**Supplementary Material**

Section 1: Journal information

**Supplementary Table 1**. shows information about the journals examined.

Section 2: Computing effect sizes from t-tests

Effect size calculation for one-sample t-test:

where *N* stands for the potential total number of participants which is df+1 for one-sample and matched t-tests.

Because, by definition:

where ‘s’ stands for the sample standard deviation.

So:

Effect size calculation for two sample t-test:

where *N* stands for the potential total number of participants which is df+2 for independent sample t-tests.

Because

Where n1 and n2 are the participant numbers in groups and:

And by definition:

Assuming that n1 = n2 = nc and letting N = n1 + n2 (ie. N = total number of participants) we get:

So:

For matched-sample t-tests:

Section 3: Power calculations

The power of t-tests can be computed from the non-central t distribution assuming an mixture of t tests (see main text for the mixture model we used).

For simpler writing we can first define the cdf for the mixture distribution where the probability of a t value depends on the proportion of one and two-sample t-tests in the data:

MixtureCdf(t,df,pr(t1),D) = pr(t1)\*nctcdf(t,df,ϴt1) + [1-pr(t1)]\*nctcdf(t,df,ϴt2)

Where

nctcdf stands for the cumulative non-central t probability density function,

t stands for a t value,

dfstands for the given degrees of freedom,

pr(t1) stands for the probability of a one-sample or matched t-test.

ϴt1 and ϴt2 are non-centrality parameters dependent on Cohen’s D (Harrison and Brady, 2004):

ϴt1 = D \* sqrt(nt1)

Where nt1 = total number of participants; i.e. nt1 = df+1.

ϴt2 = D \* ½ \* sqrt(nt2)

Where nt2 = the number of participants in one experimental group (assuming that group sizes are equal as noted above); i.e. nt2 = roundupper((df+2)/2).

Given the above notation, power for a given effect size (D), df and critical t value (t(α)) can be computed as (Harrison and Brady, 2004):

Power(t(α), df, pr(t1), D) = MixtureCdf(t(α),df,pr(t1), D) + (1 - MixtureCdf(t(α),df,pr(t1), D))

Section 4. Properties of the t-value model and extracted t-values

Based on our mixture model of one-sample and two-sample t-tests we can expect a t value distribution as depicted in **Supplementary Figure 1.** The figure illustrates themechanism of effect size exaggeration and the importance of H0:H1 odds.

In **Supplementary Figure 1A.** observe that statistically significant effects always show much larger effect sizes than non-significant effects irrespective of whether they are true positives or false positives. Consequently, a large effect size does not guarantee that results are to be trusted even if they are statistically significant. Effect size exaggeration happens because typically only statistically significant results are reported and these (by definition) must pass the statistical significance threshold. In **Supplementary Figure 1B.** observe that if H0:H1 odds are high (that is, there are many more true negative than true positive hypotheses to test) then a very high proportion of t values will come from true H0 situations. That is, the statistically significant part of the t-value distribution will include a very large number of 'false positive t values' (the 'false positive tail' of the true H0 distribution). If H0:H1 odds are high enough these false positive t-values will outnumber the statistically significant t values coming from the true H1 distribution.

The extracted t-value distribution is represented in **Supplementary Figure 2**.

Section 5: False and True Report probability

If we use nil-null Hypothesis Significance Testing (NHST) then the long run False Report Probability (FRP) is the long run probability that the null hypothesis (H0) is true when we get a statistically significant finding. The long run True Report Probability (TRP) is the long run probability that the alternative hypothesis (H1) is true when we get a statistically significant finding. Computationally, FRP is the number of statistically significant false positive findings divided by the total number of statistically significant findings. TRP is the number of statistically significant truly positive findings divided by the total number of statistically significant findings.

FRP and TRP can be computed by Bayes' theorem. If we take 'sig' to stand for 'statistically significant test outcome' then the total probability of finding a statistically significant result is:

Hence, FRP and TRP can be written as:

Considering a long run of NHST studies, the long run probability of having a significant test outcome when H0 is true is α and the long run probability of having a significant test outcome when H1 is true is Power = 1 - β. That is, and Power . Hence, **Eq.16.** and **Eq.17.** can be re-written as:

**Eq.16.** and **Eq.17.** can also be expressed in terms of pre-study odds ratios of true H0 and true H1 situations. For example, we can denote the odds of true H0 situations as ‘O’ and write:

We can express pr(H0) using the above as:

then **Eq.16.** Can be rewritten as:

This can be simplified by :

Using α and Power = 1 - β we can write:

+ = 1; e.g.:

Consequently:

and

Equivalently to the above, we can also express the odds of true H1 situations as the ratio of pr(H1) and pr(H0) and denote it with ‘R’ as in Ioannidis (2005):

Substituting **Eq.25.** into **Eq.16.**:

Simplifying:

Using α and Power = 1 - β we can write:

**Eq.28.** is equivalent to the one used by Ioannidis (2005) with a slightly different notation.

He defined PPV = ; Power = 1-β and equivalently to Eq. 28 he wrote:

Ioannidis (2005) also defined Bias, signified by ‘*u*’. On the one hand, bias results in categorizing fraction u (u\*(1 - α)) of otherwise true negative results (in case there is no bias) as positive results. On the other hand, bias results in categorizing fraction *u* (uβ = u\*(1 - Power)) of otherwise (in case there is no bias) missed true positive results as positive results. That is, bias alters **Eq.28.** as:

Using the notation of Ioannidis (2005) this can be rewritten as:

Also, notice the relation between and :

Hence,

Similarly:

Hence,

**References**

Harrison D.,A., Brady A.,R., Sample size and power calculations using the noncentral t-distribution. *The Stata journal*. 4, 142-153 (2004).

Fritz, C.,O., Morris, P.,E., Richler, J.,J, Effect size estimates: Current use, calculations and interpretation. *Journal of Experimental Psychology: General.* 141, 2-18 (2012).

Ioannidis, J.P.A. Why most published research findings are false. *PLoS Medicine*. *2,* e124. (2005).

Hunter, J.E & Schmidt F.,L. Methods of meta-analysis: Correcting error and bias in research findings. Newbury park. Sage. (1990).