Correcting for bias in the literature

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

Felix Schönbrodt, Evan Carter, Will Gervais, Joe Hilgard

This presentation is licensed under a CC-BY 4.0 license. You may copy, distribute, and use the slides in your own work, as long as you give attribution to the original author at each slide that you use.
Meta-analysis is at the top of the evidence-based medicine pyramid - the pinnacle of evidence-based medicine.

Cochrane Collaboration


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*

* simulated data

5% false positive ("significant") studies
True $H_0 +$ directional publication bias

There seem to be some studies missing!

Studies “huddle” against the significance threshold

Meta-analytic effect size estimate: $d = 0.42$

* simulated data
True $H_0$ + publication bias

Negative correlation of study size & estimated effect size:
Smaller studies have larger effects

Studies “huddle” against the significance threshold

* simulated data
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
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

14 replication studies, all n.s.
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

PET / PEESE

• Extrapolates the „small study effect“ to samples with $\infty$ sample size

• What would be the effect size if we had an infinitely large sample?

• PET: linear regression

• PEESE: squared slope

Selection models

• Explicitly model the functional form of publication bias

• Provide estimates for, e.g., \( \text{Prob}(\text{published} \mid \text{n.s.}) \)

• Three-parameter SM: \( \mu, \tau \), and \( \text{Prob}(\text{published} \mid \text{n.s.}) \)

• Four-parameter SM: \( \mu, \tau \), and \( \text{Prob}(\text{pub} \mid \text{n.s. & correct direction}) \) and \( \text{Prob}(\text{pub} \mid \text{wrong direction}) \)

Hedges, L. V. (1984)
Performance of bias correcting methods
Simulation study

Table 1
Simulation parameters

<table>
<thead>
<tr>
<th>Experimental factors</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>True underlying effect ($\delta$)</td>
<td>0, 0.2, 0.5, 0.8</td>
</tr>
<tr>
<td>Between-study heterogeneity ($\tau$)</td>
<td>0, 0.2, 0.4</td>
</tr>
<tr>
<td>Number of studies in the meta-analytic sample ($k$)</td>
<td>10, 30, 60, 100</td>
</tr>
<tr>
<td>Publication bias ($PB$)</td>
<td>None, medium, strong</td>
</tr>
<tr>
<td>QRP environment ($QRP$)</td>
<td>None, medium, high</td>
</tr>
</tbody>
</table>

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

fully crossed: 432 conditions
Results (a selection)

(A) no publication bias

(B) medium publication bias

(C) strong publication bias
Meta-Showdown Explorer

What setting describes best the analyzed research environment?

Basic settings

Severity of publication bias:
- none
- medium
- high

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

Note: The results of H0 are always displayed and compared to one H1, which is selected here.

QRP environment:
- none
- med
- high

Is there an effect or not?

Note: H0 is rejected if the p-value is < .05 and the estimate is in the expected direction.

Under H0

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

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

Bias-corrected estimate for delta = 0
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)
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.”

https://psyarxiv.com/9h3nu/