Maximizing Statistical Interactions
Part I: Preliminary help (Grants and Power)
Provided by: The Biostatistics Collaboration Center (BCC) at Northwestern University

A statistician is an asset in the design of a study even beyond the requisite power and sample size justification section that most granting agencies want to see in applications. In order for us to optimize the help we can give to you, we need to know about your proposed plan. The following guide will help you prepare for your statistical collaboration pertaining to grant applications.

1. What is your deadline?
   a. Ideally, you have been working with us from the conception of the study idea. While statisticians tend to be "known" for providing detailed power and sample size justifications and statistical methodology, statisticians can also give sound input on study design that might prove helpful (additionally, the study design will likely have an impact on the power/sample size calculation). That said, keep in mind that there will likely be some "back and forth" in study design, and the more time you provide for this, the better!
   b. If you have waited to consult a statistician, know that apart from the administrative work that grants need (time for departmental approval, signatures, budget issues), it may take some time to fully understand your proposal, potentially search for additional estimates that we need (standard deviations, intra-class correlation estimates), run a power analysis, and provide a reasonable write up for a methods section.
   c. Although it may be tempting, please don't "cut and paste" an old statistical section for us to edit (as a help). A clean slate is usually best.
2. What is your primary research question – what are your hypotheses?
   a. If we know your SOCO (single overriding communication objective), so to speak, we can think about different efficient ways to approach the problem. (Is it possible to have a cross-over design? A paired analysis? Matched by certain factors? Does randomization by a "cluster" make sense, or would it be more appropriate to randomize individuals?)
   b. If you are designing a multidimensional project (like fMRI or genetics), keep in mind that you will need to control for an overall type I error. The standard "all hypothesis tests will be run at α=0.05" will not be acceptable. In those cases, usually either Bonferroni-type corrections or false discovery rates will need to be used.
3. Who are you interested in studying?
   a. Is this a multi-center or single center study? What level of observation do you want to make inferences on? (e.g., patient, cell, department or floor, clinic, hospital?) – To follow that, it may be helpful to think of what your sampling frame may be.
   b. Inclusion criteria – who NEEDS to be in the study. This is essentially another look at your primary hypothesis from a "population" perspective.
   c. Exclusion criteria – who really should NOT be in the study. While exclusion criteria may limit the population to which you can generalize your results, there may be cases where you need to exclude people who may add too much variability to your study. There may also be ethical reasons for exclusion of specific potential participants. The exclusion criteria may prohibit the possible sample size that you can expect to recruit!

d. You may also need to consider "statistical independence" – that is, if you are recruiting from a minority population, getting someone's entire family to be in the study may not be the wisest thing to do as there will be some correlation that you would expect from family ties. This violates an assumption of many popular statistical tests.

4. What is your primary outcome (dependent variable)?
   a. How is it measured? - Is this a continuous measure? Ordinal? Binary? What are "usual" descriptive statistics for those measures (means, standard deviations, rates) that you may expect to see in your population of interest? What change in these measures might you expect to see in a population that was watched for as long as you want to study them (if at all)? What change would make you happy? What changes have other, similar studies, seen?
      - Example: If you had 40% of patients in your clinic report significant depression, you might expect that 20% would resolve (32% rate of depression) on its own over the six-month period you want to study them; with treatment you would like to see an overall rate of 10% depression at six month follow-up.
   b. When or how often will it be measured? – do you have trained personnel collecting data with a set protocol? If you are collecting follow-up measures, do you have a "window" for eligible follow-ups? Have you considered methods to minimize missing data or study dropouts?
   c. Has it been or will it be validated? If this is measured via a surrogate (like a survey) has it been validated in the population in which you are interested? Are you including any validation measures in your analysis? Is your outcome measured in an objective, standardized way?
   d. Are you interested in the measure or change in the measure? If you are looking at comparing change in a measure, it would be best to base any power calculation on the *change* (and standard deviation of the change in the measure) rather than on the measure itself. Is any information available on that?
5. If you have "groups" how are they defined (independent variables)?
   a. Are they demographic-based? – if so, do you think that misclassification could be an issue?
   b. Will you generate them (ala treatment/control) and require a randomization? Do you want to run a stratified randomization to control imbalance in specific groups?
   c. Will they be "matched" in any way? This, like stratification may slow down recruitment efforts, but may be better, statistically, for controlling extraneous variability.
6. Are there other conditions or factors for which you need to account (potential confounders/effect modifiers), or that you would like to formally test? How are those defined?

Additionally, there may be some other factors that you may need to consider:

1. Is the possible sample size restricted? (e.g., your practice only sees about five patients per month who might fit your inclusion/exclusion criteria, and you need to have everyone recruited within a year). Might you have to worry about people not wanting to participate in the study? Dropout rates? Non-compliance issues? How many study visits or laboratory assessments are feasible given your budgetary constraints?

2. Multiple “looks”? If you are considering a longitudinal clinical trial, does your data need to be analyzed several different times (for interim or DSMB reports)? There is statistical theory available for Stopping Rules (ending trials early for very significant or insignificant findings) that may be of interest. Additionally, if you are analyzing your data several times, type I errors do require adjustment.
3. Blinding? If you are considering a blinded trial, you may even need to consider an additional statistician (one who can remain blinded, and one can provide a randomization and possibly interim reports).
4. Power calculations are complex and depend on multiple inputs. The more complicated your proposed analysis is, the more information we may need you to provide. Power calculations (even for basic analyses) may be sensitive to assumptions that you may not realize you need to make! Example: When sample sizes are small, or when data is really not normally distributed in the population, a Mann-Whitney test is used rather than an Independent t-test. The following table shows different assumptions on the underlying distribution for the dependent variable (which is usually unknown to some extent) for an effect size of 0.8 (note: Assuming a Uniform distribution gives similar results as power or sample size for an independent t-test).

Distribution	N/group for 80% power	Actual power for N=20/group
Normal	28	67.0%
Double exponential	18	86.1%
Uniform	26	69.3%
Logistic	24	71.6%

5. Beyond the basics. Even with the most basic “guide” of knowing means and effect sizes, please do not be offended if a power analysis for a simpler design is presented.
   Example: a power calculation was run to test the between group difference of a two-way repeated measures ANOVA. Depending on the covariance structure (the degree to which measures further apart are correlated), the required N varied from 18 to 108. In this case, the same effect size using a paired t-test showed that a total sample size of 70 gave the similar power. The take home message is that if you do expect a complicated power analysis, you will need to provide extensive information from the literature or even preliminary data.