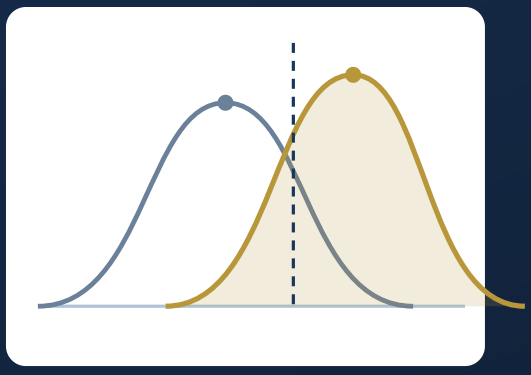
Research Gold

RESEARCH GOLD FIELD GUIDE

Sample Size and Power, Without the Panic

How to justify your sample size in a proposal, choose the right inputs, and defend the number to a reviewer or committee. A plain-language quick guide from PhD methodologists.



Written by PhD methodologists researchgold.org

START HERE

Sample size is a justification, not a guess

A reviewer who asks about your sample size is really asking one thing: can this study actually detect the effect it claims to study. The answer is not a bigger number plucked from the air. It is a transparent calculation from four inputs, decided before data collection.

PRINCIPLE

Power is the probability that your study finds a real effect if one exists. A study that is too small can miss a true effect and report nothing, wasting the participants and the time. An a priori sample size calculation is how you promise, in advance, that the study is capable of answering its own question.

The four levers, and how they move together

1. **Significance level**
   The risk you accept of a false positive, conventionally set at 0.05. Tightening it requires a larger sample.

2. **Power**
   The chance of detecting a true effect, conventionally set at 0.80 or 0.90. More power requires a larger sample.

3. **Effect size**
   How large an effect you want to be able to detect. Smaller effects need much larger samples, and this is the input reviewers scrutinise most.

4. **Sample size**
   The output you solve for, once the other three are fixed. Change any of the first three and this moves.

THE HARD INPUT

Where the effect size comes from

The effect size is the input people invent, and the input reviewers catch. A number with no source behind it is the fastest way to lose credibility on the methods.

DEFENSIBLE SOURCES FOR THE EFFECT SIZE, BEST FIRST

Source	Why it is credible	Caution
A pilot study in your population	Closest match to your actual setting	Small pilots give noisy estimates; be conservative
A meta-analysis or prior trials	Pooled, published evidence	Check the population and outcome match yours
The smallest effect that matters	Anchors the study to clinical or practical importance	Requires a defensible rationale for the threshold
Conventional benchmarks	A last resort when nothing else exists	Weakest justification; reviewers may push back

WATCH OUT FOR

Do not reverse-engineer the effect size to produce a sample you can afford. Reviewers recognise a calculation built backwards from a convenient number. Justify the effect first, then accept whatever sample it demands, or change the design honestly.

QUICK REFERENCE

What each common design needs

The calculation depends on the test you plan to run. Use this as a starting map, then confirm the exact inputs for your design.

COMMON DESIGNS AND THEIR SAMPLE SIZE INPUTS

Your comparison	Typical test	Effect size input you supply
Two group means	Independent samples t test	Standardised difference between means
Before and after, one group	Paired t test	Standardised difference, paired
Three or more group means	Analysis of variance	Effect size for the group factor
Two proportions	Test of proportions	The two expected proportions
Association between variables	Correlation	Expected correlation coefficient
Several predictors	Multiple regression	Expected variance explained and number of predictors

PRINCIPLE

Inflate the raw calculation for expected attrition. If the calculation needs a complete dataset of a given size and you expect a fifth of participants to drop out, recruit enough that the surviving sample still meets the target. State the attrition assumption explicitly.

WATCH OUT FOR

A closed-form formula only fits a simple, single-level design. The moment your data are clustered (pupils within schools, measurements within patients) the effective sample is smaller than the headcount, because responses within a cluster are correlated. Multiply the simple sample by the design effect, 1 plus (average cluster size minus 1) times the intraclass correlation, and you will often need substantially more participants than the naive number. For multilevel models, structural equation models, or interaction effects, no tidy formula exists and a simulation-based (Monte Carlo) power analysis is the honest route. Match the power method to the model you will actually fit, not to the simplest test it resembles.

THE WRITE-UP

State the calculation so a reviewer cannot fault it

A power statement that names every input is almost impossible to argue with, because the reader can check each number. Adapt the template below.

A priori sample size statement

An a priori power analysis was conducted to determine the sample size required to detect [effect size] in [the primary outcome], using [the planned test]. Assuming a significance level of [alpha] and power of [power], the analysis indicated a required sample of [N] participants. Allowing for [attrition rate] attrition, we will recruit [adjusted N]. The effect size was based on [source]. The analysis was performed using [software].

- [ ] Name the test the calculation is built on, and make sure it matches your analysis plan.
- [ ] State the significance level, the power, and the effect size as three separate, visible numbers.
- [ ] Cite the source of the effect size in the same sentence, so it is not left hanging.
- [ ] Show the attrition adjustment and the final recruitment target, not just the raw figure.
- [ ] Name the software or method used, so the calculation is reproducible.

THE DEFENCE

Answering the sample size questions you will be asked

These are the challenges that come up in proposals, vivas, and peer review. Each has an honest, specific answer.

THE QUESTION BEHIND THE QUESTION

What they ask	What they want to know	How to answer
Why this effect size?	Is the target effect realistic	Cite the source and explain why it fits your population
Your sample looks small.	Can the study detect anything	Show the a priori calculation and the power it delivers
What if the effect is smaller?	How fragile is the study	Report the smallest effect you are powered to detect
Did you adjust for dropout?	Will the analysed sample still be adequate	Show the attrition assumption and adjusted target

WATCH OUT FOR

If the data are already collected, do not compute observed power from your own result and present it as reassurance. Post-hoc power calculated from the observed effect is circular and reviewers know it. Report confidence intervals and discuss precision instead.

BEFORE YOU SUBMIT

The sample size checklist

Run your proposal or methods section through this list. Every box should be tickable before a reviewer sees it.

- [ ] The calculation is a priori, performed before data collection, not after.
- [ ] The test the calculation assumes matches the analysis you actually plan to run.
- [ ] The significance level and power are stated explicitly.
- [ ] The effect size has a named, defensible source, not a convenient guess.
- [ ] Attrition is accounted for, with a stated rate and an adjusted recruitment target.
- [ ] The software or method is named so the calculation can be reproduced.
- [ ] You can state the smallest effect the study is powered to detect.
- [ ] Clustered or multilevel designs are inflated by the design effect, or powered by simulation.
- [ ] No post-hoc power from the observed effect is used as evidence.

WHEN THE NUMBER HAS TO HOLD UP

A sample size you can defend, and the analysis that follows it

If you need an a priori power calculation matched to your exact design, written so a committee or reviewer accepts it, our PhD methodologists handle the justification and the analysis that follows: the right tests, the assumptions checked, and a results section ready to defend. You stay the author and make every final call.

Get a quote for your statistics

researchgold.org . Written by PhD methodologists . We work in writing, on your schedule.