# Studies (tests) assessing the efficacy of treatments against the alcohol deprivation effect (ADE) are underpowered

François Léonard[1], Ezio Tirelli[1]

[1] Department of psychology, ULiège

## Introduction

Many psychopharmacological and psychological treatments seemingly effective in preclinical models have been found to be ineffective in clinical trials (Ray et al., 2018). This may be partly due to a lack of prospective power in the preclinical literature. Underpowered studies are more likely to yield false negatives, false positives and inflated effect sizes (or Small-Study Effect, SSE).

We examined the prospective power, the observed effect sizes and the False Discovery Rate (FDR) (Szucs & Ioannidis, 2017) in articles assessing the efficacy of treatments against ADE in rodents.

## Methods

A literature search on PubMed yielded 154 titles that were also used in a distinct project dealing with the reporting quality of this literature. Twenty of them were excluded due to insufficient statistical information. The 48 remaining studies contained 93 between-group statistical tests (73 F tests and 20 t tests).

Effect sizes (Hedges' g) and related standard errors were extracted and synthesized with a Random Effect Meta-Analysis (REMA). Between-study heterogeneity was estimated with the restricted maximum likelihood method. We used the Knapp-Hartung adjustment to calculate the confidence interval of the Pooled Effect Size (PES). The PES obtained with the REMA was corrected with the Trim-and-Fill (Duval & Tweedie, 2000), the PET-PEESE (Stanley & Doucouliagos, 2014) and the limit meta-analysis (Rücker et al., 2011) methods.

We used these four PESs as hypothetical effect sizes to calculate the prospective power of individual studies. These prospective powers were then used to generate FDR curves (Szucs & Ioannidis, 2017).

## Results

The PES obtained with the REMA was statistically significant, whereas its values corrected for the SSE were not significant. They ranged from a very small negative effect (PET-PEESE) to an almost large effect (Trim-and-Fill), none being statistically significant (Table 1).

Table 1: Pooled effect sizes

| Method | Effect size | 95% CI lower | 95% CI upper |
|---|---|---|---|
| REMA | 1.51 | 0.47 | 2.55 |
| Trim-and-Fill | 0.77 | -0.35 | 1.90 |
| PET-PEESE | -0.05 | -2.99 | 2.89 |
| Limit | 0.36 | -0.24 | 0.96 |

Median statistical powers computed from these four PESs ranged from 0.11 to 0.88 (Table 2 and Figure 1).

Table 2: Median prospective powers derived from the pooled effect sizes

| Effect size | Median | Q1 | Q3 |
|---|---|---|---|
| REMA (g = 1.51) | 0.88 | 0.80 | 0.94 |
| Trim-and-fill (g = 0.77) | 0.34 | 0.29 | 0.41 |
| PET-PEESE (g = -0.05) | 0.05 | 0.05 | 0.05 |
| Limit (g = 0.36) | 0.11 | 0.10 | 0.12 |

The median sample size (n = 8) calculated from the 93 tests telled us that the smallest effect size detectable with a statistical power of 0.80 for a two-sided two-sample t test was g = 1.5.

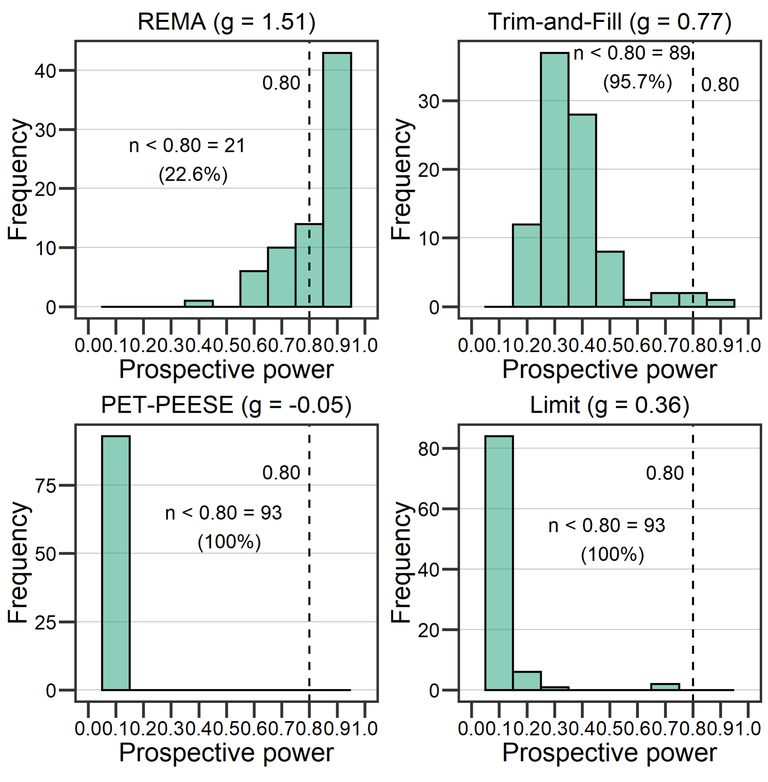


Figure 1: Histograms of statistical powers for the four effect sizes. The dashed line corresponds to the recommended minimal level of power (80%). Top left panel: prospective powers using the PES found with REMA (g = 1.51). Top right panel: prospective powers using the PES corrected with the Trim-and-Fill method (g = 0.77). Bottom left panel: prospective powers using the PES corrected with the PET-PEESE method (g = -0.05). Bottom right panel: prospective powers using the PES corrected with the limit method (g = 0.36).

The FDRs computed with our four PESs and a $H_1$ prior probability (plausibility) of 0.1 ranged from 0.339 to 0.898. With a $H_1$ prior probability of 0.5 the FDRs ranged from 0.054 to 0.495 (Figure 2)

## Discussion

The REMA yielded a promising and statistically significant effect of the treatments against the ADE (g = 1.51). When the Trim-and-Fill method was used, this PES was merely divided by two (g = 0.77) and became statistically non-significant. The PES obtained with the PET-PEESE and limit methods were even smaller. The use of the methods correcting for a SSE pointed out that effect sizes of treatments against ADE are probably smaller than those reported in most of research papers.

After correcting for the SSE at least 95.7% of the tests assessing the efficacy of the treatments against ADE appeared to be underpowered (1-bêta< 80%). This is a problem because low power increases the risk of inflated effect sizes.

The PET-PEESE method gives probably an underestimated PES, which is much smaller than the other estimated PESs. This likely results from a high between-study heterogeneity and a large number of small sample sizes.

One possibility to diminish the waste in effort and animals in the search of an effecatious ADE counteracting treatment is to conduct high-powered studies to achieve a clinically useful effect size.

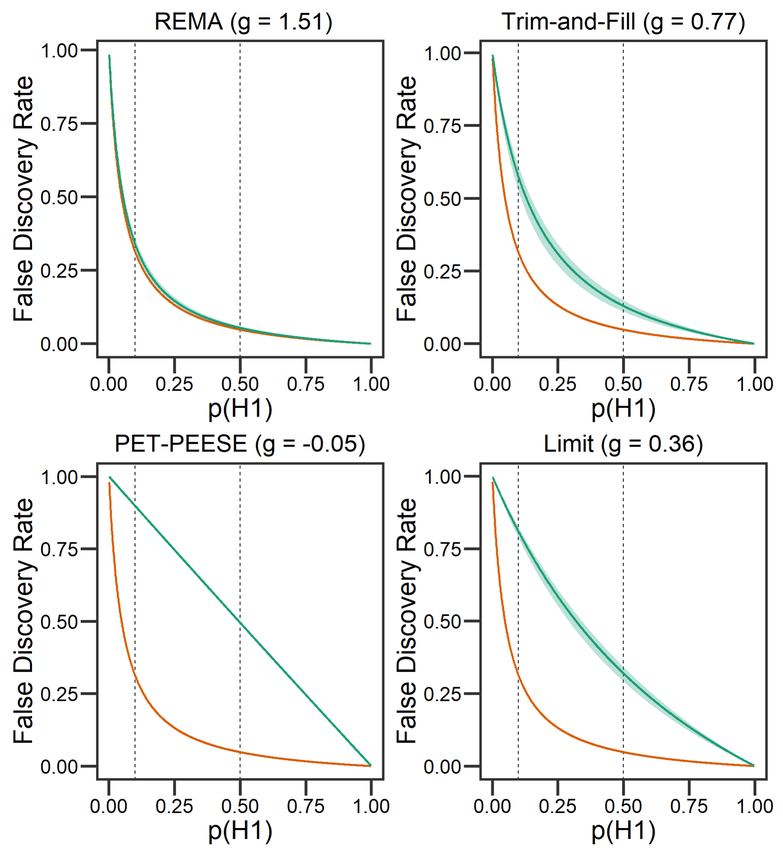


Figure 2: Relationship between the False Discovery Rate (FDR) and the prior probability of H1 (green area indicates the interquartile range). The orange curves concen FDR values derived from a statistical power of 0.99.Top left panel: FDR curve derived from the median prospective power using the PES of the REMA. Top right panel: FDR curve derived from the median prospective power calculated with the PES of the Trim-and-Fill method. Bottom left panel: FDR curve derived from the median prospective power calculated with the PES of the PET-PEESE method. Bottom right panel: FDR curve derived from the median prospective power calculated with the PES of the limit method.

## References

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-PlotBased Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Ray, L. A., Bujarski, S., Roche, D. J. O., & Magill, M. (2018). Overcoming the "Valley of Death" in Medications Development for Alcohol Use Disorder. *Alcoholism: Clinical and Experimental Research*, *42*(9), 1612–1622. https://doi.org/https://doi.org/10.1111/acer.13829

Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H., & Schumacher, M. (2011). Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*, *12*(1), 122–142. https://doi.org/10.1093/biostatistics/kxq046

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Szucs, D., & Ioannidis, J. P. A. (2017). When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00390