

A survey of the quality and transparency of statistical power analyses conducted using GPower

Robert T. Thibault & Emmanuel Zavalis
(Contributor roles are detailed before the references)

Address correspondence to robert.thibault@stanford.edu

2022-04-28

Methods

Search method and selection criteria

This is a pilot study that was conducted prior to the extraction of the actual sample that we will study in our paper. Using the search strategy 'g*power' in PubMed Central we searched for the papers that used the software GPower for their power analysis to examine their use of this tool and the handling of power calculations in general. The search was conducted on April 10th 2022 and retrieved 23927 publications.

From these a random sample of 15 articles were examined in detail using random seed 1453.

The search for the sample to find the articles and was used in PubMed Central was:

'PMC7273017 OR PMC5857919 OR PMC8005969 OR PMC5495109 OR PMC5629614 OR PMC7349576 OR PMC5609352 OR PMC6062074 OR PMC6542583 OR PMC7190572 OR PMC8244525 OR PMC8815471 OR PMC8952887 OR PMC7146211 OR PMC7547415'

We assessed the papers for eligibility first with the inclusion criteria being that the paper uses GPower to perform a power calculation. We also extracted meta-data that included what journal it was from, what impact factor the journal has, as well as the publication year. Finally we looked into whether the paper was a clinical trial defined as a study that includes human subjects and studies an intervention.

Results

Search results

We sampled 15 papers of which 13 met our inclusion criteria. Eight of these performed power calculations to solve for sample size and the power calculation appears to be conducted before running the study. One solved for power and one solved for effect size (see Supplementary Table 2 for all counts).

[Supplementary Table 1]

Of the articles we surveyed # had human participants, # had non-human animals. # (%) were protocols.

Meta-data of the articles

The impact factors of the journals the studied papers was on average 4.3 (95% CI 2.4, 6.2). The papers were published from 2017 to 2022.

Estimation of amount of articles affected

We estimate that between 13045 and 160882 published since 2017 use GPower for an a priori sample size calculation and that between 27775 and 202140 since 2017 do so for any power calculation (see Supplementary Table 2 for additional details)

[Supplementary Table 2]

Reproducibility of power calculations

Articles often did not report key values necessary to reproduce power analyses, including alpha, power, effect size, sample size, and statistical test (Table 1). We could only reproduce 2 of the power calculations and 2 of them required that we make assumptions.

[Table 1]

An aspect of error types and their rate will be brought up in the study and included in Table 2.

[Table 2]

Tables

Table 1. Complete reporting and reproducibility of GPower calculations

	% Reporting
Alpha (95% CI)	92 (62, 99.6)
0.05	85
Other level of significance	8
Power (95% CI)	77 (46, 94)
80%	23
95%	46
Other level of power	10
Effect size (95% CI)	100 (72, 100)
d	30.8
f	15.4
Other	53.8
Statistical test (95% CI)	100 (72, 100)
ANOVA	0
T-test	0
Other test	100
Sample size reported (95% CI)	77 (46, 94)
Sample size median (IQR)	53 (83.5)
Any justification for their effect size (95% CI)	15 (3, 46)
Justified: previous research	15
Justified: own pilot	0
Reproducible (95% CI)	15 (3, 46)
Reproducible with assumptions	15
Reproducible w/o assumptions	0
Adjusted for multiple comparisons	11

* Note that we have used placeholder variables for the rows statistical tests, these will be replaced with actual data

† The sums of percentages may not add up since the numbers are rounded

Table 2. Error in GPower calculations

[errors in the GPower calculations will be tabulated into a table of the errors manually and inserted into the finished manuscript]

Supplementary tables

Here we have all the supplemental tables for the study compiled

Supplementary Table 1. Type and timing of power calculations

	After running the study	Before running the study	Unsure
A priori (i.e., solves for sample size)	0	8	1
Post hoc (i.e., solves for power)	1	0	1
Sensitivity (i.e., solves for effect size)	1	0	0
Unsure	0	0	1

Supplementary Table 2. Estimates of the number of published articles that use GPower.

	Point estimate	95% confidence Interval, lower bound	95% confidence Interval, upper bound
PubMed Central included	18279	12315	20597
PubMed Central a priori	11249	5784	16393
PubMed included	41227	27775	46455
PubMed, a priori	25370	13045	36973
Dimensions included	179393	120860	202140
Dimensions a priori	110395	56764	160882

* The total number of article in each database since 2017 is: PubMed Central 3,166,809; PubMed 7,142,566; dimensions.ai 31,079,708.

Supplementary Table 3. Justification of effect size

Type of justification	No. of articles
No justification	11
Previous research	2

Supplementary Table 4. Adjustment for multiple comparisons

Does it adjust for multiple comparisons?	No. of articles
No, and the article contains multiple analyses with no clear indication of a sole primary analysis for which this power calculation is for.	4
No, and there is no reason to account for multiple comparisons (e.g., there is only one analysis, or this analysis is clearly demarcated as the primary analysis).	4
Unsure	4
Yes, reasonably so (i.e., the accounting for multiple comparisons matches the analyses conducted)	1

Supplementary Table 5. The ANOVA conundrum

Does it adjust for the ANOVA conundrum?	No. of articles
No	4
Unsure what test the power calculation was for	5
Yes, but I cannot reasonably assume which option they used.	2
Yes, but the researchers use the default option without accounting for it (e.g., powering for a "medium" effect size by entering $f = 0.25$).	2

Supplementary Table 6. The match of statistical analysis to power analyses

Does the power calculation match the statistical test?	No. of articles
Unsure (there's not enough information for me to make code yes or no).	7
Yes	6