

Maximizing Statistical Interactions: Guidelines to help you prepare for Statistical Collaboration  
Provided by: The Biostatistics Collaboration Center (BCC) at Northwestern University

The BCC is here to collaborate with you on study design and statistical aspects of your research. We can participate in study design, grant preparation, data collection, data analysis and interpretation, writing of papers, and reviewer response/critique (from papers or grants). This one-page summary highlights what we may need in order to provide help.

1. Grant application/Study design

- a. What are your primary (and secondary) hypotheses?
  - i. What is the ONE most important, specific question you want answered?
  - ii. How do you intend to collect the information?
  - iii. Do you want to address other outcomes, or the same outcome in specific subgroups, or after adjusting for other variables?
- b. How will your outcomes and independent variables be measured?
  - i. What “effect size” is clinically relevant? – What difference in groups or degree of association do you think is important to be able to say “Eureka!”?
  - ii. For power/sample size calculations - please bring estimates (relevant literature, or preliminary data) for any tests of statistical significance that are similar to the types you are proposing. The more complicated an analysis, the more estimates you will need!
- c. Agree with the statistician early on as to deadlines and timeline for grant submission

2. Data/Spreadsheet formats

If you entered your data in Excel, several basic rules are necessary so that your data may be “read into” a statistical package easily. Our website has a detailed handout, but please be aware of the following:

- a. Include unique variable names ( $\leq 10$  characters, no symbols) in the first row (ONLY) of your spreadsheet.
- b. Do not color code or have blank rows within your spreadsheet. If information is important enough to “offset,” be sure a variable is included that would indicate what you would have color-coded.
- c. Data need unique (numeric) study ids (not identifiable, per HIPAA privacy rules). Row numbers in Excel do not count. Use one row of data for each unique, independent study subject. (If you have multiple follow-ups, you may consider a “tall” (aka: vertical or long) data format – see more detailed information on our website).
- d. Do not mix “string” (text) and “numeric” data types (e.g., do not include symbols with numbers, like 120/90 for blood pressure, use separate variables; do not type “na” for missing in numeric data – leave it blank).
- e. Numeric coding is preferable (e.g. 0=no, 1 = yes) with a data codebook, or data dictionary either in a separate sheet or document.

3. Analysis/Questions

- a. Consider points made above (1a-b, and 2a-e). Many of these issues will be the same.
- b. While we could “supply” you with a requested analysis, topical questions lead to a more collaborative atmosphere (“Are these related?” Rather than, “Can you run a canonical correlation?”) – Your specific data may not “fit” assumptions required for popular analyses in the literature, and those analyses may not be able to address the question you truly want answered.