Biostatistics 3H and 5M

Michael Waltenberger

January 2021

Contents

List of Tables

List of Figures

1 Course summary

1.1 Course structure

Biostatistics is the branch of statistics applied to problems in medicine. This course will focus on three main areas of biostatistics:

- **Clinical Trials** - statistical methods for designing medical trials for assessing the efficacy (effectiveness) of new treatments.
- **Survival analysis** - statistical methods for assessing how long individuals are likely to survive.
- **Epidemiology** - statistical methods for measuring disease risk and identifying risk factors for disease in the wider population.

The course consists of a set of pre-recorded short videos that replace 20 lectures which would have been given during a teaching-on-campus term. There won't be a revision lecture, unless popular demand suggests one. The videos contain auto-generated captions. While these captions are usually of high quality, they will contain some errors and should you have any questions regarding the transcripts don't hesitate to e-mail me. The videos are complemented by 3 online-live tutorials and 2 online-live computer labs. The first lecture is an online-live welcome lecture. During the welcome lecture the course outline will be discussed - no technical material will be covered during this lecture. The welcome lecture is recorded and is posted on moodle.

1.2 How to moodle

The general information section contains the announcements forum, the welcome lecture and a booking tool for open office hours. Weeks 1, 4 and 7 have a section that contains the lecture notes, forum and supplementary materials for a three week block. In week 3, 6 and 9 tutorial and lab sheets are posted. You are expected to work through them for the live session in the following week. Online-live tutorials/labs take place in week 4, 7 and 10. Week 1-9 contain 3 common sections: slides, annotated slides and videos. Past papers are posted in week 10. You have tick boxes to keep track of what you have already covered.

1.3 Videos

Videos are released every Monday morning. The first chapter was filmed with Microsoft Team and the second chapter with Echo360. I end most videos with a short summary. Videos are edited and therefore transitions might not seem smooth; that is normal and shouldn't worry you. There are some small corrections in the videos. The slides are only provided in pdf this year. If that causes you any issues, please be in touch and we will find a solution. Annotated slides are provided as well.

1.4 Lecture Notes

Chapter 1 of the notes is the course summary. Chapter 2 is provided on Monday morning in teaching week 1 (11.01.2021). Chapter 3 is provided on Monday morning in teaching week 4 (01.02.2021). Chapter 4 is provided on Monday morning in teaching week 7 (22.02.2021). My personal recommendation is to use pdf notes, however html notes are there if you prefer them. Please let me know if you find any typos!

1.5 Tutorials and Labs (subject to change - I will confirm by 25.01.2021)

There are 3 tutorials and 2 labs, each one hour long. The second lab covers survival analysis and epidemiology and takes place in week 10. Neither the labs nor the tutorials are recorded. That is to encourage participation! You can vote for the questions you would like to go through at the start of the session. I will then demonstrate how to solve these questions. You are encouraged to interrupt me as often as you would like to ask questions. The more interactive these sessions are the better! To preare the labs, I recommend to use your own R and Rstudio installation. If you already have R/Rstudio update both of them. I will record a video explaining how to install R and R studio and add it to moodle at least one week before your first lab. A potential back option that gives you access to R *might be* WVD (Windows Virtual Desktop). We are in the process of getting that ready.

1.6 Forums

You have 4 Forums, one general announcements forum which I will use to remind you on important dates (e.g. a tutorial happening soon). You also have one forum for every chapter in the course you can use to discuss the course material.
Forums will be a crucial way of asking questions during this year of online delivery. Please make use of them! You can ask whatever question you have at any point during the course, e.g. you don't have to ask questions relating to week 4 in week 4. I would love to see students answering each others questions (a rare event, yet to be witnessed by many lecturers) rather than waiting for a reply from me. All forums are anonymous so don't be shy. I will answer every question in the forum eventually, but will on purpose let some time pass to give you a chance to help each other. If you can explain a topic to someone else, you know you have properly understood it - so while answering other people's questions in the forums, you will invariably strengthen your own understanding as well. Don't worry if some responses are not correct, the discussion forum is exactly that - a way to discuss a problem rather than getting to the correct answer with the first reply. To make sure you all have confidence in the answers given in the forum I will correct wrong answers and also confirm correct ones.
I will reply to the forum within 3 working days (i.e. weekends do not count). Please resist from chasing me up before this time frame, and be assured that I will monitor the forum and reply to all of your questions.

1.7 Open office hours

Open office hours will be provided once a week during and after term time. They take place on Wednesday from 11-12. If you can't make that time please e-mail me to arrange a zoom meeting. I will not offer any help in the week before the exam. That is, since I might receive a large number of questions in the week before the exam which makes it impossible for me to answer them all. Answering only some student's questions gives them an unfair advantage in the exam. Therefore, study continuously throughout the term and ask questions as they pop up - aim to be done with your revision one week before the exam.

1.8 How to study

Watch the videos AND go through the notes. Consult the slides if you need a concise summary. Go through tutorial sheets and lab sheets. Use past papers. In short, study everything that is on Moodle. Unless explicitly stated in one of the videos or the notes, all the material is examinable.

1.9 Sources of help

If you have any problems with the material in the course then there are a number of options for getting help.

1. Use the forum on Moodle.
2. Ask questions at the tutorial sessions, that is why we have them!

3. Use the library. You can sometimes also access e-books from home using your library account.
4. Make use of open office hours.

1.10 Assesements

The course is examined 100% by exam, which will be held in the spring exam diet. More information will follow in due course.

1.11 Further reading

There are a large number of books on biostatistics, some of which can be found in the library. Here are just a few. This course is self-contained and no further reading is required to understand the material covered. Further reading is provided for interested students who would like to learn material outside the scope of this course.

- Chalmers I, Altman DG, Systematic Reviews in Health Care, BMJ Publishing (2001).
- Clayton D and Hills M, Statistical models in epidemiology, Oxford Press (1993).
- Hosmer D, Lemeshow S and May S, Applied Survival Analysis : Regression Modeling of Time-To-Event Data, Wiley (2008).
- Le C and Eberly L, Introductory Biostatistics, Wiley (2016).
- Machin D, Campbell MJ, Tan SB, Tan SH, Sample size tables for clinical studies (4th edition), Wiley-Blackwell (2018).
- Matthews JNS, An Introduction to Randomised Controlled Trials (2nd edition) Chapman and Hall / CRC (2006).
- Pocock SJ, Clinical Trials. A Practical Approach, Wiley (1983).
- Senn S, Statistical Issues in Drug Development (2nd edition), Wiley (2007).
- Woodward M, Epidemiology: Study Design and Data Analysis (2nd edition), Chapman and Hall / CRC (2005).

1.12 Aims and intended learning outcomes

The aim of this course is to give students an introduction to biostatistics, which is statistics applied to the field of medicine.

1.12.1 Biostatistics 3H

By the end of the course students should be able to:

- Describe a range of biostatistical study designs, describe their key features, and determine the appropriateness of each one for real epidemiological investigations.
- Describe a range of summary statistics and simple statistical models used to quantify biostatistical data, and be able to apply these to real data.
- Describe measures for quantifying the impact of a covariate factor on disease risk, and compute and interpret them in real epidemiological studies.

- Describe the key features of survival data, the Kaplan Meier estimator and the proportional hazards model, and be able to apply them to real data.
- Calculate appropriate sample sizes for clinical trials and interpret them in the context of real clinical trials.

1.12.2 Biostatistics 5M

By the end of the course students should be able to:

- Describe a range of biostatistical study designs, describe their key features, and determine the appropriateness of each one for real epidemiological investigations.
- Describe a range of summary statistics and simple statistical models used to quantify biostatistical data, theoretically derive their properties, and be able to apply them to real data.
- Describe measures for quantifying the impact of a covariate factor on disease risk, describe their theoretical basis, and compute and interpret them in real epidemiological studies.
- Describe the key features of survival data, the Kaplan Meier estimator and the proportional hazards model, explain their theoretical basis, and be able to apply them to real data.
- Calculate and derive theoretically appropriate sample sizes for clinical trials and interpret them in the context of real clinical trials.

2 Clinical Trials

2.1 A brief history of clinical trials

Video2.1 - A brief history of clinical trials

We start this chapter by giving a brief historical perspective of clinical trials.

Definition A **Clinical Trial** is any form of planned experiment which involves human patients. Generally, the trial is carried out on a sample of patients (selected from the population of patients with the disease), and is designed to identify the most appropriate treatment for the population of future patients with the disease.

Note Clinical trials, and in particular Randomised Controlled Trials are now accepted as standard practice in all areas of medicine. However they are a relatively recent development, evolving mainly during the second half of the 20th century. It is worth beginning with a look at their early development, before looking at modern trials in more detail.

Historically, the evaluation of improvements in treatment for various diseases has been an inefficient and haphazard process. The same could be said about the application of treatments to individuals. For example, the treatments given to King Charles II (Figure 1) in 1685:

Figure 1: King Charles II.

'A pint of blood was extracted from his right arm, and a half-pint from his left shoulder, followed by an emetic, two physics, and an enema comprising fifteen substances; the royal head was then shaved and a blister raised; then a sneezing powder, more emetics, and bleeding, soothing potions, a plaster of pitch and pigeon dung on his feet, potions containing ten different substances, chiefly herbs, finally forty drops of extract of human skull, and the application of bezoar stone; after which his majesty died'

This is despite the fact that the need for close clinical observation and careful deduction concerning disease had been recognised since the time of Hippocrates. There were many areas where diagnosis and assessment of

prognosis was good. However, when it came to treatment there was a tendency to rely on traditional remedies - largely because there were no effective treatments for many of the major (infectious) diseases.

The need for controlled experiments was recognised by some in the 18th century e.g. Lind (Figure 2) did a comparative trial of treatments for scurvy published in 1753:

Figure 2: James Lind.

'I took twelve patients in the scurvy on board the Salisbury at sea. The cases were as similar as I could have them . . . they lay together in one place . . . and had one diet common to them all. Two of these were ordered a quart of cider a day. Two others took twenty-five gutts of elixir vitriol . . . Two others took two spoonfuls of vinegar . . . Two were put under a course of sea water . . . Two others had each two oranges and one lemon given them each day . . . Two others took the bigness of a nutmeg. The most sudden and visible good effects were perceived from the use of oranges and lemons, one of those who had taken them being at the end of six days fit for duty . . . The other . . . was appointed nurse to the rest of the sick.'

Despite this, he continued to propose 'pure dry air' as a first priority with fruit and vegetables as a secondary recommendation. He had 12 patients and 6 different treatments! It was more than 50 years later that the British navy supplied lemon juice on its ships!

However, Lind was well ahead of his time in his appreciation of the scientific method. Consider this quote from Benjamin Rush (1794) on treatment of yellow fewer by bleeding:

'I began by drawing a small quantity at a time. The appearance of the blood and its effects upon the system satisfied me of its safety and efficacy. Never before did I experience such sublime joy as I now felt in contemplating the success of my remedies . . . The reader will not wonder when I add a short extract from my notebook, dated 10th September. Thank God, of the one hundred patients, whom I visited, or prescribed for, this day, I have lost none.'

Bleeding as a treatment for various disorders was finally discredited by Louis (Figure 3) in 1835:

Figure 3: Pierre-Charles-Alexandre Louis.

Louis treated 134 patients suffering from pneumonia, erysipelas (acute infection of skin and subcutaneous tissue due to Streptococcus) or throat inflammation and found no difference in outcome for patients bled and not bled.

Controlled trials of this type were a major step forward, but they often simply discredited the traditional treatments and proposed nothing new in their place. However, there were many real advances in the 19th century.

Figure 4: Joseph Lister.

In 1870, Lister (Figure 4) looked at the effect of antiseptics on mortality rates following amputation. The mortality rates were:

- $\frac{15}{35} = 43\%$ in operations before antiseptics were introduced.
- $\frac{6}{40} = 15\%$ in operations after antiseptics were introduced.

Lister commented that 'the numbers are doubtless too small for a satisfactory statistical comparison'. Do you agree?

This simply requires us to assess if two proportions are significantly different. Let (X_1, X_2) respectively denote the number of patients that died in the groups without (X_1) and with (X_2) antiseptics respectively. Then we have that:

$$X_1 \sim \text{Binomial}(n_1 = 35, \theta_1), \qquad X_2 \sim \text{Binomial}(n_2 = 40, \theta_2).$$

Then a natural estimator of the difference is

$$\hat{\theta}_1 - \hat{\theta}_2 \;=\; \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

It is straightforward to show that the expectation and variance of this estimator are:

$$\mathbb{E}[\hat{\theta}_1 - \hat{\theta}_2] \;=\; \theta_1 - \theta_2 \qquad\qquad \mathbb{V}[\hat{\theta}_1 - \hat{\theta}_2] \;=\; \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}.$$

Thus based on a normal approximation a 95% confidence interval for the difference between the two proportions is:

$$\hat{\theta}_1 - \hat{\theta}_2 \pm 1.96 \times \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}.$$

Here we have that

$$\hat{\theta}_1 = 15/35 = 0.43 \qquad\qquad \hat{\theta}_2 = 6/40 = 0.15 \qquad\qquad \hat{\theta}_1 - \hat{\theta}_2 = 0.28,$$

and

$$\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}} \;=\; 0.1009.$$

So a 95% confidence interval for the difference is given by:

$$(0.081, 0.476),$$

which does not include zero, so there is a statistically significant difference at the 5% level.

Video2.2 - The 1954 Field Trial of the Salk Poliomyelitis Vaccine

Definition

- A **placebo** is a treatment given to a patient that is known not to work. This is done to overcome the placebo effect, which is where a patient feels better because they are taking medicine, even though the medicine has no medical effect. Thus by giving some patients the real treatment and others the placebo you can overcome this effect. As a side note, the placebo effect has been the downfall of Raj from the big bang theory.
- A **control** is a patient who does not receive the treatment, and is used for comparison purposes against patients who receive the treatment. An example of a control is a patient who receives the placebo.
- **Random allocation** is the process by which patients are allocated to each treatment or placebo at random, so no systematic differences exist between the set of patients on the treatment and the placebo.
- **Blinding** (single or double) occurs if the patient (single) or the patient and doctor (double) do not know which treatment a patient is on. This is a good idea as otherwise placebo's will not work. Also, if the doctor knows you are on the treatment they may diagnose an improvement more readily than if you are on the placebo.

2.2 Example - The 1954 Field Trial of the Salk Poliomyelitis Vaccine

We will now look in more detail at one of the largest and most expensive medical experiments ever conducted - the trial of Salk's Polio Vaccine in 1954. Jonas Salk (Figure 5) famously made the vaccine available to everyone without a patent. When questioned about it, he replied *Could you patent the sun?*

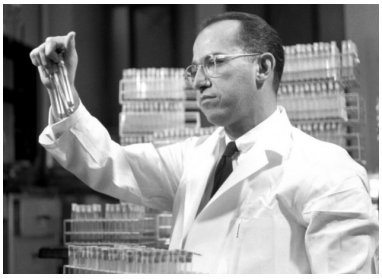

Figure 5: Jonas Salk.

Polio is known to be caused by a virus which enters the body through faecal contamination of food or water. The infection may be unnoticed, or cause minor illness, non-paralytic inflammation of the nervous system or serious and life threatening paralysis. Between 1900 and 1940, large outbreaks occurred in Scandinavia, the USA and the UK. In 1947 in the UK there was a sudden increase with 7776 paralytic cases compared to the previous high of 1489 in 1938.

In the early 1950s the Advisory Committee convened by the National Foundation for Infantile Paralysis (NFIP) decided that the killed-virus vaccine developed by Jonas Salk at the University of Pittsburgh had been shown to be both safe and capable of inducing high levels of antibody in children on whom it had been tested. This made the vaccine a promising candidate for general use, but it remained to be proven that the vaccine would actually prevent polio in exposed individuals. The assumption was that it would not be justifiable to release such a vaccine for general use without convincing proof of its effectiveness, so it was determined that a large-scale 'field trial' of the vaccine should be undertaken.

Well over a million young children participated. The experiment was carried out to assess the effectiveness, if any, of the Salk vaccine as a protection against paralysis or death from poliomyelitis. The study was elaborate in many respects:

- in the use of placebo controls (that is, children who were inoculated with simple salt solution),
- randomisation (a carefully applied chance process which gives each volunteer an equal probability of getting the test drug or placebo).
- The children were then subjected to a double-blind evaluation; that is, an arrangement under which neither the children nor the physicians who evaluated their subsequent state of health knew who had been given vaccine and who got the salt solution.

2.2.1 Three possible approaches to designing the trial

1. The **Vital Statistics** approach, with two possible designs.

   - Distribute the vaccine as widely as possible and see whether the reported incidence of polio is less than usual during the following year.
   - Use the vaccine in one area and compare the incidence rate in the following year with that of another unvaccinated area.

   This approach does not work because:

   - Large variation from year to year.
   - There might be other factors influencing polio incidence that we can't control.
   - Natural variation from area to area that might have an effect on polio incidence.

2. The **Observed Control** approach offers the vaccine to all children in second grade of schools in areas participating in the trial and compare the incidence rate of polio in those children vaccinated (i.e. the treated group) with the incidence rate in first and third grade children in the same schools (i.e. the control group). This has obvious advantages over the 'Vital Statistics' approach in that the comparison is between two groups of children from the same geographical areas in the same year. However, there are still drawbacks:

   a) Those second grade children who actually receive the vaccine are volunteers and (as was actually shown later) tend to be from better-off, better-educated families than those who did not volunteer.
   b) While there are no problems of diagnosis of cases of paralytic polio, diagnosis of mild cases of polio is more difficult. With the 'Observed Control' approach the doctor making the diagnosis will know whether the child has been vaccinated or not and this may influence his or her judgement.

If this method were employed, many epidemiologists and statisticians would remain unconvinced about the validity of the trial, and the value of the vaccine would still be uncertain. A more rigorous scientific approach is needed.

3. **Randomisation and the Placebo Control Approach**:

   - This differs from method 2 in that each volunteer is assigned to treatment with either vaccine or an ineffective salt solution (i.e. placebo) in a random manner so that each volunteer has an equal chance of receiving vaccine or placebo. This should ensure that the two groups are comparable, and overcome drawback (a) above.
   - Each vial of injection fluid is identified only by a code number so that neither the children nor their parents nor the doctors involved in the diagnosis know which treatment the child has received. This gives rise to a double-blind experiment, and overcomes drawback (b) above.

2.2.2 Results of the study

Some health departments accepted the 'Observed Control' method as satisfactory while others would only participate if a randomised, double blind trial with placebo controls were employed. The outcome was that each of these two methods was used in approximately half the areas involved. As an interesting side issue, this allowed the results from the two types of experiment to be directly compared! The results of the study in Table 1 show that the Placebo Control design is more powerful as it shows more of a difference between vaccinated and control children. Thus, careful and statistically informed experimental design contributed greatly to the success of this trial of Salk polio vaccine. As a result of this trial, the Salk vaccine was adopted for general use in the prevention of polio. However, several years later, the Salk vaccine was withdrawn after it was shown that some batches of vaccine, which had been incorrectly prepared, had actually caused polio infection in children. As a result 5 people were killed and 51 paralysed. Eventually in 1962, a new live virus vaccine was developed (Sabin vaccine) and adopted for general use. Very similar vaccines are still in use today.
You might notice that the number of polio cases per 100,000 children in the observe control group is smaller than the number of children in the placebo control group. That might seem counterintuitive at first, since children in the placebo control group are volunteers and are expected to be on average healthier than their non-volunteer counterparts in the observed control group. In this example, volunteer bias leads to the placebo control group having a lower antibody count for polio, which results in this group being more susceptible to the virus.

Table 1: Comparison between Observed Control areas method and Placebo Control areas method in the Salk polio vaccine trial. The number of polio cases per 100,000 children is shown. The placebo method appears to show a greater effect.

	Observed Control Areas	**Placebo Control Areas**
Vaccinated	25	28
Control	54	71

2.3 Design of clinical trials

Video2.3 - Design of clinical trials I

Definition Within the pharmaceutical industry, drug trials are usually classified into one of 4 phases (Pocock, 1983).

1. **Phase I - Clinical Pharmacology and Toxicity** - These are primarily concerned with drug safety, and are performed on healthy human volunteers (not patients). Dosing schedules are also established.
2. **Phase II - Initial Clinical Investigation for Treatment Effect** - These are small scale studies in patients, looking at the effectiveness and safety of the drug. Is there any evidence that the drug actually works and is non-toxic? It is becoming increasingly common for phase II trials to be randomised, but in the (fairly recent) past they were generally uncontrolled (i.e. no control group).
3. **Phase III - Full Scale Evaluation of Treatment** - After a drug is shown to be reasonably effective and worthy of a full scale investigation, it is essential to compare it with current standard treatments for the same disease (or with a placebo) in a randomised controlled trial. Phase III trials involve hundreds and often thousands of people.

4. **Phase IV - Post-marketing Surveillance** - If a drug is approved for marketing, its performance is still closely monitored for evidence of side effects and to obtain more information about its long-term effectiveness. It is still possible for treatments to be withdrawn at this stage if they produce serious side effects which are too rare to be detected in a phase III trial, and only become apparent when data are available for large numbers of treated patients.

However, things can go wrong as shown in this Guardian article. This tragic event was later made into a documentary and serves as an unprecedented example of the importance of properly conducted clinical trials. I include this link for people who are interested, it is not examinable and you do not have to watch this documentary to study for this course. Note, that in that trial there is a pre-stage to the 4 phases listed above, that of animal testing, which is a controversial topic. In addition to animal tests, drugs are also tested on cell cultures as well.

2.3.1 Control groups

Suppose a pharmaceutical company wishes to test the efficacy of a new treatment for a particular disease. Then they need something to compare the efficacy of their treatment to. This leads to the notion of a control group.

Definition As defined previously, a **control** is a patient that does not receive the treatment, and thus a **control group** is the set of patients that do not receive the treatment. Patients in the control group can receive the currently used **standard treatment** or a **placebo**.

Note, often, the clinical trial is split into 3 groups, the treatment group, the standard treatment group and the placebo group. In this case there are two control groups.

(i) Trials using historical controls For any investigator with a new treatment to test, there is a natural desire to treat all future patients with the new treatment (and thus avoid randomisation). The most obvious way to do this is to use historical controls - i.e. compare current patients on the new treatment with previous patients on the old treatment. Thus the previous patients are the historical controls. There are, however, various sources of potential bias. The two main ones are:

- how do we know that patient selection criteria have remained constant?
- the experimental environment for the new drug may lead to patients getting more and better, attention as medicine is always improving.

(ii) Trials using Concurrent Non-Randomised Controls A trial could be organised with some kind of systematic allocation to treatment such as:

- Date of admission to hospital.
- Alternate assignment of every second patient to treatment and every other to placebo.

The problem is that, if it becomes known what method of allocation is being used, then it is possible for someone to manipulate it to allocate patients to the treatment they believe is best.

Example Table 2 shows a review of all of the published clinical trials of anticoagulant (blood thinner) therapy following myocardial infarction (i.e. heart attack), published in 1977. In this instance, the apparent benefits of anticoagulant therapy are greatly exaggerated in the studies with historical controls and non-randomised controls compared to the randomised trials. In this study, the case fatality rate on anticoagulant therapy in the randomised trials (15.4%) is significantly lower than that in the controls (19.6%), indicating that anticoagulant therapy is useful, but not to the extent suggested by the surveys with historical controls.

Table 2: Comparison between clinical trials of anticoagulants using historical controls, alternatively assigned controls and randomly assigned controls in clinical trials. Both historical and alternatively assigned controls appear to overstate the effect of the the drug.

		Fatality rate (%)	
	No. of patients	**Controls**	**Anticoagulants**
Historical controls	9090	38.3	22.3
Alternatively assigned controls	3144	29.2	22.6
Randomised controls	3854	19.6	15.4

2.3.2 Randomisation

It has been shown that, on purely scientific grounds, the use of a randomised control group is preferred since alternatives may lead to bias and overly optimistic results. However, is randomisation ethical?

Definition In 1964, the World Medical Association issued a statement of *Ethical Principles for Medical Research Involving Human Subjects*. This is known as the Declaration of Helsinki. It has been modified and updated a number of times to take account of new developments in medical research. A copy of the complete declaration is on Moodle.

There are various methods of constructing a randomisation list. The following illustrations assume two treatments A and B (or a treatment and a placebo), but can be easily generalised.

Definition **Simple Randomisation** uses random number tables and assigns each patient to their treatment (A or B) based on a pre-defined rule.

Example Consider a simple randomisation algorithm where each patient is randomly assigned an integer number between 0 and 9 which correspond to treatments as:

- Assign patient to treatment A if digits 0,1,2,3,4 are drawn.
- Assign patient to treatment B if digits 5,6,7,8,9 are drawn.

This has the advantage of being completely unpredictable, and the list can be made as long as necessary. Probability theory ensures that, in the long run (i.e. as the number of study patients $N \to \infty$), the numbers of patients assigned to each treatment will not be too different. However, as Table 3 shows (from *Pocock SJ, Clinical Trials. A Practical Approach, Wiley (1983).*), it is possible to end up with a large imbalance of patients in the two groups, which would be disastrous in a small trial. So with a small sample size simple randomisation is not a good idea because equal sized groups is far from guaranteed.

Table 3: Possible imbalance in simple randomisation with two treatments. This table shows the difference in treatment numbers (or more extreme) liable to occur with probability at least 0.05 or at least 0.01 for various trial sizes; source: Pocock, 1983.

	Difference in numbers	
No. of patients	**Probablity > 0.05**	**Probablity > 0.01**
10	2:8	1:9
20	6:14	4:16
50	18:32	16:34
100	40:60	37:63
200	86:114	82:118
500	228:272	221:279
1000	469:531	459:541

Definition A **Random permuted blocks** design ensures exactly equal numbers of patients on each treatment. For a large trial involving more than 100 patients it might be reasonable to have blocks of size 20. (If the block size is too small, the sequence becomes predictable at the end of each block). Random permuted block designs can be implemented in a similar way to simple randomisation, except that it is done separately for each block and the same number of patients are enrolled in each group within each block.

Example Suppose you have treatments A and B and 100 patients and 5 blocks of 20. Suppose also you have the following assignment algorithm:

- Assign patient to treatment A if numbers 0 - 9 are drawn.
- Assign patient to treatment B if digits 10 - 19 are drawn.

Then separately for each of the five blocks create the assignment by randomly drawing the numbers 0-19. Once this is done create a second random draw of the numbers 0-19, and so on for the other blocks. So if the assignment was:

5, 17, 2, 4, 16, 19, 10, 11, 14, 7, 12, 15, 1, 18, 6, 9, 0, 3, 13, 8.

Then the patient assignment is

A, B, A, A, B, B, B, B, B, A, B, B, A, B, A, A, A, A, B, A.

for the first 20 patients. The block size can easily be varied from block to block to reduce predictability. This is especially useful if small blocks are to be used.

Definition Randomisation as described above, with or without using random permuted blocks, will ensure approximate balance between groups in terms of factors affecting a patient's outcome (prognostic factors). This is true for both known and unknown factors. However, if there are factors which are known to affect a patient's outcome and for which data are available before randomisation is carried out, it may be desirable to ensure that roughly equal numbers of patients are allocated to each treatment in each subgroup defined by the factor. This ensures that the treatment groups are balanced in terms of such a factor. This is of greater importance if the trial is small, since the chances of a large imbalance in treatment allocation decreases with increasing sample size. This process is known as **stratified randomisaton**.

Example As an example of how this works in practice consider a trial of two treatments for primary breast cancer in which two factors known to affect survival are age and number of positive axillary nodes. To avoid over-elaboration, each factor was reduced to 2 categories, defining 4 strata: Random permuted blocks of size 4 were then used in each strata giving the pattern seen in Table 4.

Table 4: An example of random permutated blocks within strata for a trial in primary breast cancer. The data has been stratified using age and the number of positive auxiliary nodes; source: Pocock, 1983.

Age / No. of +ve nodes			
$<$50 / 1-3	**$\geq$50 / 1-3**	**$<$50 / $\geq$4**	**$\geq$50 / $\geq$4**
B	B	A	B
A	B	A	A
B	A	B	A
A	A	B	B
A	A	B	A
B	A	A	B
A	B	B	B
B	B	A	A
A	B	B	B
B	A	A	B
B	A	B	A
A	B	A	A
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Definition Standard practice in a randomised trial with 2 treatments is to allocate a patient to A or B with probability $\theta = 0.5$. Statistically this provides the most efficient means of treatment comparison. However, if one of the treatments is a new treatment there is likely to be enthusiasm for its use among investigators, as well as interest in gaining experience of its general performance. Thus, it might be worth randomising a higher proportion of the patients to the new treatment, provided the loss of statistical efficiency is not great. This is known as **unequal randomisation**.

Example The power (the probability of detecting a statistically significant difference in treatment effects when one is present) of the study when equal sample sizes are used is 0.95. Figure 6 illustrates the effect on the power if the percentage of patients on the new treatment is increased. The loss of power is not great, provided that the percentage of patients on the new treatment is not increased much above 75%. Beyond this, the power decreases rapidly.

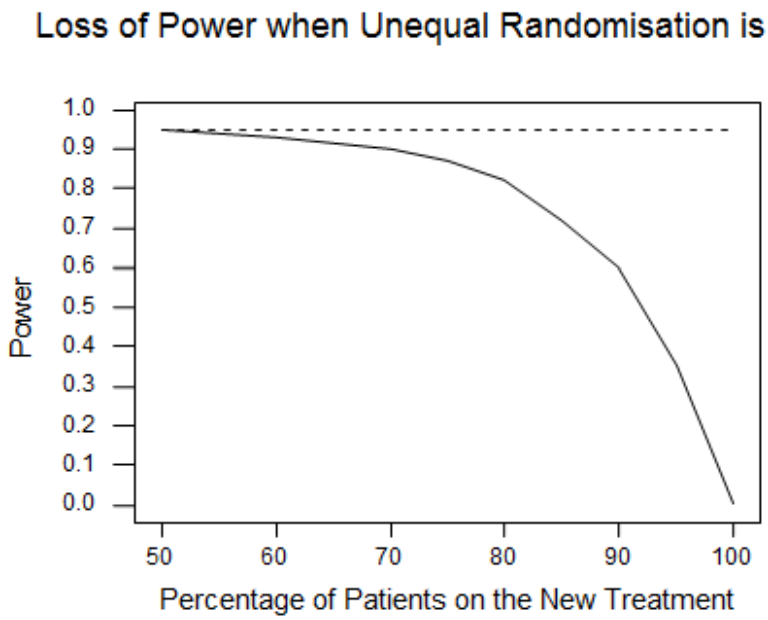

Figure 6: Graph shows how power decreases as the allocation of patients to the two treatment groups becomes uneven. The loss of power starts to decrease sharply once a 75/25 split is exceeded.

2.3.3 Blinding

As previously discussed **Blinding** (single or double) occurs if the patient (single) or the patient and treatment team (double) do not know which treatment a patient is on. In a randomised trial, the comparison of treatments may be distorted if the patient and those responsible for treatment and evaluation know which treatment is being used. This can occur as follows:

- **The Patient** - psychological effects of knowing he/she is receiving a new/standard treatment. This is obvious in treating psychiatric illness, but is also known to occur in many other areas. Notably, this fact has been used by Sheldon Cooper from the Big Bang Theory in a failed attempt to cure Amy.
- **The Treatment Team** - doctors, nurses may observe patients on a new treatment more closely, give them more attention, be more enthusiastic and thereby affect the patient's attitude and consequently their response to the treatment.
- **The Evaluators** - may be biased towards recording more favourable responses on the new treatment, especially (but not solely) with outcomes which are rated subjectively.

For these reasons, the most scientifically acceptable option is to carry out a **double-blind** study in which neither the patients nor those treating and evaluating their condition are aware which treatment the patient is receiving. Obviously this must be implemented with care and arrangements made to **unblind** the study for any patient in an emergency.

Double blinding a trial may not be possible for:

- trials involving surgery;
- trials of highly toxic drugs.

2.3.4 Types of study design

Video2.4 - Design of clinical trials II

There are a number of different clinical trial designs, and we describe the main ones here.

Definition The simplest design is a **parallel group trials**, in which individual patients are allocated to one and only one of a set of distinct treatment groups such as (treatment, placebo), or (new treatment, existing treatment, control).

Notes

- Parallel group trials are the most natural approach for diseases or conditions where the objective is to 'cure' the patient.
- There are many chronic conditions such as rheumatism, diabetes or asthma where there is (currently) no known cure and the objective of the trial is to alleviate symptoms. In this situation one could use a crossover trial.

Definition The simplest type of **crossover trial** (Table 5) comparing two treatments is the **two treatment, two period crossover design (AB / BA design)**, in which patients are randomly allocated to receive either A followed by B (Group I) or B followed by A (Group II). There is a break between the two treatment periods sometimes known as a **washout** - to enable the patients condition to return to a level uninfluenced as far as possible by the treatment used in Period 1.

The **run-in** period is optional, but it provides an opportunity for the investigator to make a baseline observation. The response variable will often be measured at the end of each of the treatment periods to allow the full effect of the treatment to be measured, but it could be the average or maximum of a series of observations taken throughout each period, or the difference in values between the start and end of the period.

Despite the apparent simplicity of this design, the interpretation of the results can be surprisingly complex - especially if there is a **carry over effect** in which the effect of the treatment given in the first period has an influence on the patients' response in the second period.

Table 5: Stages of crossover trial; source: Pocock, 1983.

	Run-in	**Period 1**	**Washout**	**Period 2**
Group I (n_1)	–	Treatment A	–	Treatment B
Group II (n_2)	–	Treatment B	–	Treatment A

Defintion A **Cluster randomised trial** is one in which groups of individuals are randomly allocated together to one of the treatments being compared. They are particularly useful in comparisons of different methods of organising and delivering treatments to patients.

Example 25 hospitals around the UK were invited to participate in a trial to compare different methods of providing feedback to ward staff about the incidence of MRSA infection in their wards. Within each participating hospital, two eligible wards were identified. One ward was randomly allocated to feedback using data presented in the form of charts; the other ward received standard (non-graphical) feedback. Data were collected on a monthly basis in each ward by observing whether or not each individual patient develops MRSA.

2.4 Sample sizes for clinical trials

Video2.5 - Sample sizes for clinical trials I

There are three main issues to be considered in determining how many patients should be entered into a clinical trial:

1. **Practical issues**:- For example: how many patients are available for study? What resources are available (time, staff, money, drugs etc).
2. **Requirements**:- The trial requires as many patients as possible to reduce the standard error of the estimated treatment effect, and therefore reduce the chances of drawing the wrong conclusion from the trial.
3. **Ethical requirements**:- The trial requires as few patients as possible to be randomised, so that if the new treatment really is more effective (or less effective) than the standard, the number of patients who are given the inferior treatment is kept to a minimum.

Clearly, (2.) and (3.) are in conflict. One key job of a statistician on a clinical trial is to compute the number of patients that should be recruited into the trial. We discuss the theory in the context of the following example.

2.4.1 Example - The Anturan Re-infarction Trial

The Anturan Re-infarction Trial was a randomised, placebo-controlled trial of a new treatment (anturan) to be given to patients who had been admitted to hospital following a myocardial infarction (heart attack). These patients were still at high risk, and it was thought that this new treatment would reduce the chances of a second heart attack and therefore improve the patients long-term survival prospects. Remember that for the trial to be ethically acceptable there must be genuine uncertainty about whether anturan is beneficial. At the start of this study there were sound theoretical reasons for believing that anturan would reduce the chances of a second heart attack, but it was not known whether this would be offset by increases in mortality from other causes due to unwanted side effects of the drug. To determine the number of subjects to include in the trial, the following are required:

1. **Purpose of the Trial**:- to investigate whether anturan is of value in preventing mortality after myocardial infarction.
2. **Primary Measure of Patient Outcome**:- death from any cause within one year of commencing treatment.
3. **Analysis of data**: - primary analysis will be by a simple comparison of the percentages of patients dying within one year on anturan and placebo.
4. **Results Expected on Standard Treatment**:- at the time of this study, there was no standard treatment, so placebo was used as a control. Based on past records, about 10% of patients who would be eligible for inclusion in a study of this type would have been expected to die within a year.
5. **Minimum Clinically Significant Difference between Treatments**:- what is the smallest difference that is of such clinical value that it would be undesirable to fail to detect it? It could be argued that any treatment benefit, no matter how small, is relevant and must be detected, but this is unrealistic since the

sample size would have to be infinite to achieve this! Realistically, if the true effect of anturan were to halve the mortality rate from 10% to 5%, we would like to have a high probability of detecting this in the trial (i.e. obtaining a statistically significant result). Thus, we wish the study to have a high power (typically 0.80 - 0.95) to detect a difference in mortality rates of 10% versus 5%.

We now derive how to calculate the sample size needed for a trial such as this one. However, before that we discuss some preliminary statistical ideas you should know.

Definition Consider a hypothesis test with a null hypothesis H_0 and an alternative hypothesis H_1. Then there are four outcomes shown in Table 6.

Table 6: Relationship between null/alternative hypothesis with power/significance and the related errors.

	Reject $\mathbf{H_0}$	**Fail to reject $\mathbf{H_0}$**
$\mathbf{H_0}$ is true	Type I error: α	Confidence: 1-α
$\mathbf{H_1}$ is true	Power $1-\beta$	Type II error: β

Thus we have that

- $\mathbf{P}(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$. This is the **significance level** of the test.
- $\mathbf{P}(\text{Reject } H_0 | H_1 \text{ is true}) = 1-\beta$. This is the **power** of the test.

Definition Consider a normal (Gaussian) random variable $X \sim \mathsf{N}(\mu, \sigma^2)$, then we know that

$$Z = \frac{X-\mu}{\sigma} \sim \mathsf{N}(0,1).$$

Furthermore, due to the symmetry of the standard normal distribution around zero, we know that:

- $\phi(r) = \mathbf{P}(Z \leq r) = \mathbf{P}(Z \geq -r) = 1 - \phi(-r)$.
- $\phi(-r) = \mathbf{P}(Z \leq -r) = \mathbf{P}(Z \geq r) = 1 - \phi(r)$.

Finally, if $\phi(r) = \mathbf{P}(Z \leq r) = C$, for some constant $C \in [0,1]$, then $r = \phi^{-1}(C)$, where ϕ^{-1} is the inverse normal function. Thus the normal functions $[\phi(), \phi^{-1}()]$ tell you:

- $\phi()$ - For a given real number r, $\phi(r)$ tells you the probability of a standard normal random variable being less than or equal to r.
- $\phi^{-1}()$ - for a given probability C, tells you the real number r that satisfies $\mathbf{P}(Z \leq r) = C$.

This now allows us to state and prove an approximate sample size result in the context of the Antruan trial.

Theorem Consider two treatments: A and B, where for example: A - Placebo and B - Anturan. Let θ_A denote the probability of a patient dying within one year having taken the placebo, and θ_B denote the probability of a patient dying within one year having taken Anturan. Then the hypotheses we wish to test are:

$$H_0 : \theta_A = \theta_B \qquad \text{vs} \qquad H_1 : \theta_A \neq \theta_B.$$

Assuming we want to have a significance level of α and power of $1-\beta$, then based on a normal approximation the number of patients we need to detect the clinical difference $\delta = \theta_A - \theta_B$ in each group is:

$$N = \frac{\theta_A(1-\theta_A) + \theta_B(1-\theta_B)}{(\theta_A - \theta_B)^2} \left[\phi^{-1}(1-\alpha/2) + \phi^{-1}(1-\beta)\right]^2.$$

2.4.2 Proof

Video2.6 - Sample sizes for clinical trials II

Suppose we have $2N$ patients, with N randomised onto Anturan and N randomised onto the placebo. Then if (X_A, X_B) denote the number of people on each treatment that die within one year, then we know that:

$$X_A \sim \text{Binomial}(N, \theta_A) \qquad \text{and} \qquad X_B \sim \text{Binomial}(N, \theta_B).$$

At this stage remember that the sampling distribution of $\hat{\theta}_A$ and $\hat{\theta}_B$ is Gaussian. Now let $\hat{\theta}_A = X_A/N$ and $\hat{\theta}_B = X_B/N$ be the estimators of the sample proportions of people who died within one year on each treatment (recall, these are the maximum likelihood estimates for (θ_A, θ_B)), therefore:

$$\hat{\theta}_A \sim \text{N}\left(\theta_A, \frac{\theta_A(1-\theta_A)}{N}\right) \qquad \text{and} \qquad \hat{\theta}_B \sim \text{N}\left(\theta_B, \frac{\theta_B(1-\theta_B)}{N}\right).$$

Therefore we have that

$$\hat{\theta}_A - \hat{\theta}_B \;\sim\; \text{N}\left(\delta = \theta_A - \theta_B, \frac{\theta_A(1-\theta_A) + \theta_B(1-\theta_B)}{N}\right).$$

Here we expect δ to be positive, as the probability of dying on the placebo should be higher than the probability of dying on Anturan. Then an appropriate test statistic is:

$$T = \frac{(\hat{\theta}_A - \hat{\theta}_B) - \delta}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}} \sim \text{N}(0,1).$$

Then to compute the test statistic we assume H_0 is true which corresponds to $\delta = \theta_A - \theta_B = 0$, and the test statistic is

$$T = \frac{(\hat{\theta}_A - \hat{\theta}_B)}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}} \sim \text{N}(0,1).$$

Now, the probability of detecting a true difference of δ between the two treatments would be

$$\mathbf{P}(\text{Reject } H_0 \text{ at significance level } \alpha \mid \text{true difference is } \delta) = 1 - \beta,$$

which corresponds to

$$\mathbf{P}\left(\left|\frac{(\hat{\theta}_A - \hat{\theta}_B)}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}}\right| > \phi^{-1}(1-\alpha/2)\right) = 1 - \beta.$$

The above equation can be re-arranged to give

$$\mathbf{P}\left(|\hat{\theta}_A - \hat{\theta}_B| > \phi^{-1}(1-\alpha/2)\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}\right) = 1-\beta.$$

The absolute value means that in general:

$$\mathbf{P}(|\hat{\theta}_A - \hat{\theta}_B| > C) = \mathbf{P}(\hat{\theta}_A - \hat{\theta}_B > C) + \mathbf{P}(\hat{\theta}_B - \hat{\theta}_A > C).$$

However, as Anturan is expected to be better than the placebo then the second term is negligible, so we simplify the equation to:

$$\mathbf{P}\left(\hat{\theta}_A - \hat{\theta}_B > \phi^{-1}(1-\alpha/2)\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}\right) = 1-\beta.$$

Now, standardising both sides by subtracting the mean and dividing by the standard deviation gives:

$$\mathbf{P}\left(Z > \frac{\phi^{-1}(1-\alpha/2)\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}} - \delta}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}}\right) = 1-\beta.$$

Now we know that $\mathbf{P}(Z \leq -r) = \mathbf{P}(Z > r)$, so we change the above equation to

$$\mathbf{P}\left(Z \leq \frac{\delta - \phi^{-1}(1-\alpha/2)\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}}\right) = 1-\beta.$$

This is now a standard normal calculation as $\phi(r) = \mathbf{P}(Z \leq r)$ and can be re-arranged to give:

$$\frac{\delta - \phi^{-1}(1-\alpha/2)\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}} = \phi^{-1}(1-\beta).$$

Thus simplifying the left hand side gives:

$$\frac{\delta}{\sqrt{\frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{N}}} - \phi^{-1}(1-\alpha/2) = \phi^{-1}(1-\beta).$$

Now adding $\phi^{-1}(1-\alpha/2)$ to both sides, squaring and a slight re-arrangement gives the result

$$N = \frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{(\theta_A-\theta_B)^2}\left[\phi^{-1}(1-\alpha/2) + \phi^{-1}(1-\beta)\right]^2.$$

as required as $\delta = \theta_A - \theta_B$.

□

2.4.3 Example cont. - The Anturan Re-infarction Trial

Video2.7 - Sample sizes for clinical trials III

Back to the Anturan example. Suppose that

- $\theta_A = 0.1$ - 10% die within one year on the placebo.
- $\theta_B = 0.05$ - 5% die within one year on the Anturan. This is the effect size we hope to be able to detect. Thus $\delta = \theta_A - \theta_B = 0.05$.
- Suppose we have a type I error of $\alpha = 0.05$ (as usual), and want a power of 90% so that $\beta = 0.1$. Then applying the formula gives:

$$N = \frac{0.1 \times 0.9 + 0.05 \times 0.95}{(0.1 - 0.05)^2} [1.96 + 1.28]^2 \approx 580.$$

Notes

- It is important to realise that this method is fairly crude: the value of N obtained depends critically on the choice of $(\alpha, \beta, \theta_A, \theta_B)$ all of which are fairly arbitrary. For example:
  - If $(\alpha = 0.05, \beta = 0.1, \theta_A = 0.1, \theta_B = 0.08)$ then $N \approx 4300$.
  - If $(\alpha = 0.05, \beta = 0.2, \theta_A = 0.1, \theta_B = 0.05)$ then $N \approx 435$.
  - If $(\alpha = 0.01, \beta = 0.05, \theta_A = 0.1, \theta_B = 0.05)$ then $N \approx 980$.
- It is often a good idea in any application of this method to show how the required sample size varies for a range of plausible values of (θ_A, θ_B). Conventionally $\alpha = 0.05$ (2 sided test) is used in combination with a β in the range 0.05 to 0.20 (i.e. between 80-95% power).
- The approach outlined above is relatively simple from a statistical point of view. There are a range of more complex methods to describe the relationship between $(N, \alpha, \beta, \theta_A, \theta_B)$ more exactly, but these seem scarcely worthwhile in view of the arbitrary nature of the whole process.
- Several extensive sets of tables have been published, giving N in terms of $(\alpha, \beta, \theta_A, \theta_B)$. The best of these, is *Machin D, Campbell MJ, Tan SB, Tan SH Sample Size Tables for Clinical Studies (3rd edition) Wiley-Blackwell 2009*, which has accompanying software.

Definition Now suppose you have a continuous response (rather than a binary die/survive), and the two treatments have mean responses (μ_A, μ_B). Finally, let σ be the standard deviation of the response within each treatment group. Then using a similar argument as before we get that

$$N = \frac{2\sigma^2}{(\mu_A - \mu_B)^2} \left[\phi^{-1}(1 - \alpha/2) + \phi^{-1}(1 - \beta)\right]^2.$$

Example A trial was planned to evaluate vitamin D supplements (versus placebo) in pregnant women, for the prevention of neonatal hypocalcaemia. The principal outcome measure was the infants serum calcium level at an age of 1 week. From routine evaluation of untreated women, it was known that $\mu_A = 9.0$ mg per 100ml and that $\sigma = 1.8$ mg per 100ml. It was thought that an increase in mean serum calcium (CA) of 0.5 mg per 100ml or more would be of clinical importance. If $\alpha = 0.05$ and $\beta = 0.10$, this gives

$$N = \frac{2 \times 1.8^2}{(9.0 - 9.5)^2} [1.96 + 1.28]^2 \approx 273.$$

2.4.4 Allowing for patients dropping out

This is extremely important and easy to overlook at the planning stages of a trial. All of the methods described above give an estimate of the numbers of evaluable patients required at the end of the trial. However, not all patients who enter a trial will complete it. Some may withdraw because of side effects: some may simply not return for their next scheduled visit; some may move away from the area. In long term studies, a substantial proportion of subjects may not complete the trial. This proportion should be estimated in advance, and the proposed sample size scaled up appropriately to allow for drop outs during the study.

2.4.5 Group sequential designs

In the previous sections, the determination of the trial size was based on a fixed number of patients to allow some minimum clinically significant difference to be detected with a specified significance level and power. However, in practice, patients do not all start the trial at the same time, but are entered sequentially. Thus the opportunity usually exists to examine the accumulating data before the trial is finished. Most large scale trials will have a data monitoring committee whose remit is to oversee the progress of the trial. It is clearly of value to monitor the progress of the trial to ensure that all aspects of the protocol (including entry criteria, randomisation and blinding) are being adhered to. One important advantage of monitoring progress is that if the difference between the treatments turns out to be larger than originally thought, there may be enough evidence to allow the trial to be stopped early if a statistically significant difference between the treatments is found.

This has important ethical advantages, but care is needed in deciding how many of these interim analyses are required, and at what time intervals they should be carried out. There is a temptation to carry out a large number of interim analyses, but clearly if all the tests are carried out at the 5% significance level, and if H_0 is really true, the probability that at least one of these tests will give a significant difference may be much greater than 5%. Table 7 illustrates the problem:

Table 7: Repeated significance tests on accumulating data for two treatments, a normal response with known variance and equally spaced analyses. Broadly similar results are true for other types of data; source: Pocock, 1983.

No. of repeated tests at 5 % lvl	**Overall signifiance**
1	0.05
2	0.08
3	0.11
4	0.13
5	0.14
10	0.19
20	0.25
100	0.37
1000	0.53
∞	1.0

This can be overcome by choosing a smaller 'nominal significance level' for each test in such a way that the overall significance level for the entire set of tests is 5% (this is analogous to the multiple comparisons problem in one way ANOVA). Table 8 illustrates how this can be achieved:

Table 8: Nominal significance level required for repeated two-sided significance testing with overall significance level $\alpha = 0.05$ and $\alpha = 0.01$ and various values of N, the maximum number of tests. These nominal levels are exactly true for a normally distributed response with known variance, but are also a good approximation for many other types of data; source: Pocock, 1983.

N	$\boldsymbol{\alpha = 0.05}$	$\boldsymbol{\alpha = 0.01}$
2	0.029	0.0056
3	0.022	0.0041
4	0.018	0.0033
5	0.016	0.0028
10	0.0106	0.0018
15	0.0086	0.0015
20	0.0075	0.0013

Example To illustrate this, consider the following trial to compare two drug combinations, CP (cytoxan + prednisone) and CVP (+ vincristine) in the treatment of non-Hodgkin's lymphoma (both of these treatment regimes have now been superseded). The outcome variable was whether or not the tumour responded to treatment. It was planned to enter about 120-130 patients and to analyse the data after every 25 patients had been evaluated. Figure 7 shows how the response rates varied during the course of the study. The arrows indicate the points at which interim analyses were carried out. Table 9 summarises the results of the interim analyses.

There is clearly no reason to think of stopping the trial after Analysis 1, 2 or 3. At Analysis 4, the results are beginning to look interesting. At the final analysis when the trial was finished anyway, the χ^2 test gave $p = 0.04$ which, strictly speaking, is not statistically significant at the level of 0.016 required for $n = 5$ analyses. From these data alone, it could be inferred that the superiority of CVP in this trial is interesting and worthy of further study, but the results are **inconclusive**.

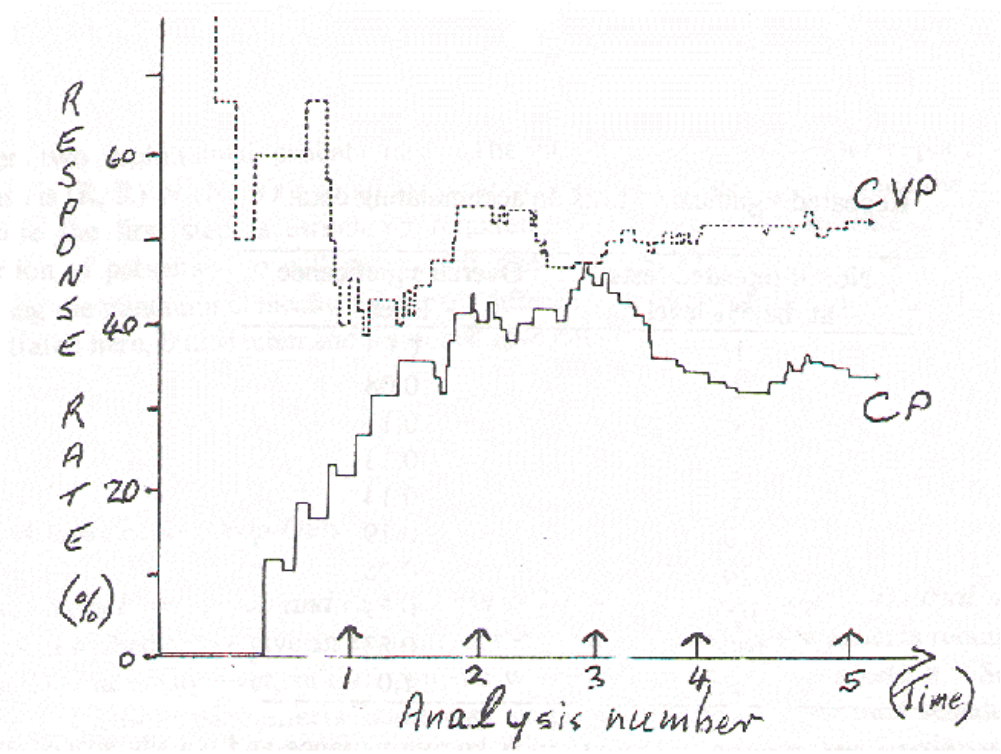

Figure 7: Response rate plottet against time with the points of interim analysis highlighted; source: Pocock, 1983.

Table 9: Results of interim analyses for a trial in non-Hodgkin's lymphoma. Response rates, chi square test statistics and p-values are shown; source: Pocock, 1983.

	Response rates			
	CV	**CVP**	χ^2	**p-value**
Analysis 1	3/14	5/11	1.64	0.20
Analysis 2	11/27	13/24	0.92	0.34
Analysis 3	18/40	17/36	0.04	0.85
Analysis 4	18/54	24/48	2.91	0.09
Analysis 5	23/67	31/59	4.24	0.04

As you can see, the interpretation of the results in this study at Analysis 5 is controversial. The problem is caused by the large difference between the unadjusted significance level of 0.05 and the threshold after adjusting for multiple tests (0.016, for 5 analyses). There is no reason why the threshold needs to remain constant over all interim analyses. Various alternatives have been proposed and some are plotted in Figure 8. These require a very low significance level for the early interim analyses and in compensation allow a significance level close to 0.05 at the final analysis. All give an overall significance level for the set of 5 interim analyses very close to 0.05.

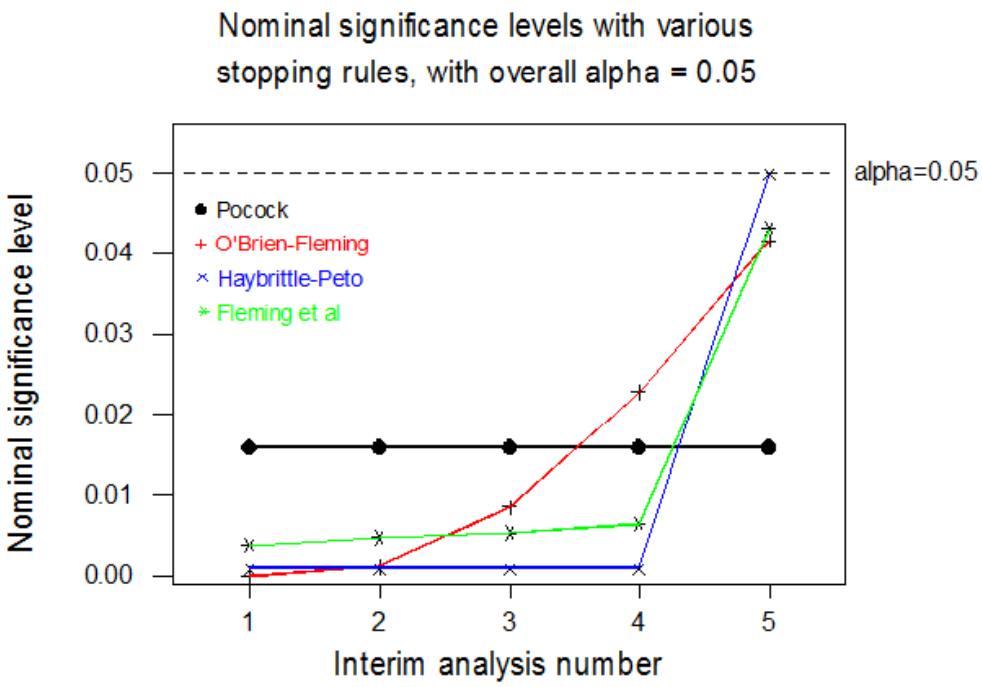

Figure 8: Comparison of 4 methods to adjust the p-value in sequential analysis of clinical trials data.

2.5 Analysing data from clinical trials

Video2.8 - Analysing data from clinical trials

Most of the methods of analysis in common use in clinical trials are covered in later parts of this course or in other courses such as linear models, regression modelling M, etc. There are, however, a few problems specific to clinical trials data that need to be considered and may not have been emphasised elsewhere.

2.5.1 Analysis of withdrawals

Any report on a trial should aim to provide an honest account of all deviations from the protocol. If the existence of patients who withdrew from therapy during the trial (or deviated from the protocol in some other way) is not mentioned this can lead to exaggerated claims about the benefits of treatment. However, should such patients be included in the main treatment comparisons or simply noted and excluded from further analysis? There are two general approaches.

Definition **Intention to treat analysis** is the approach where all eligible patients, regardless of compliance with the study protocol are included in the analysis whenever possible. This is normally preferred, since it provides an assessment of the treatment effect as it relates to actual clinical practice.

The **explanatory approach** to analysis limits the analysis to patients who received therapy according to protocol (i.e. analysis of compliers only). This allows the assessment of the treatment effect if the therapy is administered exactly according to the protocol.

Note that these two approaches to the analysis can sometimes produce different conclusions. Hardly surprising since they are addressing different questions about the treatment.

Example Figure 9 the results from the following trial of surgical versus medical therapy in bilateral carotid stenosis (this is hardening of the arteries which supply blood to one part of the brain). Each patient was assessed for a recurrent transient ischaemic attack (TIA), stroke or death in a defined period of follow-up after discharge from hospital.

Here, the explanatory approach restricts attention to compliers - i.e. patients who left hospital alive and free from stroke, as in part (a) of the table below. This analysis appears to produce a significant reduction of risk on surgical treatment, but it excludes 16 patients who died or had a stroke before leaving hospital - all but one of whom were randomised to surgical treatment.

Including these patients in an analysis by intention to treat (pragmatic approach) produces the results in part (b) of the table. The reduction in risk on surgical treatment is no longer statistically significant.

Comparison of surgical and medical therapy for bilateral carotid stenosis

Treatment	Recurrent TIA, stroke or death	
(a) Excluding deaths or strokes while in hospital		
Surgical	43/79 = 54%	$\chi^2 = 5.98$, P = 0.02
Medical	53/72 = 74%	
(b) Including all patients		
Surgical	58/94 = 62%	$\chi^2 = 2.80$, P = 0.09
Medical	54/73 = 74%	

Figure 9: Comparison of intention to treat and explantory approach in a study investigating treamtents of bilateral carotid stenosis. The results show that there can be a difference between the two approaches of analysis, with one being significant and one not; source: (Pocock, 1983).

The pragmatic (*intention to treat*) approach seems more reasonable here, as ignoring the 16 patients who died before leaving hospital were not saved by the treatment they had.

Notes

- The exclusion of withdrawals from statistical analysis does not often make such a dramatic difference to the results. Rather it creates a feeling of uncertainty and suspicion in the mind of anyone reading a report of the trial. Matthews, chapter 10, gives further examples of this.
- Any published paper reporting the results of a clinical trial must account for all patients who were eligible for inclusion in the trial and must state explicitly what approach has been taken in the analysis of the data. Since 1997, the British Medical Journal, Lancet and many other leading journals have adopted the CONSORT agreement which specifies the information which must appear in a report of a clinical trial before it will be considered for publication.

2.5.2 Multiplicity of data

In any clinical trial, even a relatively small scale one, it is possible to generate an almost overwhelming amount of data leading to a multiplicity of hypotheses for potential investigation. The main contributions to this are as follows:

- **Multiple treatments**. Some trials have more than two treatments. The number of possible treatment comparisons increases rapidly with the number of treatments.
- **Multiple end-points**. There may be many different ways of evaluating how each patient responds to treatment. It is possible to make a separate treatment comparison for each end-point.
- **Repeated measurements**. In some trials each patients progress can be monitored by recording his or her disease state at several fixed time points after start of treatment. A separate analysis for each time point could then be produced.
- **Subgroup analyses**. It is possible to record prognostic information about each patient prior to treatment. Patients may then be classified into prognostic subgroups and each subgroup analysed separately.
- **Interim analyses**. In most trials there is a gradual accumulation of data as more and more patients are evaluated. Repeated interim analyses of the accumulating data might be carried out while the trial is in progress.

Example Suppose a trial for hypertensive patients compared two different hypotensive agents each at two dose levels. Each patient had systolic and diastolic BP measured before, during and after a standard exercise test. These measurements were taken weekly over a four-month period. Patients could be classified into subgroups by age, sex and initial blood pressure readings. Interim analyses, undertaken after every 20 patients, were evaluated. This study design could generate literally thousands of hypotheses to be examined. For instance, you could compare:

a. low dose levels of treatments A and B for post-exercise systolic BP after one month for the first 30 male patients; or

b. low dose vs high dose of treatment A for pre-exercise diastolic BP after two months for the first 20 patients under age 60, etc.

In these circumstances, you need to be careful to avoid data dredging, which can lead to the identification of a large number of false positives. It is easy to find some significant treatment differences somewhere, if the data are manipulated sufficiently! How can this problem be overcome?

- In the study protocol specify in advance a limited number of analyses that you will do.
- Use standard statistical techniques to answer the questions posed.
- Treat any additional analyses you decided to do after the study has finished with caution. Use a multiple comparison adjustment procedure to be conservative in your conclusions. That is, don't overstate the significance of results.

2.5.3 Subgroup analysis

One of the easiest traps to fall into in this area is the misinterpretation of data from subgroups, where the treatment appears to have a different effect in different subgroups.

Example The results of a study of the effect of vitamin D supplements on babies serum calcium levels are shown in Table 10.

Table 10: Results of vitamin D study.

	Artificially fed		Breast fed	
	Vitamin D	**Control**	**Vitamin D**	**Control**
No. of infants	169.00	285.00	64.00	102.00
Mean calcium	9.20	8.78	9.79	9.64
SD	1.10	1.28	1.17	1.26

It seems as though the vitamin D has had a greater effect in artificially fed infants than in breast fed infants. If separate 2 sample t-tests are carried out in each subgroup, the results are:

- **Artificial**: $\overline{x}_D - \overline{x}_c = 0.42$, t = 3.70, $\mathbf{p < 0.0001}$.
- **Breast fed**: $\overline{x}_D - \overline{x}_c = 0.15$, t = 0.78, $\mathbf{p = 0.44}$.

How is this interpreted? How should these data be analysed?

- Use of separate t-tests is misleading. Cannot assume vitamin D is effective for artificially fed, but not breast fed on this basis.
- Need to compare the **difference** in sizes of effects in the two subgroups **directly**.
- Need to test for **interaction** between type of feeding and treatment in a **two way-anova**.
- There is no convincing evidence that Vitamin D is better in the artificially fed group than in the breast fed group (p = 0.22).

2.5.4 Dichotomising continuous variables

In the vitamin D trial described above, the response variable used was the infants serum calcium level (mg/100ml) one week after birth. Some of the clinicians wanted to use an alternative response variable - whether or not the infant suffered from hypocalcaemia one week after birth, which was defined as serum calcium <7.5mg/100ml. The following results were obtained from this alternative analysis and show the percentage of infants suffering from hypocalcaemia (ignoring breast / bottle feeding status for simplicity):

$$\text{Vitamin D} : 12/233 = 5.2\%,$$

$$\text{Control} : 41/387 = 10.6\%.$$

Analysis A simple confidence interval for the difference between those proportions is:

$$p_c - p_D = 5.4\%,$$

$$95\%\ \text{CI} : (1.3\%, 9.6\%),$$

How does that compare with an analysis that uses serum calcium level as a continuous response variable?

Such analysis concludes that women who receive vitamin D supplements have on average a serum calcium level that is 0.35mg/100ml higher than women who do not receive that supplement. This result is significant with a confidence interval of (0.154, 0.555). Note, these results come from a normal linear regression model, which details are omitted. Do you think dichotomising is a good idea? What are the main issues here?

- Both types of responses and tests have shown evidence that vitamin D treatment improves the situation.
- The cut-off of 7.5mg/100ml is arbitrary – there is no real scientific justification for it.
- Dichotomizing continuous variables means throwing away information and **losing power**.

2.5.5 Methods of analysis for some common designs

The methods of analysis people typically use are:

1. **Continuous response**
   - If there are no covariates and two treatment groups then a confidence interval or hypothesis test (e.g. t-test) for the difference in mean response from the two treatments is appropriate.
   - If there are more than two groups then one can fit a normal linear model with treatment group as a factor variable. Then one can compute confidence intervals for the difference in the response between each pair of treatments, allowing of course for the multiple testing problem by adjusting the significance levels appropriately (e.g. Bonferroni, Tukey's honest significant difference).
   - If there are covariate factors then these can simply be included in the linear model.
2. **Binary response**
   - If the response is binary (e.g. death / survived) and there are two treatment groups and no covariates, then a simple confidence interval or hypothesis test on the difference between the two proportions can be conducted.
   - If there are covariates, then logistic regression, an extension of the normal linear model, can be conducted. For example, consider a binary response Y_i, a binary treatment factor x_i, and a covariate z_i. Then a simple logistic regression model is given by:

$$\begin{aligned} Y_i &\sim \text{Binomial}(N_i = 1, \theta_i) \\ \log\left(\frac{\theta_i}{1-\theta_i}\right) &= \beta_0 + \beta_1 x_i + \beta_2 z_i. \end{aligned}$$

   You will learn more about logistic regression in the generalised linear models course.

2.6 Meta-analysis

Video2.9 - Meta-analysis I

2.6.1 Background

It is extremely rare for the effectiveness of any new treatment to be assessed on the basis of a single clinical trial, even one with adequate power. In fact for drug trials in the USA the FDA (Food and Drug Administration) require evidence from at least two 'well conducted' trials before they will consider the licensing of a new drug.

In general it is good scientific practice to carry out a number of trials with different investigators in different geographical regions, perhaps with slightly different protocols. If the treatment performs well 'across the board', the case for its use in routine practice is considerably strengthened. This means that the medical journals will accumulate a number of reports of clinical trials assessing the effect of any new treatment. How should this body of evidence be evaluated?

Definition **Meta-analysis** is defined as the process of applying statistical methods to the problem of combining results from different clinical trials of the same treatment. There are four main statistical objectives in any meta-analysis:

- consistent, objective display of data from different trials;
- testing of an overall null hypothesis;
- estimation of an average treatment effect;
- investigation of any statistical heterogeneity between trials.

Example Figure 10 shows the results of a meta-analysis of 33 clinical trials of streptokinase, given to patients who were hospitalised following a heart attack. The results of both conventional and cumulative meta-analysis of 33 clinical trials of intravenous streptokinase compared to placebo in the treatment of patients who had been hospitalised with acute myocardial infarction (this type of graph is sometimes known as a forest plot).

The left hand side of the figure shows that the effect of streptokinase on mortality was favourable in 25 out of 33 trials, but was only significant in 6 of these. A meta-analysis carried out after 1988 to estimate the overall effect from the combined results of all 33 trials, yielded a combined odds ratio estimate of about 0.8 (i.e. 20% reduction in the odds of mortality) which is highly significant ($z = -8.16$, $p < 0.0001$). Thus, the evidence in favour of streptokinase is very convincing. Note, you will learn more about the odds ratio in Section 2.6.2 and **??**

The right hand side of the figure shows what would have happened had the evidence been reviewed systematically as the results of each new trial became available. The resources for carrying out continuous updating of systematic reviews were not available much before 1990, so the analysis on the right hand side was not actually carried out between 1959 and 1988 - this just shows what would have happened if it had been. The effect of streptokinase was statistically significant in meta-analysis by 1971, and overwhelmingly significant by 1977. However, it was not approved officially by the U.S. Food and Drug Administration and used generally until after 1988.

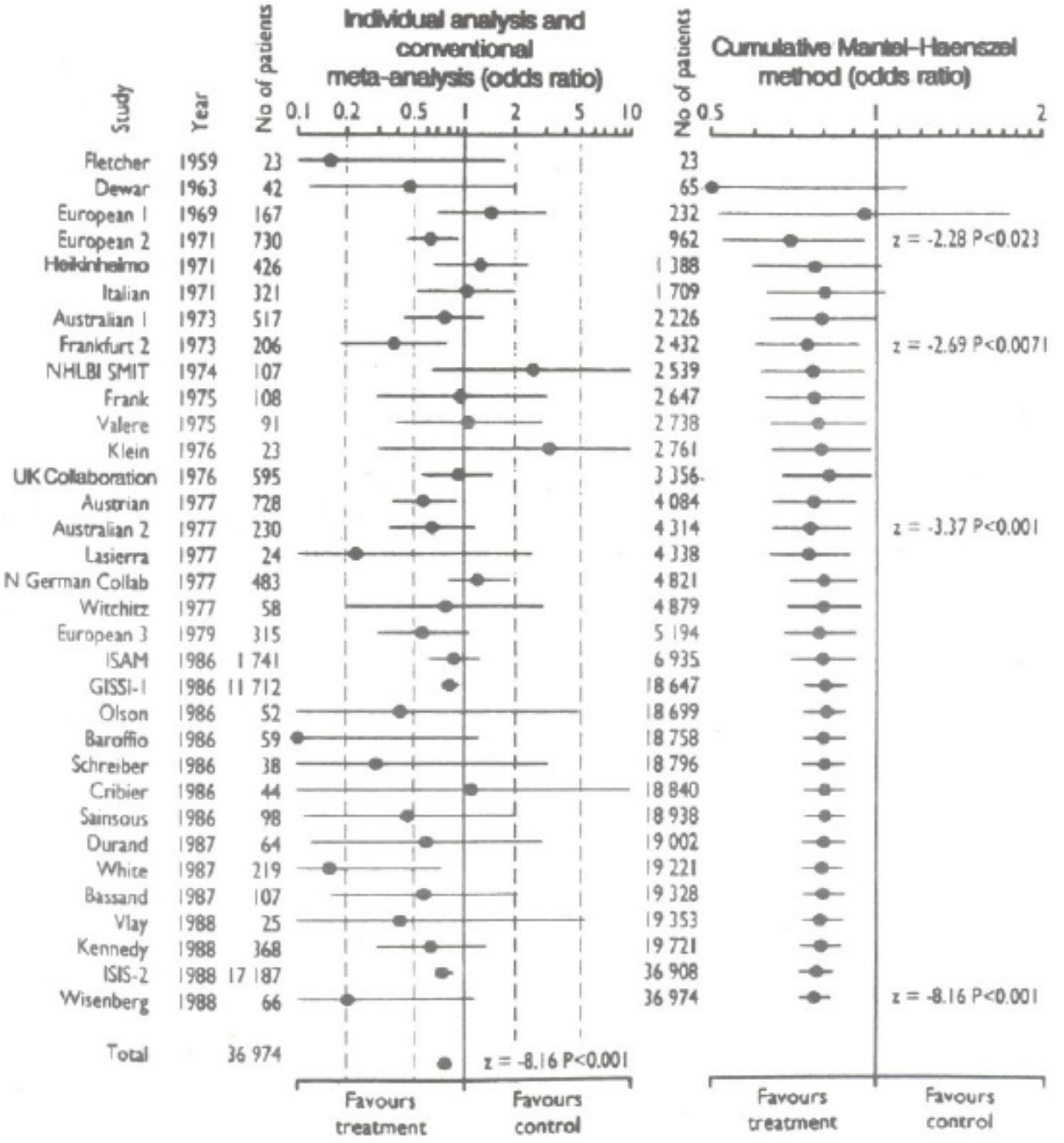

Figure Conventional and cumulative meta-analysis of 33 trials of intravenous streptokinase for acute myocardial infarction. Odds ratios and 95% confidence intervals for effect of treatment on mortality are shown on a logarithmic scale

Figure 10: Forest plot showing that streptokinase could have been approved a decade sooner, had a meta-analysis been carried out; source: Systematic Reviews in Health Care, BMJ 2001.

2.6.2 Statistical methods for meta-analysis

Ideally, the person responsible for the meta-analysis should obtain the raw data from each of the trials to be combined and work with the data at an individual patient level. In practice this is extremely difficult and the analysis is usually based on published reports of the trials' results. Thus, for each trial, we need to have a summary measure of the treatment effect size along with a measure of the variance of this estimate. Various summary measures can be used. For example,

continuous response variables

- The (standardised) difference in mean response.
- The percentage difference in mean response.

binary response variables

- The difference in failure rates.
- The ratio of the failure rates (relative risk). More information about that in Section **??**.
- The odds ratio of failure.

Odds Ratio The most commonly occurring situation in practice is the use of the odds ratio of failure for a binary response. In simple situations the data for each trial can be summarised in a 2 x 2 table as shown in Table 11.

Table 11: 2x2 table of a simple clinical trial with binary response.

	Failure	**Success**
Treatment	a	b
Control	c	d

where: a, b, c and d are rates. Then the **odds ratio** is defined by

$$\text{Odds ratio } = \frac{\text{odds of failure on treatment}}{\text{odds of failure on control}} = \frac{a/b}{c/d} = \frac{ad}{bc}.$$

Since the sampling distribution of the odds ratio is very skewed, it is usual to combine the estimates of the log odds ratio for the different trials. For each trial, the variance of the log odds ratio is

$$\mathbb{V}(\log(\text{OR})) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}.$$

2.6.3 Random and fixed effects approach

Video2.10 - Meta-analysis II

Suppose there are C trials whose results we would like to combine. Then there are two main analysis types:

- **Random effects approach** - the C trials are a sample from a larger population of studies. Attempts to model variability in effect size between trials.
- **Fixed effects approach** - assumes that all C trials are estimating the same underlying effect size.

This course concentrates on fixed effects analysis. First, define the following:

- $Y_c =$ effect size (e.g. log odds ratio, difference in mean response).
- $W_c =$ reciprocal of effect size variance, $W_c = \frac{1}{\mathbb{V}(Y_c)}$ (represents weights to weight the contribution of each study to the overall estimate).

Then the pooled estimate of the effect size is:

$$\overline{Y} = \frac{\sum_{c=1}^{C} W_c Y_c}{\sum_{c=1}^{C} W_c},$$

and its variance (assuming trials are independent) is given by:

$$\begin{aligned}
\mathbb{V}(\overline{Y}) &= \mathbb{V}\left[\frac{\sum_{c=1}^{C} W_c Y_c}{\sum_{c=1}^{C} W_c}\right] \\
&= \frac{1}{(\sum_{c=1}^{C} W_c)^2}\mathbb{V}\left[\sum_{c=1}^{C} W_c Y_c\right] \\
&= \frac{1}{(\sum_{c=1}^{C} W_c)^2}\sum_{c=1}^{C} W_c^2 \mathbb{V}(Y_c) \\
&= \frac{1}{(\sum_{c=1}^{C} W_c)^2}\sum_{c=1}^{C} \frac{W_c^2}{W_c} \\
&= \frac{1}{(\sum_{c=1}^{C} W_c)^2}\sum_{c=1}^{C} W_c \\
&= \frac{1}{\sum_{c=1}^{C} W_c}.
\end{aligned}$$

To test the hypothesis H_0 stating that the population effect size is 0:

$$z = \frac{\overline{Y}}{\sqrt{\mathbb{V}(\overline{Y})}} \sim N(0,1) \text{ under } H_0,$$

yielding a 95% CI for the population effect of:

$$\overline{Y} \pm 1.96 \sqrt{\mathbb{V}(\overline{Y})}.$$

2.6.4 Example: meta-analysis

Table 12 gives the results from 8 randomised trials comparing the effectiveness of Ranitidine (new treatment) and Cimetidine (control) in the prevention of recurrence of ulcers in patients whose ulcers have recently healed. Both treatments have since been superseded. We aim to combine the results of the 8 studies in a meta-analysis to obtain a single effect size. Figure 11 indicates that Ranitidine performs better than Cimetidine since most points lie below the line of equality.

Table 12: Data from 8 studies comparing the effectiveness of Ranitidine and Cimetidine. The number and percentage of recurrences as well as the total number of participants in each study is shown for both medications.

	Ranitidine; Recurrences			Cimetidine; Recurrences		
Author	**No.**	**%**	**Total**	**No.**	**%**	**Total**
1. Battaglia	4	22%	18	11	27%	41
2. Boyd	25	24%	106	16	29%	56
3. Van Dommelen	4	17%	24	6	27%	22
4. Walt	7	25%	28	8	24%	33
5. Gough	44	18%	243	70	29%	241
6. Silvis	7	12%	60	17	26%	66
7. Bolin	5	14%	37	3	20%	15
8. Bresci	10	18%	55	9	15%	60

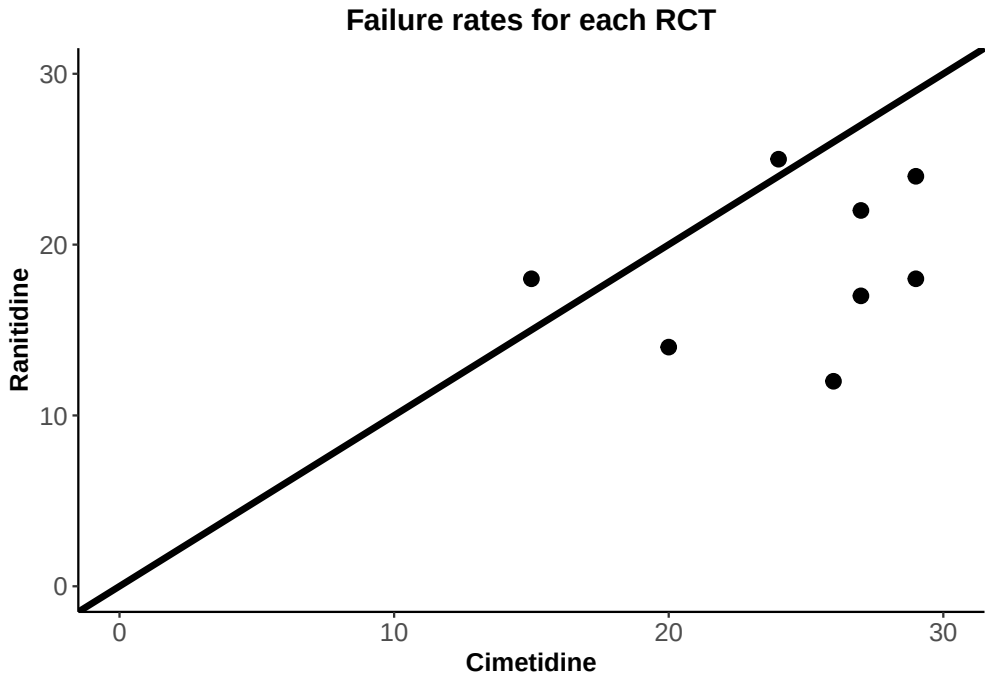

Figure 11: The percentage of recurrences using Ranitidine is plotted against the percentage of recurrences using Cimetidine. The line of equality is plotted as well. Note that most points appear to be below that line.

Let's have a closer look at the calculations needed for study 1. Table 13 displays the data of study 1 from Table 12 in the familiar form of a 2x2 table.

Table 13: 2x2 table of study 1.

	Failure	Success
Ranitidine	4	14
Cimetidine	11	30

Remember from above that we first have to compute the odds ratio (of failures) (OR_1) by dividing the odds of failing of the first drug by the odds of failing of the second drug. Then we take the logarithm to get the log odds ratio, which is normally distributed. Thanks to normality, it is possible to compute the variance of the log odds ratio $\mathbb{V}(\log((OR_1)))$ by summing the reciprocal of all the rates used in the odds ratio calculation. The weight of the first study in the computation of the overall effect size $\overline{Y}$ is given by the reciprocal of the variance of the log odds ratio:

$$\mathsf{OR}_1 = \frac{4/14}{11/30} = \frac{4 \times 30}{11 \times 14} = 0.78.$$

$$\log(\mathsf{OR}_1) = -0.25$$

$$\mathsf{V}_1 = \mathbb{V}(\log(\mathsf{OR}_1)) = \frac{1}{4} + \frac{1}{14} + \frac{1}{11} + \frac{1}{30} = 0.45$$

$$\mathsf{W}_1 = \frac{1}{V_1} = 2.24$$

Table 14 shows the results of those computations for all eight studie.

Table 14: Results of computations needed to obtain a single effect size in a meta-analysis.

Study	**OR**	**p-value**	$\boldsymbol{Y_c = \log(OR)}$	$\boldsymbol{V_c = \mathbb{V}(\log(OR))}$	$\boldsymbol{W_c = 1/V_c}$	$\boldsymbol{W_c \times Y_c}$	$\boldsymbol{W_c(Y_c - \overline{Y})^2}$
1	0.78	0.48	-0.25	0.45	2.24	- 0.56	0.09
2	0.77	0.42	-0.26	0.14	7.15	- 1.85	0.27
3	0.53	0.30	-0.63	0.53	1.89	- 1.19	0.06
4	1.04	0.59	0.04	0.36	2.81	0.11	0.68
5	0.54	0.003	-0.62	0.05	20.88	-12.86	0.56
6	0.38	0.04	-0.97	0.24	4.15	- 4.01	1.09
7	0.63	0.31	-0.47	0.65	1.54	- 0.73	0.00
8	1.24	0.42	0.23	0.25	3.95	0.91	1.84
Total					$\boldsymbol{\Sigma = 44.63}$	$\boldsymbol{\Sigma = -20.17}$	$\boldsymbol{\Sigma = 4.60}$

The total effect size and its standard deviation are given by

$$\overline{Y} = \frac{\sum W_c Y_c}{\sum W_c} = \frac{-20.17}{44.63} = -0.452,$$

$$\mathsf{se}(\overline{Y}) = \sqrt{\mathbb{V}(\overline{Y})} = \frac{1}{\sqrt{\sum W_c}} = \frac{1}{\sqrt{44.63}} = 0.15,$$

and the 95% CI for the log odds ratio is:

$$-0.452 \pm 1.96 \times 0.15 = (-0.75, -0.16).$$

Transform back by taking $\exp()$:

95% CI: (0.47,0.85), with a point estimate of 0.64.

The odds of re-bleeding are estimated to be reduced by 36% on Ranitidine, with a 95% CI from 15% to 53%. That result agrees with the exploratry plot, Figure 11. Figure 12 provides a graphical summary of the meta-analysis.

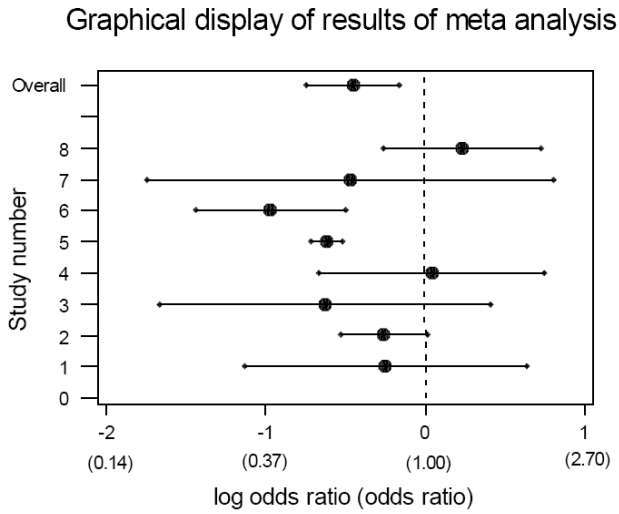

Figure 12: Forest plot showing odds ratios of individual studies as well as overall odds ratio from meta-analysis.

2.6.5 Some problems with meta-analysis

There are a number of problems that can be encountered when conducting a meta-analysis.

Different response variables used in different studies A very simple example of this would be in trials of a new treatment for high blood pressure. Some of the trials might use Systolic BP as the primary response variable while others use Diastolic BP. How might you deal with this in practice? Convert to standardised scale: standardised mean difference (SMD)

$$SMD = \frac{\text{difference in mean outcome between groups}}{\text{SD of difference}}.$$

Deciding which trials to include The existence of large computerised databases of published research means that it is relatively straightforward to identify relevant published research in journals listed on the databases. Some of the studies identified may be of poor quality and it may be sensible to include only 'good studies' - e.g. only randomised trials. It has also been suggested that studies be 'weighted by quality' in a meta-analysis, but this seems unnecessarily complex. No matter what rule is to be used, the search cannot be limited to computerised databases because of the following problem.

Publication bias It is widely recognised that clinical trials reporting large and significant treatment effects are more likely to be published in medical journals than trials with smaller, non-significant differences. There are two main reasons for this:

- investigators may be more likely to be motivated to write and submit reports of trials which give 'exciting' significant results. Studies with non-significant results are much less likely to be written up for publication;

- journal editors are more likely to publish significant results on the grounds that they are more 'interesting, and indicative of therapeutic progress'.

These are compounded by the fact that significant results are more likely to be published in major, widely read journals which are listed in computerised databases and easier to find. The bias is made even worse by the fact that many trials are of grossly inadequate size.

2.7 Example - Diamorphine for pain relief in labour

A study was conducted at the Southern General Hospital at Glasgow in 2000, and aimed to assess the effectiveness of different forms of pain relief in labour. The study aimed to assess the best way to administer diamorphine. The two methods compared are:

- injection into your leg or bottom;
- a drip which is patient controlled, allowing the women to choose the amount of pain relief to use.

The full study protocol and paper is provided on Moodle. We discuss this now in the following two videos.

Video2.11 - Protocol: pain relief study

Video2.12 - Paper: pain relief study