Weight of Statistical Evidence

Detection and Correction of Publication Bias

Servan Grüninger

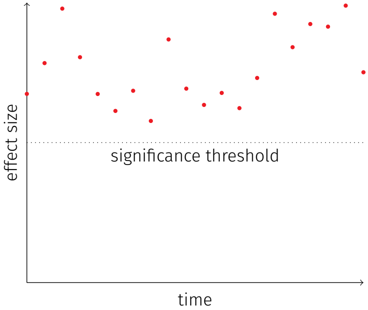
Écublens, July 9th

Master's Programme in Computational Science and Engineering

The Woozle effect

Pooh and Piglet tracking down the elusive Woozle (Image: Ernest H. Shepard)

The Woozle effect—why care?



The Woozle effect—why care?

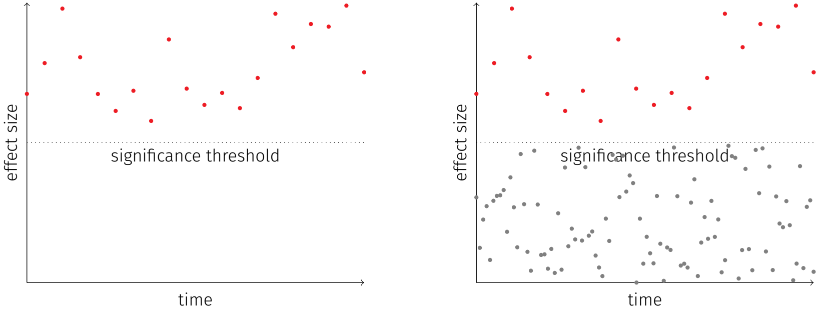


Table of Contents

Analysing and Testing Evidence

Hypothesis Testing

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

Making evidence comparable

To test hypotheses, we need a test statistic V_n.

Making evidence comparable

To test hypotheses, we need a test statistic V_n.
Ideally, test statistics fulfill the following properties Kulinskaya, Morgenthaler, and Staudte (2008):

Making evidence comparable

To test hypotheses, we need a test statistic V_n.
Ideally, test statistics fulfill the following properties Kulinskaya, Morgenthaler, and Staudte (2008):

$$E_1 : \; V_n \text{ is monotonically increasing;}$$

Making evidence comparable

To test hypotheses, we need a test statistic V_n.
Ideally, test statistics fulfill the following properties Kulinskaya, Morgenthaler, and Staudte (2008):

$$E_1: \; V_n \text{ is monotonically increasing;}$$
$$E_2: \; V_n \sim \mathcal{N}(\tau, 1);$$

Making evidence comparable

To test hypotheses, we need a test statistic V_n.
Ideally, test statistics fulfill the following properties Kulinskaya, Morgenthaler, and Staudte (2008):

$$E_1: \ V_n \text{ is monotonically increasing;}$$
$$E_2: \ V_n \sim \mathcal{N}(\tau, 1);$$
$$E_3: \ \mathrm{Var}[V_n] = 1;$$

Making evidence comparable

To test hypotheses, we need a test statistic V_n.
Ideally, test statistics fulfill the following properties Kulinskaya, Morgenthaler, and Staudte (2008):

$$\begin{aligned}
E_1 &: \ V_n \text{ is monotonically increasing;} \\
E_2 &: \ V_n \sim \mathcal{N}(\tau, 1); \\
E_3 &: \ \mathrm{Var}[V_n] = 1; \\
E_4 &: \ \mathrm{E}_\mu[V_n] = \tau(\mu) \text{ is monotonically increasing in } \mu \text{ from} \\
&\quad\ \tau(0) = 0.
\end{aligned}$$

Example case: difference in means

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ and σ^2 unknown.

Example case: difference in means

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ and σ^2 unknown.

Student's t-statistic as test statistic:

$$T_n = \frac{\sqrt{n}(\hat{\mu} - \mu_0)}{\hat{\sigma}} \overset{H_0}{\sim} t(\nu = n - 1)$$

Example case: difference in means

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ and σ^2 unknown.

Student's t-statistic as test statistic:

$$T_n = \frac{\sqrt{n}(\hat{\mu} - \mu_0)}{\hat{\sigma}} \overset{H_0}{\sim} t(\nu = n - 1)$$

with

$$\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i; \quad \hat{\sigma}^2 = s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_N)^2$$

Variance stabilising transformation

If n is large enough, T_n fulfills all properties E_1 to E_4 by virtue of the central limit theorem.

Variance stabilising transformation

If n is large enough, T_n fulfills all properties E_1 to E_4 by virtue of the central limit theorem.

For small n, however, E_2 and E_3 are violated.

Variance stabilising transformation

If n is large enough, T_n fulfills all properties E_1 to E_4 by virtue of the central limit theorem.

For small n, however, E_2 and E_3 are violated.

Solution: Transform T_n!

Variance stabilising transformation

$V_n = h_n(T_n) = \sqrt{2n}\,\sinh^{-1}(T_n/\sqrt{2n})$ yields improved approximation to standard normal distribution Kulinskaya, Morgenthaler, and Staudte (2008).

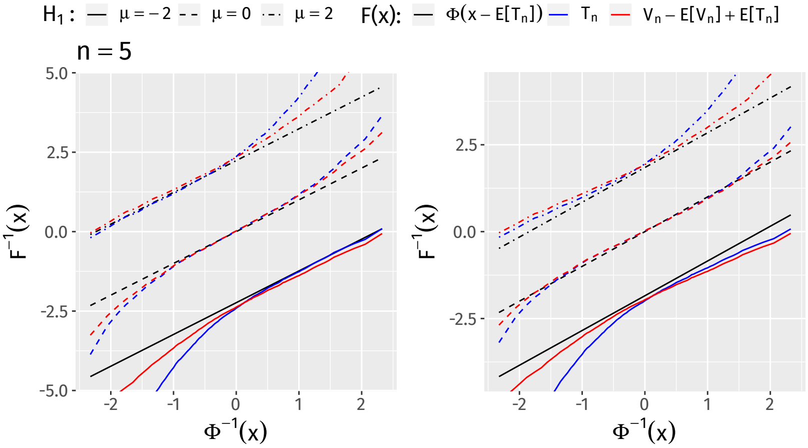
Variance stabilising transformation

$V_n = h_n(T_n) = \sqrt{2n}\sinh^{-1}(T_n/\sqrt{2n})$ yields improved approximation to standard normal distribution Kulinskaya, Morgenthaler, and Staudte (2008).

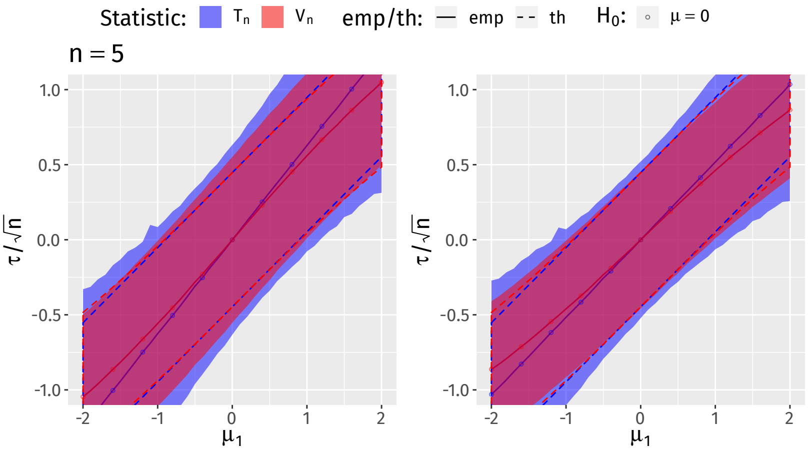
Further improvement by finite sample correction:

$$V_n^* = \frac{n-1.7}{n-1}\sqrt{2n}\sinh^{-1}(T_n/\sqrt{2n})$$

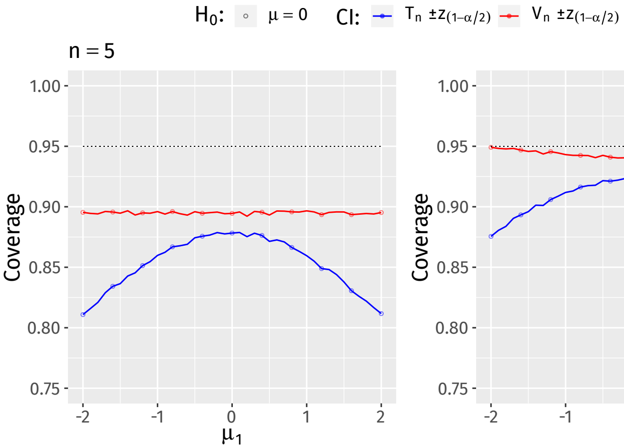
h_n improves normal fit



h_n stabilises variance



h_n improves empirical coverage probability of confidence intervals



Detecting Publication Bias

Aggregate study effects in the absence of publication bias

Without publication bias, the global effect size across k studies can be estimated by:

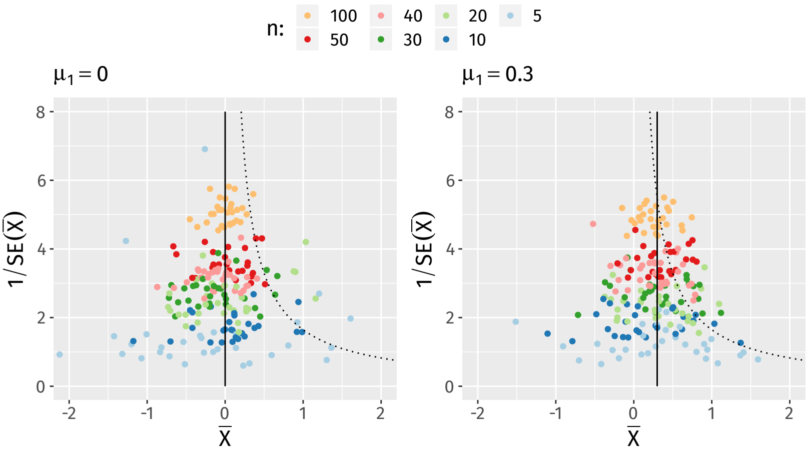
$$\bar{X}_N = \frac{\sum_{j=1}^{k} \bar{X}_{n_j} w_j}{\sum_{j=1}^{k} w_j} \sim \mathcal{N}(\mu, \sigma^2/N); \quad N = \sum_{j=1}^{k} n_j; \quad w_j = n_j/\sigma^2$$

Funnelling statistical evidence

Funnel plots: draw effect size against precision, look for gaps.

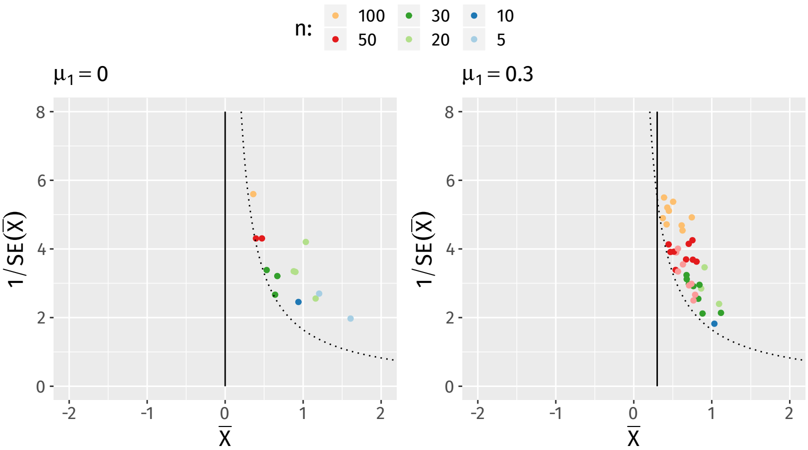
Idea proposed by Light and Pillemer (1984, p. 64–69).

Funnelling statistical evidence



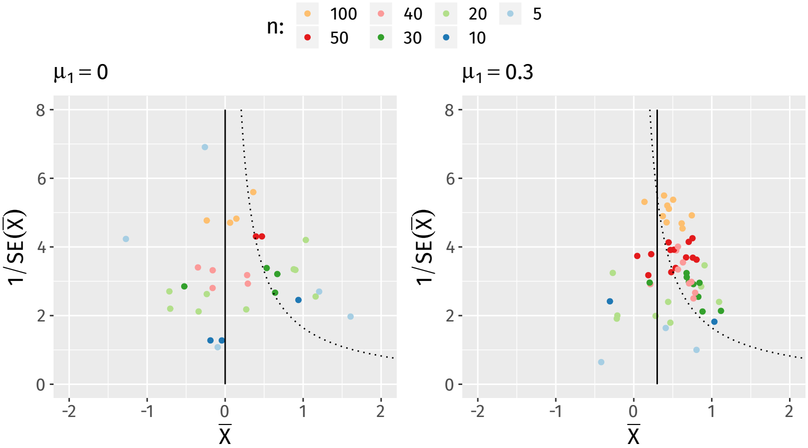
No publication bias

Funnelling statistical evidence



Significant studies only

Funnelling statistical evidence



Significant studies and 10% of non-significant studies

The file drawer problem

Many studies land in the file drawer (Image: Geckoboard)

The file drawer problem

Worst case assumption by Rosenthal (1979): Published results are Type I error only.

The file drawer problem

Worst case assumption by Rosenthal (1979): Published results are Type I error only.

Calculate number of omitted studies needed to make findings non-significant:

$$z_{(1-\alpha)} = \frac{k\bar{z}_k + o\bar{z}_o}{\sqrt{k+o}}$$

The file drawer problem

Worst case assumption by Rosenthal (1979): Published results are Type I error only.

Calculate number of omitted studies needed to make findings non-significant:

$$
\begin{aligned}
z_{(1-\alpha)} &= \frac{k\bar{z}_k + o\bar{z}_o}{\sqrt{k+o}} \\
\iff \quad o &= \frac{(k\bar{z}_k + o\bar{z}_o)^2}{z^2_{(1-\alpha)}} - k.
\end{aligned}
$$

The file drawer problem

Assume that $\bar{z}_o = 0$:

$$o = \frac{(k\bar{z}_k)^2}{z^2_{(1-\alpha)}} - k.$$

The file drawer problem

Assume that $\bar{z}_o = 0$:

$$o = \frac{(k\bar{z}_k)^2}{z^2_{(1-\alpha)}} - k.$$

More accurate is $\bar{z}_o = \mathrm{E}[Z \mid Z < z_{(1-\alpha)}]$:

The file drawer problem

Assume that $\bar{z}_o = 0$:

$$o = \frac{(k\bar{z}_k)^2}{z^2_{(1-\alpha)}} - k.$$

More accurate is $\bar{z}_o = \mathrm{E}[Z \mid Z < z_{(1-\alpha)}]$:

$$o^* = \frac{-2k\bar{z}_k\bar{z}_o + z^2_{(1-\alpha)} - z_{(1-\alpha)}\sqrt{4k\bar{z}_o^2 - 4k\bar{z}_k\bar{z}_o + z^2_{(1-\alpha)}}}{2\bar{z}_o^2}.$$

Check whether $o^* < 5k + 10$.

The file drawer problem

	k	μ_1	o	o^*	bias detected?
full sample (no bias)	200	0	-141	-181	o : no o^*: no
	200	0.3	8123	884	o : no o^*: no

Critical value: $5k + 10$

The file drawer problem

	k	μ_1	o	o^*	bias detected?
full sample (no bias)	200	0	-141	-181	o : no o^*: no
	200	0.3	8123	884	o : no o^*: no
significant studies	13	0	320	109	o : no o^*: no
	38	0.3	2705	457	o : no o^*: no

Critical value: $5k + 10$

The file drawer problem

	k	μ_1	o	o^*	bias detected?
full sample (no bias)	200	0	-141	-181	o : no o^*: no
	200	0.3	8123	884	o : no o^*: no
significant studies	13	0	320	109	o : no o^*: no
	38	0.3	2705	457	o : no o^*: no
significant studies and 10% of non-significant studies	32	0	99	43	o : yes o^*: yes
	55	0.3	3086	494	o : no o^*: no

Critical value: $5k + 10$

How many significant studies ought to be expected?

Assuming σ^2 to be known, the (overestimated) power of the jth study is:

$$1 - \beta_j = \Phi(\sqrt{n_j}\frac{\bar{X}_N}{\sigma} - z_{(1-\alpha)})$$

How many significant studies ought to be expected?

Assuming σ^2 to be known, the (overestimated) power of the jth study is:

$$1 - \beta_j = \Phi(\sqrt{n_j}\frac{\bar{X}_N}{\sigma} - z_{(1-\alpha)})$$

Expected number of significant studies in the absence of publication bias (Ioannidis and Trikalinos, 2007):

$$E = \sum_{j=1}^{k}(1 - \beta_j).$$

How many significant studies ought to be expected?

Assuming σ^2 to be known, the (overestimated) power of the jth study is:

$$1 - \beta_j = \Phi(\sqrt{n_j}\frac{\bar{X}_N}{\sigma} - z_{(1-\alpha)})$$

Expected number of significant studies in the absence of publication bias (Ioannidis and Trikalinos, 2007):

$$E = \sum_{j=1}^{k}(1 - \beta_j).$$

Compare to observed number of studies O via χ^2-test:

$$A = [(O - E)^2/E + (O - E)^2/(k - E)] \sim \chi^2_1.$$

How many significant studies ought to be expected?

	k	μ_1	$\bar{x}_N$	A	bias detected?
full sample (no bias, A)	200	0	-0.02	1.70	no
	200	0.3	0.27	0.08	no

Critical value: 3.84 (95%-quantile of χ^2_1-distribution)

How many significant studies ought to be expected?

	k	μ_1	$\bar{x}_N$	A	bias detected?
full sample (no bias, A)	200	0	-0.02	1.70	no
	200	0.3	0.27	0.08	no
significant studies (B)	13	0	0.69	11.67	yes
	38	0.3	0.61	19.94	yes

Critical value: 3.84 (95%-quantile of χ^2_1-distribution)

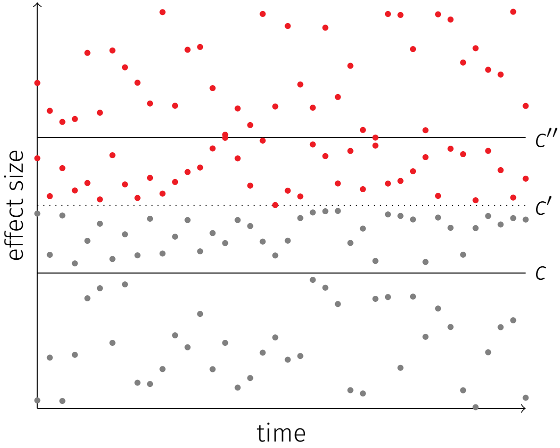
How many significant studies ought to be expected?

	k	μ_1	$\bar{x}_N$	A	bias detected?
full sample (no bias, A)	200	0	-0.02	1.70	no
	200	0.3	0.27	0.08	no
significant studies (B)	13	0	0.69	11.67	yes
	38	0.3	0.61	19.94	yes
significant studies and 10% of non-significant studies (C)	32	0	0.15	28.64	yes
	55	0.3	0.51	8.21	yes

Critical value: 3.84 (95%-quantile of χ^2_1-distribution)

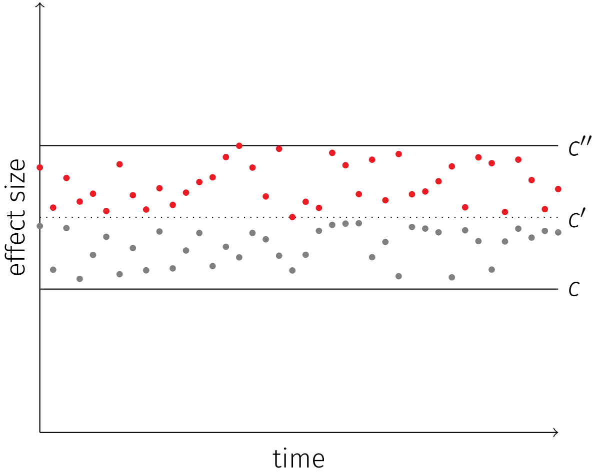
The calliper test

$$c' = z_{(1-\alpha)}; \quad c'' = c + e; \quad c = c' - e; \quad k' \in [c, c'']; \quad k'' \in [c', c''].$$



The calliper test

$$c' = z_{(1-\alpha)}; \quad c'' = c + e; \quad c = c' - e; \quad k' \in [c, c'']; \quad k'' \in [c', c''].$$



The calliper test

Proposed by Gerber and Malhotra (2006)

The calliper test

Proposed by Gerber and Malhotra (2006)

$p = \Pr[Z \in [c, c'] \mid Z \in [c, c'']]$

The calliper test

Proposed by Gerber and Malhotra (2006)

$$p = \Pr[Z \in [c, c'] \mid Z \in [c, c'']] = \frac{\Phi(c') - \Phi(c)}{\Phi(c'') - \Phi(c)}$$

The calliper test

Proposed by Gerber and Malhotra (2006)

$$p = \Pr[Z \in [c, c'] \mid Z \in [c, c'']] = \frac{\Phi(c') - \Phi(c)}{\Phi(c'') - \Phi(c)} \simeq \frac{(c' - c)\phi(c)}{(c'' - c)\phi(c)}$$

The calliper test

Proposed by Gerber and Malhotra (2006)

$$p = \Pr[Z \in [c, c'] \mid Z \in [c, c'']] = \frac{\Phi(c') - \Phi(c)}{\Phi(c'') - \Phi(c)} \simeq \frac{(c' - c)\phi(c)}{(c'' - c)\phi(c)} = \frac{c' - c}{c'' - c}.$$

The calliper test

Proposed by Gerber and Malhotra (2006)

$$p = \Pr[Z \in [c, c'] \mid Z \in [c, c'']] = \frac{\Phi(c') - \Phi(c)}{\Phi(c'') - \Phi(c)} \simeq \frac{(c' - c)\phi(c)}{(c'' - c)\phi(c)} = \frac{c' - c}{c'' - c}.$$

$K'' \sim \mathrm{Bin}(k', p = 0.5)$

The calliper test

	k	μ_1	k'	k''	$q_{(1-\alpha/2)}$	bias detected?
full sample (no bias, A)	200	0	4	3	4	no
	200	0.3	15	6	11	no

Critical value: 97.5%-quantile of Bin(k', 0.5)-distribution

The calliper test

	k	μ_1	k'	k''	$q_{(1-\alpha/2)}$	bias detected?
full sample (no bias, A)	200	0	4	3	4	no
	200	0.3	15	6	11	no
significant studies (B)	13	0	3	3	3	no
	38	0.3	6	6	5	yes

Critical value: 97.5%-quantile of Bin(k', 0.5)-distribution

The calliper test

	k	μ_1	k'	k''	$q_{(1-\alpha/2)}$	bias detected?
full sample (no bias, A)	200	0	4	3	4	no
	200	0.3	15	6	11	no
significant studies (B)	13	0	3	3	3	no
	38	0.3	6	6	5	yes
significant studies and 10% of non-significant studies (C)	32	0	3	3	3	no
	55	0.3	7	6	6	no

Critical value: 97.5%-quantile of Bin(k', 0.5)-distribution

Correcting Publication Bias

Publication probabilities and truncated distributions

- publication probability:

$$\mathrm{ppr}(S_n, \pi) = \pi + (1 - \pi)\delta(S_n)$$

Publication probabilities and truncated distributions

- publication probability:

$$\text{ppr}(S_n, \pi) = \pi + (1 - \pi)\delta(S_n)$$

- expected publication probability:

$$\text{E}[\text{ppr}(S_n, \pi)] = \pi \Pr(\text{ppr} = \pi) + \Pr(\text{ppr} = 1)$$

Publication probabilities and truncated distributions

- publication probability:

$$\text{ppr}(S_n, \pi) = \pi + (1 - \pi)\delta(S_n)$$

- expected publication probability:

$$\text{E}[\text{ppr}(S_n, \pi)] = \pi \Pr(\text{ppr} = \pi) + \Pr(\text{ppr} = 1)$$

- truncated probability density function of S_n:

$$f^*_{S_n}(s_n) = \frac{\text{ppr}(s_n, \pi)}{\text{E}[\text{ppr}(S_n, \pi)]} f_{S_n}(s_n)$$

Reweight by publication probability

Let

$$\mu_{\text{sig}} = \mathrm{E}[\bar{X}_{n_j} \mid S_{n_j} > q_{(1-\alpha)}]; \quad \mu_{\text{ns}} = \mathrm{E}[\bar{X}_{n_j} \mid S_{n_j} \leq q_{(1-\alpha)}]$$

Reweight by publication probability

Let

$$\mu_{\text{sig}} = \mathrm{E}[\bar{X}_{n_j} \mid S_{n_j} > q_{(1-\alpha)}]; \quad \mu_{\text{ns}} = \mathrm{E}[\bar{X}_{n_j} \mid S_{n_j} \leq q_{(1-\alpha)}]$$
$$\sigma^2_{\text{sig}} = \mathrm{Var}[\bar{X}_{n_j} \mid S_{n_j} > q_{(1-\alpha)}]; \quad \sigma^2_{\text{ns}} = \mathrm{Var}[\bar{X}_{n_j} \mid S_{n_j} \leq q_{(1-\alpha)}].$$

Reweight by publication probability

Let

$$\mu_{\text{sig}} = \mathrm{E}[\bar{X}_{n_j} \mid S_{n_j} > q_{(1-\alpha)}]; \quad \mu_{\text{ns}} = \mathrm{E}[\bar{X}_{n_j} \mid S_{n_j} \leq q_{(1-\alpha)}]$$
$$\sigma^2_{\text{sig}} = \mathrm{Var}[\bar{X}_{n_j} \mid S_{n_j} > q_{(1-\alpha)}]; \quad \sigma^2_{\text{ns}} = \mathrm{Var}[\bar{X}_{n_j} \mid S_{n_j} \leq q_{(1-\alpha)}].$$

Then

$$\bar{X}_{n_j} \mid S_n \sim \begin{cases} \mathcal{N}\left(\mu_{\text{sig}}\,\mathrm{ppr}_j, \sigma^2_{\text{sig}} \frac{\mathrm{ppr}_j^2}{n_j}\right), & \text{if } S_n > q_{(1-\alpha)}; \\ \mathcal{N}\left(\mu_{\text{ns}}\,\mathrm{ppr}_j, \sigma^2_{\text{ns}} \frac{\mathrm{ppr}_j^2}{n_j}\right), & \text{if } S_n \leq q_{(1-\alpha)}. \end{cases}$$

Reweight by publication probability

The biased weight $w_j^* = w_j \, \mathrm{ppr}_j$ yields biased estimator

$$\bar{X}_{\mathrm{N}} = \frac{\sum_{j=1}^{k} \bar{X}_{n_j} w_j^*}{\sum_{j=1}^{k} w_j^*} \sim \mathcal{N}(\mu^*, \sigma^*)$$

Reweight by publication probability

The biased weight $w_j^* = w_j \, \mathrm{ppr}_j$ yields biased estimator

$$\bar{X}_\mathrm{N} = \frac{\sum_{j=1}^{k} \bar{X}_{n_j} w_j^*}{\sum_{j=1}^{k} w_j^*} \sim \mathcal{N}(\mu^*, \sigma^*)$$

with

$$\mu^* = \frac{\sum_{j=1}^{s} w_j^* \mu_\mathrm{sig} + \sum_{j=s+1}^{k} w_j^* \mu_\mathrm{ns}}{\sum_{j=1}^{k} w_j^*}$$

Reweight by publication probability

The biased weight $w_j^* = w_j \, \mathrm{ppr}_j$ yields biased estimator

$$\bar{X}_{\mathrm{N}} = \frac{\sum_{j=1}^{k} \bar{X}_{n_j} w_j^*}{\sum_{j=1}^{k} w_j^*} \sim \mathcal{N}(\mu^*, \sigma^*)$$

with

$$\mu^* = \frac{\sum_{j=1}^{s} w_j^* \mu_{\mathrm{sig}} + \sum_{j=s+1}^{k} w_j^* \mu_{\mathrm{ns}}}{\sum_{j=1}^{k} w_j^*}$$

$$\sigma^* = \frac{\sum_{j=1}^{s} (w_j^*)^2 \sigma_{\mathrm{sig}}/n_j + \sum_{j=s+1}^{k} (w_j^*)^2 \sigma_{\mathrm{ns}}/n_j}{(\sum_{j=1}^{k} w_j^*)^2}$$

Reweight by publication probability

Reweight each biased weight w_j^* by inverse of the publication probability to get unbiased estimator:

$$\bar{X}_{\mathrm{N}}^* = \frac{\sum_{j=1}^{k} \bar{X}_{n_j} w_j^* \,/\, \mathrm{ppr}_j}{\sum_{j=1}^{k} w_j^* \,/\, \mathrm{ppr}_j} \sim \mathcal{N}\left(\mu, \sigma^2/N\right)$$

Reweight by publication probability

Reweight each biased weight w_j^* by inverse of the publication probability to get unbiased estimator:

$$\bar{X}_{\mathrm{N}}^* = \frac{\sum_{j=1}^{k} \bar{X}_{n_j} w_j^* \,/\, \mathrm{ppr}_j}{\sum_{j=1}^{k} w_j^* \,/\, \mathrm{ppr}_j} \sim \mathcal{N}\left(\mu, \sigma^2/N\right)$$

Inspired by Hansen and Hurwitz (1943).

Reweight by publication probability

	k	μ_1	$\bar{x}_N$	$\bar{x}_N^*$
full sample (no bias, A)	200	0	-0.02	-0.07
	200	0.3	0.27	0.15

Reweight by publication probability

	k	μ_1	$\bar{x}_N$	$\bar{x}_N^*$
full sample (no bias, A)	200	0	-0.02	-0.07
	200	0.3	0.27	0.15
significant studies (B)	13	0	0.69	0.69
	38	0.3	0.61	0.61

Reweight by publication probability

	k	μ_1	$\bar{x}_N$	$\bar{x}_N^*$
full sample (no bias, A)	200	0	−0.02	−0.07
	200	0.3	0.27	0.15
significant studies (B)	13	0	0.69	0.69
	38	0.3	0.61	0.61
significant studies and 10% of	32	0	0.15	−0.17
non-significant studies (C)	55	0.3	0.51	0.27

Maximise truncated likelihood

Truncated likelihood of μ given by

$$\mathcal{L}^*(\mu \mid v_{n_1}, \ldots, v_{n_k}) = \frac{\mathcal{L}(\mu \mid v_{n_1}, \ldots, v_{n_k})}{\mathrm{E}[\mathrm{ppr}(V_{n_j}, \pi) \mid \mu]} \prod_{j=1}^{k} \mathrm{ppr}(V_{n_j}, \pi).$$

Maximise truncated likelihood

Truncated likelihood of μ given by

$$\mathcal{L}^*(\mu \mid v_{n_1}, \ldots, v_{n_k}) = \frac{\mathcal{L}(\mu \mid v_{n_1}, \ldots, v_{n_k})}{\mathrm{E}[\mathrm{ppr}(V_{n_j}, \pi) \mid \mu]} \prod_{j=1}^{k} \mathrm{ppr}(V_{n_j}, \pi).$$

Can be maximised by grid search.

Maximise truncated likelihood

	k	μ_1	$\bar{x}_N$	$\hat{\mu}$	$\hat{\pi}$
full sample (no bias, A)	200	0	-0.02	-0.02	0.66
	200	0.3	0.27	0.27	1

Maximise truncated likelihood

	k	μ_1	$\bar{x}_N$	$\hat{\mu}$	$\hat{\pi}$
full sample (no bias, A)	200	0	-0.02	-0.02	0.66
	200	0.3	0.27	0.27	1
significant studies (B)	13	0	0.69	0.31	0
	38	0.3	0.61	0.31	0

Maximise truncated likelihood

	k	μ_1	$\bar{x}_N$	$\hat{\mu}$	$\hat{\pi}$
full sample (no bias, A)	200	0	-0.02	-0.02	0.66
	200	0.3	0.27	0.27	1
significant studies (B)	13	0	0.69	0.31	0
	38	0.3	0.61	0.31	0
significant studies and 10% of non-significant studies (C)	32	0	0.15	-0.07	0.05
	55	0.3	0.51	0.33	0.16

Conclusion

Conclusion

Variety of methods available—none is a silver bullet.

Conclusion

Variety of methods available—none is a silver bullet.

Reliable empirical data about origin of publication bias is key.

Conclusion

Variety of methods available—none is a silver bullet.

Reliable empirical data about origin of publication bias is key.

Pre-registration can both reduce publication bias and improve correction accuracy.

Next Steps

- Systematically evaluate performance of methods
- Extend assessment to additional methods
- Create ensemble models

Coming Full Circle

Fighting publication bias is like hunting the Woozle—all too often it forces you to go in circles

References i

Begg, Colin B. and Madhuchhanda Mazumdar (1994). "Operating Characteristics of a Rank Correlation Test for Publication Bias". In: *Biometrics* 50.4, p. 1088 (cit. on pp. 88–91).

Egger, M. et al. (1997). "Bias in meta-analysis detected by a simple, graphical test". In: *BMJ* 315.7109, pp. 629–634 (cit. on pp. 95–99).

Gerber, Alan and Neil Malhotra (2006). "Can political science literatures be believed? A study of publication bias in the APSR and the AJPS". In: *Annual Meeting of the Midwest Political Science Association*. CiteseerX (cit. on pp. 49–54).

Hansen, Morris H. and William N. Hurwitz (1943). "On the Theory of Sampling from Finite Populations". In: *The Annals of Mathematical Statistics* 14.4, pp. 333–362 (cit. on pp. 68, 69).

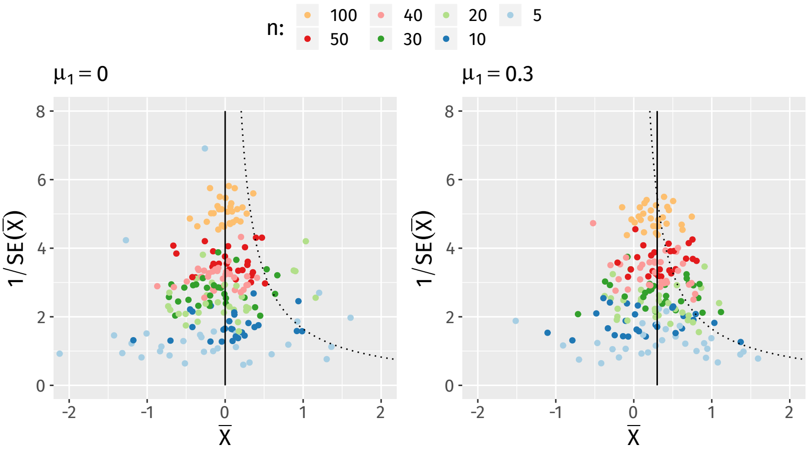
Ioannidis, J. P. and T. A Trikalinos (2007). "An exploratory test for an excess of significant findings". In: *Clinical Trials* 4.3, pp. 245–253 (cit. on pp. 41–43).

Kulinskaya, Elena, Stephan Morgenthaler, and Robert G. Staudte (2008). *Meta analysis: a guide to calibrating and combining statistical evidence*. Wiley series in probability and statistics. OCLC: 603590364. Chichester: Wiley. 260 pp. (cit. on pp. 8–13, 20, 21).

Light, Richard J. and David B. Pillemer (1984). *Summing up: The Science of Reviewing Research*. Harvard University Press (cit. on p. 27).

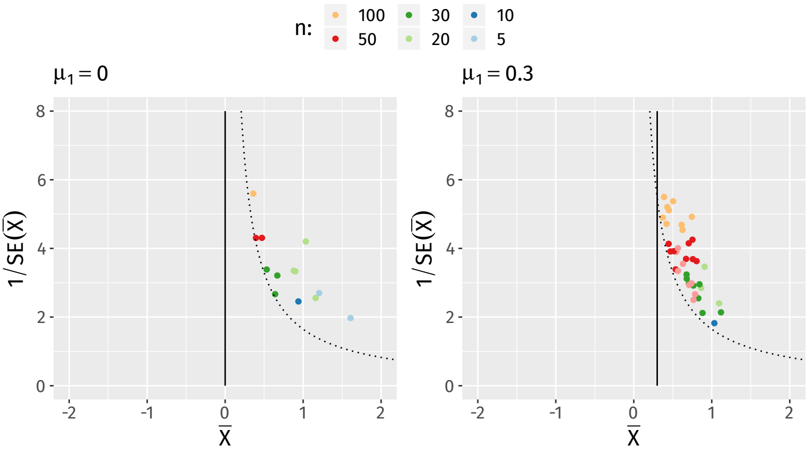
Rosenthal, Robert (1979). “The ”File Drawer Problem” and Tolerance for Null Results”. In: *Psychological Bulletin* 86.3, pp. 638–641 (cit. on pp. 32–34).

Quantify correlation between effect size and standard error



No correlation

Quantify correlation between effect size and standard error



Negative correlation

Rank correlation between effect size and standard error

Non-parametric test proposed by Begg and Mazumdar (1994):

Rank correlation between effect size and standard error

Non-parametric test proposed by Begg and Mazumdar (1994):

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0, 1);$$

Rank correlation between effect size and standard error

Non-parametric test proposed by Begg and Mazumdar (1994):

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0,1); \quad v_j = 1/w_j - 1/\sum_{i=1}^{k} w_i.$$

Rank correlation between effect size and standard error

Non-parametric test proposed by Begg and Mazumdar (1994):

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0,1); \quad v_j = 1/w_j - 1/\sum_{i=1}^{k} w_i.$$

$$Z_k = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \text{sgn}(Z_i - Z_j)\text{sgn}(v_i - v_j)}{\sqrt{(2k+5)k(k-1)/18}} \;\dot{\sim}\; \mathcal{N}(0,1).$$

Rank correlation between effect size and standard error

	k	μ_1	$\bar{x}_N$	$\|\tau/\sigma_\tau\|$	bias detected?
full sample (no bias, A)	200	0	-0.02	1.33	no
	200	0.3	0.27	0.51	no

Critical value: 1.96 (97.5%-quantile of $\mathcal{N}(0,1)$-distribution)

Rank correlation between effect size and standard error

	k	μ_1	$\bar{x}_N$	$\lvert\tau/\sigma_\tau\rvert$	bias detected?
full sample (no bias, A)	200	0	-0.02	1.33	no
	200	0.3	0.27	0.51	no
significant studies (B)	13	0	0.69	2.81	yes
	38	0.3	0.61	5.70	yes

Critical value: 1.96 (97.5%-quantile of $\mathcal{N}(0,1)$-distribution)

Rank correlation between effect size and standard error

	k	μ_1	$\bar{x}_N$	$\lvert\tau/\sigma_\tau\rvert$	bias detected?
full sample (no bias, A)	200	0	-0.02	1.33	no
	200	0.3	0.27	0.51	no
significant studies (B)	13	0	0.69	2.81	yes
	38	0.3	0.61	5.70	yes
significant studies and 10% of non-significant studies (C)	32	0	0.15	0.36	no
	55	0.3	0.51	1.60	no

Critical value: 1.96 (97.5%-quantile of $\mathcal{N}(0, 1)$-distribution)

Regress effect size against standard error

Regression-based test proposed by Egger et al. (1997)

Regress effect size against standard error

Regression-based test proposed by Egger et al. (1997)

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0, 1);$$

Regress effect size against standard error

Regression-based test proposed by Egger et al. (1997)

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0,1); \quad v_j = 1/w_j - 1/\sum_{i=1}^{k} w_i;$$

Regress effect size against standard error

Regression-based test proposed by Egger et al. (1997)

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0,1); \quad v_j = 1/w_j - 1/\sum_{i=1}^{k} w_i; \quad Z_j \sim \beta_0 + \beta_1/\sqrt{v_j}.$$

Regress effect size against standard error

Regression-based test proposed by Egger et al. (1997)

$$Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}} \sim \mathcal{N}(0,1); \quad v_j = 1/w_j - 1/\sum_{i=1}^{k} w_i; \quad Z_j \sim \beta_0 + \beta_1/\sqrt{v_j}.$$

$$T_k = \frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}} = \frac{\hat{\beta}_0}{s_{\hat{\beta}_0}} \sim t(\nu = k - 2)$$

Rank correlation between effect size and standard error

	k	μ_1	$\bar{x}_N$	$\lvert t \rvert$	$q_{(1-\alpha/2)}$	bias detected?
full sample (no bias, A)	200	0	-0.02	1.30	1.97	no
	200	0.3	0.27	0.95	1.97	no

Critical value: 97.5%-quantile of $t(\nu)$-distribution

Rank correlation between effect size and standard error

	k	μ_1	$\bar{x}_N$	$\lvert t \rvert$	$q_{(1-\alpha/2)}$	bias detected?
full sample (no bias, A)	200	0	-0.02	1.30	1.97	no
	200	0.3	0.27	0.95	1.97	no
significant studies (B)	13	0	0.69	3.77	2.20	yes
	38	0.3	0.61	8.54	2.03	yes

Critical value: 97.5%-quantile of $t(\nu)$-distribution

Rank correlation between effect size and standard error

	k	μ_1	$\bar{x}_N$	$\lvert t\rvert$	$q_{(1-\alpha/2)}$	bias detected?
full sample (no bias, A)	200	0	-0.02	1.30	1.97	no
	200	0.3	0.27	0.95	1.97	no
significant studies (B)	13	0	0.69	3.77	2.20	yes
	38	0.3	0.61	8.54	2.03	yes
significant studies and 10% of non-significant studies (C)	32	0	0.15	1.02	2.04	no
	55	0.3	0.51	4.05	2.01	yes

Critical value: 97.5%-quantile of $t(\nu)$-distribution

Maximise truncated likelihood

Grid search:

1. Define a set of candidate values $\{\mu_1, \ldots, \mu_m\}$ and $\{\pi_1, \ldots, \pi_n\}$ for μ and π, respectively.

Maximise truncated likelihood

Grid search:

1. Define a set of candidate values $\{\mu_1, \ldots, \mu_m\}$ and $\{\pi_1, \ldots, \pi_n\}$ for μ and π, respectively.
2. For each combination of μ and π, calculate the likelihood.

Maximise truncated likelihood

Grid search:

1. Define a set of candidate values $\{\mu_1, \ldots, \mu_m\}$ and $\{\pi_1, \ldots, \pi_n\}$ for μ and π, respectively.
2. For each combination of μ and π, calculate the likelihood.
3. Choose the candidate value for μ and π that yields the highest likelihood.

Publication bias—the bane of scientific publishing

Publication bias in a nutshell (Image: Hilda Bastian)

Trim-and-fill: Closing gaps in funnel plots