Tables for G*Power manuscript

Robert Thibault

2024-06-26

Proportion of PMC articles that use GPower for any power calculation, with 95% CIs: 0.65 (0.62 - 0.67)
Proportion of PMC articles that use GPower for a sample size calculation, with 95% CIs: 0.44 (0.38 - 0.49)
n articles in PMC per 1 that uses GPower for any calculation: 154
n articles in PMC per 1 that uses GPower for a sample size calculation: 229

Participants that partook in studies with reproducible or likely reproducible sample size calculations: n = 3083
Participants that partook in studies with non-reproducible sample size calculations: approximately n = 8386 (we assumed the sample size for 5 of these articles that did not report sample size)
Total number of participants: 1.1469×10^4
Proportion of participants that partook in a study with a reproducible or likely reproducible sample size calculation: 0.27

Cohen's d for height difference between men and women in the US (based on NHANES data Tables 9 and 11: https://www.cdc.gov/nchs/data/series/sr_03/sr03-046-508.pdf):
All people aged 20 and over: $d = 1.01$
All people aged 20-29: $d = 1.46$
Non-Hispanic, white aged 20-39: $d = 1.69$

Table 1. Types of power calculations.

Power calculation solved for:	N=141	Percent	(95% CI)
Sample size	95	67%	(59 - 75%)
Power	29	20%	(14 - 28%)
Effect size	13	9%	(5 - 15%)
Sample size (after completing the study)	5	3%	(1 - 7%)
Unsure	3	2%	(<1 - 5%)

The sum of the first column is greater than n = 141 because four articles performed multiple power calculations that solved for different variables. We only coded 'Sample size (after completing the study)' when an article unambiguously performed their calculation for a future study. Some articles used G*Power to calculate effect sizes after a study was complete.

Table 2. Reproducibility of sample size calculations performed using G*Power.

Transparency element	N=95	Percent	(95% CI)
Reproducible*			
Yes, without assumptions	2	2%	(<1 - 6%)
Likely, with assumptions	27	28%	(20 - 38%)
No	66	70%	(60 - 78%)
All 6 elements reported†			
Yes	13	13%	(8 - 21%)
No	82	87%	(79 - 92%)
Alpha			
0.05	80	84%	(76 - 91%)
Other	2	2%	(<1 - 6%)
Not reported	13	13%	(8 - 21%)
Power			
0.80	53	56%	(46 - 66%)
0.95	16	17%	(10 - 25%)
Other	18	19%	(12 - 27%)
Not reported	8	8%	(4 - 15%)
Effect size type			
d	15	16%	(9 - 24%)
f	15	16%	(9 - 24%)
Non-standardized	11	11%	(6 - 19%)
Other	9	9%	(4 - 16%)
Not reported	45	47%	(37 - 57%)
Effect size value			
Reported	76	80%	(71 - 87%)
Not reported	19	20%	(13 - 29%)
Statistical test			
ANOVA	14	14%	(8 - 23%)
t-test	12	12%	(7 - 20%)
Other	6	6%	(2 - 12%)
Not reported	63	66%	(57 - 75%)
Sample size‡			
Reported	90	95%	(89 - 98%)
Not reported	5	5%	(2 - 11%)

*We use the term 'likely' because we cannot be certain that all of our assumptions were correct. †These six elements are: alpha, power or beta, effect size type, effect size value, statistical test, and sample size. We considered statistical test reported if they named the general test, even if details were missing (e.g., reporting an ANOVA, but not what type of ANOVA). ‡The median sample size was 55 (IQR: 28 to 116).

Table 3. Article characteristics

Article characteristic	N=95	Percent
Unit of study		
Human	85	89%
Non-human animal	10	11%
Year of publication		
2017	4	4%
2018	8	8%
2019	20	21%
2020	29	31%
2021	22	23%
2022	12	13%
Protocol article		
Yes	8	8%
No	87	92%
Multiple sample size calculations		
Yes	5	5%
No	90	95%
Publisher		
BMC	18	19%
MDPI	14	15%
Frontiers Media SA	7	7%
PLOS	6	6%
Springer	5	5%
Wiley	5	5%
Nature Portfolio	3	3%
Other	37	39%

The median journal impact factor was 2.7 (range: 1.4 to 11). 18 articles were pubished in journals that did not have an impact factor. Year 2023 contains fewer articles than the preceeding years because we only sampled until May 31, rather than the entire year.

Table 4. Quality of the sample size calculations performed using G*Power

Quality measure	N=95	Percent	(95% CI)
Analysis match in results section			
Yes	29	30%	(22 - 40%)
No	13	13%	(8 - 21%)
Unsure	46	48%	(38 - 58%)
NA (protocol)	7	7%	(3 - 13%)
Error			
Yes	10	10%	(5 - 17%)
No	14	14%	(8 - 23%)
Unsure	71	75%	(66 - 83%)
Adjusted for multiple comparisons			
Yes	0	0%	(NA)
No, and multiple analyses are performed	71	75%	(66 - 83%)
No, but a single outcome was identified	22	23%	(15 - 32%)
Unsure	2	2%	(<1 - 6%)
Justification for chosen effect size*			
Previously published research	25	26%	(18 - 36%)
Effect size conventions	25	26%	(18 - 36%)
General reference to another study	9	9%	(4 - 16%)
Pilot data	5	5%	(2 - 10%)
Effect size of interest	4	4%	(1 - 9%)
Other	3	3%	(1 - 8%)
No justification reported	31	33%	(24 - 42%)

*Some articles provided more than one justification for their chosen sample size, and thus the sum of the percentages is greater than 100%.

Table 5. Selection of default ANOVA option

Option used	N=36	Percent	(95% CI)
Non-default option	3	8%	(2 - 19%)
Default option	18	50%	(34 - 66%)
Unsure	15	42%	(26 - 58%)

Supplementary Table 1. Estimated number of articles published between 1 Jan 2017 and 31 May 2022 that reference G*Power

	PubMed Central (95% CI)	PubMed (95% CI)
Any power calculation	21000 (20000 - 22000)	48000 (46000 - 49000)
Sample size calculation	14000 (13000 - 16000)	32000 (28000 - 36000)
ANOVA sample size calculation	3000 (2000 - 5000)	8000 (5000 - 11000)

We excluded articles that discuss G*Power, but do not report using this software for a power calculation. The table includes rows for publications that report using 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 from 1 Jan 2017 to 31 May 2022 is: PubMed Central 3,285,893; PubMed 7,318,980. Numbers are rounded to the nearest thousand to avoid suggesting a higher level of precision than our method of estimation can provide.

Supplementary Table 2. Inter-rater agreement

Variable	Cohen's k	Percent agreed	n disagreed	n total	n categories
id	1.00	100%	0	147	147
pmcid	1.00	100%	0	147	147
protocol	0.65	97%	5	147	2
include	0.72	98%	3	147	2
participants	0.82	96%	5	141	2
journal	0.73	74%	37	141	141
pub_year	0.95	96%	6	141	7
impact_factor	0.82	83%	24	141	112
power_calc_type	0.68	84%	23	141	10
multiple	0.77	89%	16	141	3
version	0.95	98%	2	91	3
version_text	0.97	98%	2	93	13
power	0.92	99%	1	91	3
power_text	0.89	94%	5	84	13
alpha	0.84	97%	3	91	3
alpha_text	1.00	100%	0	79	4
sample_size	0.21	92%	7	91	3
sample_size_text	0.87	88%	10	81	70
effect_size_type	0.82	87%	12	93	10
effect_size_value	0.80	93%	6	91	3
effect_size_value_text	0.95	95%	3	64	35
stat_test	0.66	83%	16	93	7
reproducible	0.22	69%	29	93	4
justification	0.46	59%	38	93	15
just_previous	0.38	81%	18	93	2
just_pilot	0.48	96%	4	93	2
just_convention	0.73	89%	10	93	2
just_mcid	0.39	97%	3	93	2
just_none	0.55	78%	20	93	2
just_ref	0.21	88%	11	93	2
just_other	-0.02	96%	4	93	2
mult_compare	0.23	65%	33	93	4
anova_within_between	0.51	67%	31	93	6
match	0.46	66%	32	93	5
error	0.24	70%	28	93	4

The variables are listed as they appear in the open data. See the data dictionary for a description of each variable. Cohen's kappa is mostly irrelevant for variables with a large number of categories, and can be ignored. Not all variables were relevant for all articles we coded; thus, 'n total' differs among the variables. 'justification' was coded as a multiple selection question with 7 options. We re-coded this variable into 7 binary variables and calculate the inter-rater agreement for each one.