

Contents

Preliminaries

0: Introduction

1: Background & Basic Concepts

2: Basic Trial analysis

3: Randomization

4: Size of the Trial

5: Multiplicity & Interim Analysis

6: Crossover Trials

7: Binary Response Data

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 103

- 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, 2012 Medical Statistics: Clinical Trials 104

- 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, 2012 Medical Statistics: Clinical Trials 105

- **Difference of clinical interest?**
 - ◆ Solely responsibility of clinician
 - ◆ Clinicians are overambitious
 - Start by saying '*any difference*'
 - ◆ Smaller desired detectable difference → the greater the size
- ©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 106

- Detecting a difference of size δ :
 - ◆ obtaining a significant result when there is indeed a difference δ
 - depends on significance level used
- ©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 107

- **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 (δ)
 - ◆ degree of certainty with which we wish to detect it (i.e. the **power: $1 - \beta$**)
- ©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 108



- **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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 109

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 110

- **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 = $d/(b+d)$

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 111

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 112

- **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
 - $P[\text{test stat} > \text{crit val} \mid H_0 \text{ true}] = \alpha$
 - $P[z > 1.96 \mid H_0 \text{ true}] = 0.025$

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 113

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 114



- 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 α if $z > c_\alpha$
- Then $P[z > c_\alpha \mid \delta = 0] = \alpha$
- Power when difference is δ is $P[z > c_\alpha \mid \delta]$
 - **NB:** the power depends on δ
 - (& on sample size n)

©NSF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 115

- Can calculate n for given power and δ
- Or power for given n and δ
- Or δ for given power and n

©NSF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 116

- **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 δ for this maximum sample size n

©NSF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 117

- Testing equality of two binomial proportions θ_1 and θ_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 , β or θ_2 from the other 2

©NSF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 118

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

©NSF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 119

- 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

©NSF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 120



- 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 α and β , 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 dropout rate so need to inflate sample size to allow this:
 - e.g. if 10% dropout rate expected then increase sample size by 11.1% [=1/(110%)]



- 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 σ
 - Pilot study or published data
 - $\frac{1}{4} \times$ likely range
 - Guess at maximum and minimum possible



- Testing equality of two normal means μ_1 and μ_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, SPLUS, 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
- ◆ <http://www.divms.uiowa.edu/~rlenth/Power/index.html> gives Java applets from Russ Lenth, University of Iowa



- **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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 133

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 134

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 135

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 136

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 137

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

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 138



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, 2012 Medical Statistics: Clinical Trials 139

- **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, 2012 Medical Statistics: Clinical Trials 140

- 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, 2012 Medical Statistics: Clinical Trials 141

- Practical Limitations:
 - ◆ Dropouts
 - 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, 2012 Medical Statistics: Clinical Trials 142

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 143

©NRJF, University of Sheffield, 2012 Medical Statistics: Clinical Trials 144

