

Research Design and Statistical Analysis Criteria for YCCI Protocols

The following criteria/questions should be clearly described in protocols requesting the use of YCCI resources.*

Aims: The study objectives or what you intend to do.

1. What are the primary aims?
2. What are the secondary aims?

Hypotheses: A version of the research question that provides the basis for the statistical analysis.

3. What are the primary hypotheses?
4. What are the secondary hypotheses?

Study Design:

5. Is the study observational or experimental?
6. What type of design will be employed (eg. cross-sectional, concurrently controlled trial, crossover trial)?
7. What are the study groups?
8. Who are the study observers/raters/interviewers?
9. At what timepoints will variables be assessed? (For longitudinal studies, a timeline is very helpful for explaining when interventions and assessments will be made.)
10. Will there be blinding?
11. Is it a multicenter study?

Variables: The measurements which you wish to assess.

12. What are the outcome variables and how are they measured?
13. What are the predictor variables and how are they measured?
14. What are the confounding variables and how are they measured?
15. Have the instruments used to assess outcomes and main predictors/confounders been validated in populations similar to your study population?

Subjects: The individuals that will be used to represent your target population.

16. Are subjects appropriate for study?
17. How are subjects selected?
18. Are inclusion/exclusion criteria clear?
19. Are subjects assigned randomly to groups? If not, are groups matched?

Data Analysis: The data analysis plan provides the means by which the study hypotheses will

be statistically evaluated.

20. What is the data analysis plan *for each aim*?
 - a. What statistical method/model will be used?
 - b. What are the factors included in the analysis?
 - c. How will confounders be controlled (eg. matched analysis, stratified, covaried)?
 - d. Are there corrections employed for multiple comparisons?

Sample Size/Power Calculations: These calculations are essential for judging the ability of your study to provide accurate descriptions of the target population. The essential components of a sample size calculation are:

- i. **Type I error (false positive or α error)**
- ii. **Type II error (false negative or β error)**
- iii. **n**- the number of study subjects in each group
- iv. **SD** – the standard deviation of the outcome measure
- v. **Effect size (Δ)**

Knowing 4 of these components allows you to estimate the 5th. Typically, these calculations are performed to determine the n required to detect a Δ of a particular size or to determine the power (1-type II error) to detect a particular Δ given n.

21. What is the specified type I error?
22. What is a reasonable expectation for the SD? (This can come from the literature or pilot work, but its origin should be stated in the protocol.)
23. How many subjects are needed?
24. Is the Δ clinically meaningful?
25. What is the study power?
26. What is the effect of dropouts? (Sample size calculations depend on the number of subjects available at the analysis stage, so the dropout rate should be accounted for)
27. What is the likelihood of reaching the targeted sample size? (A description of the number of eligible subjects you expect to be available should be included. Most investigators tend to overestimate this number, so when possible, empirical evidence should be shown.)

*Exploratory studies do not have the same requirements as those for hypothesis testing.

If you have any questions or concerns about your protocol, please contact Jim Dziura, YCCI Biostatistician, Suite 411, 2 Church St South.
Office extension 7-4468; E-mail: james.dziura@yale.edu; fax 785-7854.

Helpful Hints in Developing your Protocol

Aims and Hypothesis should be clearly stated and unambiguous in their intent
Limit the number of predictor and outcome variables
Avoid using response measures of uncertain utility
Describe study subjects and their selection clearly and completely
Avoid vague and sweeping statements about data collection and analysis
Show clearly how sample sizes were derived
Contact a statistician early in the protocol development process

Glossary

Outcome:	The variable(s) for which you intend to evaluate the impact of the predictors.
Predictor:	The variable(s) which you intend to manipulate or control.
Confounder:	The variable(s) associated with both predictor and outcome that may introduce bias into the predictor/outcome relation.
Reliability:	The extent to which, if you measure something, then measure it again with the same instrument, the 2 measures correspond.
Validity:	The extent to which your measuring device actually measures that which you are interested, e.g. Using a tape measure to assess length.
Type I error:	The probability of declaring populations different based on study samples when true differences <i>do not</i> exist in the populations
Type II error:	The probability of declaring no differences based on study samples when true differences <i>do</i> exist between the populations
Power:	The probability or ability to detect differences based on study samples when true differences <i>do</i> exist between the populations. (1-Type II error)
Effect Size:	The smallest clinically meaningful difference that you wish to detect
Bonferroni:	A technique used to adjust the Type I error resulting from too many comparisons.
Exploratory Analysis:	A study in which data are analyzed to uncover indications of an effect.
Planned Analysis:	A study in which hypotheses are tested to confirm the effect(s) of factor(s).