Meta-Analysis

CLPS 2908

L8 | Feb 26, 2019

Meta-Analysis: Brief Review

- **Effect size** (ES)
  = [estimated true] mean difference between two groups, correlation between two variables, etc.
  - *How can one study ever get at this?*
  - *Best estimate from multiple data points*
- **Meta-analysis**
  = combining multiple studies' results to estimate true effect sizes from sampled effect sizes.
  - *Data point* $\in$ *study* is like *study* $\in$ *meta-analysis*
- Most often for a particular topic/hypothesis
- Works for studies in your own multi-study series

Meta-Analysis: Brief History

- Early practices of combining research results: **nose counting** later became averaging of ***r* values** (with Z transformation).
- Gene Glass (1976): *meta-analysis* = "the analysis of analyses . . . the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3)

  Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3-8.
- Cohen (1977) introduced **effect size *d***, the universal effect size parameter.

$$d = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{pooled}}$$

- Other major players:
  - Rosenthal (direct *p* value aggregation, out of fashion; but long-standing promoter)
  - Hedges (inverse variance weights; commonly used)
  - Hunter & Schmidt (corrections for sampling error, unreliability, etc.)
- Core resource: Cooper, H. & Hedges, L. V. (Eds.) (1994). *Handbook of Research Synthesis.* New York: Russell Sage Foundation.

Online Meta-Analysis Resources

http://www.psychwiki.com/wiki/Meta-analysis

Intro, steps, many more resources

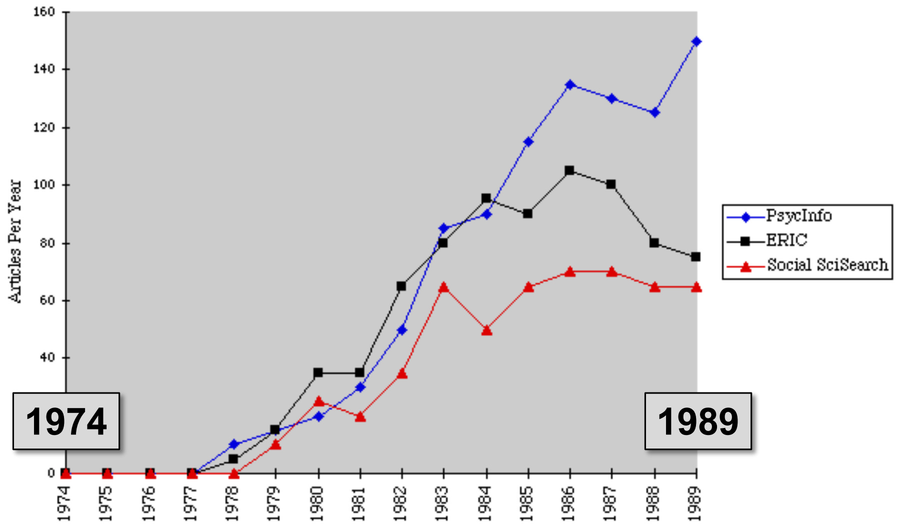
http://mason.gmu.edu/~dwilsonb/ma.html

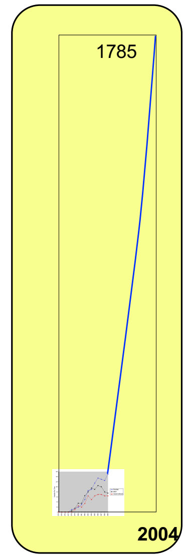
Introduction, issues, software

http://faculty.cas.usf.edu/mbrannick/meta/index.html

Including R program for actual analyses and a graphing program

Meta-Analysis in the Literature





PsycINFO **until 1993**: about 1000. **1994-2004?**

Dieckmann, Malle, & Bodner, 2009

Not a Panacea

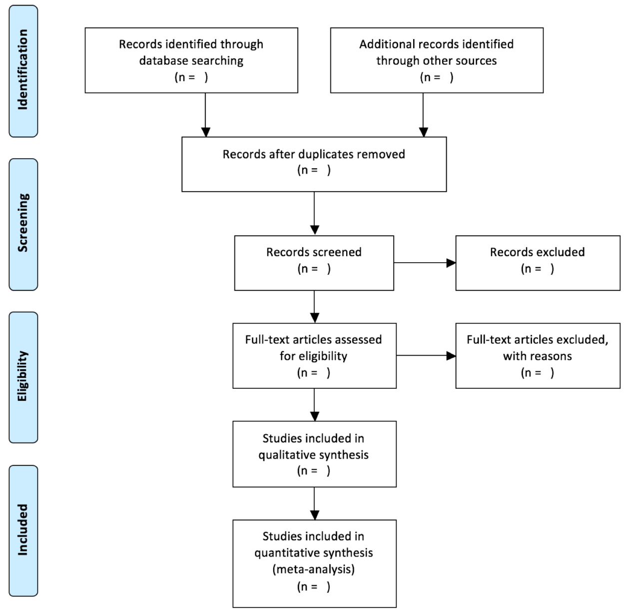
- Study selection can be biased
  - Whether published, populations sampled from, journal bias
- Quality control?

	Meta-Analyses (Psychology)	Systematic Reviews (Medical care)
Protocol or 'a priori' design	0%	43%
Comprehensive search & report	60%	90%
Duplicate study selection	50%	74%
Quality assessment of studies	40%	81%
Quality considered in conclusions	30%	60%
Publication bias assessed	30%	28%

Hilda Bastian looking at 10 recent meta-analyses in **2016**
http://blogs.plos.org/absolutely-maybe/2016/07/05/psychologys-meta-analysis-problem/

Page, M. J., Shamseer, L., Altman, D. G., Tetzlaff, J., Sampson, M., Tricco, A. C., Catalá-López, F., et al. (2016). Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. *PLOS Medicine*, *13*, e1002028.

Study Selection



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

Effect Size for Mean Differences

$$d = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_{pooled}}$$

What parameter is omitted?

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Bias correction for *d*:

$$d' = \left[1 - \frac{3}{4N - 9}\right]d$$

Another Excel file...

Hedges g, Cohen's d, corrected $g \rightarrow d$...
https://en.wikipedia.org/wiki/Effect_size

Acquiring Effect Sizes (ES)

(in decreasing precision)

- Direct calculation based on means and standard deviations
- Algebraically equivalent formulas (*t* test, *F* test)
- Transform other results (e.g., *r*, χ^2)
- Exact *p* value for a *t* test
- Construct single-df *F* from multi-df *F*
- Approximate the pooled standard deviation; the means; the correlations for w/s data and difference scores, etc.
- Reconstructing s_p from complex Anova designs
- Probability range ($p < .01$)
- *"not significant"*

$$ES = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

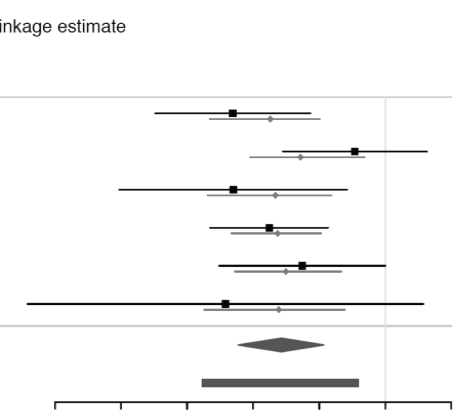
$$s_{pooled} = \sqrt{\frac{SS_B + SS_{AB} + SS_W}{df_B + df_{AB} + df_W}}$$

Jamie deCoster's guide

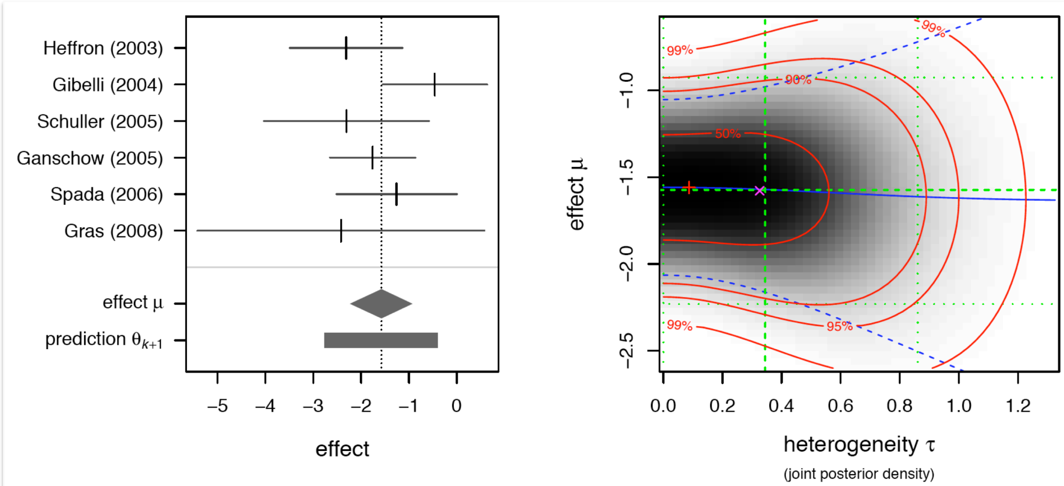
Effect Size Variations

- Sampling error of each ES (mainly due to *N*)
- Each ES is weighted by this error before combination ("fixed-effect approach")
- Alternative procedures include a random variance component (between-study variations; "random-effects approach")

study	estimate	95% CI
Heffron (2003)	−2.31	[−3.48, −1.13]
Gibelli (2004)	−0.46	[−1.55, 0.63]
Schuller (2005)	−2.30	[−4.03, −0.58]
Ganschow (2005)	−1.76	[−2.65, −0.86]
Spada (2006)	−1.26	[−2.52, −0.00]
Gras (2008)	−2.42	[−5.41, 0.58]
mean	**−1.57**	**[−2.23, −0.93]**
prediction	**−1.57**	**[−2.77, −0.41]**

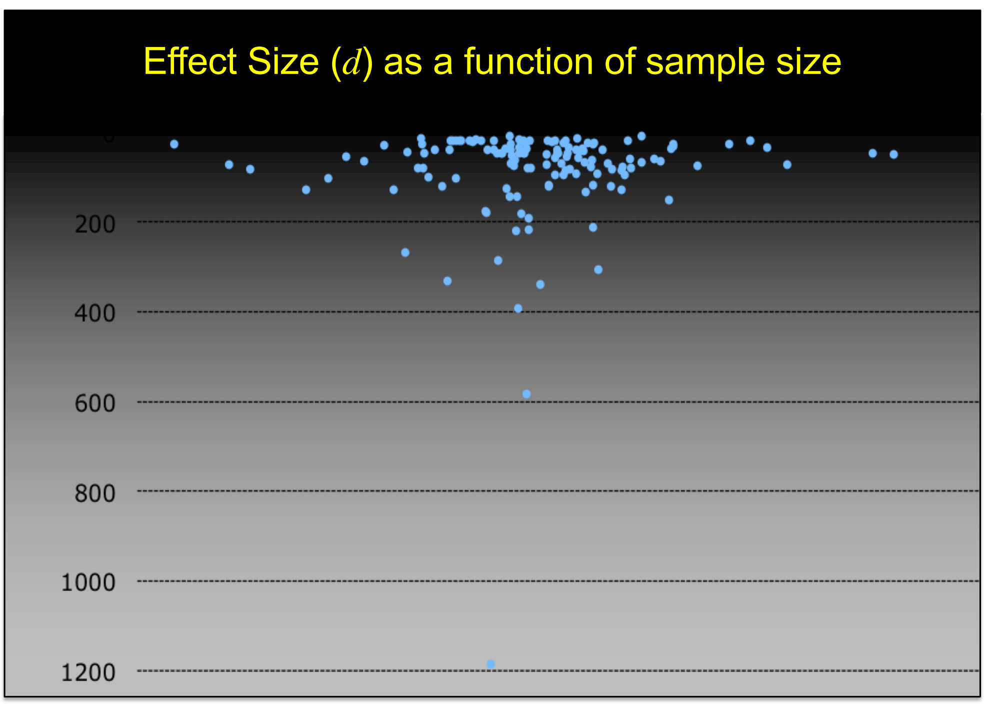


Bayesian Approach



right plot illustrates the joint posterior density of heterogeneity τ and effect μ, with darker shading corresponding to higher probability density. The red lines indicate (approximate) 2-dimensional credible regions, and the green lines show marginal posterior medians and 95% credible intervals. The blue lines show the conditional posterior mean effect $\widehat{\mu}(\tau)$ as a function of the heterogeneity τ along with a 95% interval based on its conditional standard error $\widehat{\sigma}(\tau)$ (see also Section 2.4). The red cross (+) indicates the posterior mode, while the pink cross (×) shows the ML estimate.

Visualizing Effect Size Patterns

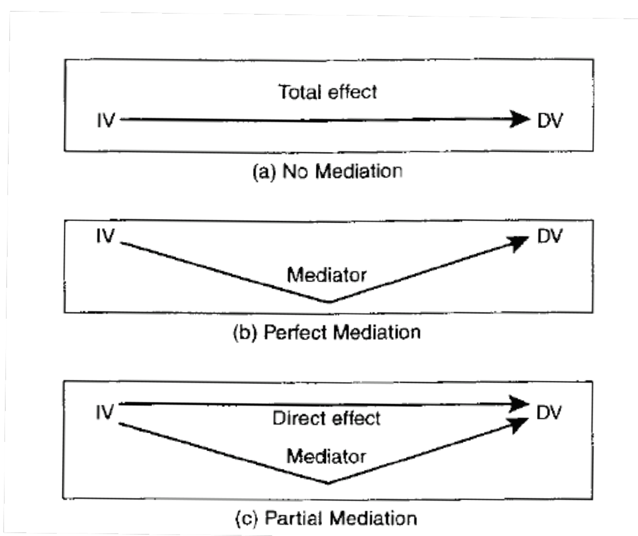


Excursion: Moderator vs. Mediator

An effect **E** (e.g., IV → DV) is moderated by variable **M**: E varies at different levels of **M**
Interaction effect!

- **M** is often a group variable (e.g., gender, manipulation order)
- **M** can be continuous
  - See MReg lecture for treatment of continuous interaction effects
  - Not a **covariate** (which cleans out DV without consideration of IV; only after clean out does IV get to predict DV)
- **Mediator** is a variable that intervenes in the IV → DV relationship: IV → M → DV

Mediation (Simple version)

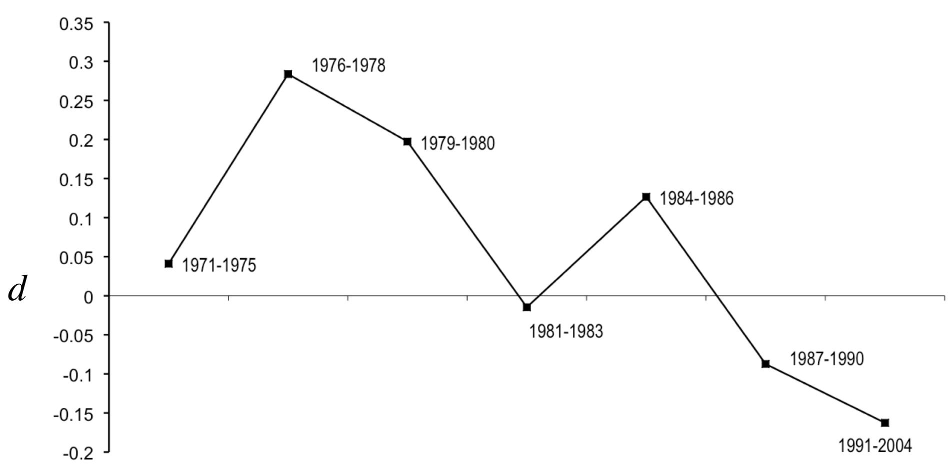


Test for Mediation:

1. Show that β(IV) ➛ DV
2. Show that β(Med)➛ DV
3. Show that β(IV) ➛ 0
When β(Med) controlled

Meta-Analysis Moderator Analyses

- Effect Size might vary as a function of…
  - Method and source of calculation (inference steps)
  - N basis
  - Variance of effect size (w/s different from b/s) $$\sigma_d{}^2 = \frac{n_e + n_c}{n_e n_c} + \frac{d^2}{2(n_e + n_c)}.$$
  - Measure used (reliability? → Hunter & Schmidt)
  - Between or within design
  - Lab or field study
  - Year of study
  - Publication venue (e.g., Impact factors)
- Enter into flexible data storage system (e.g., Excel)
  - ➢ Nesting of ES for smallest samples (conditions) and aggregating by measures, studies, articles, journals, etc.



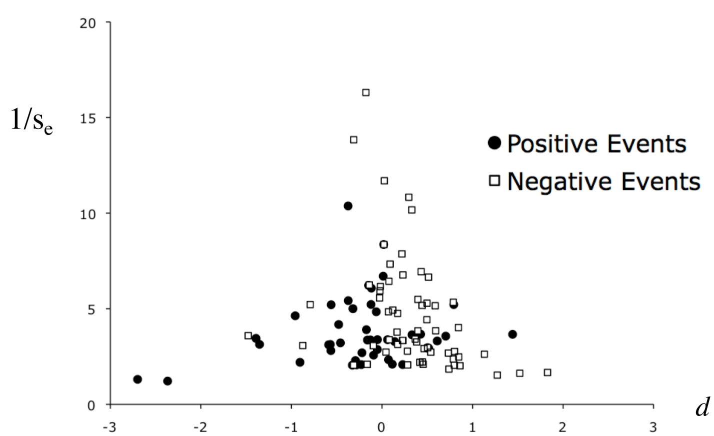
Malle, B. F. (2006). The actor-observer asymmetry in causal attribution: A (surprising) meta-analysis. *Psychological Bulletin*, *132*, 895-919.

Moderator Variables and Numbers of Studies in Each Category

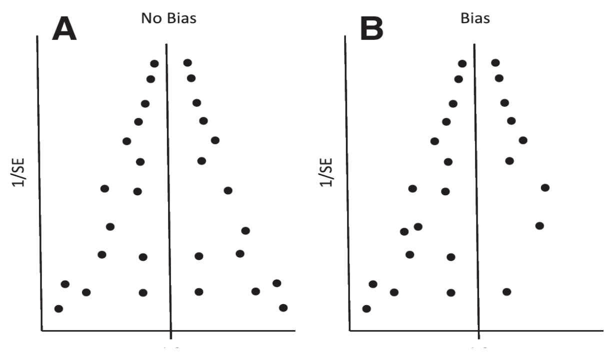
Variable and values	Number of studies/ samples	Reliability
1. Valence of event explained		Agree = 74%, κ = 0.61
Negative	61	
Neutral, mixed, or unknown	67	
Positive	45	
2. Visual perspective switched (actor sees own behavior from observer perspective and vice versa)	6	Published record
3. Involved observers (those who have a special interest in the actor's behavior)	14	Published record
4. Familiarity between actor and observer		Agree = 98%, κ = 0.95
High: romantic partner, parent, good friend	29	
Low: stranger, new acquaintance	143	
Mixed	1	
5. Age groups		Published record
Children: 5–17 years old	19	
College students: 19 years on average	113	
Adults: 20 years and older	37	
6. Standard/nonstandard methodology		Published record
Standard studies (containing no independent variable presumed to increase or decrease the asymmetry; akin to control groups)	48	
Nonstandard studies	127	
Explanations of positive or negative events	106	
Explanations of hypothetical events	24	
Presentation of fictitious base rates	7	
Time-delayed attributions (2 are within-subject)	6	
Non-Western samples	3	
7. Event type		Agree = 88%, κ = 0.75
Behavioral or psychological event (e.g., actions, emotions)	82	
Outcome (e.g., success, failure)	87	
Unknown	4	
8. Intentionality of event		Agree = 81%, κ = 0.70
Unintentional	94	
Mixed, unknown	45	
Intentional	34	
9. Realism of event		Agree = 93%, κ = 0.69
Real to both actor and observer	144	
Hypothetical/imagined to actor and observer	23	
Mixed	6	
10. Time of attribution (for real events)		Agree = 87%, κ = 0.71
Immediately after event occurred	101	
Delayed	43	
11. Source of valence		Published record
Experimentally manipulated	71	
Naturally occurring	102	
12. Attribution assessment format		Published record
Rating scales	152	
Open-ended responses	20	
Mixed	1	
13. Types of rating scales		Published record
Bipolar (I-E) scale only	22	
Unipolar scales I and E	64	
Single-item I and E ratings	44	
Multiple-item I and E ratings	20	
I and E based on ability, effort, luck, and task scales	48	
I and E based on scales with specific content	18	
14. Reference to stable traits in internal attribution measure		
Continuous percentage (range: 0–100%)		ICC = 0.68
15. Degree of inference required for effect size extraction		ICC = 0.93, κ = 0.71
0	84	
1	62	
2	24	
3	3	
16. Study design for actor–observer factor		Agree = 93%, κ = 0.84
Between-subjects	123	
Within-subject	50	
17. Study setting		Published record
Actors and observers complete questionnaires	93	
Observer observes actor	61	
Observer interacts with actor	19	
18. Did actor and observer explain the identical event?		Agree = 84%, κ = 0.69
Identical event	87	
Not identical event	86	
19. Median split by publication year		Published record
First half of studies: 1971–1981	85	
Second half of studies: 1982–2004	88	
20. Citation impact factor (range: 0.10–3.10)		Archival record

Meta-Analysis Moderator Analyses

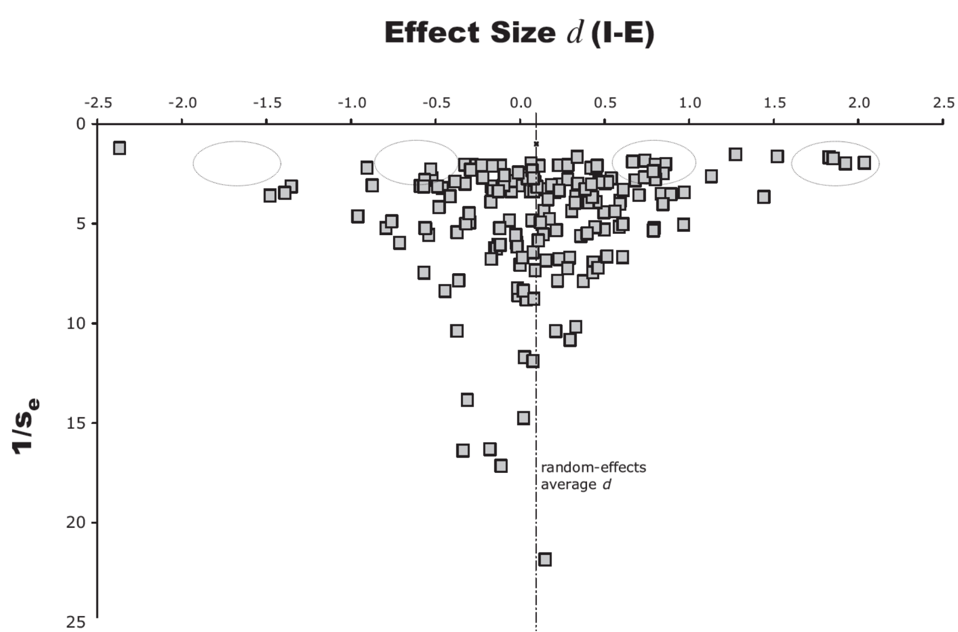
- Predict variation in total ES from moderators (use MReg)
  - Coded study features
  - Conditions under which ES derived (e.g., sample characteristics/subject population, experimental manipulations)
- Plot subsets of studies in boxplots, dotplots, etc.



Threats to Validity: Publication Bias



– Publication bias (improving??)
  - “File-drawer effect” (Rosenthal, 1979)
  - Greenwald (1975) surveyed reviewers and journal authors → *p* (published | confirming) ~ 0.50, *p* (published | disconfirming) ~ 0.30.
  - Fail-safe *N* (how many *failures to replicate* could the result tolerate)
  - Estimation methods (none of them terribly reliable)
    – “Trim and Fill” (Duval & Tweedie, 2000)



File Drawer Problem: Not Always

Psychological Bulletin
2006, Vol. 132, No. 6, 895–919

0033-2909/06/$12.00 DOI: 10.1037/0033-2909.132.6.895

Events explained	σ^2-weighted random effects	95% CI	σ^2-weighted fixed effects
All studies			
I-E	0.095*	0.032, 0.159	0.032*
I	0.062	−0.019, 0.143	−0.016
E	0.023	−0.064, 0.109	0.016
Standard studies			
I-E	−0.001	−0.105, 0.104	−0.049*
I	−0.093	−0.194, 0.008	−0.070*
E	0.007	−0.100, 0.115	−0.015