# Lies, Damned Lies, and Meta-Analysis

Katie Corker, Kenyon College

June 18, 2015

# Origins of Meta-Analysis

- Developed in mid-1970s by Glass, Rosenthal, Schmidt, and others
- Goal: To synthesize increasingly large numbers of studies
- Early issues:
  - Role of unpublished work
  - Quality of constituent studies
  - Fixed vs. random effects

# Bias in ES Estimates

- Non-significant studies more likely to be censored from literature (publication bias)
- Non-significant analyses may be swapped for more sig. analyses (*p*-hacking)
- Researchers selectively report sig. analyses as if they were predicted (HARKing)
- Small studies tend to overestimate ES (small study bias)
- ***In reality, published literature reflects an unknown combination of these biases.***

# New Techniques

- Meta-regression, esp. PET-PEESE (Stanley & Doucouliagos, 2014)
- Tests of excess significance
  - Ioannidis & Trikalinos (2007)
  - Francis (e.g., 2013)
  - Incredibility index (Schimmack, 2012)
  - R-Index & TIVA (Schimmack, 2014a; 2014b)
- *p*-curve (Simonsohn, Nelson, Simmons, 2014a; 2014b)
- ~~Trim & Fill~~

# Practical Concerns

- How should researchers decide which tools to use?
  - Only skeptics/quants use anything more than a simple average, right?!
- Different techniques:
  1. Require different meta-data
  2. Have different assumptions
- How many studies are sufficient?
- Role of heterogeneity?
- Role of covariates/more complex analyses?

# Selecting Effects for Meta-Analysis

- Keep everything (Glass, 2015)
- Address study quality (Verhagen et al., 2001)
- Top 10– Keep only the most precise 10% of studies (Stanley et al., 2010)
- Analyze only predicted effects for which $p < .05$ (Simonsohn et al., 2014a, 2014b)

# Emerging Issues

- Pre-registration
  - Scooping?
- Cumulative meta-analysis
  - curatescience.org
- Empirical options for testing meta-analytic tools (e.g., with the Many Labs data sets)

# Conclusions

- Researchers need practical tools and guidance to do high quality meta-analysis
  - Esp. for complex designs
  - *metafor* (Viechtbauer, 2014)
  - Lakens et al. (in press) paper
- Best practices for adjudicating conflicting test results

# References


Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology, 57*, 153-169.

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials, 4*, 245-253.

Lakens, D., Hilgard, J., & Staaks, J. (In press). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology.*

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17,* 551-566

Schimmack, U. (2014a). Quantifying statistical research integrity: The replicability-index. Unpublished manuscript. http://www.r-index.org/uploads/3/5/6/7/3567479/introduction_to_the_r-index__14-12-01.pdf

Schimmach, U. (2014b). The test of insufficient variance (TIVA): A new tool for the detection of questionable research practices. https://replicationindex.wordpress.com/2014/12/30/the-test-of-insufficient-variance-tiva-a-new-tool-for-the-detection-of-questionable-research-practices/

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC: Medical Research Methodology,* 7:10.

# References


Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). *p*-Curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*, 534-547.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). *p*-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*, 666-681.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5,* 60-78.

Verhagen, A. P., de Vet, H. C. W., de Bie, R. A., Kessels, M. B., Bouter, L. M., & Knipschild, P. G. (1998). The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology, 51,* 1235-1241.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

# Additional Material

# Quality Checklist: For Studies

- The Delphi List (Verhagen et al., 1998)

| Item | Options |
|---|---|
| Was a method of randomization performed? | Yes/No/Don't know |
| What the treatment allocation concealed? | Yes/No/Don't know |
| Were the groups similar at baseline regarding the most important prognostic indicators? | Yes/No/Don't know |
| Were the eligibility criteria specified? | Yes/No/Don't know |
| Was the outcome assessor blinded? | Yes/No/Don't know |
| Was the care provider blinded? | Yes/No/Don't know |
| Was the patient blinded? | Yes/No/Don't know |
| Were point estimates and measures of variability reported for the primary outcome measures? | Yes/No/Don't know |
| Did the analysis include an intention-to-treat analysis? | Yes/No/Don't know |

# Quality Checklist: For Meta-Analyses

- AMSTAR (Shea et al., 2007)

| Item | Options |
|---|---|
| Was an ‘a priori’ design provided? | Yes/No/Can’t Answer/NA |
| Was there duplicate study selection & data extraction? (i.e., 2+ coders) | Yes/No/Can’t Answer/NA |
| Was a comprehensive literature search performed? | Yes/No/Can’t Answer/NA |
| Was the status of publication (i.e., grey literature) used as an inclusion criterion? | Yes/No/Can’t Answer/NA |
| Was a list of studies (included and excluded) provided? | Yes/No/Can’t Answer/NA |
| Were the characteristics of the included studies provided? | Yes/No/Can’t Answer/NA |
| Was the sci. quality of the included studies assessed & documented? | Yes/No/Can’t Answer/NA |
| Was the scientific quality of the included studies used appropriately in formulating conclusions? | Yes/No/Can’t Answer/NA |
| Were the methods used to combine findings of studies appropriate? | Yes/No/Can’t Answer/NA |
| Was the likelihood of publication bias assessed? | Yes/No/Can’t Answer/NA |
| Was the conflict of interest stated? | Yes/No/Can’t Answer/NA |

# Lakens et al. (in press), *BMC Psych*

- Six Recommendations for Reproducibility of Meta-analyses
  1. "Future proof:" Disclose all meta-data with quotes from text
  2. Quality Control: At least 2 coders, formulas for ES conversions and calculations
  3. PRISMA: Report essential info
  4. Pre-register: Set your protocol ahead of time to distinguish confirmatory from exploratory tests
  5. Share your data and code to facilitate re-analysis
  6. Consult a statistician and a librarian

# Effect of Temperature on Pro and Antisocial Behavior

- With Dermot Lynott (U. Lancaster)
- Heat causes antisocial behavior

  (Baron, 1972; Baron & Lawton, 1972; Baron & Bell, 1975; Baron & Bell, 1976; Baron, 1976; Bell & Baron, 1976; Bell & Baron, 1977; Palmarek & Rule, 1979; Bell, 1980; Kenrick & McFarlane, 1986; Vrij & Van der Steen, 1994; Boyanowsky, 1999; Breines, 2012; Fay & Maner, 2014; Gockel et al., 2014)
- Heat primes warmth → prosocial behavior

  (Schneider, Lesko, & Garrett, 1980; Williams & Bargh, 2008; Kang et al., 2011; Kolb et al., 2012; Ijzerman et al., 2013; Story & Workman, 2013; Lynott et al., 2014; Lynott, 2014; Lynott et al., 2015)
- Write me for more info!