Sample Size Determination
Introduction

- Integral part of vast majority of quantitative studies
- Important in ensuring validity, accuracy, reliability & scientific & ethical integrity
- Don’t give in to temptations of taking a shortcut
- Highly recommended to ask a professional statistician to conduct the sample size calculation (they may even show you methods to decrease your sample size!)

- Reviewed the power of 71 published RCTs which had failed to detect a difference
- Found that 67 could have missed a 25% therapeutic improvement
- 50 could have missed a 50% improvement
Introduction

• Three main parts in its calculation
  ○ Estimation (depends on a host of items)
  ○ Justification (in the light of budgetary or biological considerations)
  ○ Adjustments (accounting for potential dropouts or effect of covariates)
Introduction

- Consequences of getting it wrong
  - Scientific
  - Ethical
  - Economical

- Problems can be approached in two ways
  - *Patients I need* approach: based on calculations of sample size for a given power, significance, and clinically meaningful difference
  - *Patients I can get* approach: based on calculations of power for a given sample size & level of significance
Pilot Studies

- It is a preliminary study intended to test the feasibility of a larger study, data collection methods, collect information for sample size calculations
- It should not be regarded as a study which is too small to produce a definitive answer
- It should be regarded as a tool in finding the answer as long as it is followed through
- Sample size calculations may not be required
Importance of Sample Size calculation

- Scientific reasons
- Ethical reasons
- Economic reasons
Scientific Reasons

- In a trial with **negative** results and a **sufficient sample size**, the result is concrete.
- In a trial with **negative** results and **insufficient power (insufficient sample size)**, may mistakenly conclude that the treatment under study made no difference.
Ethical Reasons

- An *undersized* study can expose subjects to potentially harmful treatments without the capability to advance knowledge.
- An *oversized* study has the potential to expose an unnecessarily large number of subjects to potentially harmful treatments.
Economic Reasons

- **Undersized study** is a waste of resources due to its inability to yield useful results
- **Oversized study** may result in statistically significant result with doubtful clinical importance leading to waste of resources (Cardiac Studies)
Classic Approaches to Sample Size Calculation

- **Precision analysis**
  - Bayesian
  - Frequentist
- **Power analysis**
  - Most common
In studies concerned with estimating some parameter:
- Precision
- Accuracy
- Prevalence
Power Analysis

- In studies concerned with detecting an effect
- Important to ensure that if an effect deemed to be clinically meaningful exists, then there is a high chance of it being detected

<table>
<thead>
<tr>
<th>Test for</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality</td>
<td>$H_0 : \mu_T - \mu_S = 0$</td>
<td>$H_a : \mu_T - \mu_S \neq 0$</td>
</tr>
<tr>
<td></td>
<td>$H_0 : \mu_T - \mu_S \geq \delta$</td>
<td>$H_0 : \mu_T - \mu_S &lt; \delta$</td>
</tr>
<tr>
<td>Non-inferiority</td>
<td>$H_0 : \mu_T - \mu_S \leq \delta$</td>
<td>$H_0 : \mu_T - \mu_S &gt; \delta$</td>
</tr>
<tr>
<td>Superiority</td>
<td>$H_0 :</td>
<td>\mu_T - \mu_S</td>
</tr>
<tr>
<td>Equivalence</td>
<td>$H_0 :</td>
<td>\mu_T - \mu_S</td>
</tr>
</tbody>
</table>
Factors That Influence Sample Size Calculations

- The objective (precision, power analysis)
- Details of intervention & control Rx.
- The outcomes
  - Categorical or continuous
  - Single or multiple
  - Primary
  - Secondary
  - Clinical relevance of the outcome
  - Any missed data
  - Any surrogate outcomes
    - Why
    - Will they accurately reflect the main outcome
Factors That Influence Sample Size Calculations

- Any covariates to control
- The unit of randomization
  - Individuals
  - Family practices
  - Hospital wards
  - Communities
  - Families
  - Etc
- The unit of analysis
  - Same as above
Factors That Influence Sample Size Calculations

- The research design
  - Simple RCT
  - Cluster RCT
  - Equivalence
  - Non-randomized intervention study
  - Observational study
  - Prevalence study
  - A study measuring sensitivity & specificity
  - A paired comparison
  - Repeated-measures study
  - Are the groups equal
Factors That Influence Sample Size Calculations

- Research subjects
  - Target population
  - Inclusion & exclusion criteria
  - Baseline risk
  - Pt. compliance rate
  - Pt. drop-out rate
Factors That Influence Sample Size Calculations

- Is the F/U long enough to be of any clinical relevance
- Desired level of significance
- Desired power
- One or two-tailed test
- Any explanation for the possible ranges or variations in outcome that is expected
- The smallest difference
  - Smallest clinically important difference
  - The difference that investigators think is worth detecting
  - The difference that investigators think is likely to be detected
  - Would an increase or decrease in the effect size make a sig. clinical difference
- Justification of previous data
  - Published data
  - Previous work
  - Review of records
  - Expert opinion
- Software or formula being used
Statistical Terms

- The numerical value summarizing the difference of interest (effect)
  - Odds Ratio (OR) Null, OR=1
  - Relative Risk (RR) Null, RR=1
  - Risk Difference (RD) Null, RD=0
  - Difference Between Means Null, DBM=0
  - Correlation Coefficient Null, CC=0
**Statistical Terms**

- **P-value**: Probability of obtaining an effect as extreme or more extreme than what is observed by chance.
- **Significance level of a test**: cut-off point for the p-value (conventionally it is 5%).
- **Power of a test**: correctly reject the null hypothesis when there is indeed a real difference or association (typically set at least 80%).
- **Effect size of clinical importance**.
One sided & Two sided tests of significance

- **Two-sided test**
  - Alternative hypothesis suggests that a difference exists in either direction
  - Should be used unless there is a very good reason for doing otherwise

- **One-sided test**
  - when it is completely inconceivable that the result could go in either direction, or the only concern is in one direction
    - Toxicity studies
    - Safety evaluation
    - Adverse drug reactions
    - Risk analysis
  - The expectation of the result is not adequate justification for one-sided test
Approach

- Specify a hypothesis
- Specify the significance level alpha
- Specify an effect size
- Obtain historical values
- Specify a power
- Use the appropriate formula to calculate sample size

- After the study is finished compare the variance of actual data with the one used in sample size calculations
### Table 1: Formulae for Sample Size Calculations for Comparisons Between Means

<table>
<thead>
<tr>
<th>Design</th>
<th>Hypothesis</th>
<th>$H_0$</th>
<th>$H_a$</th>
<th>Hypotheses and Sample Size Rules</th>
<th>Basic Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-sample</td>
<td>Equality</td>
<td>$\mu - \mu_0 = 0$</td>
<td>$\mu - \mu_0 \neq 0$</td>
<td>$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu - \mu_0)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>$\mu - \mu_0 \leq \delta$</td>
<td>$\mu - \mu_0 &gt; \delta$</td>
<td>$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu - \mu_0 - \delta)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>$</td>
<td>\mu - \mu_0</td>
<td>\geq \delta$</td>
<td>$</td>
</tr>
<tr>
<td>Two-sample Parallel</td>
<td>Equality</td>
<td>$\mu_1 - \mu_2 = 0$</td>
<td>$\mu_1 - \mu_2 \neq 0$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>$\mu_1 - \mu_2 \geq \delta$</td>
<td>$\mu_1 - \mu_2 &lt; \delta$</td>
<td>$n_1 = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>$\mu_1 - \mu_2 \leq \delta$</td>
<td>$\mu_1 - \mu_2 &gt; \delta$</td>
<td>$n_1 = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>$</td>
<td>\mu_1 - \mu_2</td>
<td>\geq \delta$</td>
<td>$</td>
</tr>
<tr>
<td>Two-sample Crossover</td>
<td>Equality</td>
<td>$\mu_1 - \mu_2 = 0$</td>
<td>$\mu_1 - \mu_2 \neq 0$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{2(\mu_1 - \mu_2)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>$\mu_1 - \mu_2 \geq \delta$</td>
<td>$\mu_1 - \mu_2 &lt; \delta$</td>
<td>$n_1 = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{3(\mu_1 - \mu_2)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>$\mu_1 - \mu_2 \leq \delta$</td>
<td>$\mu_1 - \mu_2 &gt; \delta$</td>
<td>$n_1 = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{3(\mu_1 - \mu_2)^2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>$</td>
<td>\mu_1 - \mu_2</td>
<td>\geq \delta$</td>
<td>$</td>
</tr>
</tbody>
</table>

### Table 2: Formulae for Sample Size Calculations for Comparisons Between Proportions

<table>
<thead>
<tr>
<th>Design</th>
<th>Hypothesis</th>
<th>$H_0$</th>
<th>Hypotheses and Sample Size Rules</th>
<th>Basic Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-sample</td>
<td>Equality</td>
<td>$\pi - \pi_0 = 0$</td>
<td>$\pi - \pi_0 \neq 0$</td>
<td>$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1-\pi)}{(\pi - \pi_0)^2}$</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>$\pi - \pi_0 \leq \delta$</td>
<td>$\pi - \pi_0 &gt; \delta$</td>
<td>$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1-\pi)}{(\pi - \pi_0 - \delta)^2}$</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>$</td>
<td>\pi - \pi_0</td>
<td>\geq \delta$</td>
</tr>
<tr>
<td>Two-sample Parallel</td>
<td>Equality</td>
<td>$\pi_1 - \pi_2 = 0$</td>
<td>$\pi_1 - \pi_2 \neq 0$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1-\pi_1)}{(\pi_1 - \pi_2)^2}$</td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>$\pi_1 - \pi_2 \geq \delta$</td>
<td>$\pi_1 - \pi_2 &lt; \delta$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1-\pi_1)}{(\pi_1 - \pi_2 - \delta)^2}$</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>$\pi_1 - \pi_2 \leq \delta$</td>
<td>$\pi_1 - \pi_2 &gt; \delta$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1-\pi_1)}{(\pi_1 - \pi_2 - \delta)^2}$</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>$</td>
<td>\pi_1 - \pi_2</td>
<td>\geq \delta$</td>
</tr>
<tr>
<td>Two-sample Crossover</td>
<td>Equality</td>
<td>$\pi_1 - \pi_2 = 0$</td>
<td>$\pi_1 - \pi_2 \neq 0$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi_1 \pi_2}{2(\pi_1 - \pi_2)^2}$</td>
</tr>
<tr>
<td></td>
<td>Non-inferiority</td>
<td>$\pi_1 - \pi_2 \geq \delta$</td>
<td>$\pi_1 - \pi_2 &lt; \delta$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi_1 \pi_2}{2(\pi_1 - \pi_2 - \delta)^2}$</td>
</tr>
<tr>
<td></td>
<td>Superiority</td>
<td>$\pi_1 - \pi_2 \leq \delta$</td>
<td>$\pi_1 - \pi_2 &gt; \delta$</td>
<td>$n_1 = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi_1 \pi_2}{2(\pi_1 - \pi_2 - \delta)^2}$</td>
</tr>
<tr>
<td></td>
<td>Equivalence</td>
<td>$</td>
<td>\pi_1 - \pi_2</td>
<td>\geq \delta$</td>
</tr>
</tbody>
</table>
Sample Size Adjustments

- Separate sample size calculation should be done for each important outcome & then use the max. estimate.
- When two variables are correlated with a factor $p$, then sample size can be reduced by a factor of $1-p^2$.
- Another option is to use Bonferroni correction for multiple outcomes.
Sample Size Adjustments

- Allowing for response rates & other losses to the sample
  - The expected response rate
  - Loss to f/u
  - Lack of compliance
  - Other losses

\[ n_{\text{new}} = \frac{n}{1 - L} \]
when \( L \) is the loss to f/u rate
Sample Size Adjustments

Adjustment for unequal group size

- Assuming \( n_1/n_2 = k \)
- Calculate \( n \) assuming equal
- Then

\[
\begin{align*}
  n_2 &= 0.5n(1+1/k) \\
  n_1 &= 0.5n(1+k)
\end{align*}
\]
Reporting Sample Size Calculations

- Clear statement of the primary objective
- The desired level of significance
- The desired power
- The statistics that will be used for analysis
- Whether the test would be one or two-tailed
- The smallest difference
  - Smallest clinically important difference
  - The difference that investigators think is worth detecting
  - The difference that the investigators think is likely to be detected
Reporting Sample Size Calculations

- Justification for prior estimates used in calculations
- Clear statements about the assumptions made about the distribution or variability of the outcomes
- Clear statement about the scheduled duration of the study
- Statement about how the sample size was adjusted
- The software or formulae that was used
- Take the reporting seriously as your documentation may be used in the future for sample size calculations
Example: Comparing Two Means

Scenario: A randomized controlled trial has been planned to evaluate a brief psychological intervention in comparison to usual treatment in the reduction of suicidal ideation amongst patients presenting at hospital with deliberate self-poisoning. Suicidal ideation will be measured on the Beck scale; the standard deviation of this scale in a previous study was 7.7, and a difference of 5 points is considered to be of clinical importance. It is anticipated that around one third of patients may drop out of treatment.
Example: Comparing Two Means

Required information

- Primary outcome variable = The Beck scale for suicidal ideation. A continuous variable summarized by means.
- Standard deviation = 7.7 points
- Size of difference of clinical importance = 5 points
- Significance level = 5%
- Power = 80%
- Type of test = two-sided
Example: Comparing Two Means

The formula for the sample size for comparison of 2 means (2-sided) is as follows:

\[ n = \left[ A + B \right]^2 \times 2 \times \frac{SD^2}{DIFF^2} \]

- where \( n \) = the sample size required in each group (double this for total sample).
- \( SD \) = standard deviation, of the primary outcome variable - here 7.7.
- \( DIFF \) = size of difference of clinical importance - here 5.0.
Example: Comparing Two Means

- $A$ depends on desired significance level (see table) - here 1.96.
- $B$ depends on desired power (see table) - here 1.28.
- Table of values for $A$ and $B$
  - Significance level $A$:
    - 5%: 1.96
    - 1%: 2.58
  - Power $B$:
    - 80%: 0.84
    - 90%: 1.28
    - 95%: 1.64

Inserting the required information into the formula gives:

$$n = \frac{(1.96 + 0.84) \times 2 \times 2 \times 7.72}{5.02} = 38$$

This gives the number required in each of the trial's two groups. Therefore the total sample size is double this, i.e. 76.

- To allow for the predicted dropout rate of around one third, the sample size was increased to 60 in each group, a total sample of **120**.
"A sample size of 38 in each group will be sufficient to detect a difference of 5 points on the Beck scale of suicidal ideation, assuming a standard deviation of 7.7 points, a power of 80%, and a significance level of 5%. This number has been increased to 60 per group (total of 120), to allow for a predicted drop-out from treatment of around one third"
“A previous study in this area recruited 150 subjects & found highly sign. Results”
- Previous study may have been lucky

“Sample sizes are not provided because there is no prior information on which to base them”
- Do a pilot study
- Standard Deviation could be estimated from range
  \[ SD = \frac{\text{max} - \text{min}}{4} \]

Number decided based on available pts alone
- Extend the length
- Consider a multi-center study
Failure to Achieve Required Sample Size

- Pt. refusal to consent
- Bad time of the study (heavy clinic study in the winter)
- Adverse media publicity
- Weak recruiting staff
- Lack of genuine commitment to the project
- Lack of staffing in wards or units
- Too many projects attempting to recruit the same subjects
Possible Solutions

- Pilot studies
- Have a plan to regularly monitor recruitment or create recruitment targets
- Ask for extension in time and/or funding
- Review your staffs commitment to other ongoing trials or other distracters
- Regular visits to trial sites
Strategies For Maximizing Power & Minimizing the Sample Size

- Use common outcomes (the power is driven more by the number of events than the total sample size)
- Use paired design (such as cross-over trial)
- Use continuous variables
- Choose the timing of the assessments of primary outcomes to be when the difference is most likely to be optimal
Recalculation of Sample Size Mid-Trial

- **Two main reasons**
  - **Changing input factors**
    - Changes in the anticipated control group outcome
    - Changes in the anticipated treatment compliance rate
    - Changing opinions regarding min. clinically important difference (MCID)
  - **Increasing accrual rates**
    - Increasing the sample size to increase the power to detect the same MCID
    - Increasing the sample size to allow smaller differences to be detected
Retrospective Sample Size Calculations

- Controversial
- Most recommend to avoid it as it really doesn’t add more information in most cases and may confuse or misguide the conclusion
General Rules of Thumb

- Don’t forget multiplicity testing corrections (Bonferroni)
- Overlapping confidence intervals do not imply non-significance (up to 1/3 can overlap even when significant)
- Use the same statistics for both sample size calculation and your analysis (superiority, equality, etc)
  - Otherwise you may alter the anticipated power
- Usually better to adopt a simple approach
- Better to be conservative (assume two-sided)
General Rules of Thumb

1. The basic rule of thumb for estimating the sample size for testing equality of two means is

   \[ n_1 = n_2 = \frac{8\sigma^2}{\delta^2}; \quad \text{where} \quad \delta = \mu_1 - \mu_2. \]

2. The basic rule of thumb for estimating the sample size for testing equality of two proportions is

   \[ n_1 = n_2 = \frac{8\pi(1 - \pi)}{(\pi_1 - \pi_2)^2}; \quad \text{where} \quad \pi = \frac{\pi_1 + \pi_2}{2}. \]

- Remember that sample size calculation gives you the minimum you require.
- If the outcome of interest is “change”, then use the standard deviation (SD) of the change and not each individual outcome.
General Rules of Thumb

- Non RCTs generally require a much larger sample to allow adjustment for confounding factors in the analysis.
- Equivalence studies need a larger sample size than studies aimed to demonstrate a difference.
- For moderate to large effect size ($0.5 < \text{effect size} < 0.8$), 30 subjects per group.
- For comparison between 3 or more groups, to detect a moderate effect size of 0.5 with 80% power, will require 14 subjects/group.
- Use sensitivity analysis to create a sample size table for different power, significance, or effect size and then sit and ponder over it for the optimal sample size.
Rules of Thumb for Associations

- **Multiple Regression**
  - Minimal requirement is a ratio of 5:1 for number of subjects to independent variables
  - The desired ratio is 15:1

- **Multiple Correlations**
  - For 5 or less predictors (m) use $n > 50 + 8m$
  - For 6 or more use 10 subjects per predictor

- **Logistic Regression**
  - For stable models use 10-15 events per predictor variable
Rules of Thumb for Associations

- Large samples are needed
  - Non-normal distribution
  - Small effect size
  - Substantial measurement error
  - Stepwise regression is used
- For chi-squared testing (two-by-two table)
  - Enough sample size so that no cell has less than 5
  - Overall sample size should be at least 20
Rules of Thumb for Associations

Factor analysis
- At least 50 participants/subjects per variable
- Minimum 300
  - N=50 very poor
  - N=100 poor
  - N=200 fair
  - N=300 good
  - N=500 very good
Sample Sizes for Time-to-Event Studies

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Examples of Associated Time-to-Event Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>Survival Time</td>
</tr>
<tr>
<td>Response (Tumor shrinkage)</td>
<td>Duration of response</td>
</tr>
<tr>
<td>Recurrence of disease</td>
<td>Time to recurrence</td>
</tr>
<tr>
<td>Relief of symptoms</td>
<td>Time to relief of symptoms/without symptoms</td>
</tr>
<tr>
<td>Quality of life 'scores'</td>
<td>Time to improvement/deterioration in scores</td>
</tr>
<tr>
<td>Toxicity</td>
<td>Time with/without toxicity</td>
</tr>
</tbody>
</table>

\[ n = \frac{2 \left( Z_{\alpha/2} + Z_{\beta} \right)^2}{(\ln(M_t/M_c))^2} \]

- Most software require the use of event-free rates (survival) and not event rates (death), because the log rank test is based on event-free rates.
- Beware if your software is giving you total number of subjects or events.
Sample Size for Cluster RCT

- Clusters or units are randomized

- Reasons
  - Logistical
    - Administrative convenience-easier than individual pt. recruitment or randomization
  - Ethical
    - Hard to randomize part of a family or community
  - Scientific
    - Worry about treatment contamination-changing behavior or knowledge during the trial
    - Plan for cluster level intervention-family physician or hospital units
    - Cluster action for an intervention-communities
There are Many Other Types of Studies

- Specialized sample size calculations
  - Cross-over design
    - Needs half as many as an RCT
    - There should be no carry-over effect of Rx.
    - Most suited for chronic conditions (pain, insomnia), not acute (death)
  - Analysis of change from baseline
  - Comparisons of means for two Poisson population
  - Testing for a Single Correlation Coefficient
  - Comparing Correlation Coefficients for Two Independent Samples
Transformations

- Most of the statistical testing is based on a Normal distribution.
- Quite often the assumed distribution may not fit the data:
  - Duration of symptoms
  - Cost
- Changing the scale of the original data (transforming) and assuming the distribution for the transformed data may provide a solution.
- Log-transformation may normalize the distribution leading to a log-normal distribution to work with.
Non-parametric Tests

- Non-parametric (also called *distribution free*) methods are designed to avoid distributional assumptions

- **Advantages**
  - Fewer assumptions are required
  - Only nominal (categorical data) or ordinal (ranked) are required, rather than numerical (interval) data

- **Disadvantages**
  - Less efficient
    - Less powerful
    - Overestimates variance
  - Do not lend themselves easily to sample size calculations and CI
  - Interpretation of the results is difficult
  - Most software don’t do them
Software for Sample Size Calculations

- nQuery Advisor 2000
- Power and Precision 1997
- Pass 2000
- UnifyPow 1998
Freeware on The Web (User beware)

http://www.stat.ucla.edu/~jbond/HTMLPOWER/index.html
http://www.health.ucalgary.ca/~rollin/stats/ssize/
http://www.stat.uiowa.edu/~7erlenth/Power/index.html
http://www.dssresearch.com/SampleSize/
http://www.stat.ucla.edu/calculators/powercalc/
http://hedwig.mgh.harvard.edu/sample_size/size.html
http://www.bobwheeler.com/stat/SSize/ssize.html
http://www.math.yorku.ca/SCS/Online/power/
http://www.surveysystem.com/sscalc.htm
http://www.researchinfo.com/docs/calculators/samplesize.cfm