

**Contents**

Preliminaries

0: Introduction

1: Background & Basic Concepts

2: Basic Trial analysis

3: Randomization

4: Protocol Deviations

**5: Size of the Trial**

6: Multiplicity & Interim Analysis

7: Crossover Trials

8: Combining Trials

9: Binary Response Data

10: Comparing Methods of Measurement

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 121

- Need to choose size of trial to give a reasonable chance of detecting differences of clinical interest
- Unethical to conduct trial with
  - ◆ Little chance of reaching conclusion
  - ◆ More patients than necessary to reach decision

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 122

**Reasonable chance?**

- ◆ 80%? 95%? 70%? 90%? ?????
- Essentially the choice of the clinician
- Higher the chance → the greater the size

▪ **80% common benchmark standard**

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 123

- **Difference of clinical interest?**
  - ◆ Solely responsibility of clinician
  - ◆ Clinicians are over-ambitious
    - Start by saying '*any difference*'
  - ◆ Smaller desired detectable difference → the greater the size

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 124

- Detecting a difference of size  $\delta$ :
  - ◆ obtaining a significant result when there is indeed a difference  $\delta$ 
    - depends on significance level used

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 125

- **Specifications required**
  - ◆ main outcome measure
  - ◆ method of analysis
    - e.g. two-sample t-test
  - ◆ result on standard treatment
    - or pilot results
  - ◆ minimum difference required to detect ( $\delta$ )
  - ◆ degree of certainty with which we wish to detect it (i.e. the **power:  $1 - \beta$** )

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 126



- **non-significant difference**  
*different from*
- **no clinically relevant difference**
  - ◆ ‘non-significant difference’
    - statistical evidence is not convincing
  - ◆ no clinically **relevant** difference
    - if any difference then not medically important

127

- **Types of Error**
  - ◆ Type I: false positive
    - treatments equivalent but result significant
      - $\alpha$  represents risk of false positive result
  - ◆ Type II: false negative
    - treatments different but result non-significant
      - $\beta$  represents risk of false negative result
      - **power = 1 –  $\beta$**
  - ◆ **c.f. sensitivity & specificity**

128

- **Sensitivity & Specificity:**

		Disease Status		
		Positive	Negative	Total
Test	Positive	a	b	a+b
	Negative	c	d	c+d
	Total	a+c	b+d	n

Sensitivity=  $a/(a+c)$  Specificity= $d/(b+d)$

129

- **Sensitivity & Specificity:**
  - ◆ Sensitivity:
    - probability test ‘correctly identifies’ null hypothesis ( $=1 - \alpha$ )
  - ◆ Specificity
    - probability test ‘correctly rejects’ null hypothesis ( $=1 - \beta$ )

130

- **General Method:**
  - ◆ Determine criterion for declaring a significant result:
  - ◆ i.e. value of test statistic > critical value
    - e.g. z-value > 1.96
  - ◆ **NB:** P[test statistic > critical value when no difference] = significance level
    - $P[z > 1.96 | H_0 \text{ true}] = \alpha$ 
      - $P[z > 1.96 | H_0 \text{ true}] = 0.025$

131

- Find probability of significant result when difference is  $\delta$ 
  - ◆ i.e. the **power** of the test when difference is  $\delta$
- This probability depends on  $\delta$  and n
- Relates **power**,  $\delta$  and n in one equation
- Given 2 of these we can calculate 3<sup>rd</sup>.

132



- e.g.  $H_0: \mu_1 = \mu_2$  vs  $H_A: \mu_1 \neq \mu_2$ 
  - ◆ Let  $\delta = \mu_1 - \mu_2$ 
    - then  $H_0: \delta = 0$
  - ◆ Test statistic  $z$ , reject  $H_0$  at level  $\alpha$  if  $z > c_\alpha$
- Then  $P[z > c_\alpha | \delta = 0] = \alpha$
- Power when difference is  $\delta$  is  $P[z > c_\alpha | \delta]$ 
  - **NB:** the power depends on  $\delta$ 
    - (& on sample size  $n$ )

Medical Statistics: Clinical Trials
133

- Can calculate  $n$  for given power and  $\delta$
- Or power for given  $n$  and  $\delta$
- Or  $\delta$  for given power and  $n$

Medical Statistics: Clinical Trials
134

- **Often  $n$  is limited by practicalities**
  - ◆ Often clinician asks about  $n$
  - ◆ Finds required  $n$  is unrealistic
  - ◆ Tells you what is a realistic  $n$
  - ◆ You can give a table of obtainable **power** and  $\delta$  for this maximum sample size  $n$

Medical Statistics: Clinical Trials
135

- Testing equality of two binomial proportions  $\theta_1$  and  $\theta_2$ 
  - (using Normal approximation)
- General formula
 
$$n \approx \frac{\theta_2(1-\theta_2) + \theta_1(1-\theta_1)}{(\theta_2 - \theta_1)^2} \{ \Phi^{-1}(\beta) + \Phi^{-1}(\alpha/2) \}^2$$
- Can use to find any one of  $n$ ,  $\beta$  or  $\theta_2$  from the other 2

Medical Statistics: Clinical Trials
136

- **N.B. both  $\Phi^{-1}(\beta)$  and  $\Phi^{-1}(\alpha/2) < 0$** 
  - ◆ Be careful if inverting formula and need to take square roots

Medical Statistics: Clinical Trials
137

- Approximation requires
 
$$\frac{\sqrt{2\theta_1(1-\theta_1)}}{\sqrt{\theta_2(1-\theta_2) + \theta_1(1-\theta_1)}} \approx 1$$
- Here = 1.14, so reasonable
  - Packages may use exact formula

Medical Statistics: Clinical Trials
138



- Term  $\{\Phi^{-1}(\beta) + \Phi^{-1}(\alpha/2)\}^2$ 
  - ◆ is the same for many different tests
  - ◆ only multiplier differs

- If we can really justify a 1-sided test (e.g. from a pilot study)
- then put  $\Phi^{-1}(\alpha/2) \rightarrow \Phi^{-1}(\alpha)$ .
- However, this **MUST** be specified in the protocol
  - ◆ Cannot decide after trial has begun
  - ◆ Definitely not after collecting data

- 1-sided testing
  - ◆ Reduces the required sample size
  - ◆ Increase the power
  - ◆ Reduces the CRD

- e.g. a p-value of 8% on a 2-sided test becomes 4% if test is converted to 1-sided
- – cannot decide to use 1-sided just because a 2-sided p-value is 8%

- For given  $\alpha$  and  $\beta$ ,  $n$  depends [inversely] on  $(\theta_2 - \theta_1)^2$
- **half** the CRD needs
  - Clinically Relevant Difference
- **fourfold** increase in trial size.

- Increase in power from 50% to 95% needs sample size  $\times 3$
- NB In practice need to allow for a drop-out rate so need to inflate sample size to allow this:
  - e.g. if 10% drop-out rate expected then increase sample size by 11.1% [=1/(1-10%) ]



- Only possible to do sample size calculations in simplified situations
  - ◆ Cannot take into account more complex statistical tests that are actually likely to be used.
    - e.g cannot do sample size calculations for linear regression or ANOVA
    - Need to use calculations from a simplified situation as a guide

- Typically sample size is limited by practical situations
- More useful to find the power of a study for a range of sample sizes and CRDs

- Continuous data
  - ◆ Same general method
  - ◆ Need to know standard deviation  $\sigma$ 
    - Pilot study or published data
    - $\frac{1}{4} \times$  likely range
      - Guess at maximum and minimum possible

- Testing equality of two normal means  $\mu_1$  and  $\mu_2$ 
  - (using Normal approx to t-distribution or known variance)

- General formula

$$n = \frac{2\sigma^2}{(\mu_2 - \mu_1)^2} \left\{ \Phi^{-1}(\beta) + \Phi^{-1}\left(\frac{\alpha}{2}\right) \right\}^2$$

- Same term as before
- Same dependence on CRD

- **Computational Note**
  - ◆ Sample size facilities in R, S-PLUS, Minitab & StatsDirect
  - ◆ These have different default choices & may give slightly different answers
  - ◆ Tables by Machin & Campbell (1997) provide a disk with SampSize (V2.0)
  - ◆ nQuery Advisor is industry standard
  - ◆ Programme POWER.EXE available through course web page

- **Implementation in R**

- Three functions

- `power.t.test()`
- `power.prop.test`
- `power.anova.test()`

- ◆ Each will find any of power, sample size or CRD from the other 2



**Example:**

```
> power.prop.test(p1=0.9,p2=0.95,n=581,sig.level=0.05)
```

Two-sample comparison of proportions power calculation

n = 581  
 p1 = 0.9  
 p2 = 0.95  
 sig.level = 0.05  
 power = 0.8999597  
 alternative = two.sided

NOTE: n is number in \*each\* group

**Find power**

**Example:**

```
> power.t.test(delta=1.6577,sd=3.6,power=0.8,sig.level=0.05)
```

Two-sample t test power calculation

n = 75.00648  
 delta = 1.6577  
 sd = 3.6  
 sig.level = 0.05  
 power = 0.8  
 alternative = two.sided

NOTE: n is number in \*each\* group

**Find n**

**Programme POWER.EXE**  
 available from course web page

**Main Menu**

- Display Test Menu
- Display Help Screen
- Quit the program

**Instructions**  
 press Uparrow,Dnarrow to move between options  
 press Enter to choose particular option

**Test Menu**

- One sample t test/paired t test
- Two sample t test
- Log Rank test
- Test for single proportion
- Two sample Test for proportions
- Significance of correlation coeff.
- The Mann-Whitney U test
- McNemars test
- Multigroup comparisons (independent groups)
- Xover trial comparisons

**Instructions**  
 press Esc for main menu  
 press Uparrow,Dnarrow to move between options  
 press Enter to choose particular test, or H for help

**Test Menu**

- One sample t test/paired t test
- Two sample t test
- Log Rank test
- Test for single proportion
- Two sample Test for proportions
- Significance of correlation coeff.
- The Mann-Whitney U test
- McNemars test
- Multigroup comparisons (independent groups)
- Xover trial comparisons

**Instructions**  
 press Esc for main menu  
 press Uparrow,Dnarrow to move between options  
 press Enter to choose particular test, or H for help

**Two Sample test for proportions**

Calculation mode: power sample size CRD  
 significance level: 0.05  
 power: 90  
 Sample size group 1: is to be calculated  
 ratio group1:group2: 1:1  
 group1 proportion: .25  
 group2 proportions: .45

**perform calculations** No Yes

**Specific Help :**  
 Press Left & RightArrow keys to move between options, Press DnArrow, Enter, or Uparrow to select mode

**Instructions:**  
 Press Esc for test menu.  
 Press Uparrow,Dnarrow & Enter to move between items on the form



```

Two Sample test for proportions
significance level 0.05
power 90
ratio group1:group2 1:1
group1 proportion .25
group2 proportions .45
Sample size group 1 118
    
```

One Value Menu: Choose from Print Result Main Menu EditForm Test Mer

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 157

- **Summary and Conclusions**
  - ◆ Samples too small
    - ⇒ little chance of conclusion
    - ⇒ exposing patients to risk with no benefit
  - ◆ Samples too large
    - ⇒ may expose too many subjects to inferior treatment

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 158

- For sample size calculation need:
  - ◆ outcome measure
  - ◆ method of analysis
    - including significance levels
  - ◆ CRD
  - ◆ power
  - ◆ results on standard treatment
    - including likely variability

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 159

- Practical Limitations:
  - ◆ Drop-outs
    - Inflate sizes to allow for this
  - ◆ Use simplified version of statistical design as a guide to sample sizes
  - ◆ Consider effects on power and CRD for practical range of sample sizes

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 160

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 161

©NRJF, University of Sheffield, 2011/12 Semester Medical Statistics: Clinical Trials 162

