Correcting for bias in the literature

A comprehensive comparison of meta-analytic methods for bias-correction

Felix Schönbrodt, Evan Carter, Will Gervais, Joe Hilgard

Felix Schönbrodt
Ludwig-Maximilians-Universität München

www.nicebread.de
www.researchtransparency.org
@nicebread303

Meta-analysis is at the top of the evidence-based medicine pyramid - the pinnacle of evidence-based medicine.

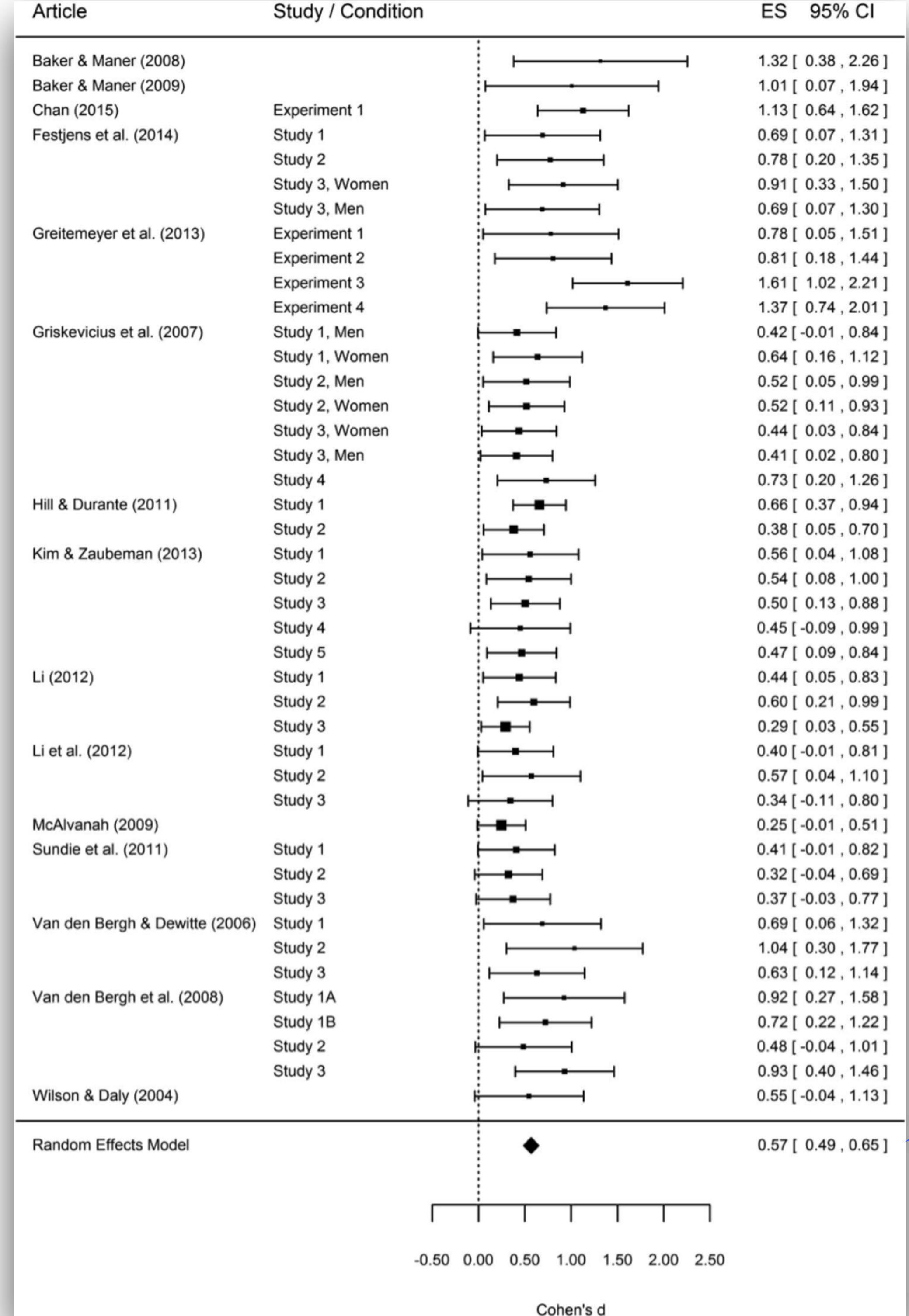
Cochrane Collaboration

https://uk.cochrane.org/news/meta-analysis-what-why-and-how

Meta-analyses are fucked.

Mickey Inzlicht

http://www.slate.com/articles/health_and_science/cover_story/2016/03/ego_depletion_an_influential_theory_in_psychology_may_have_just_been_debunked.single.html

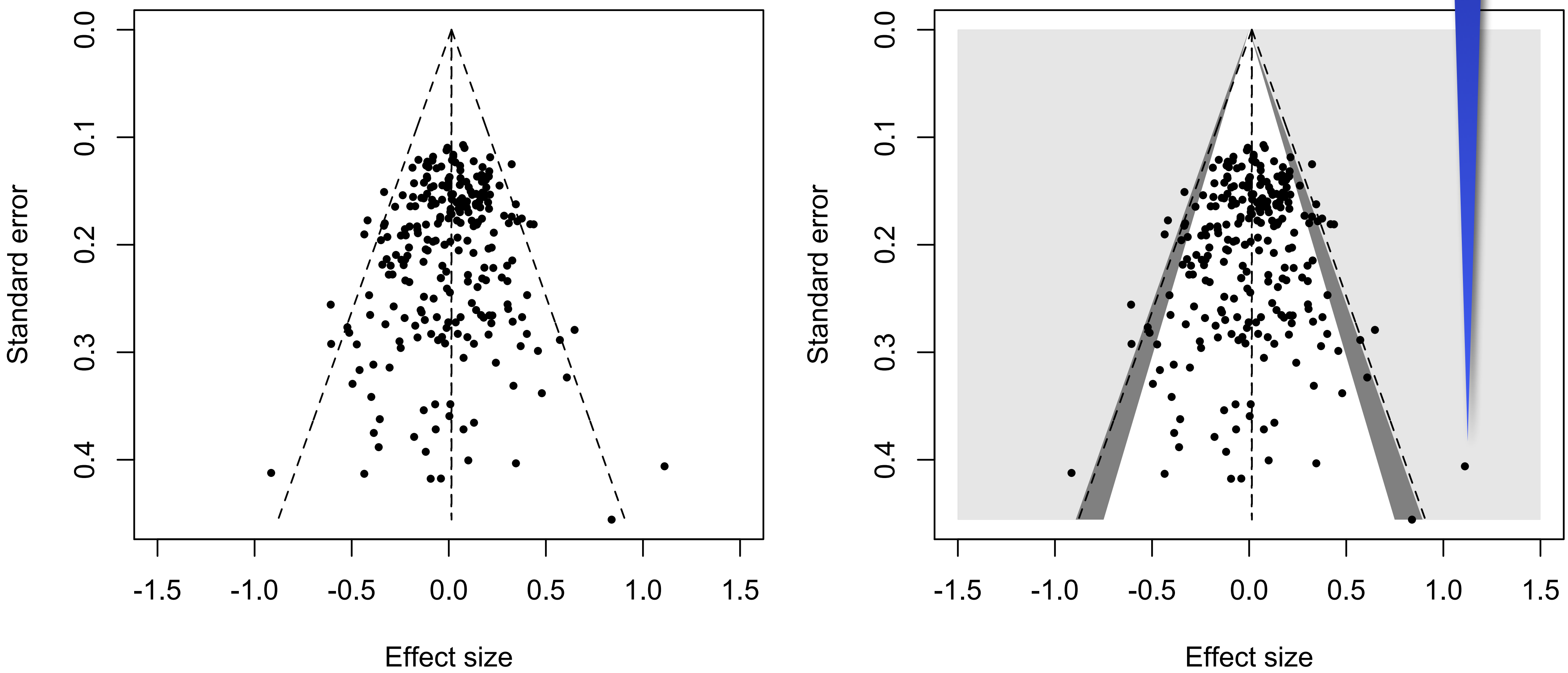


Random effects meta-analytic estimate:
d = 0.57 [0.49; 0.65]

42/43 studies are significant (98% success rate)

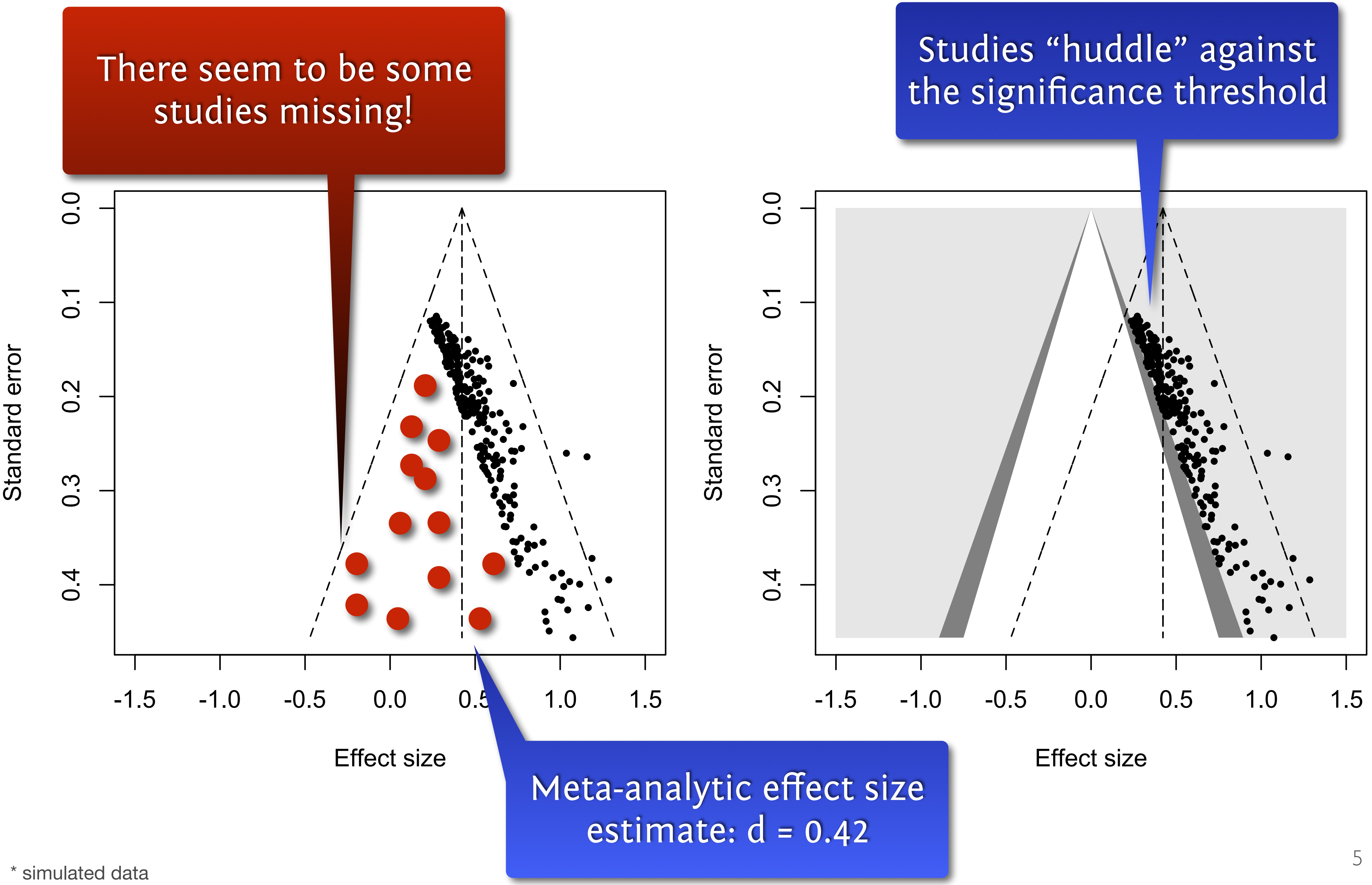
True H_0 samples*

5% false positive ("significant") studies



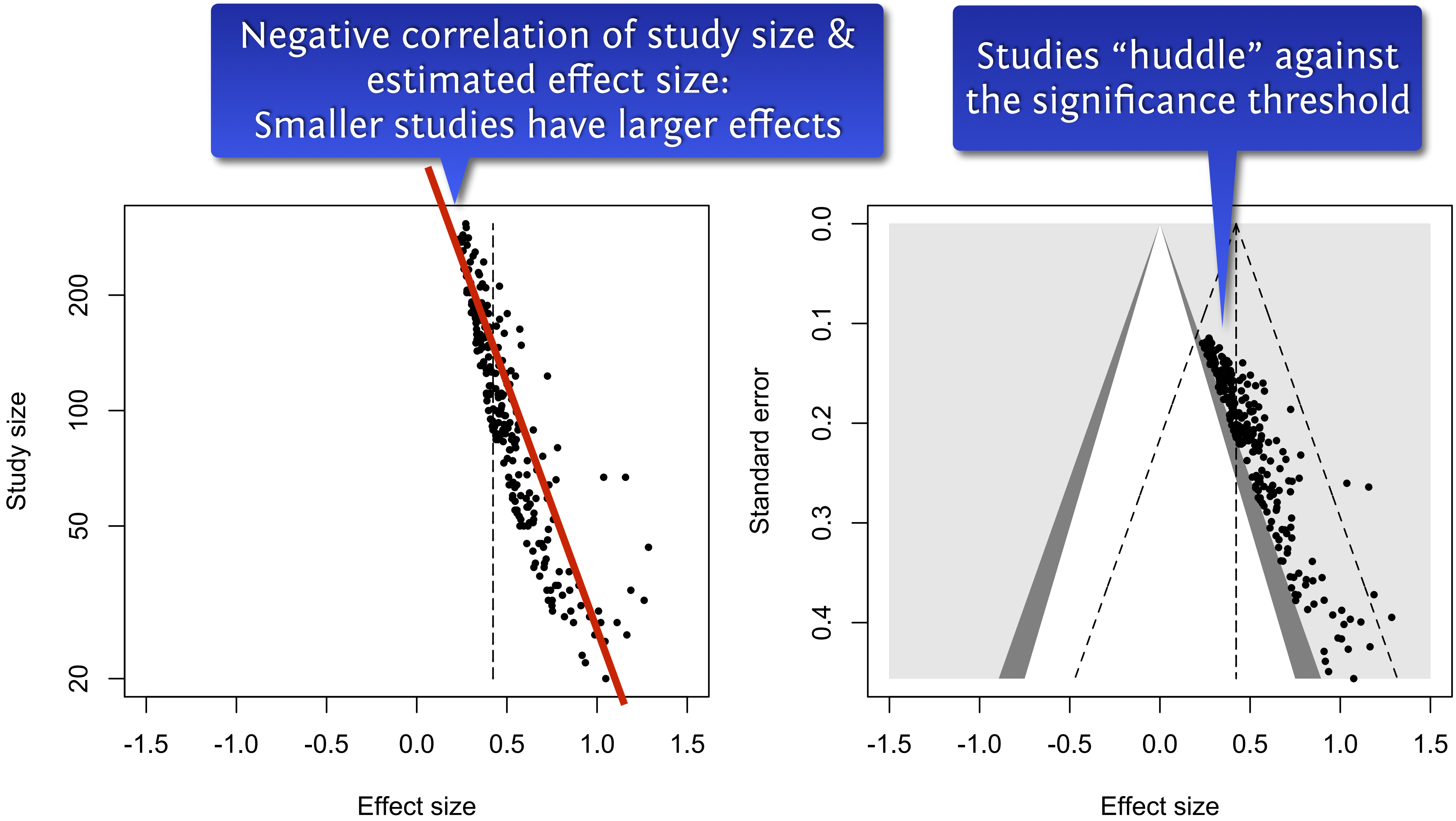
* simulated data

True H_0 + directional publication bias



* simulated data

True H_0 + publication bias



* simulated data

Journal of Experimental Psychology: General

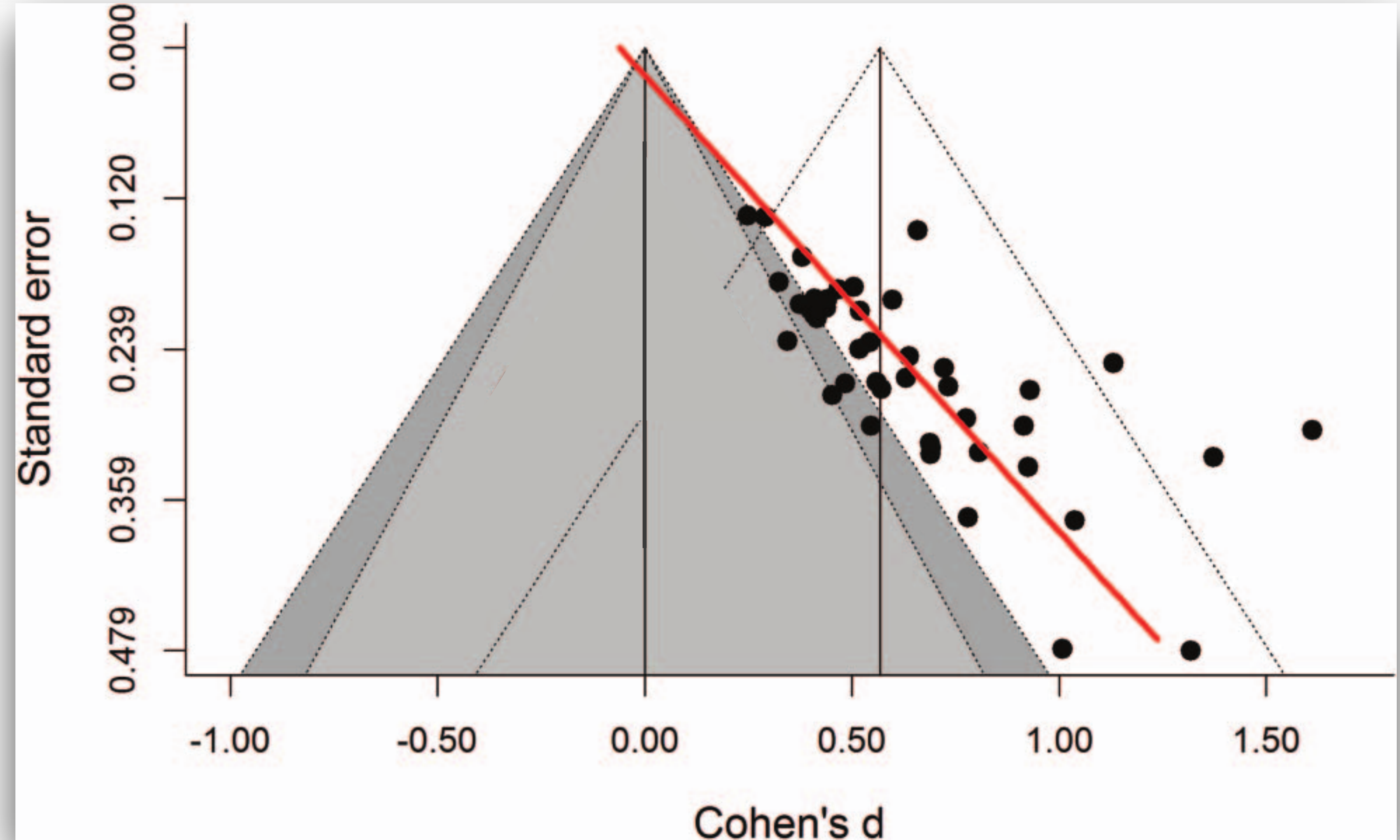
0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000116

Romance, Risk, and Replication: Can Consumer Choices and Risk-Taking Be Primed by Mating Motives?

David R. Shanks
University College London

Miguel A. Vadillo
King's College London

Benjamin Riedel, Ashley Clymo, Sinita Govind, Nisha Hickin, Amanda J. F. Tamman, and Lara M. C. Puhlmann
University College London



Journal of Experimental Psychology: General

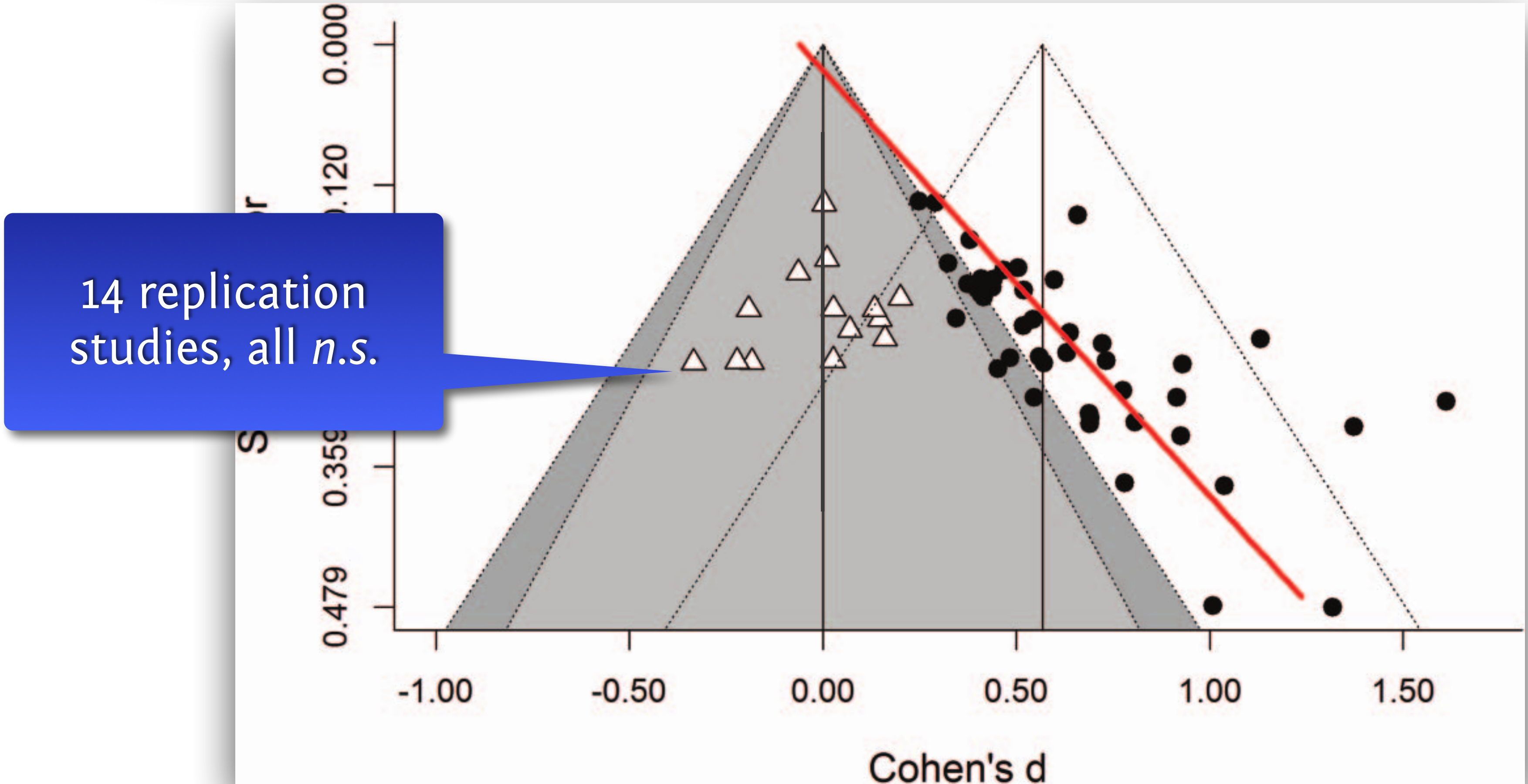
0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000116

Romance, Risk, and Replication: Can Consumer Choices and Risk-Taking Be Primed by Mating Motives?

David R. Shanks
University College London

Miguel A. Vadillo
King's College London

Benjamin Riedel, Ashley Clymo, Sinita Govind, Nisha Hickin, Amanda J. F. Tamman, and Lara M. C. Puhlmann
University College London



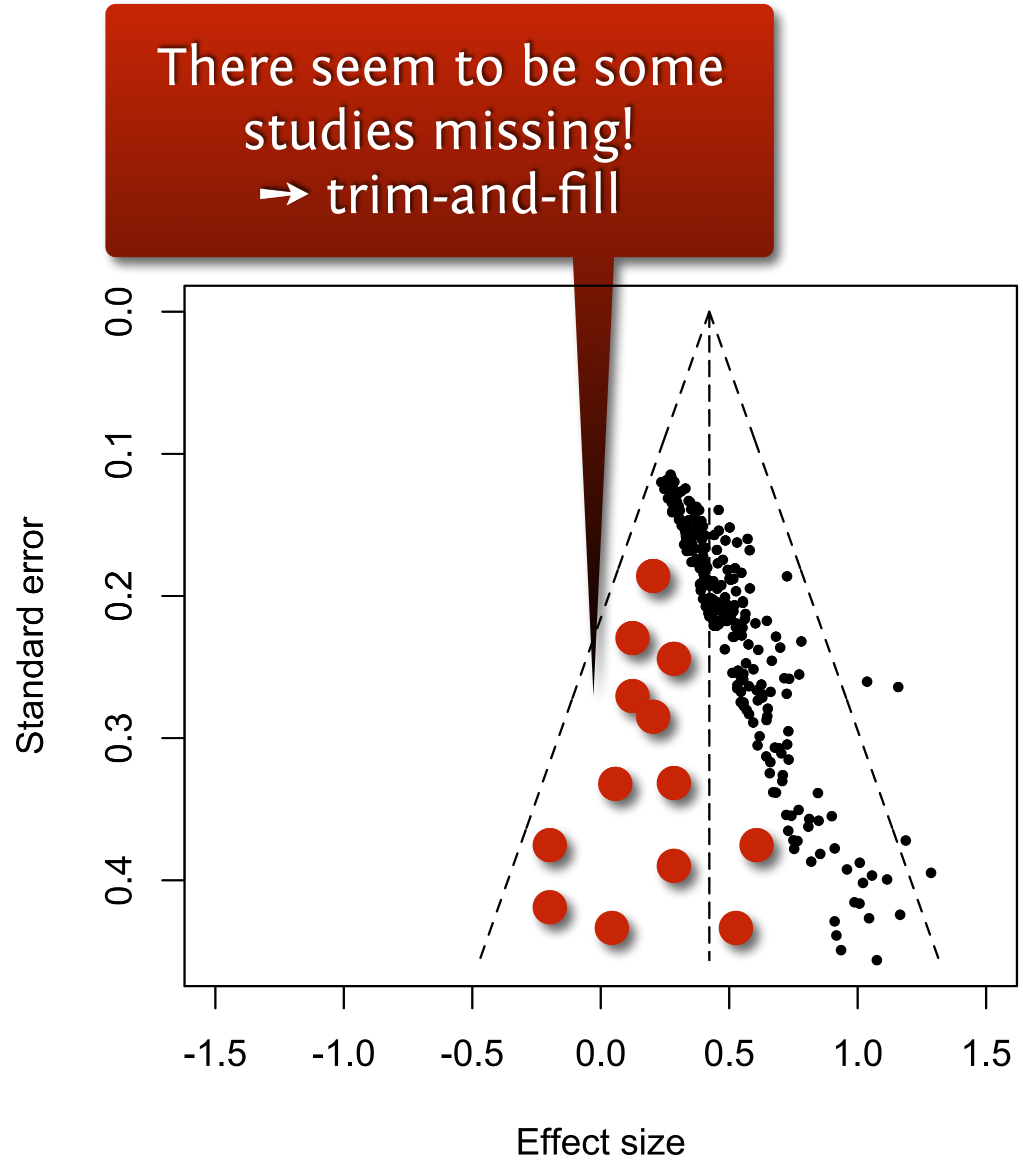
Correcting for publication bias (PB)

or

Can we clean up the mess,
if we only had the right tool?

Trim & Fill

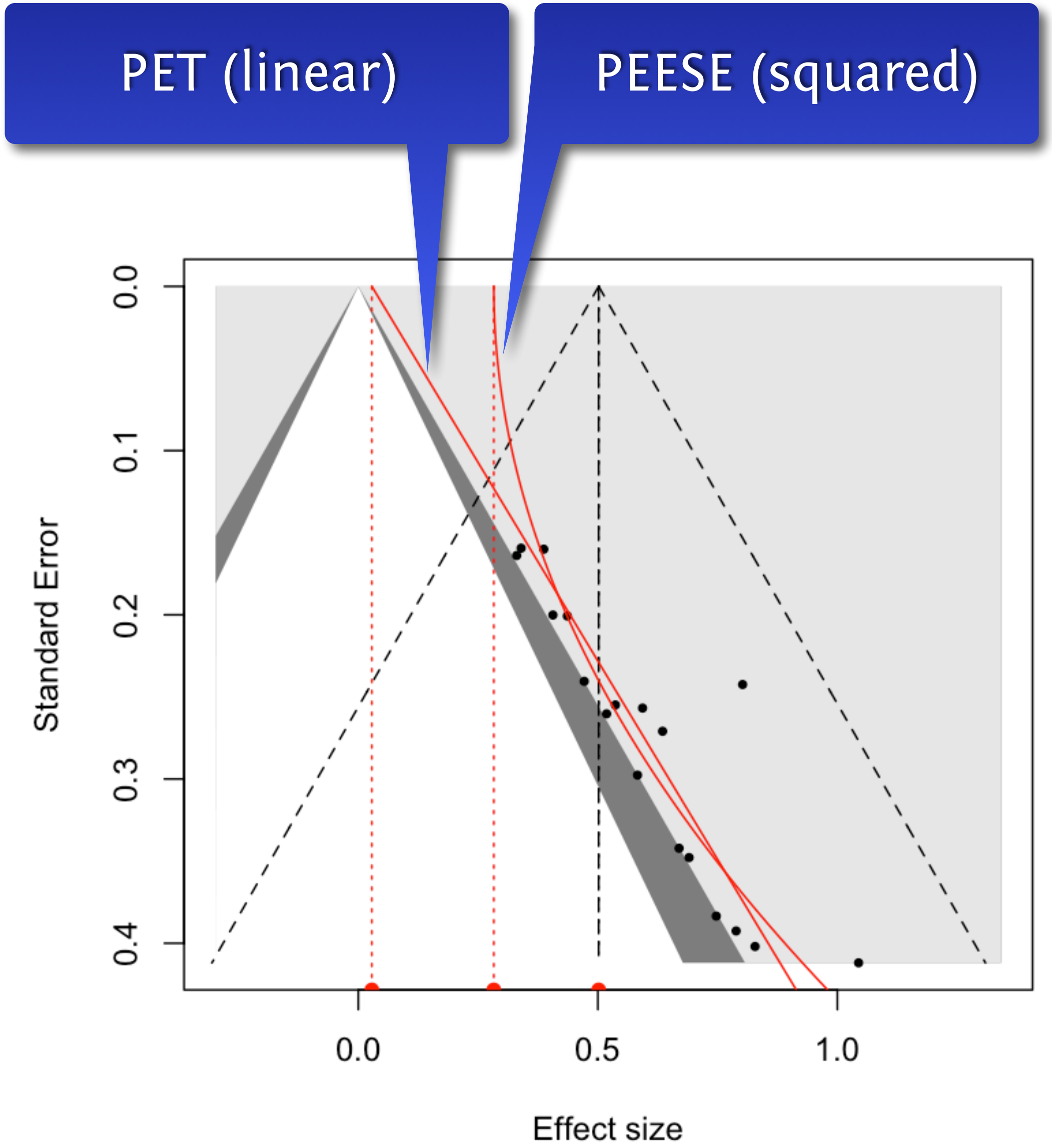
- Originally designed as a *test* for PB, but also used to *correct* for PB
- Algorithmically fill in missing studies to achieve a symmetric funnel plot
- Compute meta-analysis on the data set including imputed studies



Duval, S. & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56 (2)*, 455–463.

PET / PEESE

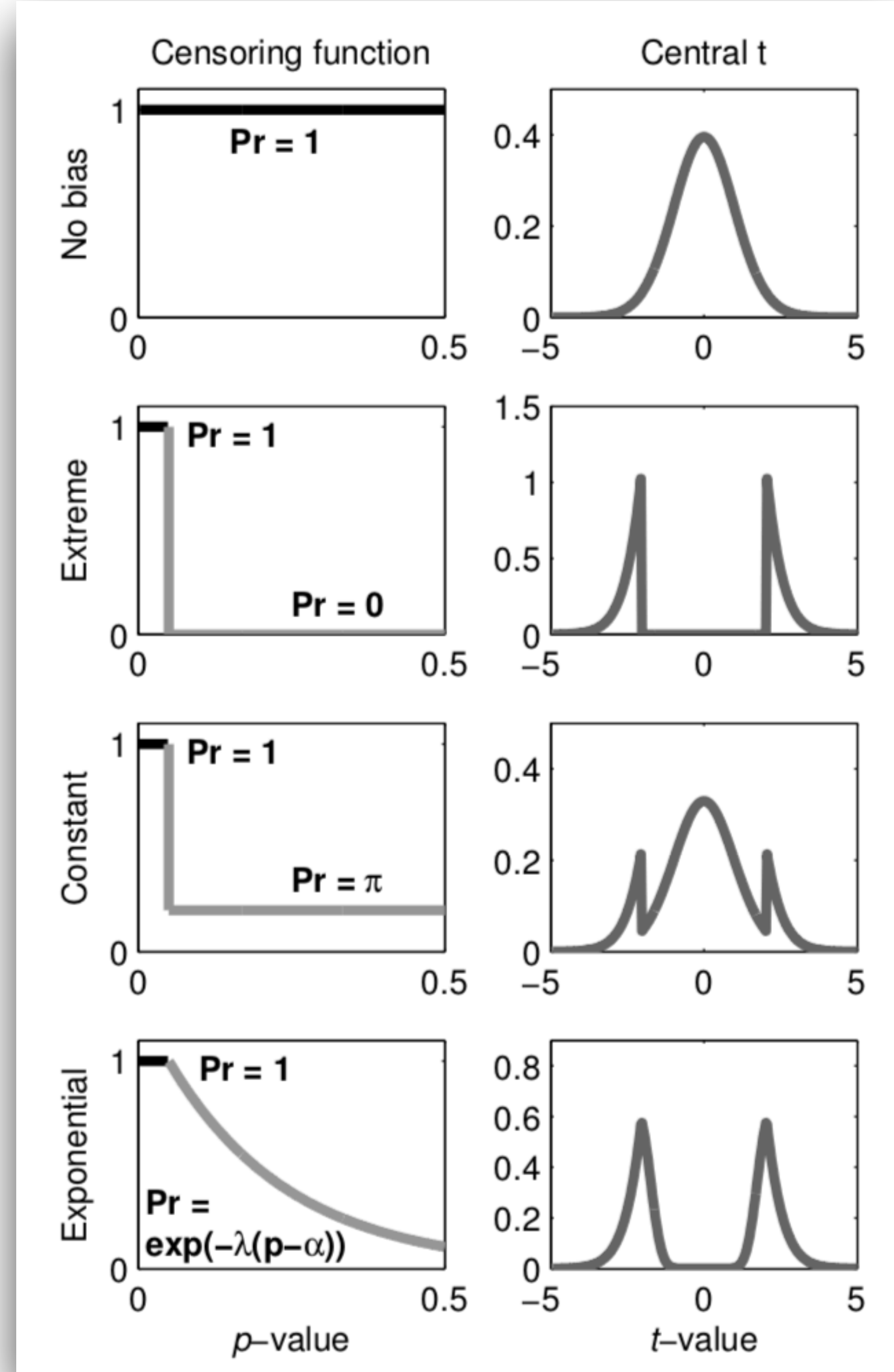
- Extrapolates the „small study effect" to samples with ∞ sample size
- What would be the effect size if we had an infinitely large sample?
- PET: linear regression
- PEESE: squared slope



Stanley, T. D., & Doucouliagos, H. (2013). Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods, 5(1), 60–78. http://doi.org/10.1002/jrsm.1095

Selection models

- Explicitly model the functional form of publication bias
- Provide estimates for, e.g., *Prob*(published | n.s.)
- Three-parameter SM: **μ**, **τ**, and *Prob*(published | n.s.)
- Four-parameter SM: **μ**, **τ**, and *Prob*(pub | n.s. & correct direction) and *Prob*(pub | wrong direction)

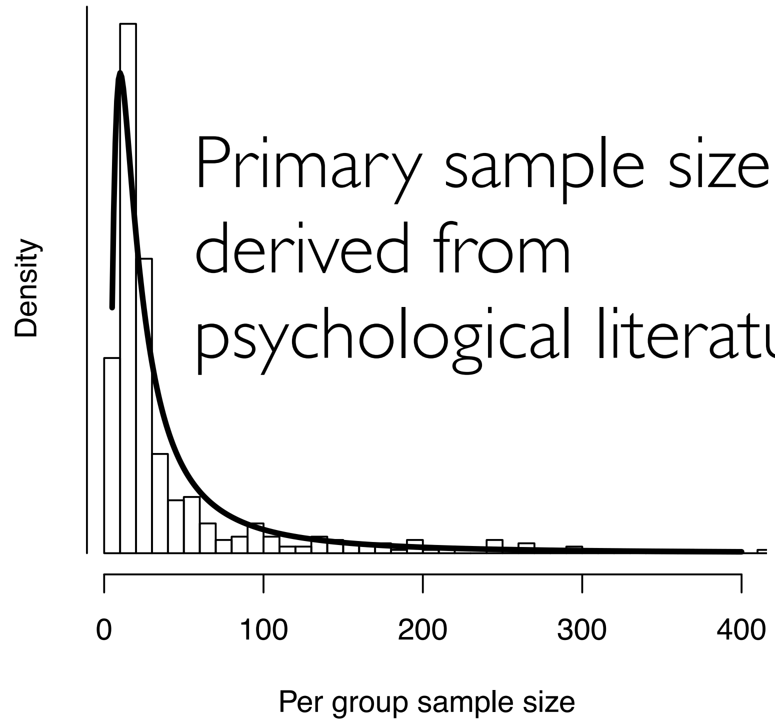


from Guan & Vandekerckhove, 2015)

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016).
Iyengar, S. & Greenhouse, J. B. (1988)
Hedges, L. V. (1984)

Performance of bias correcting methods

Simulation study



Primary sample size derived from psychological literature

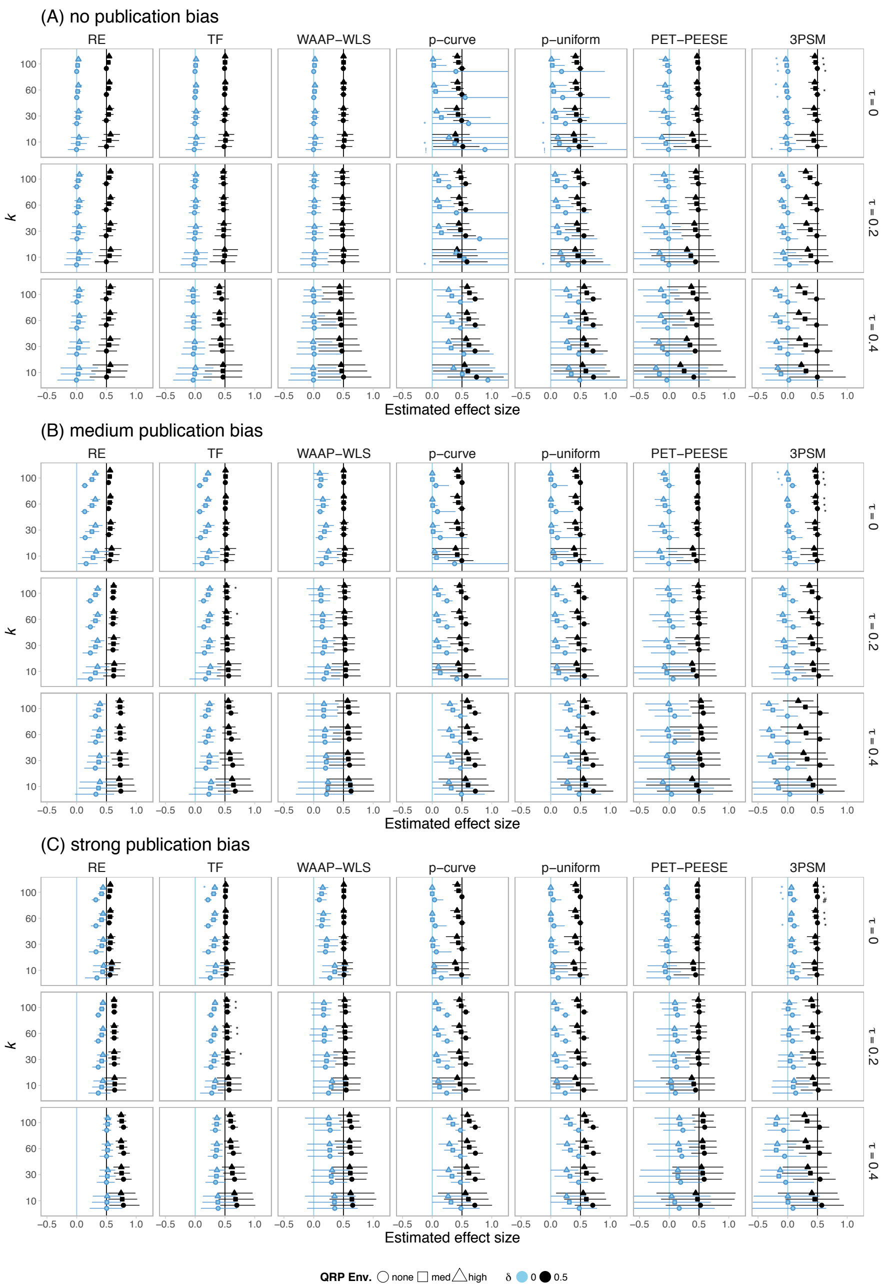
Table 1
Simulation parameters

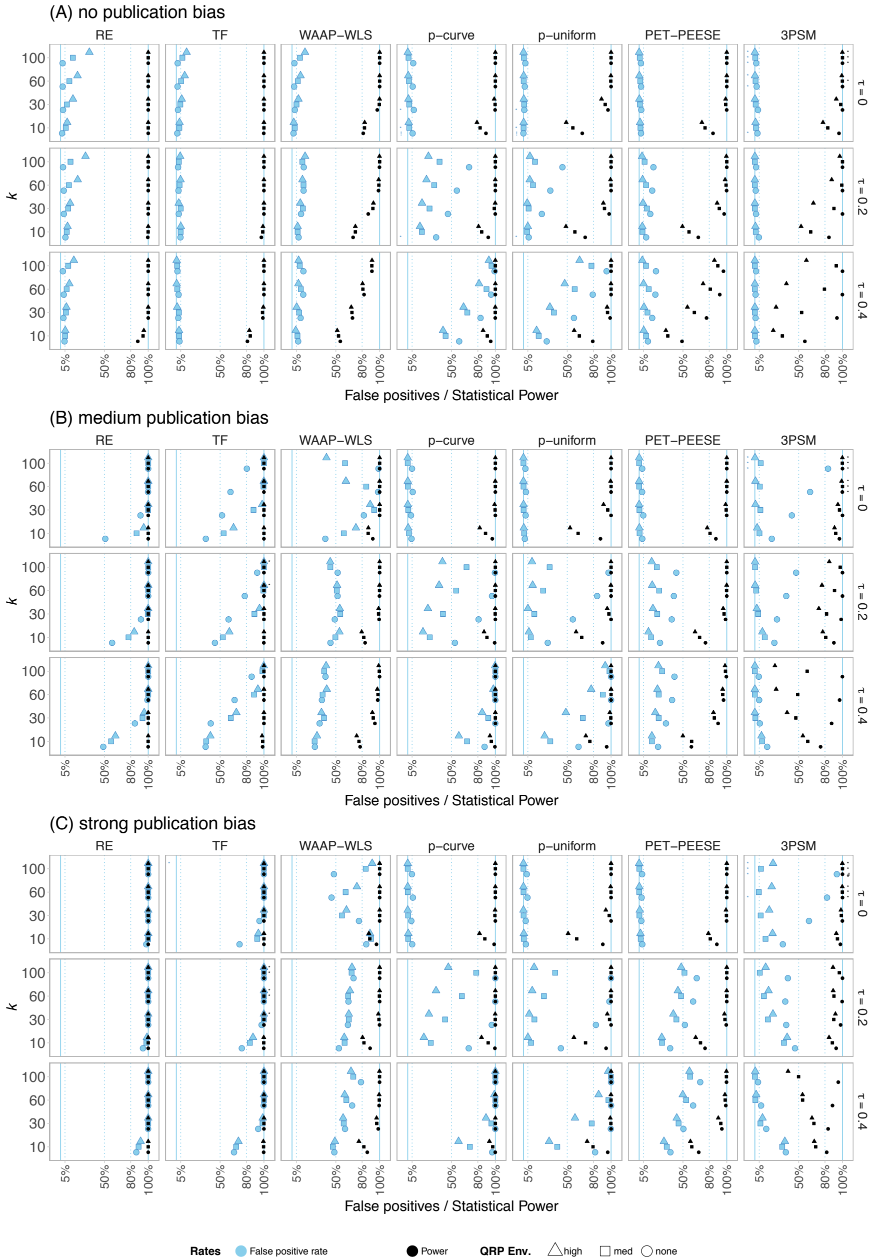
Experimental factors	Levels
True underlying effect (δ)	0, 0.2, 0.5, 0.8
Between-study heterogeneity (τ)	0, 0.2, 0.4
Number of studies in the meta-analytic sample (k)	10, 30, 60, 100
Publication bias (PB)	None, medium, strong
QRP environment (QRP)	None, medium, high

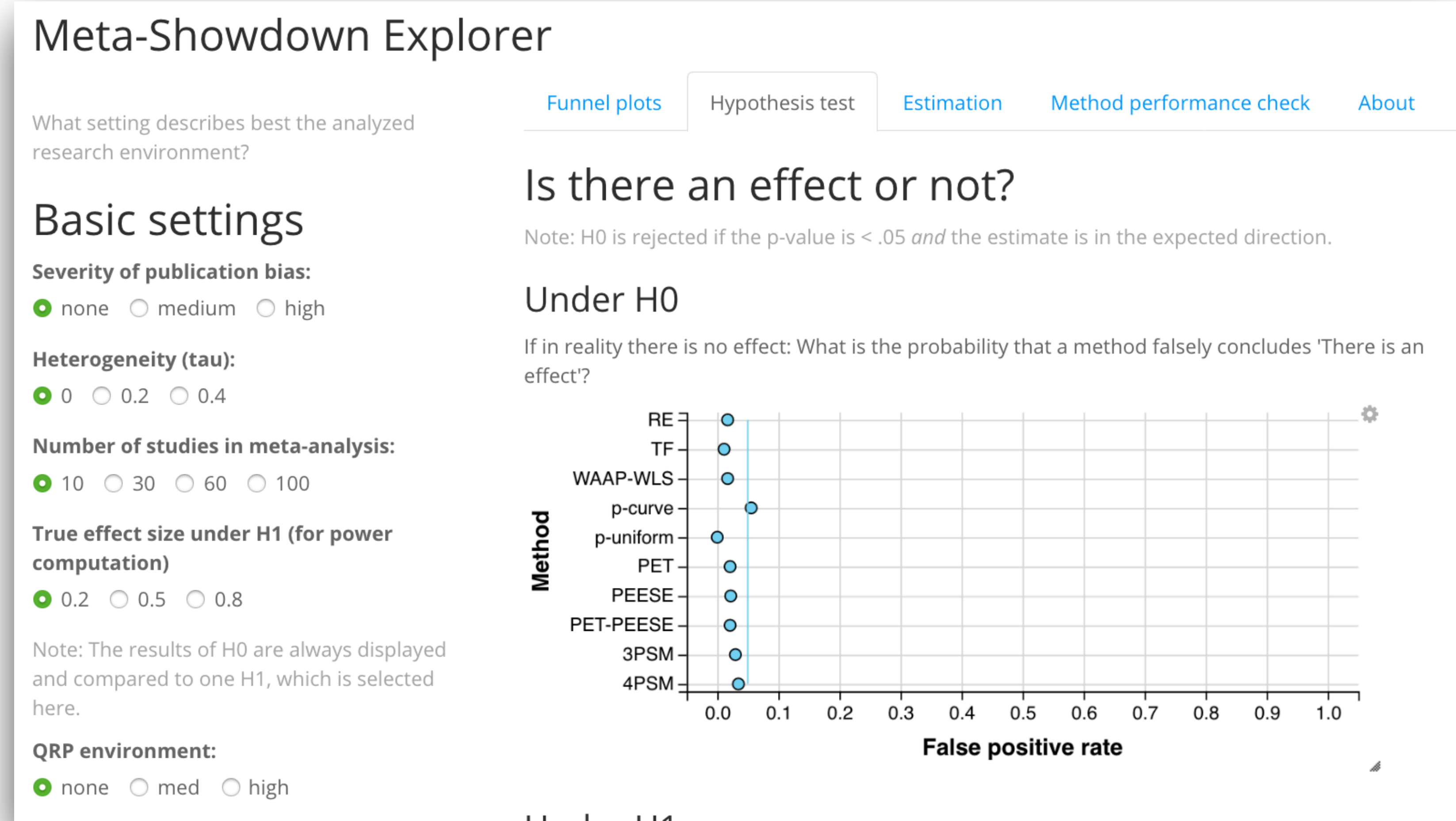
fully crossed:
432 conditions

Estimators:
(naive) Random effects meta-analysis, Trim&Fill, PET, PEESE, PET-PEESE, three-parameter selection model (3PSM), four-parameter selection model (4PSM), *p*-curve, *p*-uniform, WAAP-WLS

Results (a selection)







http://shinyapps.org/apps/metaExplorer/

Hypothesis test

How many % of original studies are submitted to publication bias?:

● 0% ○ 60% ○ 90%

Heterogeneity (tau):

● 0 ○ 0.2 ○ 0.4

Number of studies in meta-analysis:

○ 10 ● 30 ○ 60 ○ 100

True effect size under H1 (for power computation)

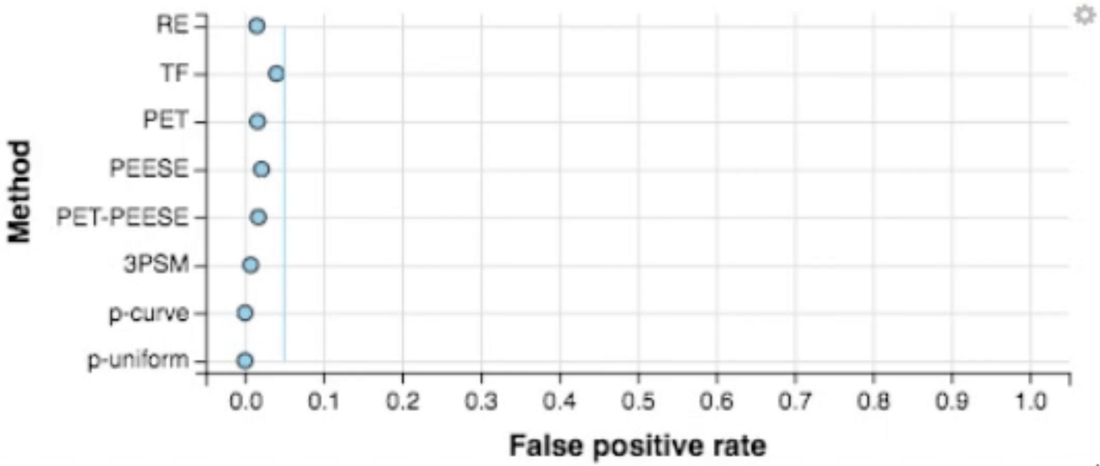
● 0.2 ○ 0.5 ○ 0.8

QRP environment:

● none ○ med ○ high

Under H0

If in reality there is no effect: What is the probability that a method falsely concludes 'There is an effect'?



Effect size estimation

Basic settings

How many % of original studies are submitted to publication bias?:

● 0% ○ 60% ○ 90%

Heterogeneity (tau):

○ 0 ● 0.2 ○ 0.4

Number of studies in meta-analysis:

○ 10 ○ 30 ○ 60 ● 100

True effect size under H1 (for power computation)

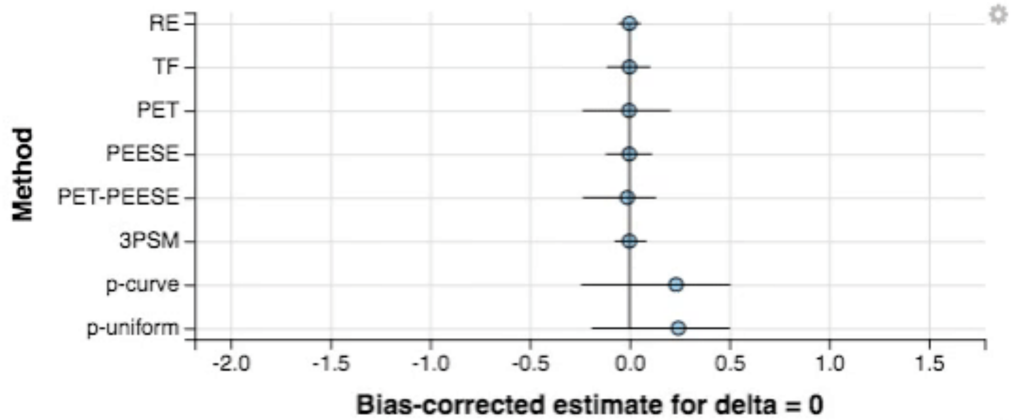
● 0.2 ○ 0.5 ○ 0.8

QRP environment:

● none ○ med ○ high

Bias-corrected estimates of the true effect

Under H0



Method performance check

- Hope that all bias-correcting methods will converge on the same value? Usually that does not happen
- ➙No vote counting - no triangulation:
  - Even if three out of four methods converge on a value this is irrelevant, when those three are known to perform badly in plausible conditions.
- Use the app to see which bias-correcting methods perform well in plausible conditions for the meta-analysis at hand
- Do a sensitivity analysis - but only including methods that passed the performance check!

Meta-analysis - the pinnacle of evidence-based research?

Meta-analyses are fucked?

- Publication bias and *p*-hacking massively distorts the evidence:
  Garbage in - garbage out.
- Even meta-analyses of many dozen significant primary studies can come from a null effect.
- Each type of bias-correction works in some conditions, but fails in other conditions.
  Problem: We do not know which condition we are in.
- Doing biased research and hoping to correct it afterward *does not work*.
- Better put efforts into improving primary studies themselves (e.g., by using registered reports which combat both *p*-hacking and publication bias)

Correcting for bias in psychology: A comparison of meta-analytic methods

Evan C. Carter*
U.S. Army Research Laboratory, Aberdeen, MD, USA

Felix D. Schönbrodt*
Ludwig-Maximilians-Universität, Munich, Germany

Will M. Gervais
University of Kentucky, Lexington, KY, USA

Joseph Hilgard
University of Pennsylvania, Philadelphia, PA, USA

Publication bias and questionable research practices in primary research can lead to badly overestimated effects in meta-analysis. Methodologists have proposed a variety of statistical approaches to correct for such overestimation. However, much of this work has not been tailored specifically to psychology, so it is not clear which methods work best for data typically seen in our field. Here, we present a comprehensive simulation study to examine how some of the most promising meta-analytic methods perform on data that might realistically be produced by research in psychology. We created such scenarios by simulating several levels of questionable research practices, publication bias, heterogeneity, and using study sample sizes empirically derived from the literature. Our results clearly indicated that no single meta-analytic method consistently outperformed all others. Therefore, we recommend that meta-analysts in psychology focus on sensitivity analyses—that is, report on a variety of methods, consider the conditions under which these methods fail (as indicated by simulation studies such as ours), and then report how conclusions might change based on which conditions are most plausible. Moreover, given the dependence of meta-analytic methods on untestable assumptions, we strongly recommend that researchers in psychology continue their efforts on improving the primary literature and conducting large-scale, pre-registered replications. We provide detailed results and simulation code at https://osf.io/rf3ys and interactive figures at http://www.shinyapps.org/apps/metaExplorer/.

Keywords: meta-analysis, publication bias, *p*-hacking, questionable research practices, bias-correction.

Statistical techniques for analyzing the results from a set of studies in aggregate—often called meta-analysis—are popular in psychology and many other scientific disciplines because they provide high-powered tests, the ability to examine moderators across studies, and precise effect size estimates that are useful for planning future studies and making policy decisions. However, just as the results from individual studies can be made completely misleading by bias (e.g., Simmons, Nelson, & Simonsohn, 2011), so too can meta-analytic results. To address this, researchers have developed statistical techniques designed to identify and correct for bias. Without having a particular preference in any specific method, we present a neutral comparison (Boulesteix, Wilson, & Hapfelmeier, 2017) of how several promising methods perform when applied to simulated data that could have plausibly been produced by research in psychology. Our goal is to help researchers in psychology know what to expect from different methods when conducting meta-analysis in the face of bias.

Meta-analysis

Meta-analytic techniques involve synthesizing a set of results from studies investigating the same empirical phenomenon (Borenstein, Hedges, Higgins, & Rothstein, 2011). Most often, the results from the individual studies take the form of effect size estimates, and because meta-analyses are usually applied to studies with dependent variables measured on different scales, effect size estimates are typically standardized. The typical goal of a meta-analysis is to produce a single summary estimate of the hypothetical true underlying effect, δ, estimated by each effect size in the dataset. This is usually called fixed-effect meta-analysis (Cooper, Hedges, & Valentine, 2009) and can be modeled as $d_i = \delta + e_i$, where d_i is the observed effect size for study i that differs from the true underlying effect, δ, by some amount of sampling er-

Correspondence concerning this article should be addressed to Evan Carter, Email: evan.c.carter@gmail.com. *These authors contributed equally to this work.

https://psyarxiv.com/9h3nu/

- „Researchers should **not expect** to produce a conclusive, **debate-ending result** by conducting a meta-analysis on an existing literature"
- „Instead, we imagine meta-analyses may serve best to draw attention to the existing strengths and/or weaknesses in a literature and these results can then inspire a careful re-examination of methodology and theory followed by, if necessary, **large-scale, preregistered replication efforts.**"