Power Calculations

March 3, 2016

Joseph Massaro
Department of Biostatistics
jmm@bu.edu
Introduction

• A statistician’s role in a study begins well before data are ready for analysis

• Statistician assists in the “design” of the study prior to study start
  - Major design component: Determining appropriate sample size for the study
    • i.e., determine sample size large enough to represent population with reasonable confidence;
    • while not having a sample size be cumbersome from cost and/or time perspective.
Suppose have two treatment groups that we wish to compare on a dichotomous outcome in a clinical trial:

\[ H_0: p_E \geq p_C \] ("Null" hypothesis)
\[ H_1: p_E < p_C \] ("Alternative" hypothesis – what we hope to prove)

where \( p_E \) and \( p_C \) are the TRUE (but unknown) event rates of a negative outcome (e.g., Major Adverse Event) for experimental and control groups IN THE POPULATION.
Introduction

• We hope the null hypothesis is false in the population.

• We base our decision about whether the null hypothesis is false (whether to reject $H_0$) on a sample from the population.

• We want a sample large enough to represent the population with reasonable confidence (there’s always a chance we’ll obtain a sample not representative of the population, but we want to minimize this chance).
**Introduction**

"TRUTH" (Unknown):

<table>
<thead>
<tr>
<th>Decision:</th>
<th>$H_o$ True</th>
<th>$H_o$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept $H_o$</td>
<td>CORRECT DECISION</td>
<td>Type II ERROR</td>
</tr>
<tr>
<td>Reject $H_o$</td>
<td>Type I ERROR</td>
<td>CORRECT DECISION</td>
</tr>
</tbody>
</table>
### Introduction

#### “TRUTH” (Unknown):

<table>
<thead>
<tr>
<th>$H_0$ True</th>
<th>$H_0$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accept $H_0$</strong></td>
<td><strong>Reject $H_0$</strong></td>
</tr>
<tr>
<td>CORRECT DECISION</td>
<td>Type I ERROR</td>
</tr>
<tr>
<td>Type II ERROR</td>
<td>CORRECT DECISION</td>
</tr>
</tbody>
</table>

\[ \beta = P(\text{Type II Error}) = P(\text{accept } H_0 | H_0 \text{ False}) \]

\[ \alpha = P(\text{Type I Error}) \]

*Controlled by significance level; usually set to one-sided 0.025 for pivotal trials.*

\[ \text{Power} = 1 - \beta \]

*(set to at least 0.80 for Confirmatory studies)*
β and Power

- As $\alpha$ decreases, $\beta$ increases (and power decreases)

- As sample size ($n$) per group increases, power increases (if the null hypothesis is truly false)
  - For a confirmatory study: choose desired power; determine $n$ per group

- Power also depends on the true treatment effect
Example

\[ H_0: p_E \geq p_C \text{ ("Null" hypothesis)} \]
\[ H_1: p_E < p_C \text{ ("Alternative" hypothesis – what we hope to prove)} \]

• Free sample size calculation software (for basic but commonly required calculations) at:

http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize#Downloading_and_Installing_the_PS_Software
Studies that are analyzed by chi-square or Fisher's exact test

**Output**

**What do you want to know?**
- Sample size

**Case sample size for uncorrected chi-squared test**

**Design**
- **Matched or Independent?**
  - Independent
- **Case control?**
  - Prospective
- **How is the alternative hypothesis expressed?**
  - Two proportions
- **Uncorrected chi-square or Fisher’s exact test?**
  - Uncorrected chi-square test

**Input**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
</tr>
<tr>
<td>$p_0$</td>
<td>0.20</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.10</td>
</tr>
<tr>
<td>$m$</td>
<td>1</td>
</tr>
<tr>
<td><strong>power</strong></td>
<td>0.80</td>
</tr>
</tbody>
</table>

**Description**

[Description field]

[Calculate] [Graphs]
We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.2. If the true failure rate for experimental subjects is 0.1, we will need to study 199 experimental subjects and 199 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.
Studies that are analyzed by chi-square or Fisher's exact test

Output
What do you want to know? Sample size
Case sample size for uncorrected chi-squared test 330

Design
Matched or Independent? Independent
Case control? Prospective
How is the alternative hypothesis expressed? Two proportions
Uncorrected chi-square or Fisher's exact test? Uncorrected chi-square test

Input
\(\alpha\) 0.05  
\(p_0\) 0.175  
\(p_1\) 0.10  
\(m\) 1  

Calculate

Graphs

Description
We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.175. If the true failure rate for experimental subjects is 0.1, we will need to study 330 experimental subjects and 330 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.
Studies that are analyzed by chi-square or Fisher's exact test

**Output**

**What do you want to know?**

- Sample size

**Case sample size for uncorrected chi-squared test**

- 441

**Design**

- **Matched or Independent?**
  - Independent

- **Case control?**
  - Prospective

- **How is the alternative hypothesis expressed?**
  - Two proportions

- **Uncorrected chi-square or Fisher's exact test?**
  - Uncorrected chi-square test

**Input**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
</tr>
<tr>
<td>$p_0$</td>
<td>0.175</td>
</tr>
<tr>
<td><strong>power</strong></td>
<td>0.90</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.10</td>
</tr>
<tr>
<td>$m$</td>
<td>1</td>
</tr>
</tbody>
</table>

**Description**

We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.175. If the true failure rate for experimental subjects is 0.1, we will need to study 441 experimental subjects and 441 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.9. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.
Sample Size

- How do we come up with our assumptions of the “true” mean effect (i.e., true mean difference)?
  - Past Studies
  - Literature review (maybe someone else did similar work and published it)
  - Best guess?
  - What we “hope” is the effect (even though we may not have scientific evidence yet)

- Assumption should be “clinically” meaningful
Underpowering

• We do not want to “underpower” our study
  – I.e., we don’t want a sample size that is too small.

• This can happen if, when powering the study, we assume the effect size (e.g., treatment difference) is larger than it truly is
Underpowering

- E.g., if we powered previous example assuming an experimental treatment 10% MAE rate and a control treatment 20% MAE rate, we need 199 per treatment group.

- If, in reality (and unbeknownst to us) the true MAE rates are 10% and 17.5%, then in reality 199 per group yields only 58% power.
  - But at least we still have some power.
Evaluable vs. All Randomized

• The above sample size calculations yield number of evaluable subjects
  – i.e., subjects who will not prematurely withdraw and hence be available for follow-up

  – Add a certain % of patients to number of evaluable to obtain number of patients who should be enrolled

  – e.g., if 10% dropout rate is anticipated, enroll 
    \[ \frac{199}{0.90} = 222 \] patients/group.
Other Sample Size Software (not free)

- **SAS (PROC POWER)**
- **PASS (“Power and Sample Size”)**
- **nQuery**
- **Power and Precision**
  - [http://www.power-analysis.com/home.htm](http://www.power-analysis.com/home.htm)
Continuous Outcome

• A study is conducted to assess the effect of a new stem cell treatment on LVEF improvement at 4 months following acute MI

• Control group is a placebo (sham control procedure)

• $H_0: \mu_T \leq \mu_{PL}$ vs. $H_1: \mu_T > \mu_{PL}$

  where $\mu_T$ and $\mu_{PL}$ are mean improvement in LVEF % from baseline to 4 months for experimental treatment and placebo
Continuous Outcome

• Assumptions made in powering study:
  – True $\mu_T - \mu_{PL} = 5\%$
  – Standard deviation of change in LVEF = 8\% in each group
  – Desire power of 80\%
  – One-sided significance level of 0.025

• Note: We don’t need assumption of the true means; just the difference in means

• Note: We need an assumption of the variability (the more variability, the larger the sample size needed)
We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 8. If the true difference in the experimental and control means is 5, we will need to study 41 experimental subjects and 41 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this
<table>
<thead>
<tr>
<th>Objective</th>
<th>Methodology</th>
<th>Assumptions Needed for Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show superiority of one treatment over another on binary outcome rate</td>
<td>Chi-square test (or Fisher's Exact test)</td>
<td>Event rate in each treatment</td>
</tr>
<tr>
<td>Show superiority of one treatment over another on mean of outcome</td>
<td>Two-sample t-test</td>
<td>Difference between means; sd per treatment</td>
</tr>
<tr>
<td>Show that increase in a variable X increases/decreases the odds of the outcome event</td>
<td>Univariate Logistic Regression</td>
<td>Odds Ratio, binary outcome rate when X=0 (if X is binary) or at mean of X (if X is continuous)</td>
</tr>
<tr>
<td>Show that increase in a variable X increases/decreases the odds of the outcome event, after adjusting or other covariates</td>
<td>Adjusted Logistic Regression</td>
<td>Odds Ratio, binary outcome rate when X=0 (if X is binary) or at mean of X (if X is continuous), R-squared of X with other covariates</td>
</tr>
<tr>
<td>Show that increase in a variable X increases/decreases the average outcome of Y</td>
<td>Simple Linear Regression</td>
<td>Slope (average change in Y per 1 unit increase of X), sd of X, sd of Y or correlation of X and Y</td>
</tr>
<tr>
<td>Show that increase in a variable X increases/decreases the average outcome of Y, after adjusting for other covariates</td>
<td>Multiple Linear Regression</td>
<td>Slope (average change in Y per 1 unit increase of X), sd of X, sd of Y or correlation of X and Y, R-squared of X with other covariates</td>
</tr>
</tbody>
</table>
Multiple Primary Analyses

• Suppose there are two primary hypotheses; e.g.:
  \[ H_0: p_{E,SAF} \geq p_{C,SAF} \]
  \[ H_0: p_{E,EFF} \leq p_{C,EFF} \]
  \[ H_1: p_{E,SAF} < p_{C,SAF} \]
  \[ H_1: p_{E,EFF} > p_{C,EFF} \]
  (SAFETY) (EFFICACY)

• Suppose we need to reject BOTH nulls for the study to be considered successful
  – Each can be tested at a 0.05 level of significance
  – Power each endpoint at 80% could yield only 64% power to reject BOTH null hypotheses
  – We need to power each null with at least 90% power to have an 80% chance of rejecting BOTH null hypotheses
Multiple Primary Analyses

• Suppose there are two primary hypotheses; e.g.:

\[ H_0: p_{E,SAF} \geq p_{C,SAF} \quad \text{H}_0: p_{E,EFF} \leq p_{C,EFF} \]
\[ H_1: p_{E,SAF} < p_{C,SAF} \quad H_1: p_{E,EFF} > p_{C,EFF} \]

(SAFETY) (EFFICACY)

• Suppose study is a success if we reject AT LEAST ONE null hypothesis
  – Powering each hypothesis at 80% at alpha = 0.025 (=0.05/2) yields 80% power to reject at least one null hypothesis