

What you need to know before talking to your statistician about sample size

Sharon D. Yeatts, Ph.D.

Associate Professor of Biostatistics

Data Coordination Unit

Department of Public Health Sciences

Medical University of South Carolina

yeatts@musc.edu



Disclosures

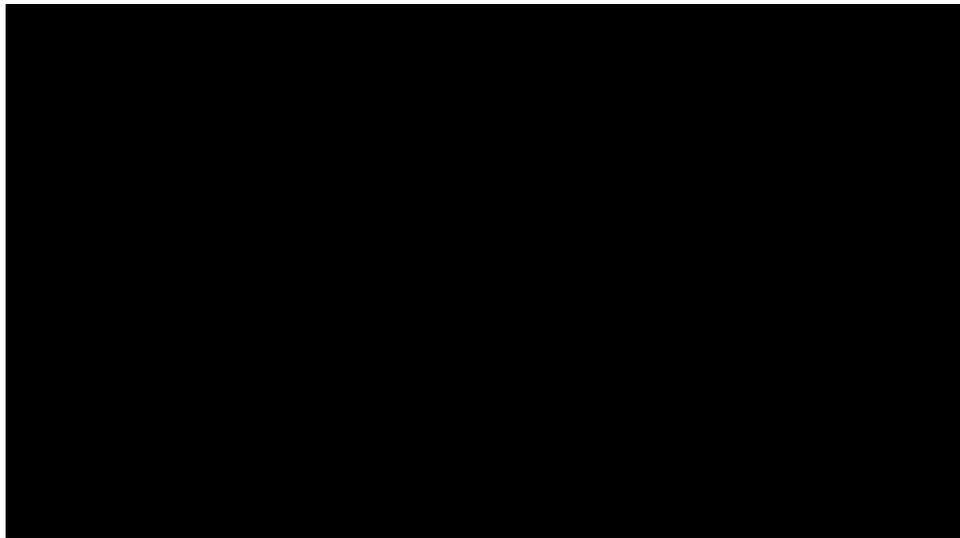
PRISMS (A Phase 3b, Double-blind, Multicenter Study to Evaluate the Efficacy and Safety of Alteplase in Patients with Mild Stroke: Rapidly Improving Symptoms and Minor Neurologic Deficits)

- I receive consultant fees from Genentech for my role on the Steering Committee.

This presentation will not include information on unlabeled use of any commercial products or investigational use that is not yet approved for any purpose.

Conversations between Biostatisticians and Scientists

- https://www.youtube.com/watch?v=Hz1f_yhVOjr4
- <https://www.youtube.com/watch?v=kMYxd6QeAss>



Disclaimer

This talk is geared towards the design of a superiority trial:

- a two arm, randomized trial, where the intention is to show that one treatment is better than the other with regard to the primary outcome
- where the deliverable is the pvalue from a hypothesis test

There are many other trial designs, where

- the objective of the trial might be something other than a demonstration of superiority (futility, non-inferiority, etc)
- the endgame might be something other than a hypothesis test (estimate of precision, selection, dose-finding)
- certain factors described here will be more/less/not relevant at all

Why Worry about Power/Sample Size?

- Provides assurance that the trial has a reasonable probability of being conclusive
- Allows one to determine the sample size **necessary**, so that resources are efficiently allocated
- Ethical Issues
 - Study too large implies some subjects needlessly exposed, resources needlessly spent
 - Study too small implies potential for misleading conclusions, unnecessary experimentation

Factors in Determining Sample Size

- Level of significance α (pre-specified)
- Power (targeted)
- Minimum Scientifically Important Difference
- Variability in the response
- Experimental Design



First things first ...

- What's the research question?
- Should be stated in the form of testable hypotheses.
 - What's the population?
 - What's the response variable?
 - What are you comparing?
 - At what timepoint?



2 states of nature: 2 hypotheses

- Alternative hypothesis (H_A)
 - The statement we hope to be able to conclude (ie, the intervention affects outcome).
 - The statement about a population that is true if the null hypothesis is not true.
- Null hypothesis (H_0)
 - Nothing unusual is occurring (ie, the intervention does not affect outcome).
 - The statement we hope to contradict with data.



Proof by Contradiction

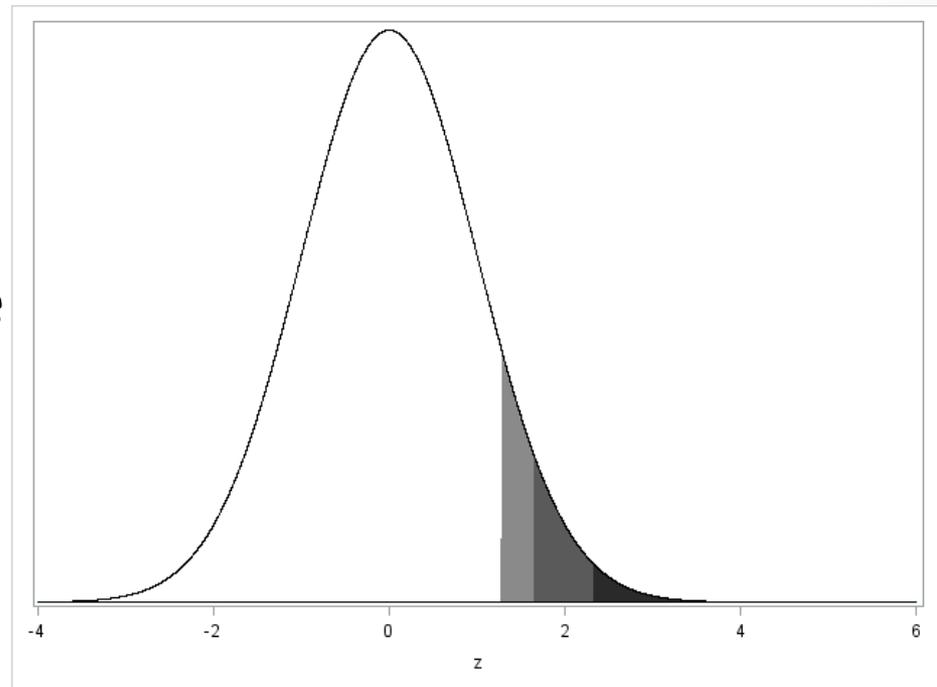
- We conceptualize nature as being in one state or the other.
 - Null hypothesis
 - Alternative hypothesis
- We collect data from the real world.
 - If the data clearly contradict the null hypothesis, we trust the data and conclude that the alternative must be true.
 - If random noise, measurement error or chance can account for observations, then there is no need to formulate a more complicated explanation.

Four Possible Outcomes

	True State of the Null Hypothesis	
Decision	H_0 TRUE (intervention doesn't work)	H_0 FALSE (intervention works)
Reject H_0 (and conclude H_A , that the intervention works)	Type I error Level of Significance α	Power
FTR H_0 (there is insufficient evidence to conclude H_A , that the intervention works)		Type II error β Related to Power

Level of Significance α

- **Interpretation: Even when the null hypothesis is true , we're going to conclude the alternative $\alpha(100)\%$ of the time.**
- Prespecified in trial design
- How often are you willing to accept a false positive?
 - One-sided α 0.10
 - One-sided α 0.05
 - One-sided α 0.01



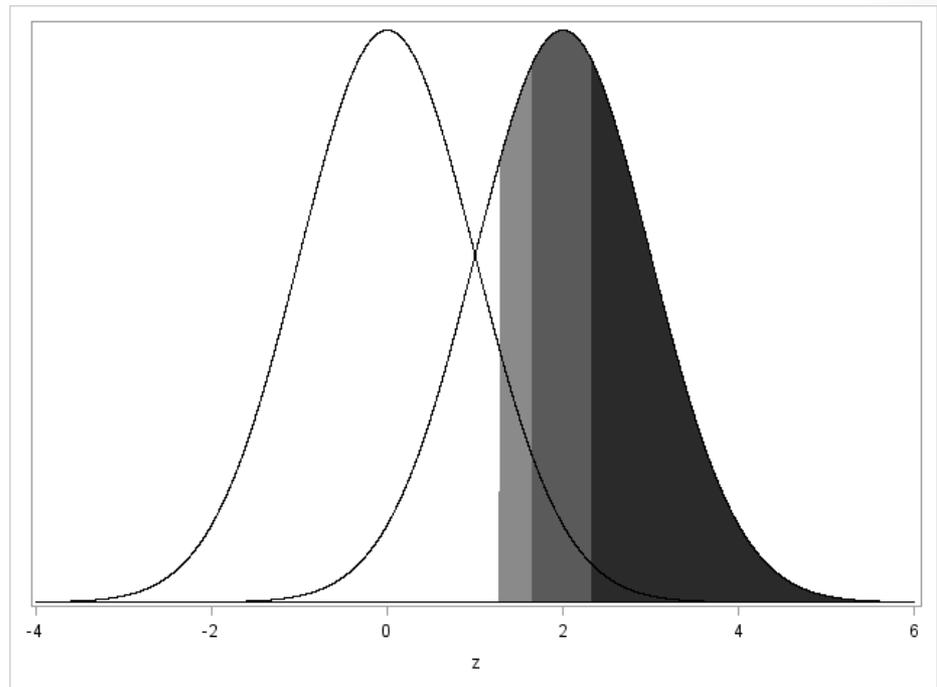


Power

- The probability that a statistical test will reject H_0 when H_0 is false
 - Power = $1 - \beta$
- **Interpretation:**
For a given value under the alternative hypothesis, we're going to reject the null in favor of the alternative $(1 - \beta)100\%$ of the time.
- **Related to the sample size**

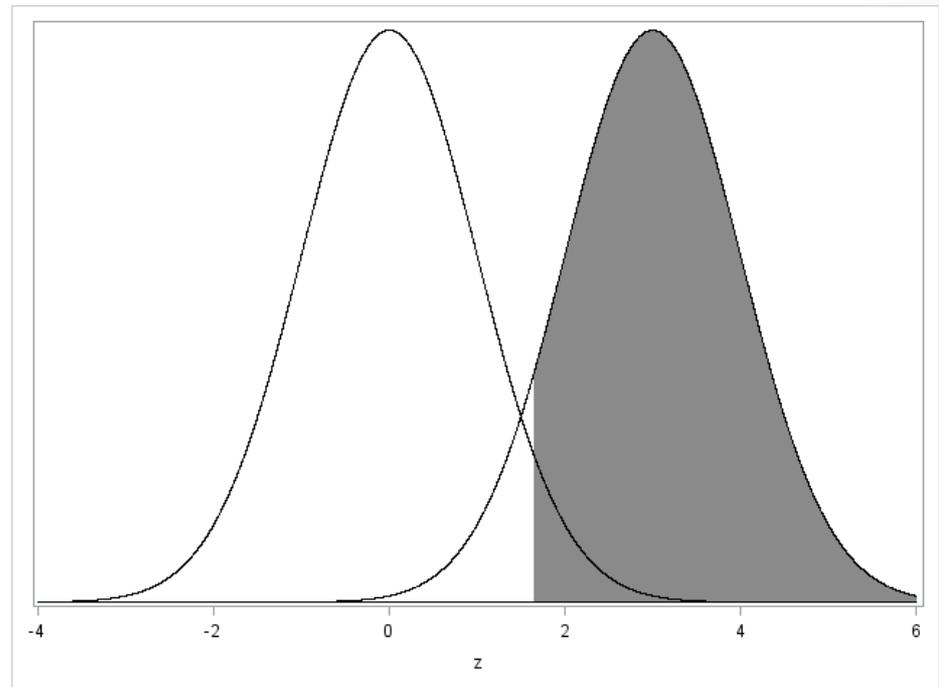
Impact on Power

- Level of Significance
 - One-sided α 0.10
 - One-sided α 0.05
 - One-sided α 0.01



Minimum Scientifically Important Difference

- the smallest difference which would mandate, in the absence of serious side effects and/or excessive cost, a change in scientific practice/ understanding
- “Larger the difference, smaller the sample size”

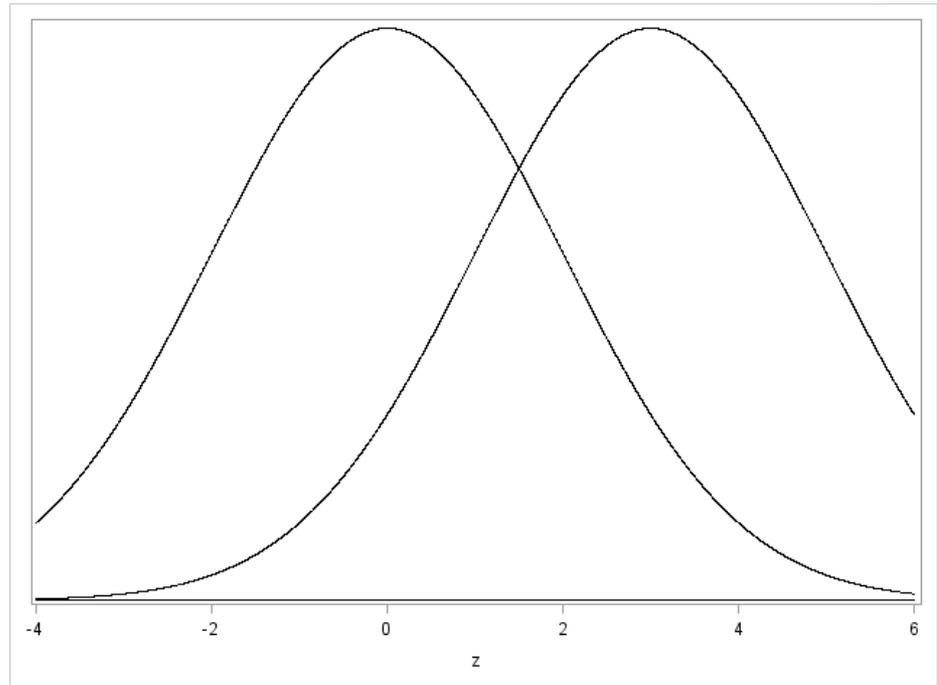


Variability

- Is the outcome continuous or categorical?
- Continuous
 - Need estimate of standard deviation/variance
- Dichotomous
 - Need estimate of control rate

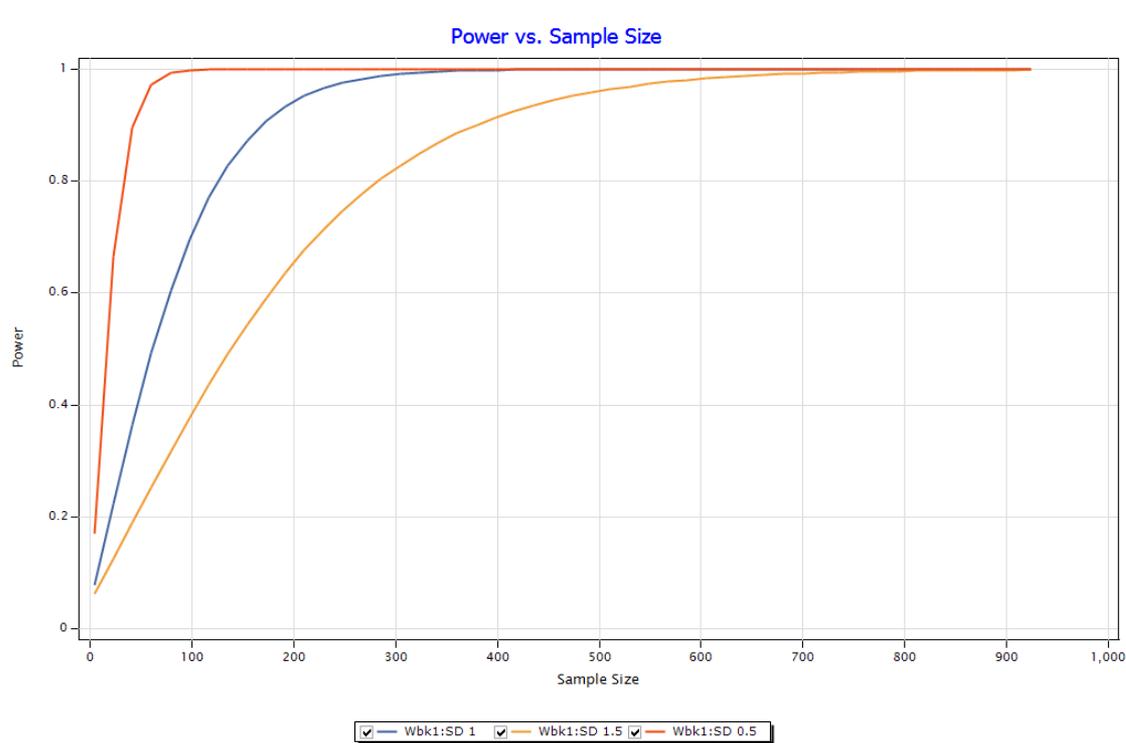
Variability

- “Larger the difference, smaller the sample size” ignores contribution of variability



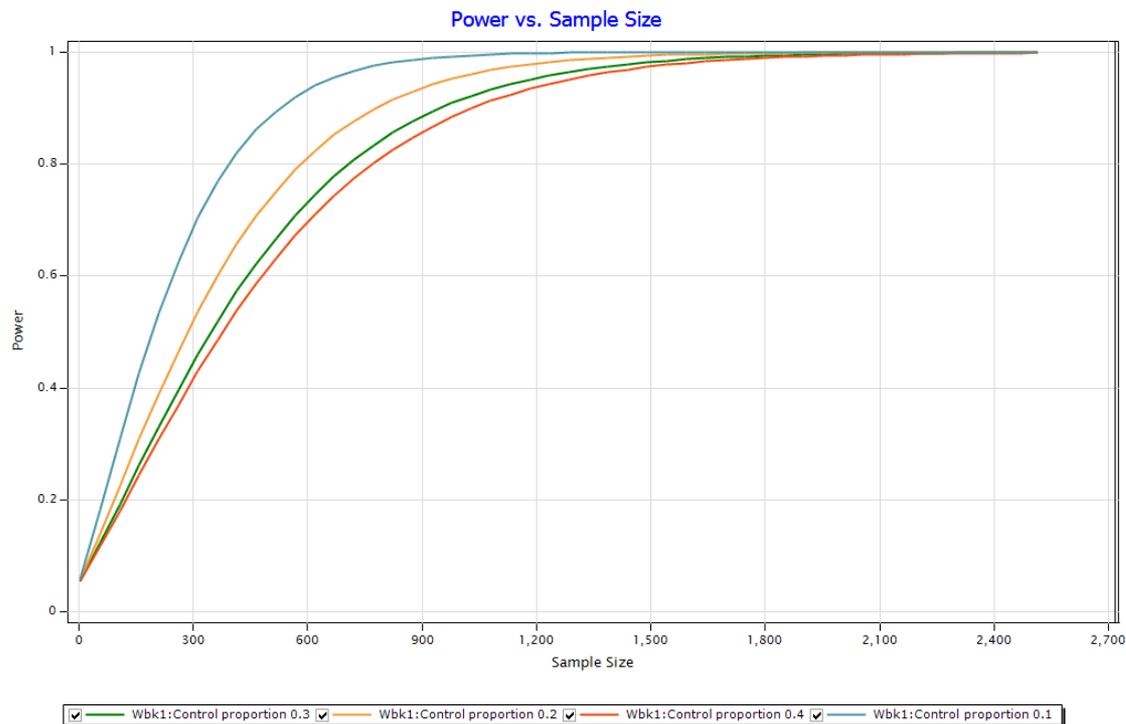
Example: Continuous Outcome

- Phase 2 clinical trial in subjects with acute ICH
- Intervention intended to impact health-related quality of life, as assessed by NeuroQOL
- Assume MCID 0.5 units



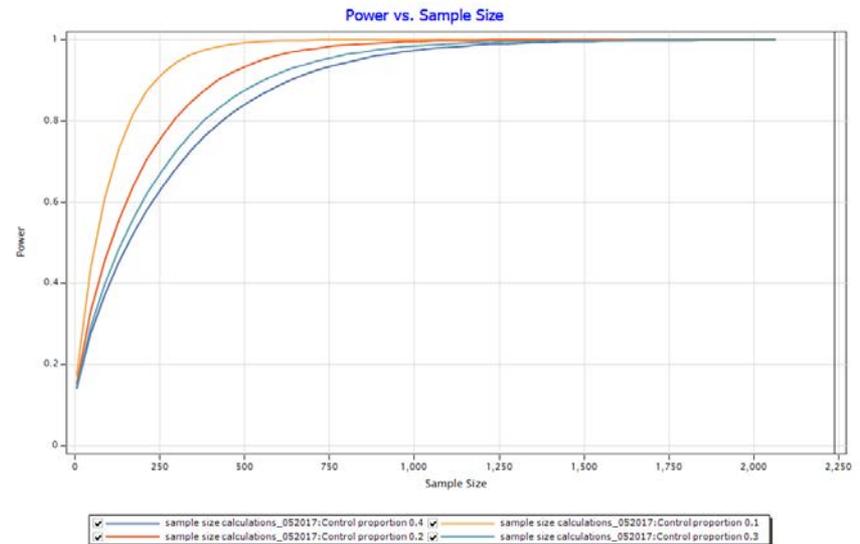
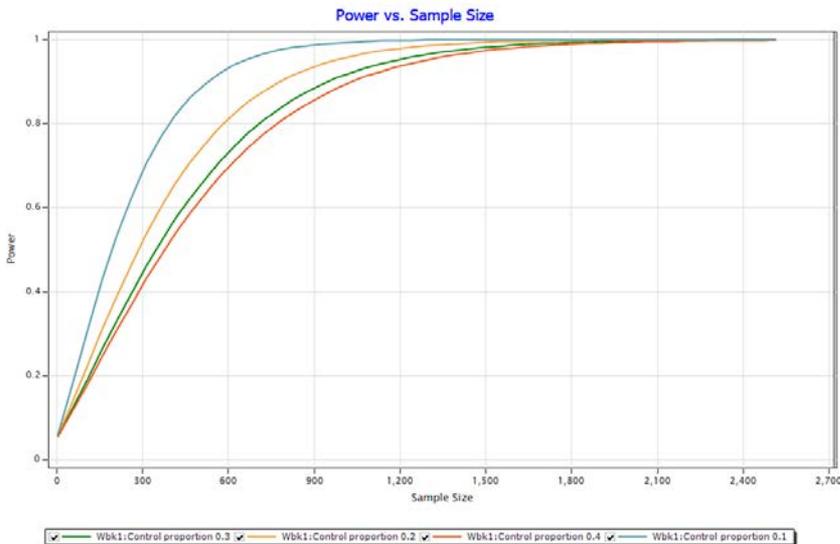
Example: Binary Outcome

- Phase 2 clinical trial in subjects with acute ICH
- Intervention intended to impact functional independence (mRS 0-2)
- Assume MCID absolute 10%



Example: Binary Outcome

- Phase 2 clinical trial in subjects with acute ICH
- What if we want to show that the effect is less than 10%, in order to discard the intervention if it is futile?



Before asking about sample size ...

... be prepared to talk about ...

- Level of significance α (set)
- Power (target)
- Expected variability in response
 - based on relevant clinical literature
 - Better yet, a range of plausible values
- Minimum Scientifically Important Difference
 - what's the smallest difference which will change practice?
 - If the sample size proves to make the trial not feasible, there's room for compromise.
- Experimental Design

Experimental Design Considerations

- Controls
 - Do you need a concurrent control arm?
 - Can you make use of historical controls?
 - Can subjects serve as their own control?
- Are there multiple questions which can be answered in the same design?
- Is a hypothesis test the best way to achieve your goal?
 - Dose-finding
 - Selection
- Sample size calculation depends on the method of analysis



Justifying Sample Size

- How many subjects?
- Is the sample size adequate for testing the hypothesis?
- How was the sample size determined?
- Do we need more?
- Can we answer questions with fewer?



Sample Size: Keeping It Small

- Study continuous rather than binary outcome (if variability does not increase)
 - change in tumor size instead of recurrence
- Study surrogate outcome where effect is large
 - cholesterol reduction rather than mortality



Sample Size: Calculation

- Calculate N
 - specify difference to be detected
 - specify variability (continuous) or control %
 - Specify target power

OR

- Calculate detectable difference:
 - specify N
 - specify target power
 - specify variability (continuous) or control %

Sample Size Software

- Freeware on the Web
 - <http://www.stat.uiowa.edu/~rlenth/Power/>
 - http://hedwig.mgh.harvard.edu/sample_size/size.html
 - <http://www.bio.ri.ccf.org/power.html>
 - R
- Purchase Software
 - <http://www.powerandprecision.com/>
 - Nquery: www.statsolusa.com
 - PASS
- Statistical collaborators will have lots of options available to them.

Sample Size Software

From the Iowa website:

I receive quite a few questions that start with something like this:
"I'm not much of a stats person, but I tried [details...] – am I doing it right?"

Please compare this with:

"I don't know much about heart surgery, but my wife is suffering from ... and I plan to operate ... can you advise me?"

Folks, just because you can plug numbers into a program doesn't change the fact that if you don't know what you're doing, you're almost guaranteed to get meaningless results -- if not dangerously misleading ones. Statistics really is like rocket science; it isn't easy, even to us who have studied it for a long time. Anybody who think it's easy surely lacks a deep enough knowledge to understand why it isn't! If your scientific integrity matters, and statistics is a mystery to you, then you need expert help. Find a statistician in your company or at a nearby university, and talk to her face-to-face if possible. It may well cost money. It's worth it.

Before asking about sample size ...

... be prepared to talk about ...

- Level of significance α (set)
- Power (target)
- Expected variability in response
 - based on relevant clinical literature
 - Better yet, a range of plausible values
- Minimum Scientifically Important Difference
 - what's the smallest difference which will change practice?
 - If the sample size proves to make the trial not feasible, there's room for compromise.
- Experimental Design
- Logistics

Logistics

- Anticipated recruitment
- Treatment crossovers
- Protocol nonadherence
- LTFU/consent withdrawal

Famous Last Words

- "You can't fix by analysis what you bungled by design."
 - Light, Singer and Willett
- To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of.
 - RA Fisher