Maximizing Your Time with a CBHDS Biostatistician
Grant Proposal Preparation

Most granting institutions require the inclusion of a statistician as a co-investigator or key personnel on submitted proposals. Collaborating with a statistician early on in your proposed study is recommended for a successful and long-term collegial relationship, and ultimately results in improved funding, productivity, and stronger science. Typically, investigators will schedule an initial meeting with us so that we may learn about the study, discuss timelines, budget, and power/sample size needs. The following guide will allow for you to prepare for collaborating with a CBHDS biostatistician on a grant proposal.

1. Please consider your timing and internal submission deadline when seeking support.
    - In the ideal situation, clinical investigators work collaboratively with their statistician from study conception through dissemination. While many investigators think of the statistician as the provider of detailed power and sample size justifications and statistical methodology sections, statisticians can also provide sound input on efficient and appropriate study design. Allowing your statistician to contribute to study design considerations is important, as study design has a direct impact on power/sample size estimates. Also note that arriving at the final study design typically is an iterative process, so the more time you allow for this, the better.
    - Please avoid delaying conversations with a CBHDS biostatistician on your proposal, as a significant amount of time is required for administrative tasks (departmental approval, signatures, budget preparation, bio, scope of work, etc), understanding your proposal, potentially performing literature searches for additional power estimates (standard deviations, intra-class correlation estimates, etc), preparing a power analysis, and writing up a statistical methodology section.
    - Please use the following lead time guidelines, as grant deadlines tend to be accompanied by many requests for support, and thus we must prioritize:
        - For larger grants (R01, U series, P series, etc), please submit your request for support at least 8 weeks prior to the external funding source due date, and at least 6 weeks before the grant is due to Virginia Tech's Office of Sponsored Projects.
        - For smaller grants (R03, R21, K series, etc), please submit your request for support at least 6 weeks prior to the external funding source due date, and at least 4 weeks before the grant is due to Virginia Tech's Office of Sponsored Projects.

2. How the collaboration typically unfolds.
    - An investigator requests support for preparing a grant proposal with appropriate lead time using the CBHDS Collaboration Request Form.
    - The investigator reviews the guidance documents on the CBHDS website and brings as much knowledge and information that he/she has with them to the initial meeting with a CBHDS biostatistician.
    - At the initial meeting, the biostatistician is provided with the specific study aims, study design and sufficient information for power/sample size calculations. Additionally, an agreement should be made on salary coverage for CBHDS. Per Welty et al (2014), a minimum of 20% effort per year should be reserved for CBHDS to support your research project, with a ratio of 1:2 for doctoral-level to master's-level personnel. For example, if the PhD-level statistician receives support for 15%, the master's level statistician would receive 30%. Any modifications to the final budget should be approved by CBHDS.
    - The biostatistician conducts a power/sample size analysis, and the budget is drafted and approved within Virginia Tech's Office of Sponsored Projects.
    - Clinical investigators work on the science, and the CBHDS biostatistician writes up the analysis plan and power section of the grant.
    - Prior to submission, the biostatistician is provided the opportunity to review the grant in its entirety, realizing that this step is critical to ensure consistency across all sections and to maximize technical merit.
3. What is your primary research question? What are your hypotheses?
    - Once you have arrived at your primary study objective and/or research question, we can consider various ways to approach the problem efficiently through design (i.e., cross- over, pre-post, difference-in-difference, matched pairs, cluster randomized, stepped wedge, randomized clinical trial design).
    - If you are designing a multidimensional project (e.g., fMRI or genetics), keep in mind that you will need to control for an overall type I error, as the standard "all hypothesis tests will be run at α=0.05" will not stand up under review. In those cases, Bonferroni-type corrections or false discovery rates will typically be applied.
4. Who do you wish to study? Who is your study population (accessible population) and who is your target population (the population you wish to generalize to)?
    - Is this a multi-center or single center study? What is your unit of measure or level of observation

you wish to make inferences on (e.g., patient, cell, department, floor, clinic, hospital, etc)? A discussion of your sampling frame may make sense when considering the unit of measure.

- What is your inclusion criteria? In other words, who must be included in your study to satisfy your defined target population?
- What is your exclusion criteria? That is, who should NOT be in the study? While exclusion criteria may limit the population to which you can generalize your results, there may be participants who must be excluded due to the excessive variability they introduce. There may also be ethical reasons for exclusion of specific participants. Pay careful attention to your inclusion/exclusion criteria, as these may limit the pool from which you can recruit participants or expect to recruit, along with your required sample size and potential enrollment period.
- You may need to pay attention to the assumption of "statistical independence." For example, if you are recruiting from a minority population and are not relying on random selection or assignment, enrolling an entire family or a circle of close friends into the study introduces correlation based on "like" members. An analysis without consideration of such clusters of correlated participants violates the main assumption of independence in many widely-used statistical tests.

5. What is your primary outcome measure(s) (dependent variable)?
    - How is it measured? Is it a continuous, ordinal, nominal, or binary measure? In your field, what are the typical descriptive statistics presented for this measure (mean, standard deviation, median, interquartile range, frequency and percentages, rates) in studies involving your population of interest? What is a reasonable expectation for change in this measure within a population observed for a period of time similar to the length of your study? What change do you consider clinically meaningful or associated with important other clinical measures? What changes have other studies observed that are similar in nature? For example, suppose 40% of the patients in your clinic report significant depression and it is expected that 20% of them will resolve on their own over the 6-month period of your study (resulting in 32% with depression after 6 months). For your study, you might propose a clinically meaningful depression rate of 15% among those treated for depression at 6-month follow-up.
    - When and how often will your outcome measure 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? How will you minimize missing data and/or study dropout?
    - Is the instrument used to measure your outcome valid and reliable? Has it been validated in the

population in which you are studying? Is your outcome measured in an objective, standardized way?

- Are you interested in the outcome measure at a specific time point or the change in the measure over time? If you are interested in studying change, power calculations should be based on the *change* (and standard deviation of the change in the measure), rather than on the measure itself. Please provide expectations on change based on previous data or estimates. Finally, carefully consider and be prepared to defend your choice of timing for the proposed assessments.

6. If your study involves the comparison of groups, how are the groups (independent variables) defined?
    - Are the groups based on demographic characteristics? If so, will misclassification be a problem?
    - Will your study involve the comparison of groups that you will create (e.g., treatment and control)? Will treatment be assigned randomly? Does stratified randomization make sense to maximize balance in specific subgroups?
    - Will participants be "matched" in any way? Like stratification, this may slow down recruitment efforts, but may be desirable for controlling extraneous variability.
7. Are there other conditions or factors that require consideration, such as potential confounders or effect modifiers? Are there subgroups of interest that should be examined for differences in effect? How are these subgroups defined?

Additionally, there may be other things worthy of consideration:

1. Do you think that recruitment could be an issue? For example, if your required sample size is 100 participants, your practice sees about five patients per month who satisfy your inclusion/exclusion criteria, and your enrollment period is one year, then you may have to consider a multi-site study or modify your current study. Do you have reason to believe that eligible participants will not want to enroll in your study? Do you have reason to believe that participants in your study will be prone to drop-out, and perhaps drop-out differentially by treatment group? Do you have reason to believe that participants will be non-compliant to study protocol procedures? How many study visits or laboratory assessments are feasible given your budgetary constraints?
2. Do you have to consider multiplicity? If you are proposing a longitudinal clinical trial, does it make sense to perform interim analyses and consider stopping for futility? If so, there are many methods for defining "stopping rules" or ending the study early due to very significant or insignificant findings. When making multiple inferential comparisons, type I errors will require adjustment.

3. Does your study involve blinding? Are you using single, double or triple blinding (when the statistician analyzing the data is blinded of group assignment, in addition to the patient and clinician performing outcome assessments)? Often times, a study will specify a statistician who remains blinded and one who provides the randomization scheme and possibly interim reports.
4. Power calculations are complex and depend on multiple inputs. The more complicated the proposed analysis, the more information that must be provided for power estimation. Even in simple cases, power calculations may be sensitive to assumptions that you may not realize you need to make (e.g., assumptions related to the underlying distribution of the dependent variable).

For grants, analyses, and papers/presentations that we support outside of direct grant funding or contractual agreements with CBHDS, please acknowledge that your work was supported by core grant funding through the iTHRIV CTSA using the following language:

"Research reported in this publication/presentation/work was supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR003015. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health."

These guidance documents are modified versions of documents prepared by our colleagues at the Biostatistics Collaboration Center (BCC) at Northwestern University. We are grateful for Dr. Leah Welty and her team for their guidance and input.