Study Design and Sample Size

Professor Leslie Daly
9th Sept, 2010
If only I had a cent for every time I was asked How many patients do I need for my study?
Advice to a Statistical Consultant:

There is no such thing as a sample size problem. Sample size is but one aspect of study design. When you are asked to help determine the sample size a lot of questions must be asked and answered before you get to that one......You may often end up never discussing sample size because there are other matters that override it in importance.

Russell Lenth (2001)
It seems an easy question, like

How much money should I take on my holidays?
Why is the Sample Size Important?

- **Ethical reasons:**
  - Proper use of scarce resources
  - You don’t want a study that won’t show anything because it’s too small
  - You don’t want a study that is too big when a smaller one would have done

- **Funding agencies require it**
How many?

- Use a similar number to previous research in the same area
  - Maybe a guideline
  - But much research is under-powered

- Use a scientific approach, but remember
  - Sample size calculations are not a substitute for careful study planning
Sample Size

Reality: resources

Scientific Validity: Sample Size Formulae
Resources

- Number of available patients
- Laboratory resources
  - Diagnostic tests etc. - if needed
- Time you have available
  - Set by funding agency
  - Set by your career trajectory
- Funds and personnel
Study Timelines

Set-up, Ethical approval etc.

Patient recruitment

Duration of Follow-up (minimum)

Analysis, Write-up

SAMPLE SIZE

Last patient recruited

Last patient Follow-up completed
Non-influences on Sample Size

- Sample size required does NOT depend on the population size
Influences on Sample Size

- Study objectives
- Study design
- Endpoint
- Type of sampling
- Type of statistical analysis
Influences on Sample Size: Study Objectives

- **Prevalence study**
  - To **estimate** a mean or percentage
  - No comparisons

- **Comparative Study**
  - Superiority study
    - To show differences between groups/treatments
  - Equivalence or non-inferiority study
    - To show that groups/treatments are similar
Influences on Sample Size

Study Design

- Cross-sectional prevalence survey
- Comparative study design
  - Cross-sectional or follow-up
  - Number of groups
    - Reduce to two groups for simplicity
  - Independent, paired or matched groups
    - Parallel or cross-over designs
    - Matched control groups
  - Other designs
    - Factorial and sequential designs
Influences on Sample Size

Endpoints

- Must relate to your objectives
- Must be measurable and well defined
  - Preferable standardized
  - Don’t reinvent the wheel
- Mean or percentage of something usually
  - Sample sizes tend to be smaller for means than percentages
Endpoints

- Need prior estimates of variability
  - Requires estimates of standard deviation and expected percentage
  - Pilot study or literature review
- If you can’t decide on a single important endpoint you will have to calculate a sample size for each and take the biggest
Sample Size for Prevalence Studies

- No comparisons involved
  - Just an estimation in a cross-sectional survey

- How precise do we want our answer to be?
  - Want the percentage of CF patients with liver disease to ± 3%
  - Want the mean serum cholesterol to within ±0.5 mmol/L

- Use the 95% confidence interval
  - Easy formulae available
Need prior estimate (rough idea) of
- Standard deviation (σ) when estimating a mean
- Where the percentage (p) might be when estimating a percentage

Use pilot study or published data

\[ \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \quad \quad \quad p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \]
Precision goes up by $\sqrt{n}$

- Doubling $n$ reduces CI width by 30%
- Multiplying $n$ by 4 reduces CI width by 50%

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>95% CI given by*</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>±12.7%</td>
</tr>
<tr>
<td>100</td>
<td>±9.0%</td>
</tr>
<tr>
<td>500</td>
<td>±4.0%</td>
</tr>
<tr>
<td>1000</td>
<td>±2.8%</td>
</tr>
</tbody>
</table>

*For estimating a percentage around 30%
Superiority Studies: Looking for Differences between Groups

- At a particular significance level
  - Usually set to 5% 2-sided
  - $P < 0.05$

A statistically significant result is one not likely to be due to chance
Superiority Studies: Looking for Differences between Groups

Looking for a difference in what?
- If there is **no** follow-up the endpoint is a measure taken at the time of the study
- If there is follow-up the endpoint is likely to relate to a measure at end of follow-up
  - Survival / mortality
    - Need special methods
  - Biochemical or measured
    - Absolute levels or changes in levels
Measured Endpoints in Follow-up Studies (1)

- **Absolute levels at follow-up**
  - Ignores baseline differences
  - May be valid for a RCT

- **Average change (% change) from baseline**
  - Calculate change for each person and average changes in each group
    - Measure change or ask about change?
  - Difficult to get prior estimates of variability
Measured Endpoints in Follow-up Studies (2)

- **Analyse absolute levels at follow-up**
  - Adjusting for baseline measures using ANCOVA
  - Most efficient
    - You can allow for it in getting sample size

- **Replicate measures pre and post treatment**
  - Can reduce sample size
  - Useful for RCTs
  - For biochemical endpoint
  - Requires efficient statistical analysis
    - ANCOVA rather than mean change from baseline
Measured Endpoints in Follow-up Studies (3)

- Your sample size is the number of patients NOT the number of replicates
  - Increase patients and reduce replicates
    - 10 measures on 100 patients is better than 100 measures on 10 patients

- Reduce number of replicate measurements
  - Relevant if the measurement is the expensive or limiting factor

- If you are measuring response over time
  - Fewer samples needed when value is stable
  - Analyse summary end point not values at each time
Superiority Studies: Looking for Differences between Groups

- How big a difference?
  - The ‘minimal worthwhile difference’
  - A clinical question not a statistical one
  - ‘What would you like to find?’
  - The effect size
The Effect Size for the Endpoint ($\Delta$)

- If the difference (the effect) is really this big:
  - I don’t want to miss it
  - I want to get a significant result
  - I want a good chance of getting a significant result

- If the effect is smaller:
  - I don’t mind missing it
  - I don’t mind a smaller chance of getting a significant result
Power of a study (1 - $\beta$)

- $\beta$ is the chance of not getting a significant result when the real effect is as big as postulated
  - Chance of a Type II error
  - Often set at $4 \times$ (significance level)
- The chance of getting a significant result when the real effect is as big as postulated
  - $1 - \beta$
  - Set at 80% for 5% significance
  - Set at 95% for 1% significance
Sample Sizes for Superiority Studies: Inter-relationships

- Sample size
- Power
- Significance level
- Effect size
Other Calculations

- Can calculate power for required difference and fixed sample size (and significance)
  - Give consequences of going ahead
  - Formulae difficult – use software

- Can calculate detectable difference for given sample size, power and significance
  - May help decision on an appropriate level to set
Sample Size for Comparing Groups increases with $\frac{1}{\Delta^2}$

- Halving effect size multiplies sample size by 4

<table>
<thead>
<tr>
<th>Sample size in each group</th>
<th>To detect a difference between groups of</th>
<th>Difference to be detected ($\Delta$)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>60% versus 20%</td>
<td>40%</td>
</tr>
<tr>
<td>50</td>
<td>60% versus 33%</td>
<td>27%</td>
</tr>
<tr>
<td>100</td>
<td>60% versus 41%</td>
<td>19%</td>
</tr>
<tr>
<td>500</td>
<td>60% versus 51%</td>
<td>9%</td>
</tr>
</tbody>
</table>

*Power 80%; 2-sided 5% significance
Free Sample Size Software

- Epi Info
  - http://www.cdc.gov/Epiinfo/

- WinPepi
  - http://www.brixtonhealth.com/pepi4windows.html

- Power and Sample Size
  - http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize
“Power and Sample Size”

![Power and Sample Size Program: Main Window](image)
“WinPepi”
Check whether it is the total sample size or the sample size in each group
- Common error

Try software or formula against an answer you know
- One of the tables in this presentation
- Use two different programmes

Allow for non-response/loss to follow-up

Play around with your calculations
Usual Formulae are for Equal Sample Sizes

- Designing a study with equal sample sizes in two comparison groups is most efficient.

- When might you have unequal n?
  - Cross-sectional survey
    - Comparing smokers to non-smokers
  - Limited availability of patients
    - Use up to 4 controls per case in a case-control study
  - Cost issues
    - More expensive to recruit into one of your groups

- Easy adjustment of ‘equal n’ calculation
# Equivalent Sample Sizes

*(giving same power to detect a given difference)*

<table>
<thead>
<tr>
<th>Ratio A:B</th>
<th>Sample size in group A</th>
<th>Sample size in Group B</th>
<th>Total Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>1:2</td>
<td>75</td>
<td>150</td>
<td>225</td>
</tr>
<tr>
<td>1:3</td>
<td>67</td>
<td>198</td>
<td>265</td>
</tr>
<tr>
<td>1:4</td>
<td>62</td>
<td>259</td>
<td>321</td>
</tr>
<tr>
<td>1:5</td>
<td>60</td>
<td>300</td>
<td>360</td>
</tr>
</tbody>
</table>
Influences on Sample Size

- **Study objectives**
  - Prevalence, Superiority or Equivalence

- **Study design**
  - Cross-sectional or follow-up; Matched or independent; Factorial or sequential designs

- **Endpoint**

- **Type of sampling**

- **Type of statistical analysis**
Factorial Designs – Two for the Price of One

- Comparing two treatments and a placebo
  - Three groups is inefficient

- Treatment A
- Treatment B
- Placebo
Comparing two treatments and a placebo

- Use four groups
- Two independent comparisons
  - Unless there are interactions
- Total sample size that of a two-group study

'\text{B}' versus 'No \text{B}''
Equivalence / Non-inferiority Studies

- When trying to show two groups (treatments) are the same
  - ‘Non-significance’ is not the same as ‘no difference’
    - Absence of evidence is not the same as evidence of absence
  - ‘Non-significance’ is ambiguous
    - There might be a difference or there might not be
  - Do you want
    - Equivalence or non-inferiority?
  - Easier to show a difference than sameness
    - Easier to disprove something than prove it
Equivalence or Non-inferiority Studies

- Must set a non-inferiority margin (equivalence limit)
  - The largest difference (between the treatments) that is clinically acceptable so that a difference bigger than this would matter in practice
Endpoint of BP Difference (BP on Treatment B minus BP on A)

Treatment B better

Treatment A better

BP Difference

-\Delta \quad 0 \quad +\Delta

Treatment A statistically **superior** to B

Treatment A clinically **equivalent** to B

Treatment A clinically **non-inferior** to B

For a significant result the 95% CI for BP difference must lie totally within the red bars

$\pm \Delta$ defines the equivalence/non-inferiority limits
Equivalence /Non-inferiority Studies: Issues

- Setting equivalence/non-inferiority limit is difficult
- For non-inferiority trials there are arguments about use of one or two-sided significance levels
- Sample sizes tend to be very large
- Software is not widely available
  - And is very difficult to use or understand
Influences on Sample Size

- Study objectives
- Study design
- Endpoint
- Type of sampling
- Type of statistical analysis
Influences on Sample Size: Type of Sampling

- Theory:
  - Our patients are a random sample from a population

- Reality:
  - We rarely work with random samples anyway
  - Either we assume we have OR
  - We adjust for sampling method
Complex Sampling

- **Cluster sampling or cluster randomisation**
  - Sampling unit is a ‘group’ of subjects
    - Randomising general practices
    - Sampling school children by class

- **Stratified sampling**

- **Sample size estimation**
  - Can adjust usual formula with a ‘design effect’ or ‘variance inflation factor’
    - Usually extremely difficult to determine
  - Often educated guess-work dressed up in scientific language
Accounting for Statistical Analysis Methods

- **Interim analyses**
  - increases sample size

- **Using regression to adjust for confounders**
  - Variance inflation/deflation factors are available
    - But usually require unobtainable prior estimates
    - No agreement on acceptable approach
  - Standard two-group sample size methods usually used
    - Probably the best practical solution at the moment

- **Using ANCOVA to adjust for initial levels**
  - Can reduce sample size required
Influences on Sample Size

- Study objectives
- Study design
- Endpoint
- Type of sampling
- Type of statistical analysis
A Final Rule: Maximise Use of Available Patients

- Increase response rate and avoid missing values
  - Avoid invasive procedures
  - Shorten your questionnaire
  - Spend time on the recruitment process

- A small sample with a high response rate or few missing values is better than
  - A large sample with a low response rate or a lot of missing data

- Use all available patients
  - Perform some measures on a subset of patients
The question of the most appropriate study size is not a technical decision to be determined by calculation, but a judgement to be determined by experience, intuition and insight.

*Modern Epidemiology*, Ken Rothman - 1986
Resources

- **Steven Julious**
  - Sample sizes for clinical trials with Normal data
- 65 page article
- Formulae for every situation
  - Comparing means only