R01 Boot Camp Biostatistics for Grants: Clinical

Jamie Myles

March 19, 2013

University of Michigan Health System

My Background

- James Myles Ph.D.
- Studied Biostatistics at Washington
- Cancer Cooperative group and Department of Epidemiology
- Pharmaceutical Industry
- University of Michigan

RDC Team in Action

www.michr.umich.edu/services/researchdevelopment

Research Development Core

The Research Development Core (RDC) provides free services and consultation to strengthen study design and grant proposals.

RDC SERVICES

Consultation

RDC's most popular service is an in-person one-hour consultation meeting intended to provide investigators with supportive advice about their proposal or research idea. Each meeting is customized to include RDC faculty and staff with backgrounds that best meet the needs of the PI. Services may include matching ideas with funding sources, developing research plans and submission strategies, identifying collaborators, grant editing, and guidance on future career direction.

RDC Can Help

RDC can help investigators when they …

- *Have an idea for a research project*
- *Want to make an existing proposal better*
- *Need funding source ideas or career direction advice*

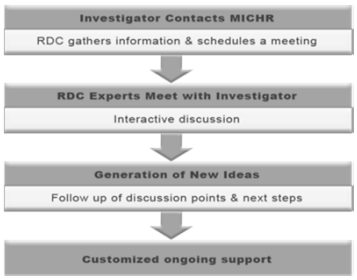
RDC Can Help

RDC will…

- Help at no cost "pre-award" and point to more MICHR services "post-award"
- Link investigators to funding sources, collaborators and other resources
- Provide study design guidance and scientific review, offer advice for resubmissions
- Provide editing support

RDC Services

Consultation Process



Customer Feedback

I was stuck, and you helped me get unstuck.
~ Suzanne Cole, PhD
Lecturer III
Environmental Health Sciences, SPH

I am constantly amazed at the wealth of resources at U. Your efforts with MICHR are a prime example of how opportunities are created here.
~ Mia Woodward, MD
Clinical Lecturer
Ophthalmology

This experience with MICHR is the best example of the "Michigan Difference" and I can't thank you enough for all you do.
~*Thomas Gardner,* MD
Professor
Ophthalmology, Molecular and Integrative Physiology and Internal Medicine

I think your service is wonderful and I will be recommending it highly to my SPH colleagues.
~ Daniel Eisenberg, PhD
Associate Professor
Health Management and Policy, SPH

I am confident that working with this group is going to allow me to succeed.
~ Matthew Greenhawt, MD
Assistant Professor
Int Med-Allergy

educate • fund • connect • support

Research Development Core

For more information

Email: michr-rdc@umich.edu

Or call: Tamara Havermahl, RDC Manager, 734.763.1715

Objective

- What are the main areas where R01's fail due to statistics?
- What can you do about it?

High-level Overview

- Excellent (vs. good) statistics are unlikely to substantially improve your likelihood of success in your R01
- Inadequate statistics are very likely to substantially reduce your likelihood of success in your R01

Statistics Targets in R01

1. No statistics section
2. Statistics not matching the aims
3. Not having a statistician on the grant when you need 1
4. Probable "Bias"
5. Inadequate description of power / sample size
6. Data unlikely to be collected / cleaned properly
7. Not able to "recruit" the subjects
8. Missed discussing an obvious problem
9. Does not look like you know what you are doing

No Statistics Section

Why Does Biostatistics Matter? (1)

- Rigorous design of studies:
  - what is the question?
  - why does this question matter?
  - how will doing this study answer this question?
  - appropriate basic design
  - appropriate endpoints
  - appropriate measurement techniques
  - appropriate population
  - sample size

Why Does Biostatistics Matter? (2)

- Appropriate analysis
  - valid approach
  - efficient: makes best use of the existing data
  - done correctly

Bad analyses can be redone

Bad designs cannot be fixed

http://vincehci.blogspot.com/2011/02/book-reading-1-design-of-everyday.html

Why Does Biostatistics Matter? (3)

- Appropriate conclusions:
  - valid conclusions based on the results of the study
  - interpretation incorporates uncertainty in the results

Components

- Design and Setting
- Study Sample
  - Inclusion Criteria
  - Exclusion Criteria
  - Availability of Subjects
- Data Collection / Procedures
- Outcomes
- Intervention / Controls (if applicable)
- Sample Size Calculations / Data Analysis (more on this later)

Design and Setting

- Describe study design in detail
  - Schema (figure) is very useful
- Describe randomization procedure, if applicable
- Describe blinding methods, if applicable
- Describe selection of cases and controls, if applicable for observational study
- Describe representativeness of target population, selection bias issues

Study Sample

- Describe and justify choice of study sample
  - Should avoid potential biases
  - Should avoid being non-representative
- NIH requires inclusion of women, minorities and children
  - Justify their inclusion / exclusion
- Provide estimates of number of potential subjects available per site
  - Pilot data particularly useful
  - Show you can get the required sample

Outcomes

- Detailed description, including operational definition and specification, of each study outcome (endpoint)
  - Consider blinding (bias), validity, reliability
  - Describe performance characteristics of the measures used for each outcome (see previous section)
- Should be adequately powered to evaluate all of the primary outcomes

Interventions / Controls

- Describe intervention strategy and its implementation
  - Consider adherence/compliance, adequate dosing/potency, standardization
- Describe interventionists, their training, and any further training required for the study
- Describe quality control

Statistics not matching the aims

- "Common critiques from reviewers are that the specific aims and hypotheses are poorly focused, underdeveloped, or overly ambitious"

- An Evidence-Based Guide to Writing Grant Proposals for Clinical Research. Sharon K. Inouye and David A. Fiellin. *Ann Intern Med.* 2005;142:274-282.

How does this happen?

- Study design does not answer the question
- Analysis is not correct for the data

Study Designs

- Observational or experimental
  - No intervention vs. influence
- Prospective or retrospective
  - Collect data from study start forward or looking back
- Cross Sectional or longitudinal
  - A snap shot vs. following people over time

Study Design

- Case-Control
- Cohort
- Cross-sectional study
- Randomized Clinical Trial

Some examples

- Framingham, Massachusetts Heart Study
- Census
- Outbreak of cholera in 1850's London
- CAST
- Nurses' Health Study
- NSABP Breast Cancer Surgery

Things to Consider

- Bias
- Blinding
- Replication
- Sample Selection
- Randomization
- Sample Size

Not having a Statistician When You Need One

- Colleague and I were discussing what type of clinical R01 would not need a statistician
- Thoughts?

Finding help from Statisticians

- MICHR
  - Help pre-award with design and grant writing
  - Possibility of writing into the grant if needed
- CSCAR
  - 1 hour free consults at any time
  - Possibility of writing into the grant
- Biostat Department
  - Some young faculty who are looking for collaborators
  - Not always easy to connect with, but great if you can

Participation in the initial stages of experiments in different areas of research leads to a strong conviction that too little time and effort is put into the planning of experiments. The statistician who expects that his contribution to the planning will involve some technical matter in statistical theory finds repeatedly that he makes a much more valuable contribution simply by getting the investigator to explain clearly why he is doing the experiment, to justify the experimental treatments whose effects he proposes to compare, and to defend his claim that the completed experiment will enable its objectives to be realized. For this reason the remainder of this chapter is devoted to some elementary comments on the subject of planning. These comments are offered with diffidence, because they concern questions on which the statistician has, or should have, no special authority, and because some of the advice is so trite that it would be unnecessary if it were not so often overlooked.

Cochran WG, Cox GM. Experimental Design. 2nd edition. New York: John Wiley & Sons, Inc. 1957, page 10.

"Probable" Bias

http://nukezilla.com/2010/05/24/unconvinced-pete-the-gamers-guide-to-cognitive-bias-part-1/

Bias

***Bias*:** the results observed reflect other factors in addition to (or even instead of) the effect of the treatment

Bias

- there are multiple potential sources of bias
- it is impossible to completely eliminate the possibility of bias
- it is possible to minimize some of the major biases with careful planning
- the accusation that a bias ***may*** exist is often sufficient to cause the validity of a study to be generally questioned

Blinding

- One of the most important bias-avoiding techniques in clinical trials
  - Can't always blind investigator or patient to Rx regimen
  - But almost always can blind end point determination

Blinding

- Single-blind: Patient does not know treatment assignment, but investigator does
- Double-blind: Neither patient nor investigator knows the treatment assignment
- Often in an industry-sponsored trial, no one at the site knows the treatment assignment

Replication

- If the item you are measuring is highly variable then measuring it more than once will help make your measure more precise
  - Blood pressure, peak expiratory flow rate
- Often the mean value is used

Sample Selection

- Is the sample representative of the group you want to study?
- A random sample of the population is considered the best but not always possible to obtain
- 1936 political poll predicted Landon would beat Roosevelt.
- Generally people in clinical trials have better outcomes

Randomization

- The process by which each subject's treatment assignment is determined by a random mechanism
- Neither the subject nor investigator should know the treatment assignment before the subject's decision to enter the study

Why Randomize?

- Removes investigator bias
- Tends to produce groups that are comparable with respect to known or unknown prognostic (risk) factors
- Guarantees the validity of statistical tests of hypotheses

Randomization Details

- Often perform stratified randomization
  - Randomize to treatments within site, age, or other prognostic factor(s)
- Usually do random permuted blocks to guarantee equal number of subjects in each treatment group for every 4 (say) subjects
  - E.g., block size of 4 for 2 treatment groups
    - Block 1: A B B A
    - Block 2: B B A A
    - Block 3: etc.

Randomization Details

Coordinator perspective

Subject ID	Randomization #
01-0B01	385
01-0B03	283
01-0B04	888
01-0B08	102

Research Pharmacist perspective

Subject ID	Rando #	Treat-ment
01-0B01	385	A
01-0B03	283	B
01-0B04	888	B
01-0B08	102	A

Randomization: Baseline / Entry Issues

- Day of randomization is time when groups are comparable (on average)
- All "baseline" evaluations should be completed at time of randomization or before (not after)
- All information necessary to establish eligibility must be available at time of randomization

Registration / Randomization

1. Obtain patient identifier (e.g., initials, hospital number)
2. Obtain informed consent
3. Check all inclusion/exclusion criteria
4. If above 3 criteria are met, assign study identifier for patient (subject ID)
5. Randomize – provide treatment code or authorize pharmacy to release blinded medications
6. Initiate intervention without delay (to avoid death or withdrawals before treatment starts)

Note: "Intent-to-treat" analysis requires comparison of groups regardless of what actually happens after randomization

Power and Sample Size

http://relevantinsights.com/tag/sample-size

The Art of Sample Size Considerations

- From a 6-month investigation for an adaptive clinical trial in cardiovascular disease to identify the dose of a new drug that is equivalent to a comparator
- To a 5-minute calculation for a placebo-controlled randomized clinical trial in asthma

Inadequate description of power / ss

- Not usually enough to say
  - "Everybody else uses 12, so we did too"
  - "We have enough power" without the details

Sample size

- replicates: need to be clear whether and how many you are using
- subjects: need to be clear what the number used in each group will be (if you know)
- should provide some rationale for this number
- the standard approach is based on a power analysis

Elements of a Power Statement

- effect size
- variability -- usually reported as standard deviation
- statistical characteristics:
  - alpha (two-sided)
  - power
- which gives you the magic number N

Sample size

- need to account for potential problems with the study
  - Dropouts or loss to follow-up
  - Missing measurements

Why is Sample Size Important?

- Statistical characteristics of your study depend upon N
  - precision of estimates
  - power (true negative rate) of comparisons
- Credibility, both from lay & scientific community
- Cost

Two Common Study Objectives

- Estimation
  - How many subjects with colon cancer do I need to estimate the prevalence of KRas?
  - How many US images and radiologists (reviewers) do I need to show the reliability of a new staging method for liver disease in cystic fibrosis subjects?
- Hypothesis Testing
  - How many subjects do I need to study to have sufficient power to detect whether early LVAD implantation is better than OMM for DT CHF patients?

Approach differs depending on purpose of the trial

- Phase I: First in Human
  - N often chosen without much statistical justification, usually quite small studies
  - Goal is often to find the Maximum Tolerated Dose

Approach differs depending on purpose of the trial

- Phase II: Is the drug active? What's the correct dose?
  - Goal often estimation of efficacy endpoint, so focus on degree of precision
  - Sample size often chosen so that a confidence interval for a probability of response is a particular width

Estimation: Example

- Fetal sonographer wants to determine the mean crown-rump length in a group of pregnancies
- Wants the limits of the 95% CI to be no more than 1 mm above or 1 mm below the mean crown-rump length of the group
- Previous studies provide some important information, SD of measurements is 3 mm
- Using this information, 35 fetuses should be examined

Approach differs depending on purpose of the trial

- Phase III: Is the new treatment efficacious?
  - Confirmatory studies that compare two (or more) interventions
  - Focus in this workshop:
    - Two treatment arms
    - One primary endpoint
    - Two-sided inference
    - Superiority (experimental treatment is better than placebo or better than active comparator)

Hypothesis Testing

- We want to be able to state with confidence that a treatment comparison is "real" and not due to chance
- The hypothesis testing framework allows us to make a decision regarding whether there is sufficient evidence from our data
- We set up a null and alternative hypothesis

Hypothesis Testing

- Null hypothesis
  H_0: true mean score for treatment A = true mean score for treatment B
- Alternative hypothesis
  H_A: true mean score for treatment A ≠ true mean score for treatment B

Hypothesis Testing Framework

" … the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis."

R.A. Fisher, 1935

Hypothesis Testing

Statistical Decision	True State of the Null Hypothesis	
	H_0 True	H_0 False
Reject H_0	Type I error (α)	Correct (1-β)
Do not reject H_0	Correct	Type II error (β)

Hypothesis Testing

- Sample size is important primarily because of its effect on statistical *power*
  - probability that a statistical test will indicate a significant difference when there truly is one

Ingredients for Sample Size

- Type I error, α
- Power = 1 – Type II error, 1 - β
- Treatment difference (and SD for continuous outcome), δ (and σ)
- Sample size, N per group

$$N = (Z_{1-\alpha/2} + Z_{1-\beta})^2 (2\sigma/\delta)^2$$

where Z_x = critical value from the standard normal distribution

Additional Issues

- Equivalence and non-inferiority trials
- Multiplicity (multiple comparisons issues)
  - Multiple primary endpoints
  - Multiple treatment groups (>2)
  - Multiple analyses → interim analyses, sample size re-estimation
- Stratification
- Drop-out
- Compliance

How to Minimize Sample Size*

- Use continuous measurements instead of categories
- Use more precise measurements
- Use paired measurements
- Expand minimum expected difference

- Browner WS, Newman TB, Cummings SR, Hulley SB. Estimating sample size and power. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing clinical research: an epidemiologic approach. 2nd ed. Philadelphia, Pa: Lippincott Williams & Wilkins, 2001; 65–84.

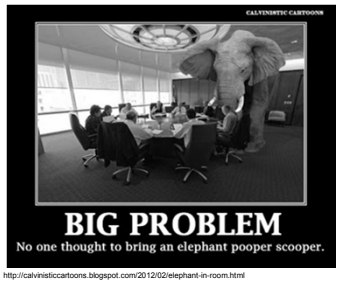
Data Collection / Procedures

- Want to show that data will be collected in standardized, reliable and valid way
- Table useful (schedule of evaluations), list all study variables (and all should be used in analyses)
- Validated instruments used?
- Biomarkers / assays have sufficient sensitivity, specificity, reliability?
- Include training and standardization, as well as quality assurance measures

Data Collection / Procedures

- If your study is going to need an IND or IDE then then how the data will be collected is more important
- Health Insurance Portability and Accountability Act (HIPAA)
- Title 21 CFR Part 11 compliance
- Good Clinical Practice (GCP)

Missing an Obvious Problem



http://calvinisticcartoons.blogspot.com/2012/02/elephant-in-room.html

Examples

- Missing data
- Multiple tests
- Bias
- etc.

Data Analysis

- if possible, mention plotting your data
  - mention a couple of specific plots as examples
- looking at data is always the way to begin
- be simple and straight forward
  - use techniques that you can do yourself
- be brief
  - the more details you give the more chance you have of making a mistake

Data Analysis

- do not get too fancy
  - this raises concerns, unless you have a statistician on the grant
- if you attempt to fake it, it will probably be obvious to the reviewer
  - never ever bluff

What You Need to Remember

Remember (1)

- **Hypothesis testing** is a formal method to make *statistical inferences* about the results
- identifies "signals" from "noise"
- there are two different ways of making an error in hypothesis testing: deciding there is a signal when there is not one, or deciding that there is no signal when there is one
- based on the P-value

Remember (2)

- P-values
- are based on the ratio: “effect” / “noise” assuming that the null hypothesis were true
- assumes that there are no biases in the results
- makes many other assumptions as well
- even more assumptions involved when the results are based on modeling

Statistical significance does not imply clinical importance.

Clinical importance does not imply statistical significance.

Remember (3)

- ask someone to read the grant who has not read it before
- take their comments seriously
- give yourself enough time to
  - get over the annoyance that someone raises issues
  - fix the grant before submitting it

Remember (4)

- Grant must be clear
- what you are doing
- why you are doing it
- why what you are doing answers the question

Remember (5)

- if using people (or samples) the number needs to be clear, and have a reasonable justification
- if doing data analysis, it should be straightforward unless you are working with a statistician (or are yourself an expert)

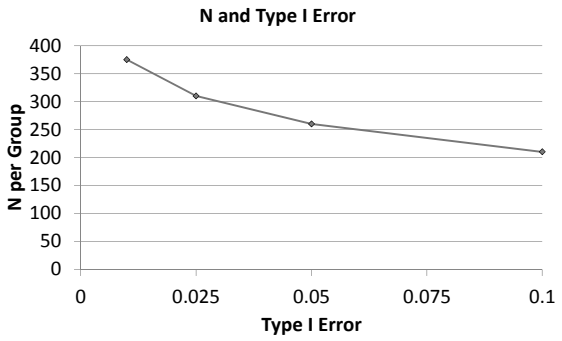
Focus on the study design

Example

- Trial of two treatments for fecal incontinence after 3rd-4th degree anal sphincter tears
- *Endpoint:* prevalence of FI 6 months post-partum
- *Treatment Difference:*
  - Control group: 17% based on CAPS pilot study
  - Want to detect a 50% reduction in prevalence for the new treatment intervention
- *Type I error: .01, .025, 0.05, 0.10*
- *Power: 50% to 95%*
- *N varies based on other considerations*

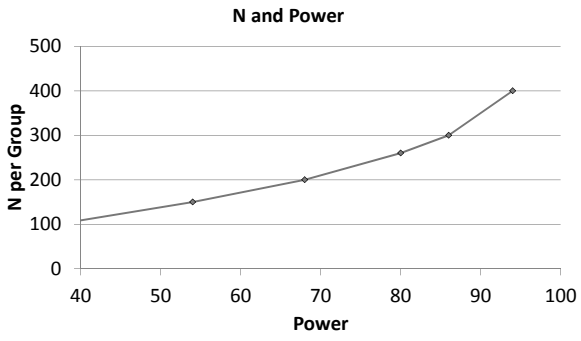
Phase III trials: sample size considerations

- Type I error: what's an acceptable false positive rate?
  - α = Pr(reject H_0 of no rx diff| H_0)
  - Usually α = 0.05 (2-sided) or 0.025 (1-sided)
  - The smaller the level, the larger the N



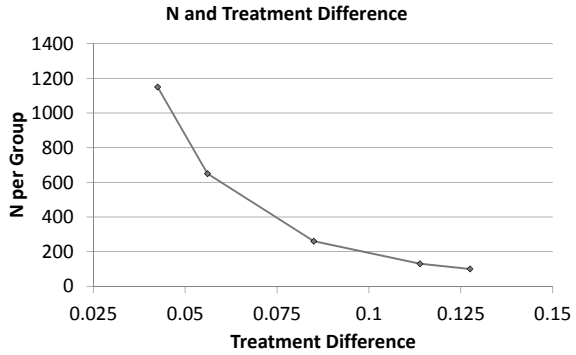
Phase III trials: sample size considerations

- Power: with what certainty do you want to detect a treatment difference?
  - 1-β = Pr(reject H_0 of no Rx diff| H_A)
  - Usually power = 80% or 90%; unethical to design a study with low power?
  - The greater the power, the larger the N



Phase III trials: sample size considerations

- Treatment difference: expected or desired difference between two treatments
  - δ = smallest difference likely to be of importance for clinical practice
  - Often called MID = minimally important difference or MCID = minimal clinically important difference
  - The smaller the Rx difference, the larger the N



Phase III trials: sample size considerations

- Often there is a reasonable estimate of the treatment effect in the control arm, based on previous placebo-controlled RCTs or on previous studies of the active comparator
  - Then focus is on what magnitude of benefit is needed to justify changing clinical practice to a new agent
  - Note: this decision is influenced not only by efficacy considerations, but by safety (adverse events), cost, health impact of the disease, etc.
- Assess the literature to double-check your estimates! If not great basis for your estimates, consider a pilot study

Phase III trials: sample size considerations

- For studies with continuous outcomes, an estimate of the variance is also needed
  - If it's difficult to come up with treatment difference (δ), it's even more difficult to justify an estimate of the standard deviation, σ!
  - The greater the variability, the larger the N

Phase III trials: sample size considerations

- Recommendation:
  - Perform comprehensive literature review (meta-analyses can be our friend)
  - Use CIs for estimates of treatment effects and their variability
  - Explore sensitivity of N to a range of δ (and σ), since α and power are often set

Sensitivity of N to Relative Risk & Control Arm Prevalence