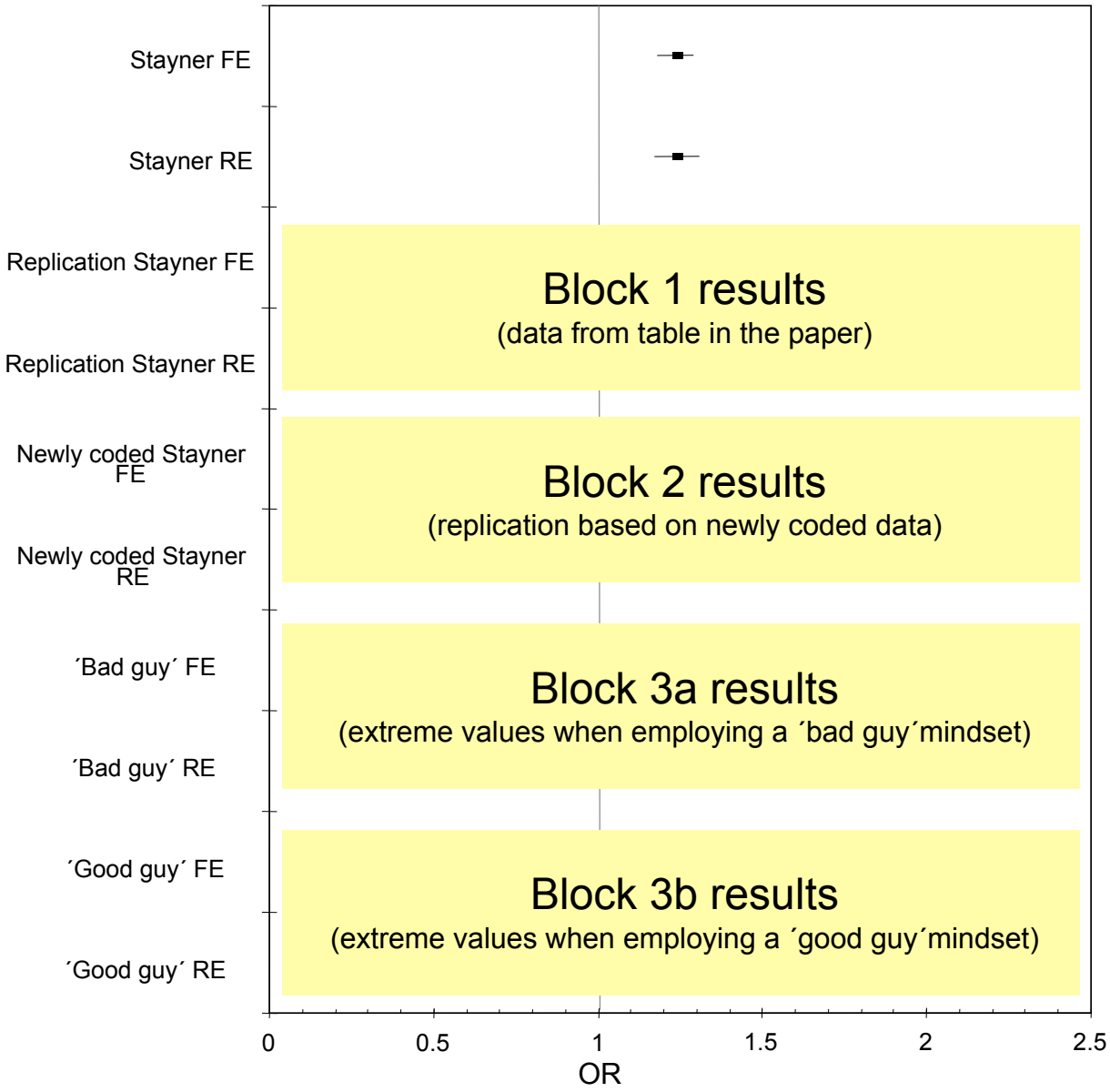
Results : Block 1

Block 1 results:

FE model:
OR= 1.27
95% CI= 1.17, 1.40

RE model:
OR= 1.28
95% CI= 1.16, 1.41

… based on 25 ES.



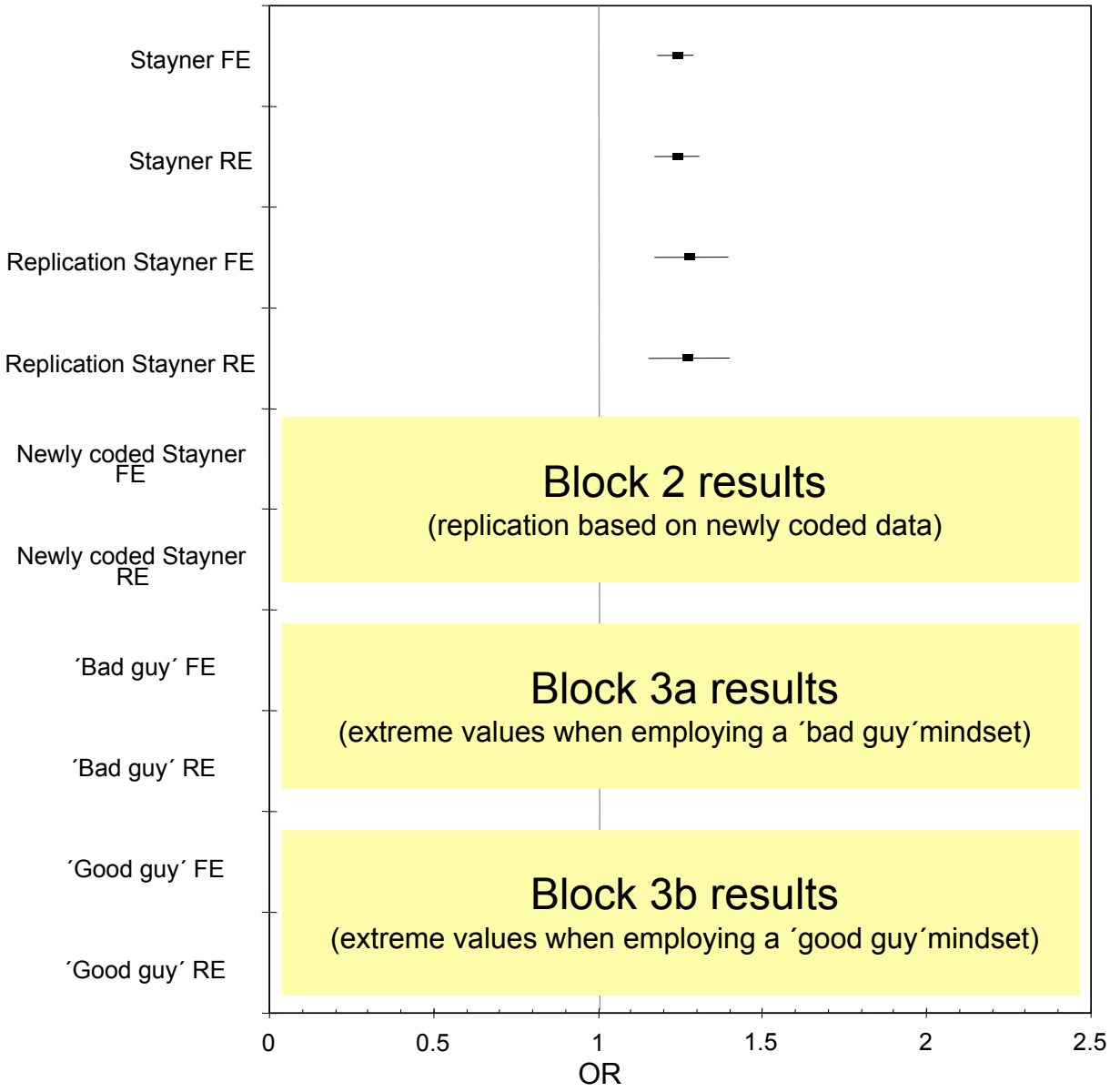
Results : Block 2

Block 2 results:

FE model:
OR= 1.22
95% CI= 1.12, 1.35

RE model:
OR= 1.23
95% CI= 1.10, 1.37

… based on 23 ES.



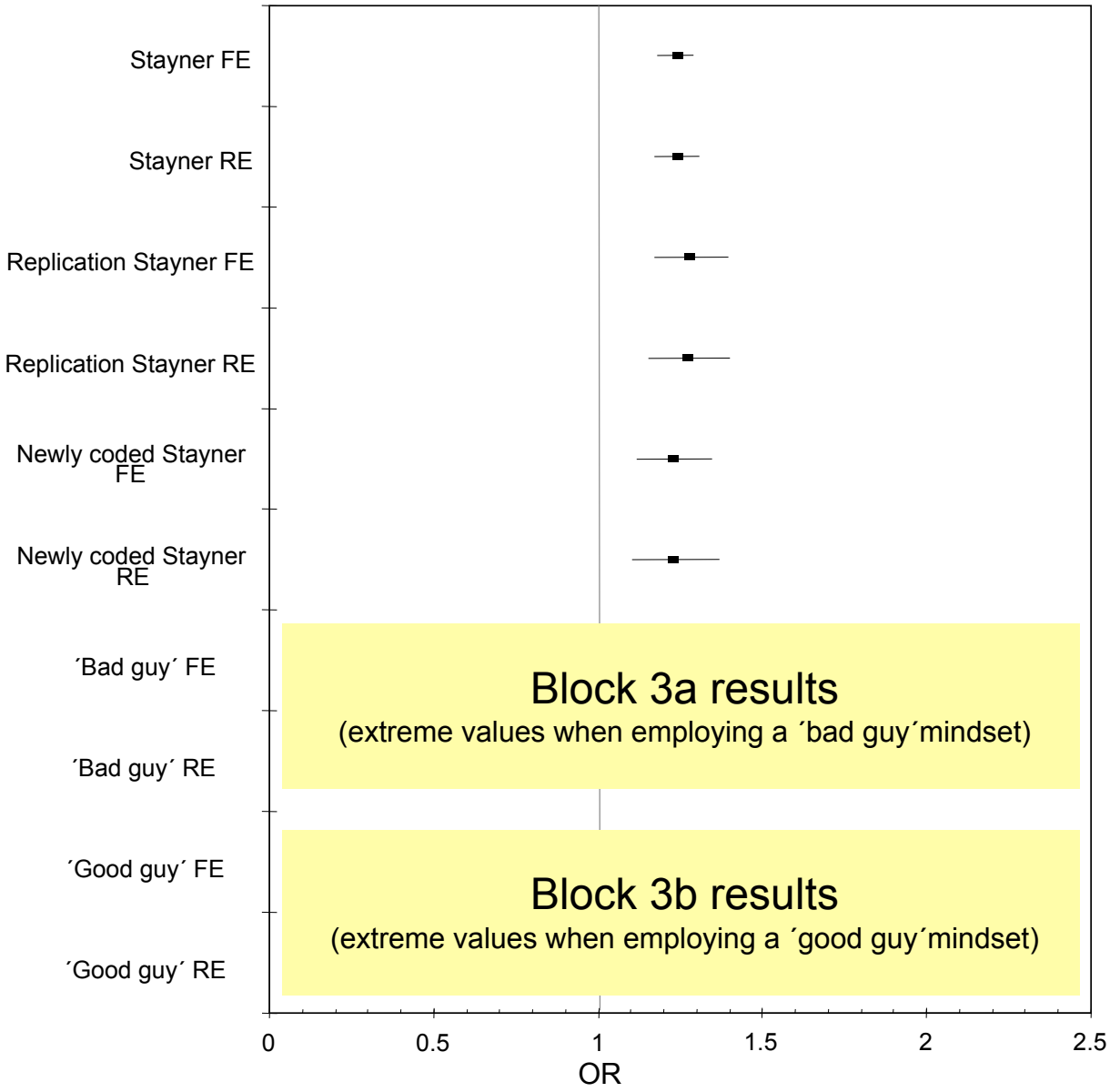
Results : Block 3a

Block 3a results:

FE model:
OR= 0.90
95% CI= 0.84, 0.93

RE model:
OR= 1.00
95% CI= 0.87, 1.14

… based on 22 ES.



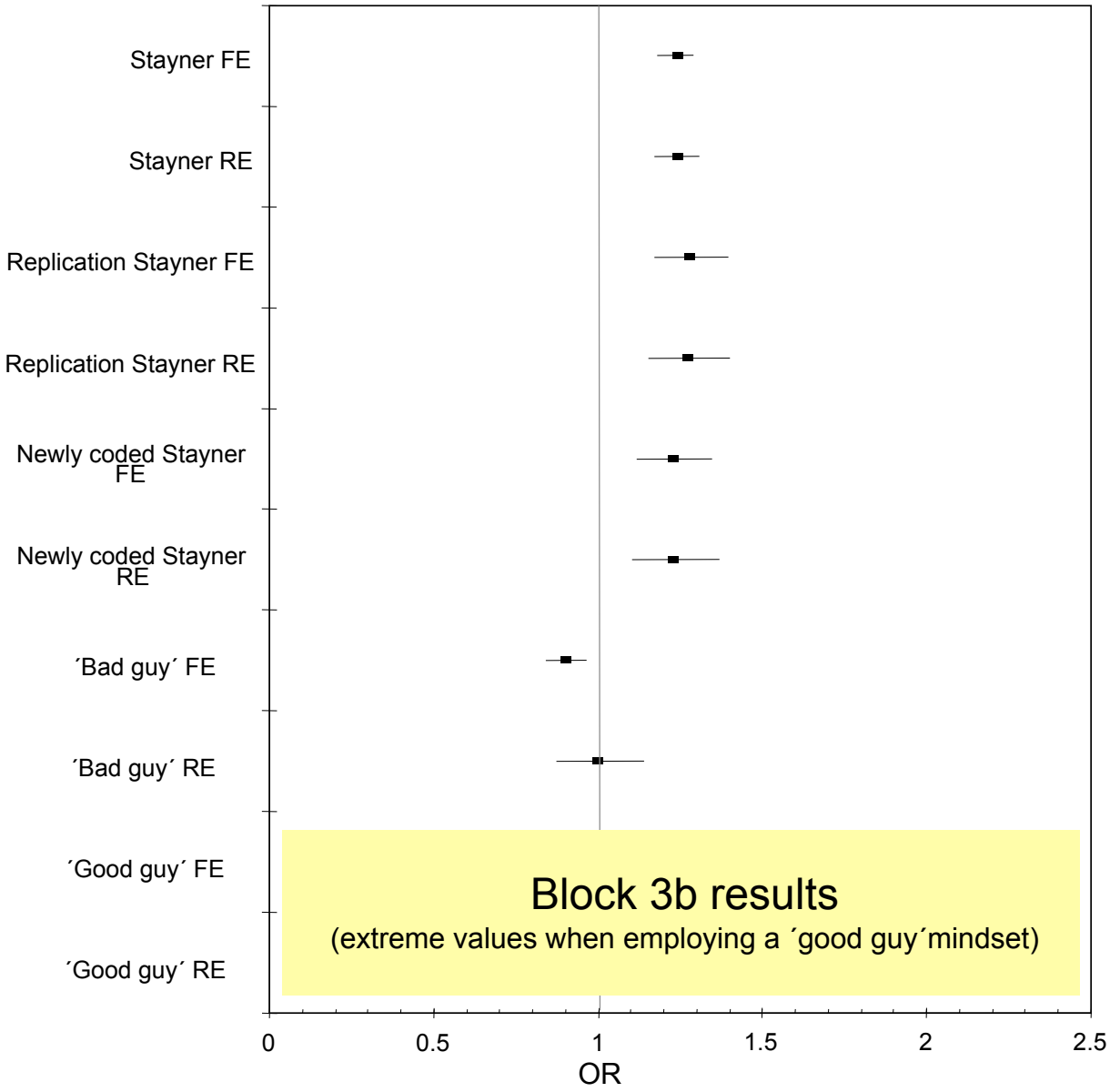
Results : Block 3b

Block 3b results:

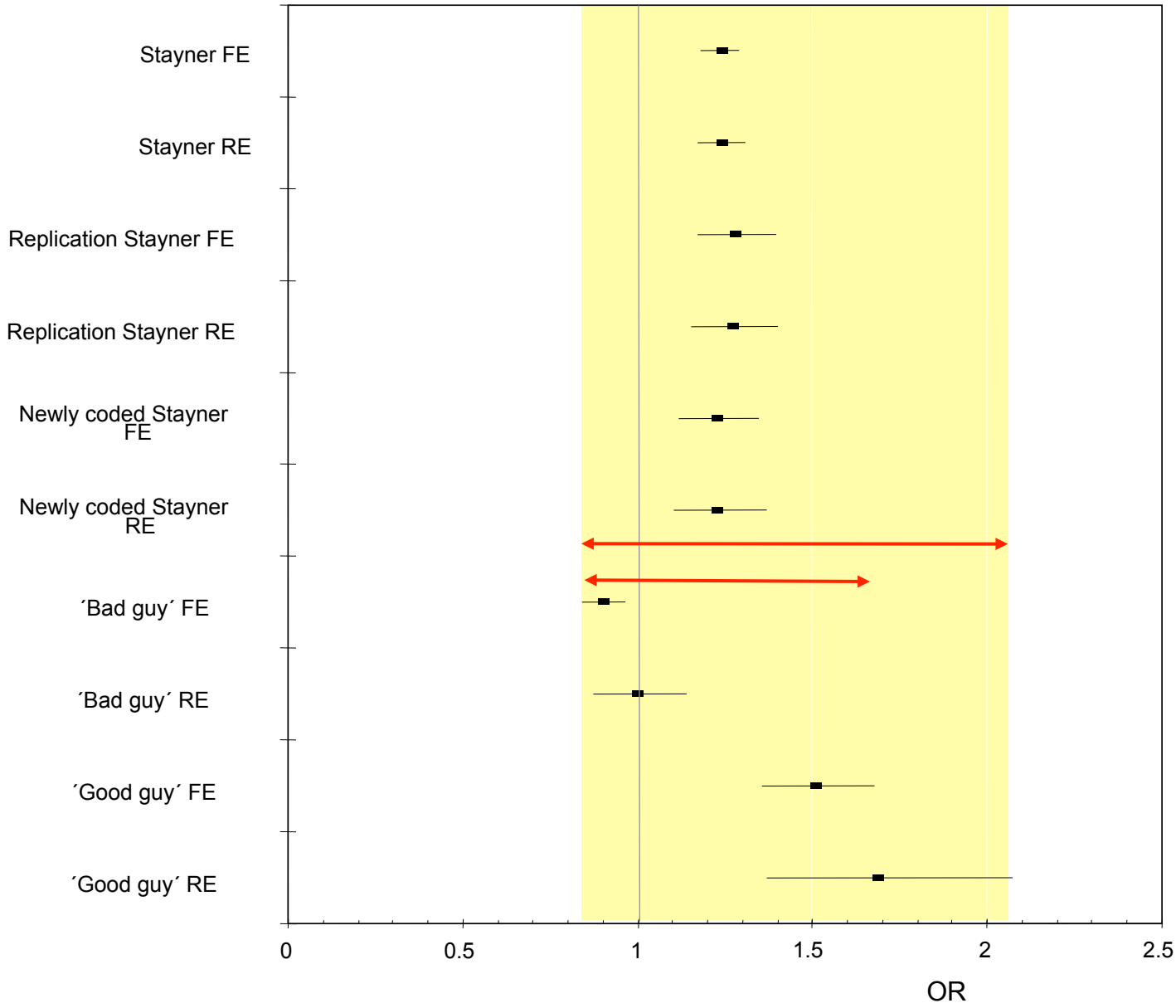
FE model:
OR= 1.51
95% CI= 1.35, 1.68

RE model:
OR= 1.69
95% CI= 1.36, 2.07

… based on 22 ES.



Overall Range *(with only 3 CR variables)*



Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - **Errors in published meta-analyses**

- Overview of threats to validity of meta-analytic conclusions

Summary of Errors Found in Trappey (1996)

- Focus here: Clear-cut errors in past meta-analyses, not ´subjective´decisions
- Examples (from Trappey, 1996, replication):
  - exclusion of one negative effect because of theoretically ´implausible result´
  - at least one primary study listed for which effect could not be computed from published paper
  - wrong control group(s) used in ES computation
  - wrong ES transformation(s) r<>d
  - etc.

- Replications of past meta-analysis needed!

Agenda

- Criticism of meta-analytic techniques

- Special issues and problems
  - Simpson´s paradox
  - (Stochastically) dependent effect sizes
  - Publication bias (File drawer problem)
  - Model-based meta-analysis (MASEM)
  - Modeling ´subjective´ decisions during the meta-analytic process (Stayner et al., 2007, replication)
  - Errors in published meta-analyses

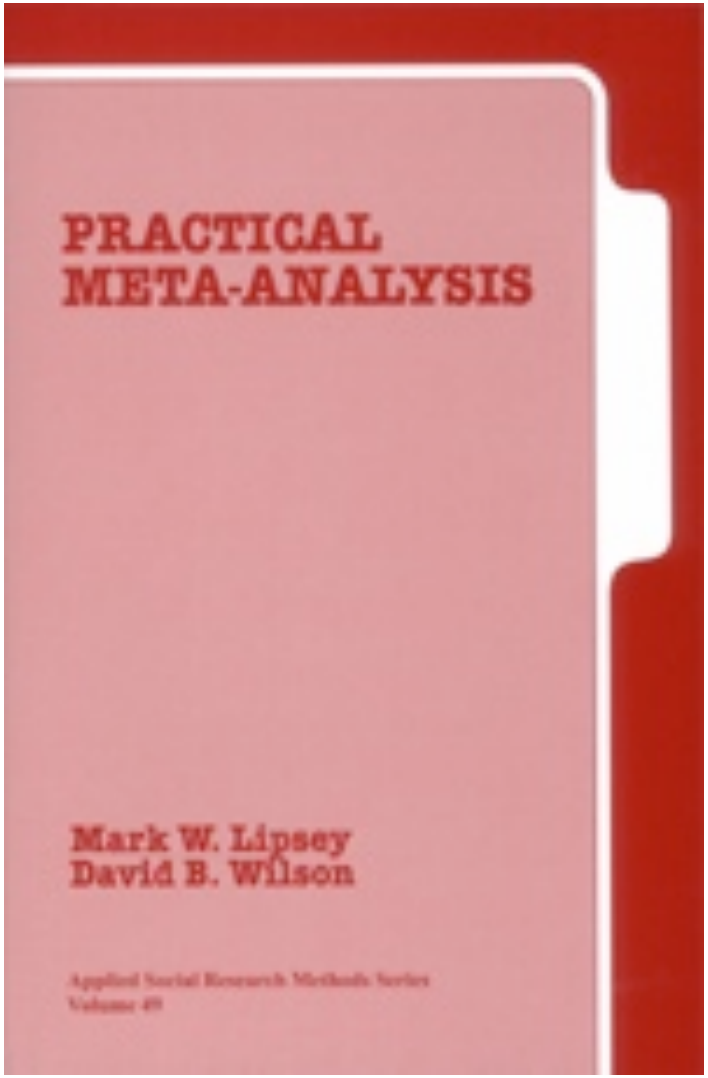
- **Overview of threats to validity of meta-analytic conclusions**

Threats to validity of meta-analytic conclusions

See supplement:

Cooper_2010_ThreatsToValidity_Tables_9_1_thru_9_7.pdf

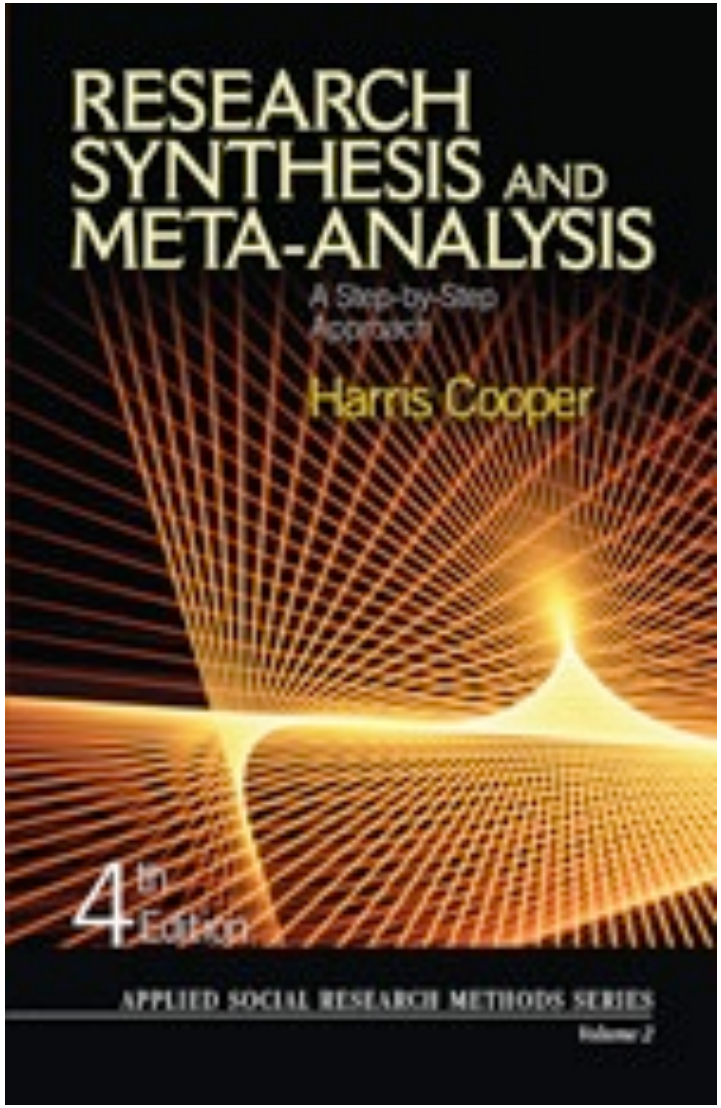
Lipsey & Wilson (2001)



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

- Chapter 1: Introduction
- Chapter 3: Selecting, Computing, and Coding the Effect Size Statistic
- Chapter 6: Analysis issues and strategies

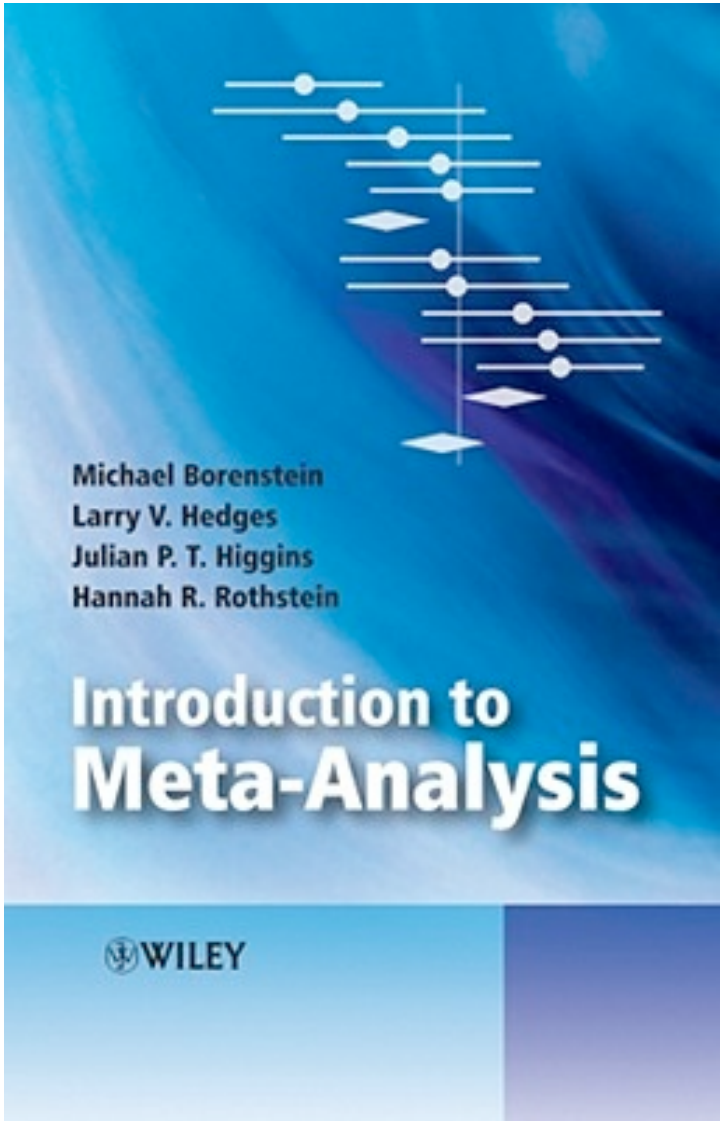
Cooper (2010)



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

- Chapter 9: Conclusion: Threats to the Validity of Research Synthesis Conclusions

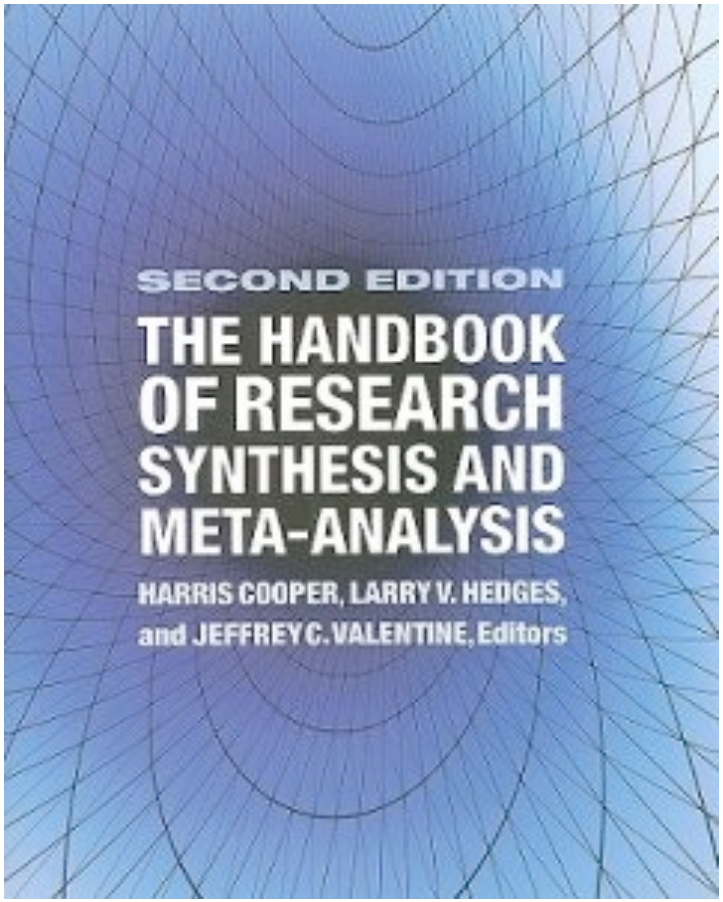
Borenstein et al. (2009)



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

- Part 5: Complex data structures:
  - Chapter 23: Independent subgroups within a study
  - Chapter 24: Multiple outcomes or time-points within a study
  - Chapter 25: Multiple comparisons within a study
- Chapter 33: Simpson´s paradox
- Chapter 43: Criticisms of meta-analysis

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.

- Part IV: Data interpretation
- Chapter 28: Threats to validity of generalized inferences
- Chapter 29: Potentials and limitations