Are the studies (tests) assessing the efficacy of the treatments of the alcohol-deprivation effect underpowered?

Léonard F.[1], Tirelli E.[1]

[1] University of Liège

Purpose of this protocol

While adequate power and sample size are indispensable for the detection (by the statistical test) of a meaningful effect size (ES), many published studies in psychology (practically all sub-fields) do not describe sample-size calculation (SSC), a practice that can reflect a real weakness of the study methodological quality (Macleod et al., 2015).

Omission of SSC is often associated with a lack of prospective power, which can be computed from (several) theoretical ES and the published sample size. This often results in a exaggeration of the observed ES in comparison with the true ES (Gelman & Carlin, 2014; Szucs & Ioannidis, 2017) and an increase of the False Report Probability (FRP) (Szucs & Ioannidis, 2017), thereby jeopardizing results reproducibility. FPR refers to the probability that statistically significant effects are true and can be computed incorporating the pre-test probability that H1 is true (H1 plausibility or H1/H0 odds). Simulations of effect size exaggeration and FRP are presented respectively in the left-hand panel and in right-hand panel of Figure 1.

Our purpose is to investigate using the above-mentioned tools, to which extent the practice of neglecting the determination of adequate sample size (via power analysis) and the lack of power concern the literature assessing the efficacy of the psychological and pharmacological treatments of the "alcohol-deprivation effect," a relapse-like phenomenon operationalized in animal model (rodents).

Methods

We will firstly select articles published from 1993 to 2020 using the database PubMed and check whether they mention a SSC.

We will then classify the articles which mention a SSC into five components of a complete description of SSC (for instance power analysis with or without details such as exact prospective power, hypothetical and observed ES). We will also check whether the hypothetical ES (used to determine sample size) is justified and the observed ES clearly interpreted (discussion). The results will be expressed in terms of percentages for each category.

In order to assess a possible ES exaggeration in the selected literature, we will examine the relation between the observed ES and the sample sizes derived from each article, the results being presented as in the left-hand panel of Figure 1. Thereupon, we will compute the implied power-to-detect (at a alpha risk of 5%) of each relevant statistical test, using (at least) a conventionally small, medium or large ES (Cohen classifications) and the sample size indicated in the article. The resulting values of the implied powers will be presented in a table (which will also include a category relating to the journals in which the articles have been published). The individual values of the implied powers will be inserted in the graph of right-hand panel of Figure 1 in the form of a vertical bar corresponding to the sample size of the article. The median and the mean of these values will be similarly presented in a separate graph. We will also plot individual power curves in three additional graphs (power versus sample size for each of the three external ES).

Finally, we will compute the FRP and the complementary True Report Probability (TRP) using a risk alpha of 1, 2.5 or 5%, the median (implied) power-to-detect and a representative range of pretest probability values of the alternative hypothesis (from 0.01 to 0.99). The resulting individual curves (FRP/TRP versus H1 probability for each implied power) will be plotted as in the graphs of Figure 2, and the curves corresponding to the median of FRP and TRP curves will be added to the graphs of Figure 2.

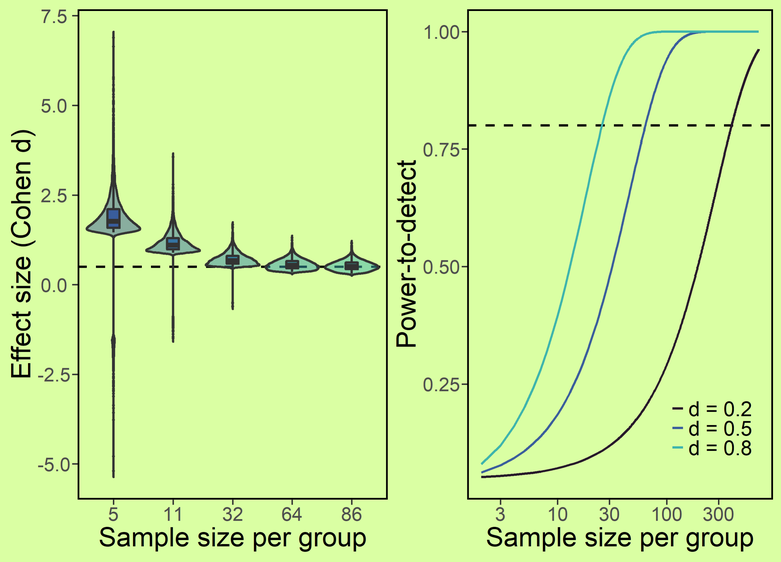

Figure 1: **Left-hand panel:** Relations between ES (Cohen's d) and the sample size derived from simulated t tests for independent groups. Each violin represents the distributions of 10 000 significant ES for five sample sizes (corresponding to 10, 20, 50, 80 and 90% prospective power). The dashed line represents the hypothetized ES (d = 0.5). With relatively small sample sizes (low power), t tests yield many significant ES greater than true ES (type M error) and some significant ES in the opposite direction of the hypothetized effect (type S error). **Right-hand panel:** Relation between the prospective power (curve) to detect the three representative ES of the Cohen 's classification (curves) and the sample size per group in two-sample t test. The dashed line represents the conventionally accepted minimal power-to-detect of 80%.

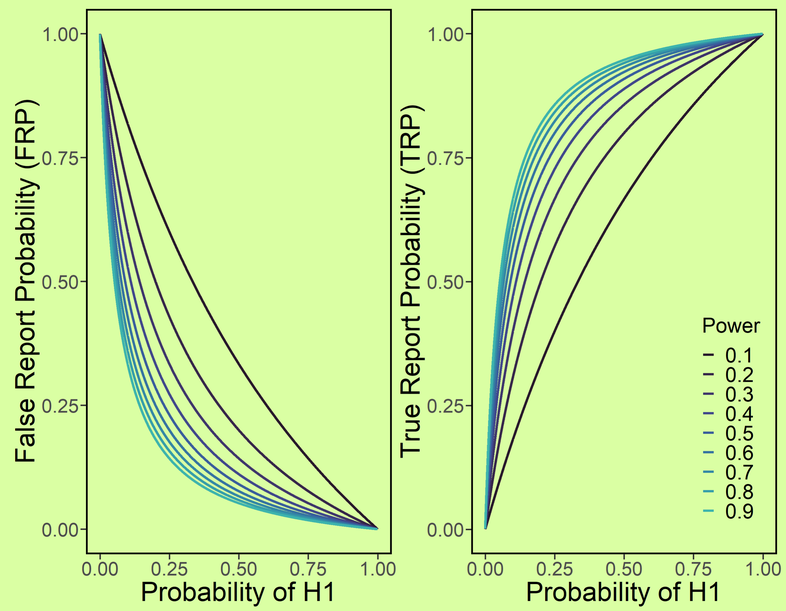

Figure 2: **Left-hand panel:** Relation between False Report Probability (FRP) and the pretest probability of H1 for various values of prospective power **Right-hand panel:** Relation between True Report Probability (TRP, complement of FRP) and the pretest probability of H1 for various values prospective power.

References

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., Hirst, T., Hemblade, R., Bahor, Z., Nunes-Fonseca, C., Potluru, A., Thomson, A., Baginskitae, J., Egan, K., Vesterinen, H., Currie, G. L., Churilov, L., Howells, D. W., & Sena, E. S. (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLOS Biology*, *13*(10), e1002273. https://doi.org/10.1371/journal.pbio.1002273

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