Sample Size and Power

Laura Lee Johnson, Ph.D.
Statistician
National Center for Complementary and Alternative Medicine
johnslau@mail.nih.gov
Fall 2010

Misconceptions

• P-value = inferential tool? Yes
  – Helps demonstrate that population means in two groups are not equal

• Smaller p-value → larger effect? No
  – Effect size is determined by the difference in the sample mean or proportion between 2 groups

Misconceptions

• A small p-value means the difference is statistically significant, not that the difference is clinically significant. YES
  – A large sample size can help get a small p-value. YES, so do not be tricked.

• Failing to reject H₀ means what?
  – There is not enough evidence to reject H₀ YES
  – H₀ is true! NO NO NO NO NO!
Objectives

- Calculate changes in sample size based on changes in the difference of interest, variance, or number of study arms
- Understand intuition behind power calculations
- Recognize sample size formulas for the tests
- Learn tips for getting through an IRB

Take Away Message

- Get some input from a statistician
  - This part of the design is vital and mistakes can be costly!
- Take all calculations with a few grains of salt
  - “Fudge factor” is important!
- Round UP, never down (ceiling)
  - Up means 10.01 becomes 11
- Analysis Follows Design

Vocabulary

- Arm = Sample = Group
- Demonstrate superiority
  - Detect difference between treatments
- Demonstrate equally effective
  - Equivalence trial or a 'negative' trial
  - Sample size required to demonstrate equivalence larger than required to demonstrate a difference
- Demonstrate non-inferiority
  - Lots of issues
Superiority vs. Equivalence

Non-Inferiority

Vocabulary (2)

• Follow-up period
  – How long a participant is followed
• Censored
  – Participant is no longer followed
  • Incomplete follow-up (common)
  • Administratively censored (end of study)
• More in our next lecture
Outline

- **Power**
  - Basic Sample Size Information
  - Examples (see text for more)
  - Changes to the basic formula
  - Multiple comparisons
  - Poor proposal sample size statements
  - Conclusion and Resources

---

**Power Depends on Sample Size**

- Power = 1-β = P( reject H₀ | H₁ true )
  - “Probability of rejecting the null hypothesis if the alternative hypothesis is true.”
- More subjects ⇒ higher power

---

**Power is Affected by.....**

- Variation in the outcome (σ²)
  - ↓ σ² → power ↑
- Significance level (α)
  - ↑ α → power ↑
- Difference (effect) to be detected (δ)
  - ↑ δ → power ↑
- One-tailed vs. two-tailed tests
  - Power is greater in one-tailed tests than in comparable two-tailed tests
Power Changes

• 2n = 32, 2 sample test, 81% power, δ=2, σ = 2, α = 0.05, 2-sided test

• Variance/Standard deviation
  – σ: 2 → 1 Power: 81% → 99.99%
  – σ: 2 → 3 Power: 81% → 47%

• Significance level (α)
  – α: 0.05 → 0.01 Power: 81% → 69%
  – α: 0.05 → 0.10 Power: 81% → 94%

Power Changes

• 2n = 32, 2 sample test, 81% power, δ=2, σ = 2, α = 0.05, 2-sided test

• Difference to be detected (δ)
  – δ: 2 → 1 Power: 81% → 29%
  – δ: 2 → 3 Power: 81% → 99%

• Sample size (n)
  – n: 32 → 64 Power: 81% → 98%
  – n: 32 → 28 Power: 81% → 75%

• Two-tailed vs. One-tailed tests
  – Power: 81% → 88%

Power should be….?

• Phase III: industry minimum = 80%

• Some say Type I error = Type II error

• Many large “definitive” studies have power around 99.9%

• Proteomics/genomics studies: aim for high power because Type II error a bear!
Power Formula

• Depends on study design
• Not hard, but can be VERY algebra intensive
• May want to use a computer program or statistician

Outline

✓ Power
➢ Basic Sample Size Information
  • Examples (see text for more)
  • Changes to the basic formula
  • Multiple comparisons
  • Rejected sample size statements
  • Conclusion and Resources

Basic Sample Size

• Changes in the difference of interest have HUGE impacts on sample size
  – 20 point difference → 25 patients/group
  – 10 point difference → 100 patients/group
  – 5 point difference → 400 patients/group
• Changes in difference to be detected, \( \alpha \), \( \beta \), \( \sigma \), number of samples, if it is a 1- or 2-sided test can all have a large impact on your sample size calculation

Basic 2-Arm Study’s
TOTAL Sample Size = \( 2N = \frac{4(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\delta^2} \)
Basic Sample Size Information

- What to think about before talking to a statistician
- What information to take to a statistician
  - In addition to the background to the project

Nonrandomized?

- Non-randomized studies looking for differences or associations
  - Require larger sample to allow adjustment for confounding factors
- Absolute sample size is of interest
  - Surveys sometimes take % of population approach

Take Away

- Study’s primary outcome
  - Basis for sample size calculation
  - Secondary outcome variables considered important? Make sure sample size is sufficient
- Increase the ‘real’ sample size to reflect loss to follow up, expected response rate, lack of compliance, etc.
  - Make the link between the calculation and increase
- Always round up
  - Sample size = 10.01; need 11 people
Sample Size in Clinical Trials

- Two groups
- Continuous outcome
- Mean difference
- Similar ideas hold for other outcomes

Sample Size Formula Information

- Variables of interest
  - type of data e.g. continuous, categorical
- Desired power
- Desired significance level
- Effect/difference of clinical importance
- Standard deviations of continuous outcome variables
- One or two-sided tests

Sample Size & Data Structure

- Paired data
- Repeated measures
- Groups of equal sizes
- Hierarchical or nested data
- Biomarkers
- Validity (of what) studies
Sample Size & Study Design

- Randomized controlled trial (RCT)
- Block/stratified-block randomized trial
- Equivalence trial
- Non-randomized intervention study
- Observational study
- Prevalence study
- Measuring sensitivity and specificity

Outline

- Power
- Basic sample size information
- Examples (see text for more)
  - Changes to the basic formula
  - Multiple comparisons
  - Rejected sample size statements
  - Conclusion and Resources

Phase I Dose Escalation

- Dose limiting toxicity (DLT) must be defined
- Decide a few dose levels (e.g. 4)
- At least three patients will be treated on each dose level (cohort)
- Not a power or sample size calculation issue
Phase I (Old Way)

• Enroll 3 patients
• If 0 out of 3 patients develop DLT
  – Escalate to new dose
• If DLT is observed in 1 of 3 patients
  – Expand cohort to 6
  – Escalate if 0 out of the 3 new patients
do not develop DLT (i.e. 1/6 at that
dose develop DLT)

Phase I (cont.)

• Maximum Tolerated Dose (MTD)
  – Dose level immediately below the
    level at which ≥2 patients in a cohort
    of 3 to 6 patients experienced a DLT
• Usually go for “safe dose”
  – MTD or a maximum dosage that is
    pre-specified in the protocol
Number of pts with DLT | Decision
---|---
0/3 | Escalate one level
1/3 | Enroll 3 more at current level
0/3 + 0/3 (To get here a de-escalation rule must have been applied at the next higher dose level) | STOP and choose current level as MTD
1/3 + 0/3 | Escalate one level (unless a de-escalation rule was applied at next higher level, in which case choose current level as MTD)
1/3 + 1/3 or 2/3 or 3/3 | STOP* and choose previous level as MTD (unless previous level has only 3 patients, in which case treat 3 more at previous level)
2/3 or 3/3 | STOP and choose previous level as MTD (unless previous level has only 3 patients, in which case treat 3 more at previous level)

**Phase I Note**

- Implicitly targets a dose with Pr (Toxicity) ≤ 0.17; if at 1/3+1/3 decide current level is MTD then the Pr (Toxicity) ≤ 0.33
- Entry of patients to a new dose level does not occur until all patients in the previous level are beyond a certain time frame where you look for toxicity
- Not a power or sample size calculation issue

**Phase I**

- MANY new methods
- Several randomize to multiple arms
- Several have control arms
- Several have 6-15 people per arm
Phase II Designs

- Screening of new therapies
- Not to prove ‘final’ efficacy, usually
  - Efficacy based on surrogate outcome
- Sufficient activity to be tested in a randomized study
- Issues of safety still important
- Small number of patients (still may be in the hundreds total, but maybe less than 100/arm)

Phase II Design Problems

- Might be unblinded or single blinded treatment
- Placebo effect
- Investigator bias
- Regression to the mean

Phase II: Two-Stage Optimal Design

- Seek to rule out undesirably low response probability
  - E.g. only 20% respond (p0=0.20)
- Seek to rule out p0 in favor of p1; shows “useful” activity
  - E.g. 40% are stable (p1=0.40)
Phase II Example: Two-Stage Optimal Design

• Single arm, two stage, using an optimal design & predefined response
• Rule out response probability of 20% (H0: p=0.20)
• Level that demonstrates useful activity is 40% (H1: p=0.40)
• α = 0.10, β = 0.10

Two-Stage Optimal Design

• Let α = 0.1 (10% probability of accepting a poor agent)
• Let β = 0.1 (10% probability of rejecting a good agent)
• Charts in Simon (1989) paper with different p1 – p0 amounts and varying α and β values

Table from Simon (1989)
Phase II Example

- Initially enroll 17 patients.
  - 0-3 of the 17 have a clinical response then stop accrual and assume not an active agent
- If ≥ 4/17 respond, then accrual will continue to 37 patients

Phase II Example

- If 4-10 of the 37 respond this is insufficient activity to continue
- If ≥ 11/37 respond then the agent will be considered active
- Under this design if the null hypothesis were true (20% response probability) there is a 55% probability of early termination
Sample Size Differences

- If the null hypothesis ($H_0$) is true
- Using two-stage optimal design
  - On average 26 subjects enrolled
- Using a 1-sample test of proportions
  - 34 patients
  - If feasible
- Using a 2-sample randomized test of proportions
  - 86 patients per group

Phase II

- Newer methods are available
- Many cite Simon (thus, why we went through it)

Phase II Historical Controls

- Want to double disease X survival from 15.7 months to 31 months.
- $\alpha = 0.05$, one tailed, $\beta = 0.20$
- Need 60 patients, about 30 in each of 2 arms; can accrue 1/month
- Need 36 months of follow-up
- Use historical controls
Phase II: Historical Controls

- Old data set from 35 patients treated at NCI with disease X, initially treated from 1980 to 1999
- Currently 3 of 35 patients alive
- Median survival time for historical patients is 15.7 months
- Almost like an observational study
- Use Dixon and Simon (1988) method for analysis

Phase II Summary

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 arm</td>
<td>Small n</td>
<td>No control</td>
</tr>
<tr>
<td>1 arm 2-stage</td>
<td>Small n, stop early</td>
<td>No control, correct responder/non responder rules</td>
</tr>
<tr>
<td>Historical controls</td>
<td>Small n, some control</td>
<td>Accurate control ?</td>
</tr>
<tr>
<td>2(+) arm</td>
<td>Control</td>
<td>Larger n</td>
</tr>
<tr>
<td>8 arm</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Phase III Survival Example

- Primary objective: determine if patients with metastatic melanoma who undergo Procedure A have a different overall survival compared with patients receiving standard of care (SOC)
- Trial is a two arm randomized phase III single institution trial
Number of Patients to Enroll?

- 1:1 ratio between the two arms
- 80% power to detect a difference between 8 month median survival and 16 month median survival
- Two-tailed $\alpha = 0.05$
- 24 months of follow-up after the last patient has been enrolled
- 36 months of accrual
Phase III Survival

- Look at nomograms (Schoenfeld and Richter). Can use formulas
- Need 38/arm, so let's try to recruit 42/arm – total of 84 patients
- Anticipate approximately 30 patients/year entering the trial

Non-Survival Simple Sample Size

- Start with 1-arm or 1-sample study
- Move to 2-arm study
- Study with 3+ arms cheat trick
  - Calculate PER ARM sample size for 2-arm study
  - Use that PER ARM
  - Does not always work; typically ok
1-Sample N Example

- Study effect of new sleep aid
- 1 sample test
- Baseline to sleep time after taking the medication for one week
- Two-sided test, $\alpha = 0.05$, power = 90%
- Difference = 1 (4 hours of sleep to 5)
- Standard deviation = 2 hr

Sleep Aid Example

- 1 sample test
- 2-sided test, $\alpha = 0.05$, 1-$\beta = 90$
- $\sigma = 2$ hr (standard deviation)
- $\delta = 1$ hr (difference of interest)

\[ n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\delta^2} = \frac{(1.960 + 1.282)^2 \cdot 2^2}{1^2} = 42.04 \approx 43 \]

Short Helpful Hints

- In humans $n = 12-15$ gives somewhat stable variance
  - Not about power, about stability
  - 15/arm minimum good rule of thumb
- If $n < 20-30$, check t-distribution
- Minimum 10 participants/variable
  - Maybe 100 per variable
Sample Size: Change Effect or Difference

- Change difference of interest from 1 hr to 2 hr
- n goes from 43 to 11

\[ n = \frac{(1.960 + 1.282)^2 \cdot 2^2}{2^2} = 10.51 \approx 11 \]

Sample Size: Iteration and the Use of t

- Found n = 11 using Z
- Use t_{10} instead of Z
  - t_{n-1} for a simple 1 sample
- Recalculate, find n = 13
- Use t_{12}
- Recalculate sample size, find n = 13
  - Done
- Sometimes iterate several times

Sample Size: Change Power

- Change power from 90% to 80%
- n goes from 11 to 8
- (Small sample: start thinking about using the t distribution)

\[ n = \frac{(1.960 + 0.841)^2 \cdot 2^2}{2^2} = 7.85 \approx 8 \]
Sample Size: Change Standard Deviation

- Change the standard deviation from 2 to 3
- n goes from 8 to 18

\[ n = \frac{(1.960 + 0.841)^2}{3^2} = 17.65 \approx 18 \] 

Sleep Aid Example: 2 Arms
Investigational, Control

- Original design (2-sided test, \( \alpha = 0.05 \), 1-\( \beta \) = 90%, \( \sigma = 2 \) hr, \( \delta = 1 \) hr)
- Two sample randomized parallel design
- Needed 43 in the one-sample design
- In 2-sample need twice that, in each group!
- 4 times as many people are needed in this design

\[ n = \frac{2(Z_{\alpha/2} + Z_{\beta/2})^2 \sigma^2}{\delta^2} = \frac{2(1.960 + 1.282)^2 \cdot 2^2}{1^2} = 84.1 \approx 85 \rightarrow 170 \text{ total!} \]
Aside  5 Arm Study

• Sample size per arm = 85
• 85*5 = 425 total
  – Similar 5 arm study
  – Without considering multiple comparisons

Sample Size: Change Effect or Difference

• Change difference of interest from 1hr to 2 hr
• n goes from 170 to 44

\[ n = \frac{2(1.960 + 1.282)^2}{2^2} = 21.02 \approx 22 \rightarrow 44 \text{ total} \]

Sample Size Change Power

• Change power from 90% to 80%
• n goes from 44 to 32

\[ n = \frac{2(1.960 + 0.841)^2}{2^2} = 15.69 \approx 16 \rightarrow 32 \text{ total} \]
Sample Size: Change Standard Deviation

• Change the standard deviation from 2 to 3
• n goes from 32 to 72

\[ n = \frac{2(1.960 + 0.841)^2 \cdot 3^2}{2^2} = 35.31 \approx 36 \rightarrow 72 \text{ total} \]

Conclusion

• Changes in the difference of interest have HUGE impacts on sample size
  – 20 point difference → 25 patients/group
  – 10 point difference → 100 patients/group
  – 5 point difference → 400 patients/group

• Changes in difference to be detected, \( \alpha \), \( \beta \), \( \sigma \), number of samples, if it is a 1- or 2-sided test can all have a large impact on your sample size calculation

2-Arm Study’s TOTAL Sample Size = \( 2N = \frac{4(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\delta^2} \)

Live Statistical Consult!

• Sample size/Power calculation: cholesterol in hypertensive men example (Hypothesis Testing lecture)
• Choose your study design
  – Data on 25 hypertensive men (mean 220, s=38.6)
  – 20-74 year old male population: mean serum cholesterol is 211 mg/ml with a standard deviation of 46 mg/ml
Example

• Calculate power with the numbers given
• What is the power to see a 19 point difference in mean cholesterol with 25 people in
- Was it a single sample or 2 sample example?

Sample Size Rulers

JAVA Sample Size
Put in 1-Sample Example #s

- 1 arm, t-test
- Sigma (sd) = 38.6
- True difference of means = 220-211=9
- n=25
- 2 sided (tailed) alpha = 0.05
  - Power=XXXX
- 90% power
  - Solve for sample size n=XXXX

Move the Values Around

- Sigma (standard deviation, sd)
- Difference between the means

Put in 2-Sample Example #s

- 2 arms, t-test
- Equal sigma (sd) in each arm = 2
- 2 sided (tailed) alpha = 0.05
- True difference of means = 1
- 90% power
- Solve for sample size
Other Designs?

Sample Size: Matched Pair Designs

- Similar to 1-sample formula
- Means (paired t-test)
  - Mean difference from paired data
  - Variance of differences
- Proportions
  - Based on discordant pairs
Examples in the Text
• Several with paired designs
• Two and one sample means
• Proportions
• How to take pilot data and design the next study

Cohen's Effect Sizes
• Large (.8), medium (.5), small (.2)
• Popular esp. in social sciences
• Do NOT use
  – Need to think
• 'Medium' yields same sample size regardless of what you are measuring

Take Home: What you need for N
• What difference is scientifically important in units – thought, disc.
  – 0.01 inches?
  – 10 mm Hg in systolic BP?
• How variable are the measurements (accuracy)? – Pilot!
  – Plastic ruler, Micrometer, Caliper
Take Home N
- Difference (effect) to be detected ($\delta$)
- Variation in the outcome ($\sigma^2$)
- Significance level ($\alpha$)
  - One-tailed vs. two-tailed tests
- Power
- Equal/unequal arms
- Superiority or equivalence

Outline
- Power
- Basic sample size information
- Examples (see text for more)
- Changes to the basic formula/Observational studies
  - Multiple comparisons
  - Rejected sample size statements
  - Conclusion and Resources

Unequal #s in Each Group
- Ratio of cases to controls
- Use if want $\lambda$ patients randomized to the treatment arm for every patient randomized to the placebo arm
- Take no more than 4-5 controls/case

$$n_2 = \lambda n_1 \rightarrow \lambda \text{ controls for every case}$$

$$n_1 = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 (\sigma_1^2 + \sigma_2^2 / \lambda)}{\delta^2}$$
K 1 Sample Size Shortcut

- Use equal variance sample size formula: TOTAL sample size increases by a factor of \(\frac{(k+1)^2}{4k}\)
- Ex: Total sample size for two equal groups = 26; want 2:1 ratio
  \[26 \times \frac{(2+1)^2}{4 \times 2} = 26 \times \frac{9}{8} = 29.25 \approx 30\]
  - 20 in one group and 10 in the other

Unequal #s in Each Group: Fixed # of Cases

- Case-Control Study
- Only so many new devices
- Sample size calculation says \(n=13\) cases and controls are needed
- Only have 11 cases!
- Want the same precision
- \(n_0 = 11\) cases
- \(kn_0 = \#\) of controls

How many controls?

\[k = \frac{n}{2n_0 - n}\]
- \(k = \frac{13}{2 \times 11 - 13} = \frac{13}{9} = 1.44\)
- \(kn_0 = 1.44 \times 11 = 16\) controls (and 11 cases) = 27 total (controls + cases)
  - Same precision as 13 controls and 13 cases (26 total)
# of Events is Important

- Cohort of exposed and unexposed people
- Relative Risk = \( R \)
- Prevalence in the unexposed population = \( \pi_1 \)

Formulas and Example

- Risk of event in exposed group
- Risk of event in unexposed group
- \( \eta_1 = \frac{(Z_{\alpha,1} + Z_{\beta})^2}{2(\sqrt{R - 1})^2} \) = # of events in unexposed group
- \( n_2 = RN_1 \) = # events in exposed group
- \( n_1 \) and \( n_2 \) are the number of events in the two groups required to detect a relative risk of \( R \) with power \( 1 - \beta \)
- \( N = n_1 / \pi_1 \) = # subjects per group

# of Covariates and # of Subjects

- At least 10 subjects for every variable investigated
  - In logistic regression
  - No general theoretical justification
  - This is stability, not power
  - Peduzzi et al., (1985) unpredictable biased regression coefficients and variance estimates
- Principal component analysis (PCA) (Thorndike 1978 p 184): \( N \geq 10m + 50 \) or even \( N \geq m^2 + 50 \)
Balanced Designs: Easier to Find Power / Sample Size

• Equal numbers in two groups is the easiest to handle
• If you have more than two groups, still, equal sample sizes easiest
• Complicated design = simulations
  – Done by the statistician

Outline

✓ Power
✓ Basic Sample Size Information
✓ Examples (see text for more)
✓ Changes to the basic formula
  ➢ Multiple comparisons
  • Rejected sample size statements
  • Conclusion and Resources

Multiple Comparisons

• If you have 4 groups
  – All 2 way comparisons of means
  – 6 different tests
• Bonferroni: divide $\alpha$ by # of tests
  – $0.025/6 \approx 0.0042$
  – Common method; long literature
• High-throughput laboratory tests
DNA Microarrays/Proteomics

• Same formula (Simon et al. 2003)
  – $\alpha = 0.001$ and $\beta = 0.05$
  – Possibly stricter
• Simulations (Pepe 2003)
  – based on pilot data
  – $k_0 = \# \text{genes going on for further study}$
  – $k_1 = \text{rank of genes want to ensure you get}$
P[ Rank (g) $\leq k_0$ | True Rank (g) $\leq k_1$]

Outline

✓ Power
✓ Basic Sample Size Information
✓ Examples (see text for more)
✓ Changes to the basic formula
✓ Multiple comparisons
   ➢ Rejected sample size statements
• Conclusion and Resources

Me, too! No, Please Justify N

• "A previous study in this area recruited 150 subjects and found highly significant results ($p=0.014$), and therefore a similar sample size should be sufficient here."
  – Previous studies may have been 'lucky' to find significant results, due to random sampling variation
**No Prior Information**

- "Sample sizes are not provided because there is no prior information on which to base them."
  - Find previously published information
  - Conduct small pre-study
  - If a very preliminary pilot study, sample size calculations not usually necessary

**Variance?**

- No prior information on standard deviations
  - Give the size of difference that may be detected in terms of number of standard deviations

**Number of Available Patients**

- "The clinic sees around 50 patients a year, of whom 10% may refuse to take part in the study. Therefore over the 2 years of the study, the sample size will be 90 patients."
  - Although most studies need to balance feasibility with study power, the sample size should not be decided on the number of available patients alone.
  - If you know # of patients is an issue, can phrase in terms of power
Outline

✓ Power
✓ Basic Sample Size Information
✓ Examples (see text for more)
✓ Changes to the basic formula
✓ Multiple comparisons
✓ Rejected sample size statements
➢ Conclusion and Resources

Conclusions: What Impacts Sample Size?
• Difference of interest
  – 20 point difference → 25 patients/group
  – 5 point difference → 400 patients/group
• $\sigma$, $\alpha$, $\beta$
• Number of arms or samples
• 1- or 2-sided test

Total Sample Size 2-Armed/Group/Sample Test

\[
2N = \frac{4(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\delta^2}
\]

No Estimate of the Variance?
• Make a sample size or power table
• Make a graph
• Use a wide variety of possible standard deviations
• Protect with high sample size if possible
Top 10 Statistics Questions
10. Exact mechanism to randomize patients
9. Why stratify? (EMEA re: dynamic allocation
8. Blinded/masked personnel
   ➢ Endpoint assessment

Top 10 Statistics Questions
7. Each hypothesis
   ➢ Specific analyses
   ➢ Specific sample size
6. How / if adjusting for multiple comparisons
5. Effect modification

Top 10 Statistics Questions
4. Interim analyses (if yes)
   ➢ What, when, error spending model / stopping rules
   ➢ Accounted for in the sample size?
3. Expected drop out (%)
2. How to handle drop outs and missing data in the analyses?
Top 10 Statistics Questions

1. Repeated measures / longitudinal data
   - Use a linear mixed model instead of repeated measures ANOVA
   - Many reasons to NOT use repeated measures ANOVA; few reasons to use
   - Similarly generalized estimating equations (GEE) if appropriate

Analysis Follows Design

Questions → Hypotheses →
Experimental Design → Samples →
Data → Analyses → Conclusions

- Take all of your design information to a statistician early and often
  - Guidance
  - Assumptions

Resources  General Books

- Rosenthal (2006) *Struck by Lightning: The curious world of probabilities*
Resources: General/Text Books

• Simon et al. (2003) *Design and Analysis of DNA Microarray Investigations*. Springer Verlag

Sample Size Specific Tables

• Categorical data: Lemeshow et al. (1996) *Adequacy of sample size in health studies*. Wiley
• Sequential trials: Whitehead, J. (1997) *The Design and Analysis of Sequential Clinical Trials, revised 2nd. ed. Wiley*

Resources Articles

Resources Articles


Resources FDA Guidance

• http://www.fda.gov/cdrh/ode/odeot476.html (devices, non-diagnostic)
• http://www.fda.gov/cdrh/osb/guidance/1620.html (diagnostics)
• And all the ones listed before

Resources URLs

• Sample size calculations simplified
  – http://www.tufts.edu/~gdallal/SIZE.HTM
• Stat guide: research grant applicants, St. George’s Hospital Medical School (http://www.sgu.ac.uk/depts/cha/cha_research/stat_guide/guide.cfm)
  – http://tinyurl.com/2mh42a
• Software: nQuery, EpiTable, SeqTrial, PS (http://biostat.mc.vanderbilt.edu/wiki/index.php/Main/PowerSampleSize)
  – http://tinyurl.com/zoysm
• Earlier lectures
Questions?