Study Protocol Code for – An evaluation of reproducibility and errors in statistical power calculations performed using GPower

Robert T. Thibault, Emmanuel Zavalis, Hugo Pedder

Address correspondence to robert.thibault@stanford.edu

2022-05-31

Monte Carlo sample size calculations:
Sample size for all articles: 95
Sample size for articles powering for an ANOVA: 60

Sample size calculations using the more common method with normal approximation. For comparison purposes only:
Sample size require for all articles: 97
Sample size require for articles powering for an ANOVA: 62

Example of results based on *SIMULATED DATA*

We sampled 120 articles, of which 110 included a power calculation for a study in that article and 95 of these articles performed power calculations to solve for sample size. n articles included a power calculation that solved for power and n for effect size (see Supplementary Table 1 for all counts). n (%) articles we surveyed included human participants and n (%) included non-human animals. n (%) were protocols. Sampled articles were published in 2017 (n = n), 2018 (n = n), 2019 (n = n), 2020 (n = n), 2021 (n = n), and 2022 (n = n). The median Journal Impact Factor of included articles was n (IQR n-n).

We estimate that between 35000 and 42000 articles indexed by PubMed and published since 2017 use G*Power for a sample size calculation and that between 9000 and 16000 do so for a sample size calculation for an ANOVA (see Table 1 for additional details).

To calculate the total number of articles using GPower in PubMed versus PMC, we simply multiplied the estimates by 2.24, which is the number of articles indexed in PubMed from 2017 onwards divided by the number of articles indexed in PMC from 2017 onwards. This calculation assumes that all PMC articles are indexed in PubMed. If we want to take a conservative estimate and assume that 50% fewer articles that are indexed in PubMed, but not indexed in PMC, use GPower, then we would need to multiply the PMC estimates by 1.62 (i.e., 1 + (2.24 - 1)*0.50) or the PubMed estimates by 0.72 (i.e., 1.62/2.24). For the rest of this article we will assume that the frequency of use of GPower in PMC and PubMed is the same.

Example calculation for the rightmost columns of Table 2 and Table 3 (see attached R code for details):

We estimate the total number of irreproducible power calculations using G*Power, published since 2017 and indexed in PubMed to be 33000 (95% CI: 28000-37000)

Table 1. Estimates of the number of published articles that use GPower

	Any power calculation	Sample size calculation	ANOVA sample size calculation
PubMed Central	20000 (19000 - 21000)	17000 (16000 - 19000)	5000 (4000 - 7000)
PubMed	45000 (42000 - 47000)	39000 (35000 - 42000)	12000 (9000 - 16000)

The table is divided into articles that use G*Power for: any power calculation related to any statistical test (Any power calculation), a power calculation for any statistical test that solves for sample size (Sample size calculation), and a power calculation for an ANOVA that solves for sample size (ANOVA sample size calculation). The total number of articles in each database since 2017 is: PubMed Central 3,246,604; PubMed 7,264,911.