Statistical Considerations and Methodologies

C. Michael Gibson, M.S., M.D.

Harvard Medical School

Chairman, PERFUSE Study Group

Founder and Chairman, WikiDoc & WikiPatient, The World's Open Source Textbook of Medicine Viewed 896 Million Times A Year

Decisions have been made

- You have defined your objectives
- You have chosen your study design
- What are the next steps?

Set up your Hypotheses

- Hypothesis testing is a method of determining
  - if what you are seeing is a meaningful difference
  - or are you seeing a difference due to chance

Hypothesis Testing

- Why do we care about hypothesis testing?
  - This is how we can make statistical conclusions about our data.
  - FDA requires it for safety or effectiveness analysis.

Hypothesis Testing Overview

- What is involved in setting up a hypothesis test?
  - An objective is needed (what do you want to prove)
  - α and β (Significance and Power levels)
  - Critical Difference
  - Measure of the variability (Standard Deviation) for continuous data (noted: σ)

Components of the Hypothesis

- H_0 and H_A
- H_A : The alternative hypothesis, this is what you want to prove
- H_0 : The null hypothesis, the opposite of what is stated in H_A

Alpha and Beta Errors

State of Nature	Decision: Reject H_0	Decision: Do not Reject H_0
H_0 is True	Type I Error (: False +)	Correct
H_0 is False	Correct	Type II Error (: False -)

Alpha and Beta Errors

- Alpha (α) is typically set at 0.05 resulting in 95% confidence.
  - Z_α is the notation used in the sample size calculations
- 1-β is called Power. Power is the probability of rejecting H_0 given H_0 is false.
  - β is typically set at 0.2 resulting in 80% power. A minimum of 80% power is desired.
  - Z_β is the notation used in the sample size calculations

Two-sided vs. One-sided

- Two-sided tests are used:
  - When you aren't sure of how the treatment will affect the patient
  - When you are trying to show improvement and you need to make sure that you are doing no harm as well
- One-sided tests are used:
  - When you want to be no worse than the control

Safety Objective

- Case Study of a Safety Objective
  - There is a new pacemaker and you want to make sure that it is as safe as the previous marketed pacemakers.
  - One way of testing for safety is to look at the complication rate a 3 months of the new pacemaker vs. a standard

Safety Objective

- The complication rate from the previous technology may be considered an Objective Performance Criteria (OPC).
- This means that the control group for this objective is a percentage at 3 months.

Safety Objective

- What do we know about the previous pacemaker studies (Historical Control)?
  - The overall observed rate in the Historical Control is 92%.
  - Of the individual studies in the Historical Control, the worst rate is 85%.
  - Our critical difference will then be
    92%-85%= 7%

Safety Objective

- Objective: To demonstrate the safety of the New Pacemaker.
- Hypothesis: The freedom from complications experienced with the New Pacemaker will be clinically equivalent (within 7%) to the Historical Control rate at 3 months.

H_0: $S_{\text{New Pacemaker}} \leq 85\%$ at 3 months

H_A: $S_{\text{New Pacemaker}} > 85\%$ at 3 months

Safety Objective

- Alpha and Power.
- We will use the standard of 95% significance and 80% power.
- The points from the Normal Distribution that correspond to the appropriate level of significance are:
  $Z_{0.95}$= 1.645 (one-sided)
  $Z_{0.8}$ = 0.842

Sample Size Calculations (Comparison to a Standard)

- Comparing a proportion to a standard

$$N = \frac{Z \sqrt{p_0(1 \quad p_0)} + Z \sqrt{p_a(1 \quad p_a)}^{\,2}}{\left(p_0 \quad p_a\right)^2}$$

- From our example:

$$N = \frac{\left[1.645\sqrt{0.85(1 \quad 0.85)} + 0.842\sqrt{0.92(1 \quad 0.92)}\right]^2}{(0.85 \quad 0.92)^2} = 136$$

Effectiveness Objective

- Your company has invented a new pacing lead that is supposed to reduce battery longevity by increasing the impedance of the lead.
- You want to prove that the impedance is clinically higher for the experimental product over the control.

Effectiveness Objective

- This study will be a randomized study with an experimental and control group.

- It has been determined that a clinically meaningful increase in longevity is 6 months. This translated to a 300 ohm difference (d).
- A conservative standard deviation (σ) from the control group is 400 ohms.

Effectiveness Objective

- Objective: To demonstrate the effectiveness of the pacing impedance.
- Hypothesis: The New Lead pacing impedance at 3 months will be clinically superior (by 300 ohms) to the control impedance at 3 months.

H_0: Mean $_{\text{New Lead Impedance}}$ = Mean $_{\text{Control Impedance}}$
H_A: Mean $_{\text{New Lead Impedance}}$ ≠ Mean $_{\text{Control Impedance}}$

Effectiveness Objective

- Alpha and Power.
- We will use the standard of 95% significance and 80% power.
- The points from the Normal Distribution that correspond to the appropriate level of significance are:
  $Z_{0.95}$= 1.96 (two-sided)
  $Z_{0.8}$ = 0.842

Comparing Two Samples

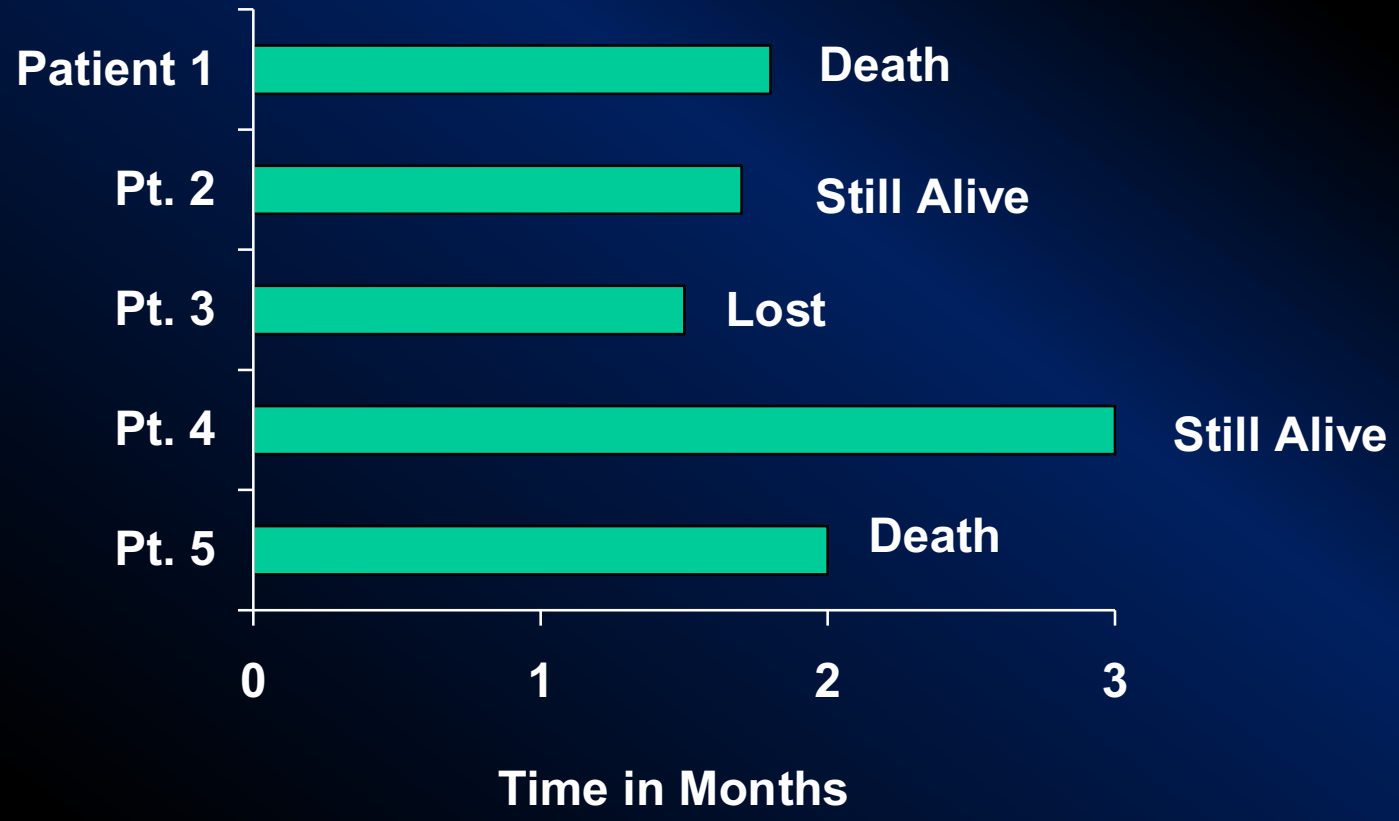
- Comparing Two Means (Total sample size)

$$N = \frac{4 \; Z_{/2}^{2} + Z^{2}}{d^{2}}$$

- From our example:

$$N = \frac{4(1.96 + 0.842)^{2} \, 400^{2}}{300^{2}} = 56 \;\; (28 \, per \; group)$$

Time to Event Statistics (Survival Analysis)



Survival Analysis

- The following is needed for each individual:
  - The start date and last known follow-up date
  - The date of an event (if it has occurred)
  - Status (lost to follow-up date, death date)
- From this information the follow-up time is calculated
- The survival probability is 1- the event probability.

Survival Analysis

- The cumulative survival for a certain data point is calculated as:
  - a percentage for each event
  - then multiplied together to obtain the cumulative percentage.
- The denominator that is used is the number remaining at that point in time.

Survival Analysis Example

Months	Survival	Cumulative Number Failed	Number Left
0	1.00	0	22
2	0.9545 1/22=0.0455	1	21
3	0.9091 1/21=0.0476 (0.9524) 0.9545*0.9524= 0.9091	2	20
4	0.9091	2	19

Stopping Rules

- There may be some instances where you may want to conduct interim analyses
- Repeated testing increases the overall significance level to > 0.05
- Repeated testing comes with a price as total required sample size increases

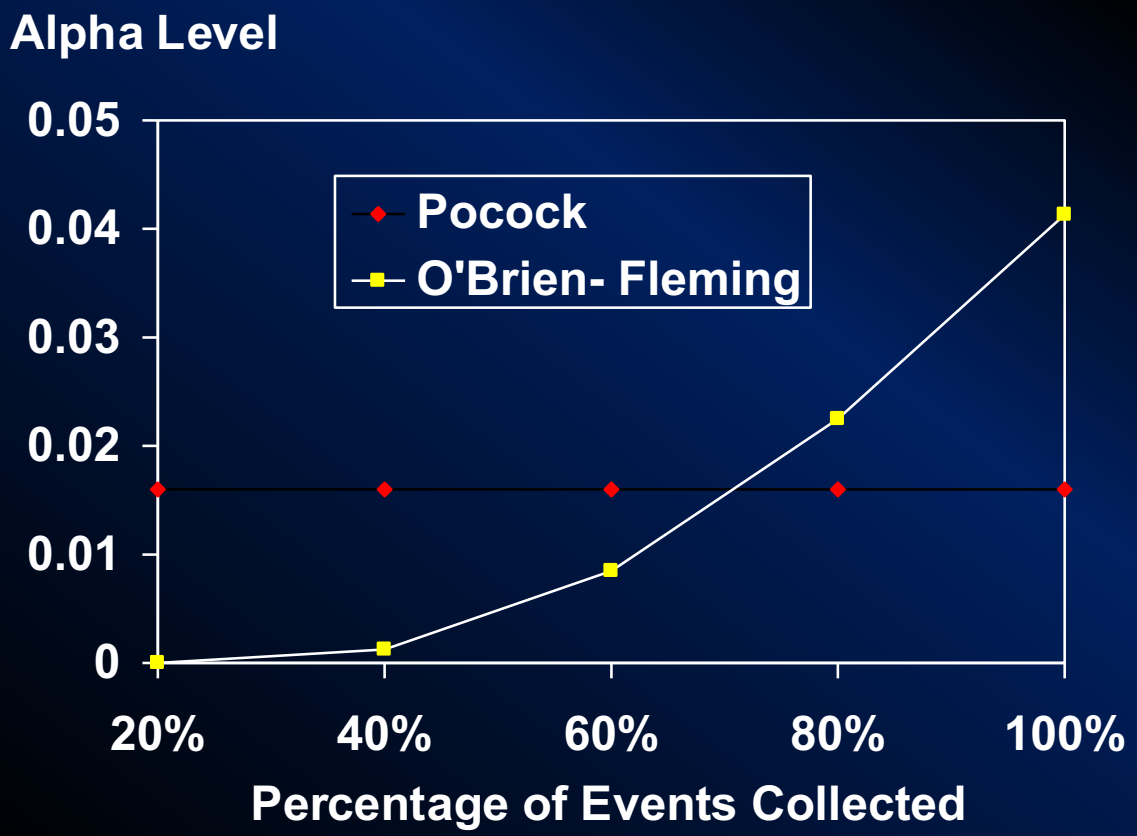
Stopping Rule Methods

- Statistical tests are conducted along interim points in the study and then the p-value is compared to an appropriate significance level.
- There are a number of methods available to “spend” your alpha level.

Spending Functions

- One way to look at the data would be 5 times, when you have 20%, 40%, 60%, 80%, and 100% of your events collected.
- Pocock would suggest the same significance level of 0.016 for each of the 5 tests.
- O'Brien Fleming would suggest a range of extremely small to 0.0413.

Spending Functions



PERFUSE

Meta Analysis

- A method used to combine data from various sources.
- This data is then used as a comparison group, “control”, for your study.

Conclusions

- Define what you are clinically wanting to prove.
- Choose your study design appropriately
- Use the objective to write a hypothesis
- Calculate the sample size using a clinically meaningful difference

Reference: Sample Size Calculations

- Comparing a proportion to a standard

$$N = \frac{Z_{/2}\sqrt{p_0(1-p_0)} + Z\sqrt{p_a(1-p_a)}}{(p_0 - p_a)}^2$$

- Comparing a mean to a standard

$$N = \frac{Z_{/2} + Z}{d}^2 2$$

Comparing Two Samples

- Comparing Two Proportions

$$N = \frac{2\ Z_{/2}\sqrt{p_0(1\ \ p_0)} + Z\ \sqrt{(p_c(1\ \ p_c) + p_e(1\ \ p_e)}^{\ 2}}{\left(p_c\ \ p_e\right)^2}$$

$$\textit{Where}\quad p_0 = \frac{(p_c + p_e)}{2}$$

- Comparing Two Means (Total N)

$$N = \frac{4\ \ Z_{/2}^{\ 2} + Z^{\ 2}}{d^2}$$

Equivalence for Two Samples (Blackwelder's Formula)

- Equivalence for two proportions (each group)

$$N = \frac{(Z_1 \quad + Z_1 \quad)^2 \left[p_0(1 \quad p_0) + p_1(1 \quad p_1) \right]}{\left(p_0 \quad p_1 \right)^2}$$

- In equivalence testing, usually $p_0 = p_1$ unless there is a known expected difference. The δ is the equivalence difference.

Equivalence for Two Samples (Blackwelder's Formula Adapted)

- Equivalence for Two Means (each group)

$$N = \frac{(z_1 \quad + z_1 \quad)^2 [\quad {}_1^2 + \quad {}_2^2]}{(\ \ )^2}$$

  - In equivalence testing, δ is the equivalence difference.

Book References

- Friedman, L.M., Furberg, C.D. and DeMets, D.L., Fundamentals of Clinical Trials, PSG Publishing Company, Inc. (1985)
- Pocock, S.J. Clinical Trials- A Practical Approach, John Wiley & Sons, Inc., (1983)
- Blackwelder, W.C., " 'Proving the Null Hypothesis' in Clinical Trials", *Controlled Clinical Trials*, 3:345-353, (1982)