Medical Statistics: Clinical Trials

Dr Nick Fieller

Department of Probability & Statistics

University of Sheffield

MAS6012/MAS461/MAS361

2011/2012

Contents

Contents

Organization of Course

- Two Components
  - **Clinical trials**
    - Experiments on human (and animal) subjects
    - Ethical issues, efficient use of subjects, etc
  - **Survival Analysis**
    - (analyzing data on length of lifetimes, e.g. times of remission in leukaemia)
- Approximately 10 lectures each

Organization of course material

- Two sets of lecture notes (Clinical & Survival)
- Clinical Chapters 1 – 10 (~ 1 per lecture)
- Survival Chapters 1 – 4 main part of course
- Appendix 0: background maths
  - Maximum Likelihood Estimation
    - (but used only in a couple of places)
- Appendix 1 use of computer packages
  - SAS, SPSS, Minitab, S-PLUS
- Exercises & Task Sheets are in Course Booklet
  - Solutions follow later as appropriate

Task Sheets & Exercises

- Task sheets:–
  - ~ each week
  - simple quick short exercises/reading
  - reinforce / consolidate lecture material
- Exercises:–
  - 3 sets during semester in weeks 5,8,10
  - Work submitted within 2 weeks will be marked and returned
- **See Study Guide**
  - **recommendations on time to spend**

- **Task Sheets & Exercises**
  - Task sheets:–
    - are designed for you to test **your own understanding** of the course material
    - *you are responsible* for your own learning on the course — task sheets help you in self-assessment
  - Exercises:–
    - **Prime route for individual feedback**
    - Task sheets often provide guide for exercises
  - Unacceptable reasons for not submitting anything
    - I did not have enough time
    - I knew I could do them so I did not need to submit
    - I could not do anything so did not think it was worth it

- **Solutions to Task Sheets & Exercises**
  - Exercises:–
    - Solutions available on web soon after submission
      - Printed solutions will be provided to those who submit
  - Task sheets:–
    - are designed for you to test your own understanding of the course material
    - if necessary go back to lecture notes (etc) & re-read relevant sections
      - (and if necessary re-read again & … …)
    - Solutions will be provided on web pages in due course (for revision etc)
      - but ***deliberately*** these will not appear very quickly

- Course web page

  http://nickfieller.staff.shef.ac.uk/
  - Click on Teaching & then on MAS6012/MAS461/MAS361 Medical Statistics
- Lecture notes, task sheets, solutions & data sets available here after distribution in lectures
  - (I don't keep back copies)

Purpose of Lectures

- There are 'complete' printed notes
  - These are not a textbook
    - some explanations are omitted
  - They are intended to allow you to concentrate on understanding & for me to cover some material very quickly
- Some lectures will be very close to the printed notes
  - This is **intended**
- Other lectures will fill in details & provide examples & **R** demos

- **Books**
  - **Campbell, M. J. (2001)** *Statistics at Square Two.* BMJ
  - **Matthews, J. N. S. (2000)** *An Introduction to Randomized Controlled Clinical Trials.* Arnold

Contents

- **Objectives**
  - The objective of this course is to provide an introduction to some of the statistical methods and statistical issues that arise in medical experiments which involve, in particular, human patients. Such experiments are known collectively as ***clinical trials***.

- **New Techniques**, e.g.
  - Sample size calculations
  - McNemar's test
  - Mantel-Haenszel test
  - logistic regression
  - crossover trails
- **Issues in medical statistics**
  - ethics
  - design

- **Outline**
  - **Background**
    - historical development
    - placebo effect
    - blindness
    - phases of clinical trial

- **Outline (ctd)**
  - Basic trial analysis
    - parallel group
    - in series designs
    - factorial designs
    - sequential designs

- **Outline (ctd)**
  - Randomization
    - Objectives of randomization.
    - Simple
    - Restricted
    - Stratified

- **Outline (ctd)**
  - Protocol deviations
    - intention to treat
    - per protocol
  - Size of trial
    - sample sizes needed to detect clinically relevant differences with specified power.

- **Outline (ctd)**
  - Multiplicity and interim analyses
    - multiple significance testing
    - subgroup analysis
    - Bonferroni corrections.
  - Crossover trials
    - treatment, period and carryover effects

- **Outline (ctd)**
  - Combination of trials
    - pooling trials and meta analysis
    - Simpson's paradox
    - Mantel-Haenszel test
  - Binary responses
    - matched pairs and McNemar's test
    - logistic regression.

Contents

- **Definition of clinical trial**
  - Any form of planned experiment which involves patients and is designed to elucidate the most appropriate treatment of future patients under a given medical condition

- **Notes**
  - Planned experiment
  - Inferential Procedure
    - want to use results on limited sample of patients to find out best treatment in the general population of patients who will require treatment in the future.

- **Historical Background**
  - 1537: Treatment of battle wounds:
  - 1741: Treatment of Scurvy
  - 1948: Streptomycin trial
  - **1954: Field Trial of Salk Polio Vaccine**
    - comparison between a randomized controlled double-blind clinical trial and a non-randomized open trial

- **Types of Trial**
  - Phase I trials: pharmacology & toxicity.
    - n=10 – 50
  - Phase II trials: safety & efficacy
    - n= 50 –100
  - **Phase III trials: treatment comparison**
    - **n= 100 – 1000**
  - Phase IV trials: Post-marketing surveillance
    - n= as many as possible

- Randomized Controlled Trials
  - Comparative
    - e.g. new *vs.*standard
  - Removes bias
    - Conscious or unconscious
  - Control group
    - As similar as possible to treated group

- **Placebo Effect**
  - **Blindness of trials**
    - Double blind
      Both patient & evaluator blind
    - Single blind
      either patient or evaluator blind

- **Ethical Considerations**
  - Treaty of Helsinki (1960+ammendments)
  - competition between individual and collective ethics
  - **unethical** to conduct research which is badly planned or executed.
    - Only compare treatment A with treatment B if we are genuinely unsure whether A or B is better

- **Local ethics committees**
  - Licence all clinical trials in their area
    - (hospital/city/region)
  - Informed Consent
  - unethical to perform a trial which has little prospect of reaching any conclusion
    - insufficient numbers
    - poor design

- **Publication Ethics**

  See BMJ Vol 323, p588, 15/09/01. (http://www.bmj.com)

  - **Concern** where authors have
    - not participated in design of study
    - had no access to raw data
    - little role in interpretation of data
    - not had control over publishing

- **Publication Bias**
  - Not all trials that take place are published
  - Only 20% of **all** cancer trials are published
    - http://theoncologist.alphamedpress.org/cgi/content/abstract/theoncologist.2008-0133v1 (The Oncologist, 15 September 2008)
  - Only 6% of cancer trials run by *commercial industry* are published

- Contributors must now sign to declare:
  - full responsibility for conduct of study
  - had access to data
  - controlled decision to publish

Evidence Based Medicine

- EBM
  - Objective is to consider critically all evidence that a treatment is effective
    - can never ***prove*** effectiveness or otherwise
    - formal hypothesis test can only assess *strength of evidence*
    - statistical test can only add to overall evidence
  - Randomized controlled trials are often regarded as the 'gold standard'
    - However RCTs often criticised, e.g. AIDS
    - Sometimes inappropriate, even for 'intervention studies'

- British Medical Journal
  - Smith & Pell (2003), **327**, 1459 – 1461
    - *Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials*

- **Results**
  - Our search strategy did not find any randomised controlled trials of the parachute.
- **Meta-analysis**
  - Our statistical approach was to assess outcomes in parachute and control groups by **odds ratios** and quantified the precision of estimates **by 95% confidence intervals**. We chose **the Mantel-Haenszel test** to assess heterogeneity, and **sensitivity and subgroup analyses and fixed effects weighted regression** techniques to explore causes of heterogeneity. We selected a **funnel plot** to assess publication bias visually **and Egger's and Begg's tests** to test it quantitatively. **Stata** software, version 7.0, was the tool for all statistical analyses.

- **Discussion**
  - **Evidence based pride & observational prejudice**
    - It is a truth universally acknowledged that a medical intervention justified by observational data must be in want of verification through a randomised controlled trial
    - We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute.
      - http://www.bmj.com/content/327/7429/1459.full.pdf

Evidence Based Medicine

- Although RCTs are generally the best for evaluating interventions they are not the only way
  - How do we evaluate quality of other types of evidence (especially observational studies)?
- Bradford-Hill Criteria
  - Sir Austen Bradford-Hill, FRS: statistical epidemiologist, co-author of first smoking & lung cancer study

- Bradford-Hill Criteria
  - help to assess cause or influence
    e.g. of medication or environmental factor
    - Temporality
    - Consistency
    - Coherence
    - Strength of association
    - Biological gradient
    - Specificity
    - Plausibility
    - Freedom from bias
    - Analogies

- **Summary & Conclusions**
  - Clinical trials are ***planned*** experiments from which wider ***inferences*** are to be drawn involving human subjects
  - Randomized controlled trials are the only effective type of clinical trial
    - they conform to the Bradford-Hill criteria
  - Bradford-Hill criteria provides method of validating conclusions from other types of study when RCTs are inappropriate

- **Summary & Conclusions** (ctd)
  - 4 phases of Clinical Trials
    - Phase III is comparative
  - Blind trials (double or single)
    - preferable to reduce bias
  - Placebo effects assessed by controls
  - Ethics are statisticians' responsibility

My work in medical statistics

- **Consulting**
  - Analysis of stage 3 clinical trials for pharmaceutical companies (with SSU)
    - One **complete** trial lasting ~5 years (on alpha-beta blockers for hypertension)
    - Complete:- design, sample size, data entry, management & auditing, analysis,
  - Advice on design & analysis at statistical clinics
  - Advice on sample size calculations
    - Most common problem everybody encounters
- **Research**

My work in medical statistics

- **Research**
  - Analysis of restricted randomization trials *(with Elsa Valdez-Marquez)*
    - (see '**minimization**' methods later)
    - Efficient designs for small trials (~50) with subject with several prognostic factors
  - Non-inferiority trials *(with Nor Afzalina Azmee)*
    - Trials to test whether a new drug is 'no worse' than a standard drug (& better than a placebo)
  - Biomarkers

My work in medical statistics

- **Research**
  - Biomarkers *(with Lu Zu, Quintiles, AstraZeneca)*
    - 'Personalised medicine'
      - Treatment given to those most likely to benefit
    - "Characteristic that is measured & evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention"
      - » ***Biomarkers Definitions Working Group 2001***
- Other work in ScHARR MSG:
  - Sample size calculations, biomarkers, quality of life measures, etc etc

Contents

- **Parallel Group Designs**
  - Each patient receives 1 treatment
  - Comparisons are '**between**' patients
    - Average difference between groups needs to be much larger than between the patients

- 2 groups
  - t-test
  - Mann-Whitney
- >2 groups
  - 1-way ANOVA
  - Kruskal-Wallis

- **Aside on two-sample t-tests:**
  - Two versions: pooled & separate variance
  - Separate: equal variances **not** assumed
    - Default in some packages (e.g. **R** & Minitab)
  - Pooled: equal variances assumed
    - Default in some packages (e.g. S-PLUS)
    - **SPSS** gives both
  - Always best to use separate variances
    - (equal variances **not** assumed)
  - If sample sizes equal then t-statistic same value
    - but p-value slightly different
    - unequal variance p-value bigger (i.e. less significant)

- Reason:–
  - model is $X \sim N(\mu_X, \sigma^2_X)$ & $Y \sim N(\mu_Y, \sigma^2_Y)$
    - estimate σ^2_X and σ^2_Y by s^2_X and s^2_Y
      - i.e. estimated variance separately
    - leads to a 'separate variance 2-sample t-test'
  - if we assumed σ^2_X and σ^2_Y were equal
    - would estimate variance differently & get a different t-test (on n_x+n_Y-2) d.f.
      - 'equal variance 2-sample t-test' (or 'pooled variance')

- **UNFORTUNATLEY**
  - 'equal variance' or pooled t-test is the default option in S-PLUS
    - even worse:–
      - S-plus calls it a 'Standard Two-Sample t-Test'
  - need to uncheck 'assume equal variance'
  - in many cases little numerical difference between two versions
    - if there is a big difference then the 'equal variance' value is **wrong**
    - so Good Statistical Practice (GSP) is **ALWAYS** to use separate variances

- **R** example:-

```
> hourssleep
PERIOD1 PERIOD2 GROUP  sum diff
      9       0     1  4.5    9
     11      14     2 12.5   -3
      7       3     2  5.0    4
     12       8     2 10.0    4
      8       8     1  8.0    0
     11       1     1  6.0   10
      4       4     1  4.0    0
      3       4     2  3.5   -1
     13       2     1  7.5   11
      7       3     2  5.0    4
      1       2     1  1.5   -1
     13       1     1  7.0   12
      6       3     1  4.5    3
      5       6     2  5.5   -1
      6       8     2  7.0   -2
      3       7     2  5.0   -4
> attach(hourssleep)
> t.test(PERIOD1,PERIOD2)
```

Two separate columns (unstacked)

- **R** example:-

```
> attach(hourssleep)
> t.test(PERIOD1,PERIOD2)
        Welch Two Sample t-test

data:  PERIOD1 and PERIOD2
t = 2.1422, df = 29.965, p-value = 0.04042
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
 0.1310748 5.4939252
sample estimates:
mean of x mean of y
   7.4375    4.6250
```


- **R** example:-

```
> hourssleep
PERIOD1 PERIOD2 GROUP  sum diff
      9       0     1  4.5    9
     11      14     2 12.5   -3
      7       3     2  5.0    4
     12       8     2 10.0    4
      8       8     1  8.0    0
     11       1     1  6.0   10
      4       4     1  4.0    0
      3       4     2  3.5   -1
     13       2     1  7.5   11
      7       3     2  5.0    4
      1       2     1  1.5   -1
     13       1     1  7.0   12
      6       3     1  4.5    3
      5       6     2  5.5   -1
      6       8     2  7.0   -2
      3       7     2  5.0   -4
> attach(hourssleep)
> t.test(PERIOD1~GROUP)
```

One column + group indicator (stacked)

- **R** example:-

```
> hourssleep
PERIOD1 PERIOD2 GROUP  sum diff
      9       0     1  4.5    9
     11      14     2 12.5   -3
      7       3     2  5.0    4
> attach(hourssleep)
> t.test(PERIOD1~GROUP)
 Welch Two Sample t-test

data:  PERIOD1 by GROUP
t = 0.7163, df = 13.183, p-value = 0.4863
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
 -2.766425  5.516425
sample estimates:
mean in group 1 mean in group 2
          8.125           6.750
```


SPSS Example:

Summary statistics for variable 'diff'

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
diff	1	8	5.50	5.529	1.955
	2	8	.13	3.357	1.187

Independent Samples Test

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
diff	Equal variances assumed	2.350	14					
	Equal variances not assumed	2.350	11.543	.037	5.375	2.287	.370	10.380

Concentrate on this

ignore this bit

t-value and p-value

- Separate variance:

$$t_r = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

  - Used as default in **R** & Minitab
    - Take [e.g.] r=min(n_1,n_2) (or use Welch formula)
    - **R** uses Welch formula

- Pooled variance

$$t_r = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

  - Default in S-Plus & available in **R,** Minitab:- $r=n_1+n_2-2$

- **In series designs**
  - Comparisons here are **'within'** patients
    - Differences between patients do not affect differences between treatment

- 2 groups
  - paired t-test (≡ t-test on differences)
  - Wilcoxon signed rank test
- >2 groups
  - 2-way ANOVA
  - Friedman's test

- Advantages
  - Patients can state preferences
  - Maybe apply treatments simultaneously
  - Comparisons **within** patients
    - reduces variability so more precise comparison

- Disadvantages
  - Trend in results
  - Treatments may persist
  - Not suitable if treatment **cures** subject! i.e. only suitable for chronic conditions
  - Withdrawals complicate analysis

- **Crossover Designs**
  - Similar to *in series design* but different groups have treatments in different orders
    - (see later)
  - Similar advantages and disadvantages

Factorial Designs

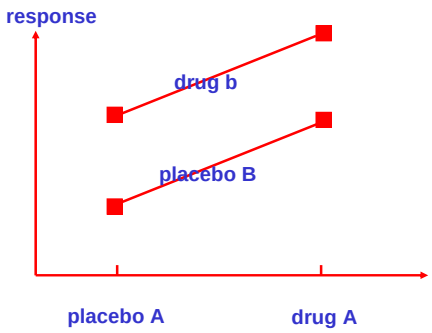
- Patients receive combinations of treatments

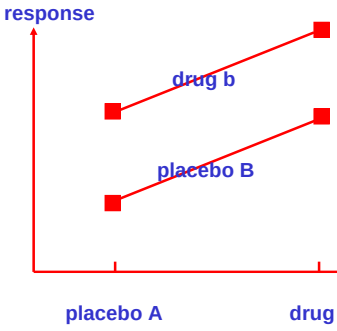
Placebo (10)	A (10)	**Not B (20)**
B (10)	A+B (10)	**B (20)**
Not A (20)	**A (20)**	**40**

Interaction

- No interaction
  - Drug A increases response by same amount irrespective of whether patient is also taking B or not

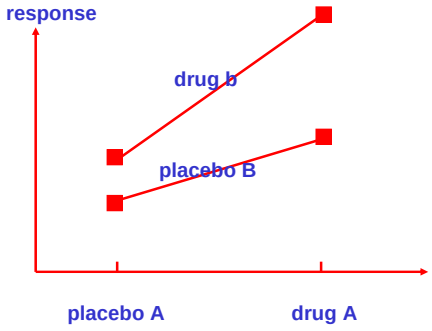
- Quantitative
  - effect of A more marked when patient is also taking B
- Qualitative
  - A increases response when alone, but decreases response when with B

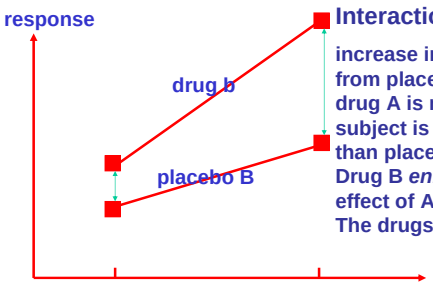




No Interaction:

increase in response from placebo A to drug A is same whether subject is on placebo B or drug B

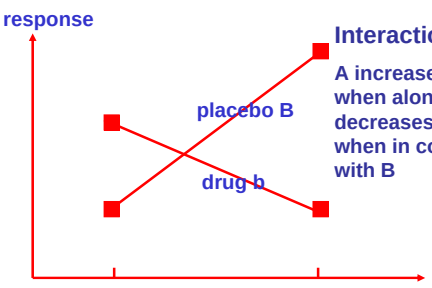




Interaction:

increase in response from placebo A to drug A is more when subject is on drug B than placebo B.
Drug B *enhances* effect of A.
The drugs *interact*

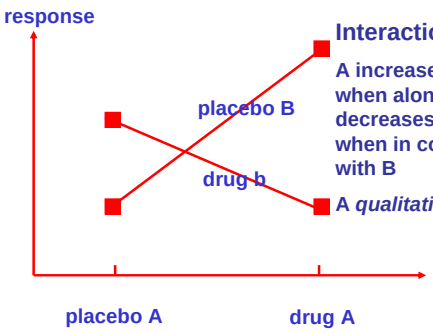
Quantitative interaction



Interaction:

A increases response when alone, but decreases response when in combination with B

Qualitative interaction



Interaction:

A increases response when alone, but decreases response when in combination with B

A ***qualitative interaction***

- **Sequential Designs**
  - Assess results after each few subjects and decide either
    - One treatment superior so stop trial
      Or
    - Test more subjects

- **Advantages**
  - Detect large differences quickly
  - Avoids ethical problem of fixed size designs
    - no patient should receive treatment known to be inferior

- **Disadvantages**
  - complicated statistical design and analysis
  - Responses needed quickly (before next group of patients arrive)
  - Drop-outs cause difficulties
  - Constant surveillance necessary
  - Requires grouping of patients

- **Summary & Conclusions**
  - Parallel group designs
    - different groups receive different treatments
    - comparisons ***between*** patients
  - In series designs
    - patients receive all treatments in sequence
    - comparisons ***within*** patients
  - Crossover designs
    - patients receive all treatments in different orders
    - comparisons ***within*** patients

- Factorial designs
  - some patients receive combinations of treatments simultaneously
  - difficulties if ***interactions***
    - quantitative or qualitative
  - comparisons are *between* patients but more available than in series designs

- Sequential designs
  - suitable for rapidly evaluated outcomes
  - minimizes numbers of subjects when clear differences between treatments

- **Efficient design of clinical trials is a crucial ethical element contributed by statistical theory and practice**