

Sample Size in Grants: What You Need to Know

Jennifer Thompson, MPH

Department of Biostatistics
jennifer.l.thompson@vanderbilt.edu
<http://biostat.mc.vanderbilt.edu/JenniferThompson>

Grants & Sample Size: What Do I Need?

What is typically included in a grant submission?

- Background & clinical relevance
- Specific aims
- Plans for enrollment and data collection
- Budget justification
- **Sample size justification & statistical analysis plan**

What will we talk about today?

- What questions need to be answered *before* discussing sample size?
- What information is necessary to calculate appropriate sample size(s)?
- How do I get the right numbers?
- How do I put those numbers into my grant?

What's the Point of Sample Sizes, Anyway?

- *Ideally*: To explain to the funding agency how many patients it needs to pay for you to enroll, so that you can find the results you're looking for.
- *Realistically*: Sample size is often at least somewhat affected by how much funding you can reasonably expect, by your ability to enroll patients, or even by ethical considerations.
- Therefore, **a sample size section needs to justify the funding you're asking for, while balancing statistical needs with feasibility**. A good sample size section is much more involved than a cut-and-pasted paragraph.
- 42% of R01s examined in one review paper were criticized for their sample size justifications or analysis plans. *Inouye & Fiellin, "An Evidence-Based Guide to Writing Grant Proposals for Clinical Research", Annals of Internal Medicine, 142.4 (2005): 274-282.*

First Things First: The Research Question

- We all start out with a "big picture" question. For example, do patients sedated using dexmedetomidine experience better sleep in the ICU than patients on GABA-agonists?
- But how do you conduct analysis on that question?
 - Do you assign which sedative patients will receive?
 - How do you account for differences in patients, especially in length of ICU stay or admission diagnosis?
 - How do you define "better?"

First Things First: Research Question

What is a more specific question?

*I hypothesize that dexmedetomidine will improve sleep quality by increasing sleep scored as stages N2 and N3, as well as increase total sleep time and decrease sleep scored as atypical, when compared to GABA agonists. My **primary outcome** will be the difference in percent time in normal sleep between dexmedetomidine and GABA agonist groups.*

First Things First: Analysis Plan

How do you plan to collect and analyze that data? All of these questions are very important in determining how to analyze the data, and therefore how to calculate an appropriate sample size.

- Study type?
 - Randomized trial (How will you randomize?)
 - Cohort (How long will you follow up?)
 - Case-control (Will you match patients, and if so, how?)
- How many measurements per person?
 - Single measurement
 - Baseline and followup (Is the raw amount of change important, or do you need to account for differences at baseline?)
 - Many measurements (Are you interested in raw values, or a trajectory?)
- Do you think other variables may affect your results?

What Else Do I Need?

Aside from an analysis plan, there are several other questions which need to be answered.

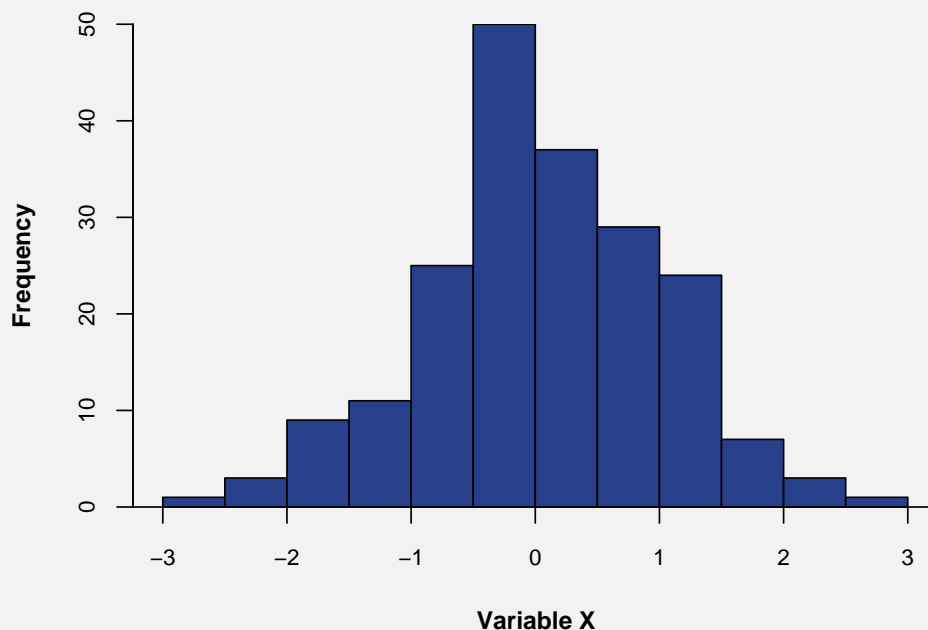
- How many patients are realistic?
 - How much are you likely to be funded?
 - How many patients can you feasibly enroll?
 - Is there enough potential harm that you need to limit enrollment?
- What effect can you reasonably expect to see?
 - Are you trying to detect a difference between groups?
 - Are you trying to estimate a quantity (such as a proportion)?
- What does your pilot data tell you?

Speaking of Pilot Data...

- ▷ Where do I get pilot data?
 - Previous studies you or close colleagues have been involved with (easy access to raw data)
 - Data from published studies, procured from peers
 - Summary data from published literature, most importantly standard deviation of primary outcome
- ▷ Points to consider:
 - Pilot data needs to be as similar as possible to what you expect to collect.
 - If continuous, is your primary outcome normally distributed? If not, you may need to consider techniques like transformations and/or nonparametric analyses.

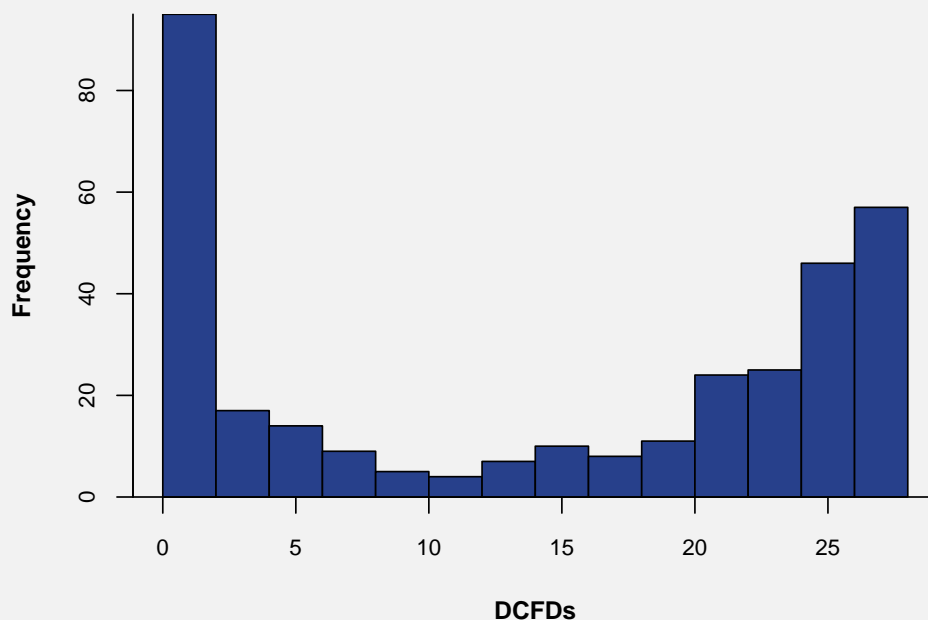
Normally Distributed Data

Many statistical tests - a t-test, for example - assume your data looks like this.



Non-Normally Distributed Data

Many statistical tests - a t-test, for example - are inappropriate for this data.



Response Rate & Patient Dropout

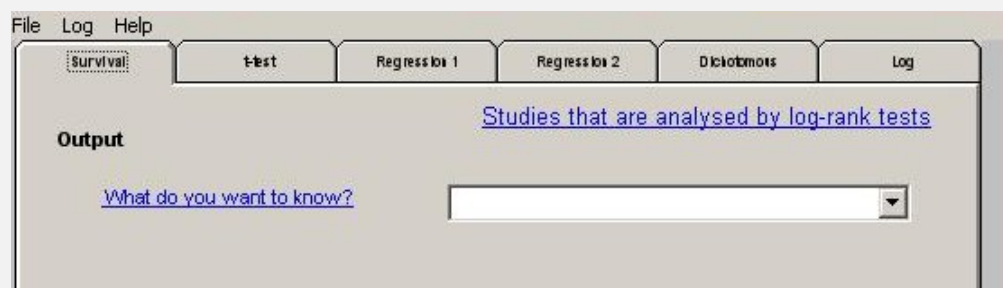
Don't forget to account for dropout and response rate. Previous studies may help address how much response and dropout you should expect.

- An RCT with a highly toxic chemotherapy drug may expect a significant amount of dropout. So might a cohort study with a very long and involved followup.
- Minimally invasive observational studies may have a very low expected dropout rate.
- Shorter surveys have higher response rates than long, involved surveys.
- If previous studies have observed a general dropout rate of 15%, add 15% or more to your final sample size and explain why.

Getting the Numbers

For standard analyses, PS software is very helpful.

- Free software; available at <http://biostat.mc.vanderbilt.edu/PowerSampleSize>
- Available for Windows, and can run on Linux using Wine (no Mac version)
- Can do sample size calculations for several common analyses



Getting the Numbers

For more complex analyses:

- nQuery is available in CRC; can do more complex calculations, such as nonparametric tests
- Simulations can be done for any statistical technique; most valuable for very complex analyses, such as mixed effects or GEE models
- Rules of thumb for various types of regression

Rules of Thumb

- Degree of freedom: In this context, a piece of information accounted for in a model.
 - Continuous, linear variables take up one df.
 - Categorical variables take up one df per category, minus one.
 - A model including age, race (white/black/other), sepsis diagnosis, and treatment group (intervention v. control) would require five df.
- Linear regression: Need approximately 15 (10-20) observations per degree of freedom.
 - Example: You are interested in predicting cognitive functioning in a group of ICU survivors, using a standard test score, while adjusting for age, education, hospital length of stay and whether or not cognitive therapy was administered after a hospital stay.
 - You have a total of four degrees of freedom, so you need approximately 60 patients.

Rules of Thumb

- Logistic regression: Divide the minimum of (events) and (non-events) by approximately 15 (10-20) to get number of df allowed.
 - “Event” is the outcome of interest - for instance, disease status.
 - Example: You are interested in predicting coronary artery disease. You expect 30% of your patients to have CAD, and want to examine as risk factors age, sex, race (W/B/O), mean blood pressure and total cholesterol.
 - Six df * 15 observations per df = 90. The minimum of “events” (has CAD) and “non-events” (no CAD) is expected to be 30%; 90 is 30% of 300. So based on pilot data or literature, you would need approximately 300 patients.

Rules of Thumb

- Cox regression (time to event): Divide the number of events by approximately 15 (10-20) to get number of df allowed.
 - “Event” is again the outcome of interest, for example, number of deaths or recurrences.
 - Example: Pilot data shows a mortality rate of about 45% in your study population, and you want to examine mortality adjusting for age, sepsis admission diagnosis, severity of illness (APACHE II), and a biomarker level (continuous and linear) - four df total. Four df times 15 observations is 60, which is 45% of 133, so you need about 133 patients.

What Goes In My Grant?

Key components for sample size justification:

- Number of patients you need (accounting for dropout)
- Power you expect to achieve with those patients
- Quantity you intend to measure
- Data on which you based those assertions

Most important: Show the reviewers that you have solid reasoning behind your calculations (as well as your analysis plan). Acknowledge it if your study will be a pilot or feasibility study.

Remember: Pictures speak louder than words. Figures can be very helpful, and can show that you realize some variability is inherent in the research process.

Example 1: Detecting a Difference Between Groups

We want to determine whether patients receiving the new sedative dexmedetomidine experience more restorative sleep in the ICU than patients receiving traditional sedatives benzodiazepine and propofol.

- Patients will be assigned to receive either dex or GABA agonists upon their enrollment, soon after ICU admission. (There will be a washout period before any measurements are taken, so that any sedatives already given can leave patients' systems.)
- Polysomnography (PSG) will be conducted on each patient for up to 96 hours or until ICU discharge. PSG data will be scored by sleep technicians and classified into sleep stages, or as "atypical" sleep.
- Quantity of measurement: Percent time in "normal," restorative sleep over a 24-hour period.

Example 1: Detecting a Difference Between Groups

We want to determine whether patients receiving the new sedative dexmedetomidine experience more restorative sleep in the ICU than patients receiving traditional sedatives benzodiazepine and propofol.

- Analysis will take advantage of repeated measurements for each patient, which will add power. However, due to time constraints, sample size will be calculated conservatively, without accounting for repeated measurements.
- We already know that this is a pilot study, and we can feasibly enroll about 30 patients.
- Pilot data shows a mean of 1.2% normal sleep in ICU patients (who are mostly on GABA agonists), with a standard deviation of 3%.

Example 1: Detecting a Difference Between Groups

The screenshot shows the 'Power and Sample Size Program: Main Window <2>' interface. The window has a menu bar with 'File', 'Log', and 'Help'. Below the menu bar are tabs for 't-test', 'Regression 1', 'Regression 2', 'Dichotomous', and 'Log'. The 't-test' tab is selected. The interface is divided into three main sections: 'Output', 'Design', and 'Input'.
Output: A link 'What do you want to know?' points to a dropdown menu set to 'Detectable alternative'. Below this, two input fields for 'Detectable alternative' both contain the value '3.1801'.
Design: A link 'Paired or independent?' points to a dropdown menu set to 'Independent'.
Input: A checkbox is checked next to a field containing '0.05'. Below it, a field for 'n' contains '15'. To the right, a field for 'x' contains '3'. Below 'n', a field for 'power' contains '0.80'. To the right of 'power', a field for 'm' contains '1'. There are 'Calculate' and 'Graphs' buttons to the right of the input fields. At the bottom left, a status bar says 'Logging is enabled.' and there is an 'Exit' button at the bottom right.

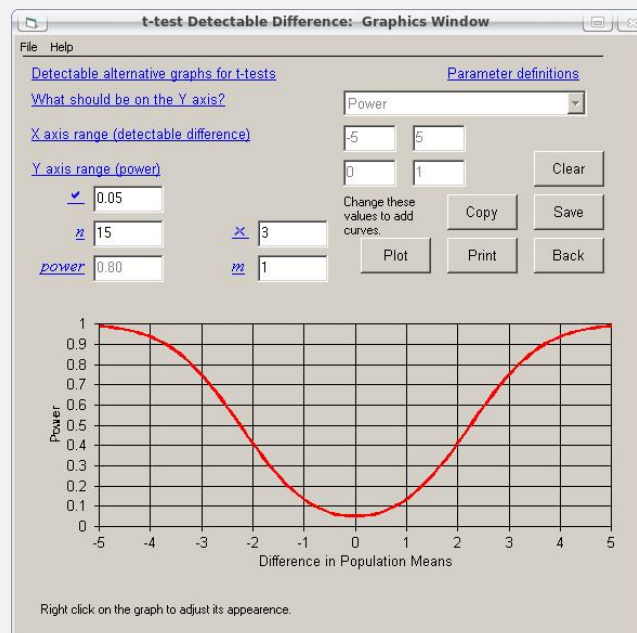
Example 1: Detecting a Difference Between Groups

Points to remember when using PS:

- α levels are for a *two-sided* test
- n is the number of patients *per group*
- PS asks for *standard deviation*, not standard error
- m is the *ratio* of patients in the intervention group to the control, so for equal groups, $m = 1$

Example 1: Detecting a Difference Between Groups

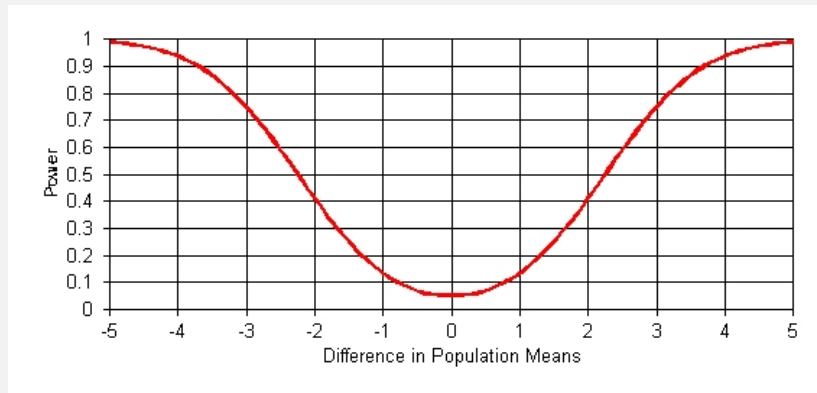
Remember: Figures can help your message.



Example 1: Detecting a Difference Between Groups

Final sample size justification:

This is a pilot study and sample size will be limited by available resources, with 30 patients felt to be a realistic, achievable enrollment. Preliminary power analysis showed that 30 patients will result in 80% power to detect a 3.2% difference between treatment groups in normal sleep. These calculations are based on prior ICU sleep studies from the Vanderbilt MICU that showed a mean of 1.2% normal sleep, with a SD of +/- 3.0%.



Example 2: Estimating a Quantity

We want to estimate the proportion of ICU patients who develop neuromuscular disease during their ICU stay. We have very inconclusive pilot data, as this area has not been studied in great depth, but a systematic review of the literature estimates the proportion at about 57%.

In this case, we can say we would like to estimate the proportion within a certain percentage - say, we are willing to have a margin of error of $\pm 7\%$.

(warning: math ahead)

Example 2: Estimating a Quantity

We want to estimate the proportion of ICU patients who develop neuromuscular disease during their ICU stay. We have very inconclusive pilot data, as this area has not been studied in great depth, but a systematic review of the literature estimates the proportion at about 57%.

The quantity that determines the width of the confidence interval of a proportion (at a typical 95% significance level) can be written:

$$1.96\sqrt{\frac{p(1-p)}{n}}$$

where p = the proportion of interest and n = the number of patients.

Example 2: Estimating a Quantity

So, given our pilot data, we need to solve the following equation to get the number of patients needed for our widest margin of error:

$$1.96\sqrt{\frac{0.57(1-0.57)}{n}} = 0.07$$

This gives us an n approximately = 193.

Point to remember: As your estimated proportion gets closer to 50%, your required sample size will go up.

Example 2: Estimating a Quantity

Example sample size justification:

This pilot study aims to estimate the proportion of patients who develop ICU-acquired neuromuscular disease. Published data estimates this proportion to be approximately 57%, with substantial variability between studies. In order to measure this proportion within a 7% margin of error, we estimate that we will need 193 patients to undergo testing and examination. Due to the somewhat invasive nature of the diagnosis instruments, we estimate a 15% dropout rate. Therefore, we plan to enroll 220 patients.

Example 3: Adjusting for Covariates

We want to determine whether baseline levels of endothelial progenitor cells (EPCs) are associated with the number of days alive and free of acute brain dysfunction in the ICU. We think this relationship may be affected or hidden by:

- Patient age (continuous)
- Comorbidities (Charlson index, continuous)
- Severity of illness (APACHE II, continuous)
- Sepsis at ICU admission (dichotomous)
- Cognitive impairment at study enrollment (IQCODE, cont.)
- ICU type (medical or surgical, dichotomous)
- APOE genotype (dichotomous)

Example 3: Adjusting for Covariates

Continuous variables are allowed to have a nonlinear relationship with the outcome. So, including EPC level, we have 13 degrees of freedom.

How many patients do we need?

13 degrees of freedom * 15 observations per df = **195 patients**

Example 3: Adjusting for Covariates

*Sample size calculation is based on the primary outcome variable of days alive and free of delirium and coma, modeled using linear regression. Continuous variables will be allowed to have a nonlinear association with delirium, using restricted cubic splines. The primary model will consist of EPC level on day 1 (continuous) plus the following seven covariates:... Continuous variables require two degrees of freedom for non-linearity, while dichotomous variables require one degree of freedom. Therefore, the minimum degrees of freedom required for the model will be 13. Assuming that one degree of freedom requires about 15 patients, a multivariable model with a complexity of 13 degrees of freedom can be reliably fitted when the effective sample size is at least $13 * 15 = 195$ subjects. We will, therefore, enroll a total of 200 patients, resulting in adequate power for this analysis.*

Take-Home Message

- Sample size justification is an essential part of every grant.
- Calculating a sample size is rarely as simple as plugging numbers into a formula.
- **Statisticians are here to help!** Involve them in the grant from the beginning, and they will help make your grant better (and more likely to be funded).

Resources & Acknowledgments

Resources

- Department of Biostatistics collaboration plan
- Daily clinics, at noon in MCN D-2221 (except Thursdays); see biostat.mc.vanderbilt.edu/Clinics

Acknowledgements

- Ayumi Shintani
- Cathy Jenkins
- Terri Scott & Nate Mercaldo
- Matt King, Mike Hooper & Chris Hughes
- Wes Ely & ICU Delirium group