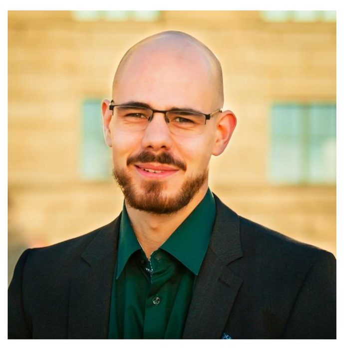
Weight of Statistical Evidence: Detection and Correction of Publication Bias



Servan Grüninger[1], Supervisor: Prof. Dr. Stephan Morgenthaler[2]

[1] Master's Programme in Computational Science and Engineering, EPFL, servan.grueninger@gmail.com
[2] Chair of Applied Statistics, EPFL

Testing Evidence in Single Studies

Creating a Test Statistic

Many scientific endeavours come down to estimating an effect size μ associated with a certain statistical population. For example, let $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ and σ^2 unknown and assume that we want to know whether μ is larger than some reference value μ_0. The corresponding hypotheses can be stated as

$$H_0\colon \mu \le \mu_0$$

$$H_1\colon \mu > \mu_0$$

To test these hypotheses, we can use the well-known Student's t-statistic

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

with $\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat{\sigma}^2 = s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ and n denoting the sample size.

Making Evidence Measures Comparable

Ideally, a test statistic is constructed so that comparisons between multiple studies are easily possible. This is achieved by constructing a new statistic V_n which fulfils the following properties:

E_1: V_n is monotonically increasing;

E_2: $V_n \sim \mathcal{N}(\tau, 1)$;

E_3: $\mathrm{Var}[V_n] = 1$;

E_4: $E_\mu[V_n] = \tau(\mu)$ is monotonically increasing in μ from $\tau(0) = 0$.

If n is large enough, T_n fulfils all these properties by virtue of the central limit theorem. For small n, however, the t-statistic follows a Student's t-distribution.

We can transform T_n into a test statistic that fulfils properties E_1 to E_4 by applying the following transformation

$$V_n = h_n(T_n) = \sqrt{2n}\,\sinh^{-1}(T_n/\sqrt{2n}).$$

Figure 1 (left column) shows the improved coverage probabilities of confidence intervals (A), normal approximation (B) and variance stabilisation (C) of V_n compared to T_n. An additional improvement is achieved by applying the following finite sample correction (Figure 1, right column):

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

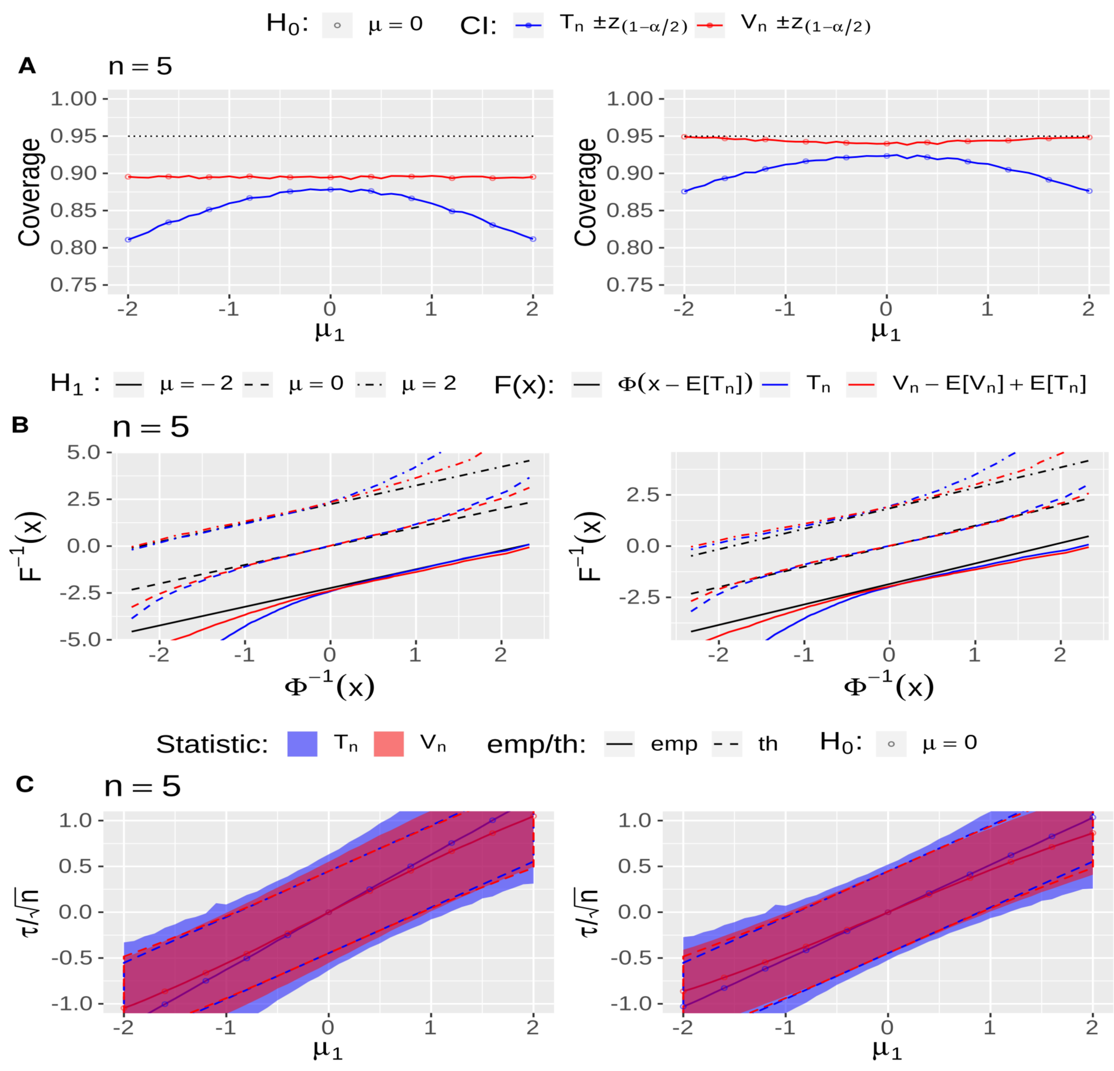


Figure 1: Empirical coverage probabilities of confidence intervals (A), quantile-quantile plots (B) as well as mean and variance of evidence measures (C) with (left column) and without (right column) finite sample correction. Solid lines indicate empirical values, dashed lines denote theoretical values.

Significance: The Fickle Gatekeeper of Scientific Publishing

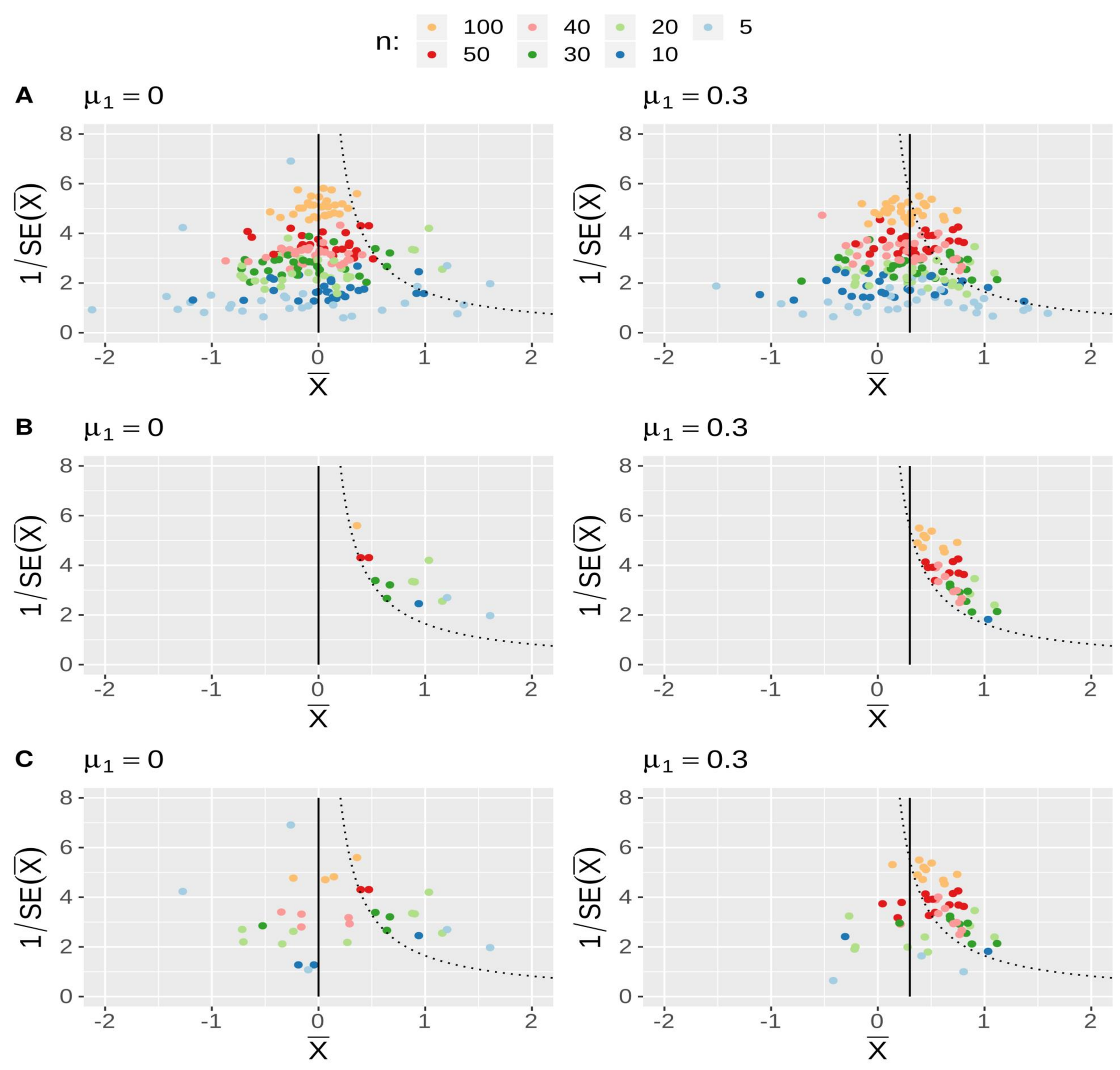


Figure 2: Funnel plots of 200 randomly simulated studies with sample sizes $n \in [5, 10, 20, 30, 40, 50, 100]$. Row A shows the complete sample, row B shows only significant studies and row C shows significant studies plus 10% of non-significant studies. The solid black line on the vertical axis denotes the true population mean whereas the dotted black line denotes the significance threshold at different levels of precision.

Finding Global Estimates

Suppose we want to estimate μ by aggregating the sample mean $\hat{\mu}_j = \bar{X}_{n_j} \sim \mathcal{N}(\mu, \sigma^2/n_j)$ across a set of k studies with n_j denoting the sample size of the jth study.

As Figure 2A shows, studies with smaller sample sizes tend to have large standard errors. Hence, it is sensible to weight each study by the inverse of the variance of each sample mean ($w_j = n/s_j^2$) to calculate the global estimate

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

Very often, however, scientific studies are more likely to be published if they report significant findings. Non-significant studies might not published at all (Figure 2B) or are only published along with a few significant results (Figure 2C). In both cases, the estimator given above overestimates the true parameter value μ.

Publication Probabilities and Truncated Distributions

Without publication bias, the distribution of published findings is given by the sampling distribution of the corresponding summary statistic, for example the sample mean $S_{n_j} = \bar{X}_{n_j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_i$. In these cases, the global estimate $\bar{X}_N$ is unbiased.

In the presence of publication bias, however, non-significant results have a lower publication probability so that the global estimate is biased upwards. Let

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

be the publication probability of S_{n_j} with π the probability of a non-significant finding to be published and δ a decision function which returns 1 if S_{n_j} is significant and 0 otherwise. If π is constant, the expected publication probability is given by

$$\mathrm{E}\left[\mathrm{ppr}(S_{n_j}, \pi)\right] = \pi \Pr(\mathrm{ppr} = \pi) + \Pr(\mathrm{ppr} = 1).$$

Finally, the truncated probability distribution of S_{n_j} in the presence of publication bias is

$$f^*_{S_{n_j}}(s_{n_j}) = \frac{\mathrm{ppr}(s_{n_j}, \pi)}{\mathrm{E}\left[\mathrm{ppr}(S_{n_j}, \pi)\right]} f_{S_{n_j}}(s_{n_j}).$$

Detecting and Correcting Publication Bias

Detecting Publication Bias

A variety of methods can detect publication bias:

1. Calculate the number of potentially omitted studies o assuming the null hypothesis is true
$$o = \frac{-2k\bar{z}_k\bar{z}_o + z^2_{(1-\alpha)} - z_{(1-\alpha)}\sqrt{4k\bar{z}_o^2 - 4k\bar{z}_k\bar{z}_o + z^2_{(1-\alpha)}}}{2\bar{z}_o^2}$$
with $\bar{z}_o = \mathrm{E}\left[Z \mid Z < z_{(1-\alpha)}\right]$ and $\bar{z}_k = \frac{1}{k}\sum_{j=1}^k z_j$.
2. Calculate the expected number of significant studies $E = \sum_{j=1}^k (1 - \beta_j)$ and compare them with the observed number of studies O using a χ^2-test-statistic
$$A = \left[(O-E)^2/E + (O-E)^2/(k-E)\right] \sim \chi^2.$$
3. Test for correlation between the standardised effect size $Z_j = \frac{\bar{X}_{n_j} - \bar{X}_N}{\sqrt{v_j}}$ and $\mathrm{Var}\left[\bar{X}_{n_j} - \bar{X}_N\right] = v_j$ using the rank correlation statistic
$$Z_k = \frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^k \mathrm{sgn}(Z_i - Z_j)\,\mathrm{sgn}(v_i - v_j)}{\sqrt{(2k+5)k(k-1)/18}} \sim \mathcal{N}(0,1).$$
4. Regress the standardised effect size against the corresponding standard error ($Z_j \sim \beta_0 + \beta_1/\sqrt{v_j}$) and test whether the intercept equals zero
$$T_k = \frac{\hat{\beta}_0}{s_{\hat{\beta}_0}} \sim t(\nu = k-2).$$
5. Check for excess results just below significance
$$K'' \sim \mathrm{Bin}(k', p = 0.5)$$
with k' denoting the number of studies in the interval $[c, c']$, K'' equals the number of studies in $[c, c'']$ and $\mathrm{p} = \Pr\left[Z \in [c, c'] \mid Z \in [c, c'']\right] \simeq \frac{c'-c}{c''-c}$ where $c' = z_{(1-\alpha)}$, $c = c' - e$, $c'' = c' + e$ and e small.

Correcting Publication Bias

Biased effect size estimates can be corrected by maximising the truncated likelihood along μ and π

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

Sample Estimate Correction

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

Table 1: Results for the effect size corrections and publication probability estimates based on the maximisation of the truncated likelihood. The correction was applied to the studies shown in Figure 2.

References

- Andrews and Kasy (2017). 'Identification of and correction for publication bias'.
- Gerber and Malhotra (2006). 'Can political science literatures be believed? A study of publication bias in the APSR and the AJPS'.
- Ioannidis and Trikalinos (2007). 'An exploratory test for an excess of significant findings'.
- Kulinskaya et al. (2008). Meta analysis: a guide to calibrating and combining statistical evidence.
- Rothstein et al.(2005). Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments.