A MANUAL ON HOW TO DO RANDOMISED CONTROLLED TRIAL

Dr. Lim Teck Onn MRCP(UK),
M.Med.Statistics(Newcastle)

Published by Clinical Research Centre
Ministry of Health
2013

WORKSHOP ON RANDOMISED CONTROLLED TRIAL (RCT)

Contents

STEPS IN RANDOMISED CONTROLLED TRIAL

Step 0 : Objections to trial : should you do the trial?
Types of trial : what kind of trial are you doing?

Step 1 : Define purpose of trial : How to state hypotheses?

Step 2 : Design the trial : How to design a trial (protocol)?

a. Study population, patient selection and sample size
b. Treatment and covariates
c. Evaluation of patient response
d. Design techniques : experimental design, randomisation, blinding and placebo.
e. Interim analysis and stopping rules
f. Ethical issues

Step 3 : Conduct the trial : How to organise a trial?

a. Planning : proposal, funding and approval.
b. Preparatory : protocol & manual, organisation, recruitment, data issues, pilot, training.
c. Trial operations : recruitment, eligibility screen, registration, informed consent, run in period, revealing randomisation, follow-up and close out..

Step 4 : Analyse the data : How to analyse trial data?

a. Preliminary
b. Summarising treatment response and effect
c. Further issues in data analysis
d. Statistical inference

Step 5 : Draw conclusion : How to report a trial?

a. Drawing conclusion and data interpretation
b. Writing trial report

1. Definition.

Clinical trial : any form of planned experiment which involves patient and designed to elucidate the most appropriate treatment of future patients with a given medical condition. (*Pocock, Clinical Trials chp 1 , 1984*)

Randomised controlled trial (RCT): a clinical trial with these 2 design features.

1. It is comparative: it compares the experience of a group of patients on the new treatment with a control group of similar patients receiving standard or no treatment.
2. It involves randomisation : this means patients are assigned to new or standard/no treatment randomly.

Bibliography : sources of material on RCT design and conduct.

None of the material in this manual is original. They are abstracted from the texts and articles published in the journals listed below. All references are given. Selected photocopies of the references cited can be found in the companion volume to this manual "Further readings on RCT "

General :

1. Pocock SJ. Clinical trials : a practical approach. Wiley 1984

2. Friedman. Fundamentals of clinical trials PSG inc 1985

3. Mienert. Clinical trials : design, conduct and analysis OUP 1986

4. Fleiss The design and analysis of clinical experiments Wiley 1986

Specialist texts :

1. Salsburg. The use of restricted significance test in clinical trial Springer

2. Whitehead The design and analysis of sequential clinical trials Halstead 1992

3. Jones, Kenward Design and analysis of crossover trials C&h 1981

4. Sen Cross-over trials in clinical research Wiley 1993

Journals :

1. Controlled clinical trial (CCT) : available from UH library
2. Statistics in Medicine (Stat med)
3. Statistical methods in medical research
4. Biometrics : most university libraries have it.
5. Journal of Chronic disease (JCD)/ Journal of Clinical Epidemiology (JCE): most university libraries have it.
6. American Journal of Epidemiology : most university libraries have it.
7. General medical journals like BMJ, New England journal of Medicine, Lancet, JAMA etc publish occasional articles on various aspects of the design and conduct of RCT.

2. 2. Objections to RCT : should you do the trial?

There are 3 potential objections to doing a trial. You need to consider these to decide whether you should do the trial.

1.Ethical objections.
2.Practical objections.
3.Scientific objections.

1. Ethical objections : the trial may not be ethically acceptable.

a. Lack of equipoise : this means genuine uncertainty regarding the comparative therapeutic merits of each arm in a trial (*Freedman NEJM 1987*); so that one may be indifferent whether patient receive the treatment or not. Note that collective equipoise may be present (collective uncertainty as manifest by practice variation) but individually each believed in the correctness of their own practice. Uncertainty can be measured. (*Lifford BMJ 1994*)

b. Conflict of obligation, violate doctor-patient relationship: doctors sacrifice the interests of current patients to the benefits of future patients when they are supposed to act on the best interest of their patients?

c. Universal ethical standards for all cultures and countries, developed and developing? See controversy over trial on treatment for HIV in developing countries (*Lurie NEJM 1997, Varmus NEJM 1997, Angell NEJM 1997, Clarke Lancet 1998, Levine Lancet 1993)*

References :

1. *Freedman. Equipoise and he ethics if clinical reasearch NEJM 1987*
2. *Hellamn Of mice but not men: problems of the randomised clinical trial NEJM 1991*
3. *Passamani Clinical trials-are they ethical? NEJM 1991*
4. *Schafer The ethics of RCT. NEJM 1982*

2. Practical objections : the trial may not be feasible :

a. Rare outcome event and therefore huge (>100,000) sample size needed.
   - do multicentre trial (*Black BMJ 1996*) or use surrogate outcome measure.

b. Rare disease and therefore takes too long to recruit enough patients.
   - Bayesian approach may work. (*Lilford BMJ 1995*) .

c. Outcome event requires prolonged observation, say >10 years.
   - use surrogate outcome measure.

d. Intervention that requires subject's active participation, which in turn depends on subject's beliefs and preferences. Examples are health promotion, psychosocial intervention. (*Black BMJ 1996*). Therefore difficult to randomise and to blind..
   - patient preference trial.

e. Intervention that cannot be randomised individually eg. health service organisation, health promotion.
   - group randomisation ie community trial

3. Scientific objections : the trial may not be valid or generalisable.

a. Setting : centre effect (centre of excellence), hospital vs primary care setting
b. Patient selection : strict criteria to ensure homogenous group in explanatory trial vs loose criteria resulting in heterogeneous group in pragmatic trial. (*Schwartz J Chron dis 1967*) Also in mega trial vs mini trial (*Topol Br Heart J 1992*)
c. Effect of trial participation : Hawthorne effect(behaviour under close observation differ from usual), trial participants (both patients and clinicians) are atypical; entry into trial associated with better survival (*Stiller Br J Cancer 1994)*
d. Treatment : standardised treatment(including ancillary care and co-treatment) per protocol in explanatory trial vs individualised treatment with considerable clinician discretion and judgement in pragmatic trial. Level of compliance, treatment supervision amd monitoring may also differ from usual.
e. Outcome that depends on characteristics (skills, attitudes and beliefs) of providers or setting. Eg surgery, physiotherapy, psychotherapy, nursing care.
f. Ignore patients' as well as clinicians' beliefs and preferences (basis for randomisation). If these are not ignorable, will impact on outcome.
g. Extrapolating average response in a group to individual patient in practice : beware of treatment effect variation by patient's risk and also qualitative or cross-over interaction.

References :

1. *Kramer. Scientific challenges to randomised trials JAMA 1984*
2. *Schwartz Explanory and pragmatic attitudes in trials JCD 1967*
3. *Black. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996.*
4. *Lilford Clinical trials and rare disease BMJ 1995*
5. *Russell Evaluating new surgical procedures BMJ 1995*
6. *Knipschild Trial and error BMJ 1993*
7. *Yusuf Why do we need large simple trial? Stat med 1984*
8. *Topol. Answers to complex questions cannot be derived from 'simple' trials Br Heart J 1992*

However, there is no question that we need RCT (*Pocock CT chp4, 1984, Urbela CCT 1981, Sibbald BMJ 1998)*

Example of fiasco without trial : Retrolental fibroplasia and oxygen therapy (*Lamman JAMA 1984)*

Nevertheless you will still need to be able to justify the trial :

1. is it ethically acceptable?
2. is it feasible?
3. is it scientifically sound?
4. is it useful? : for applied research, findings can help solve current problem; and for basic research, extend knowledge of subject.

Literature review is therefore crucial:
ie searching and summarising/synthesising existing information on research problem so that;

a. proposed research is not redundant, information does not exist.
b. know current state of knowledge and satisfy reviewer that your level of knowledge for research is adequate.
c. obtain clues for possible methodological strategies and to forewarn potential problem.
d. prepare an Investigator's brochure : a collection of data consisting of all the relevant information known prior to the onset of a clinical trial including chemical, pharmaceutical data, toxicological, pharmacokinetic/dyanamic data in animals and results of earlier clinical trials.

And adherence to current guidelines on trial design, conduct and reporting is important to assure trial is ethically and scientifically sound :

see Appendix D :

1. Declaration of Helsinki : International guidelines for biomedical research involving human subjects

2. CPMP. Good clinical practice for trials on medicinal products in the European community. *Toxicology and Pharmacology 1990*

3. CPMP. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products *Statistics in Medicine 1995*

4. Begg C, Cho M, Eastwood S, Horton R et al. Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA 1996*

3. 3. Types of clinical trials : what kind of trial are you doing?

There are many options in trial design. Decision should consider the type of trial one is doing.

1. Phases of experimentation in drug trials :
   I. Phase I trials : first trials of a new active ingredient in man, perform on 20-80 human volunteers to determine safety, pharmacokinetic/dyanamic profile
   II. Phase II trials : therapeutic pilot studies, perform on 100-200 patients to demonstrate activity, safety , dose ranges and regimens , with very close monitoring; with view to selection for full phase III trial.
   III. Phase III trials : full scale rigorous and extensive evaluation of treatment.
   IV. Phase IV trials : post marketing studies, eg post marketing surveillance, determine therapeutic strategies. Note that new indications, new methods of administration or new combinations require full scale rigorous trial.

2. Comparative versus equivalence trial : (*Jones BMJ 1996*)
   - implications for sample size planning and analysis

3. Explanatory versus pragmatic trial : (*Schwartz JCD 1967 , Roland BMJ 1998, Topol Br Heart J 1992*)
   - strict selection, complete protocolisation of patient management, comprehensive measurements including pathophysiology versus wide selection, flexible protocol and minimal measurements.

4. Specific disease and treatment under trial pose particular challenges :
   the standard trial is the drug trial for common disease with common outcome event; other disease or treatment posed special challenges:
   - rare disease : *see Lilford BMJ 1995*
   - surgical trial : *see Russell BMJ 1995*
   - Behavioral intervention /psychotherapy, educational intervention : *see Bradley Diab care 1993.*.
   - Non therapeutiv or community trial; health service provision, health promotion, disease prevention : *see Buck & Donner 1982*
   - Diagnostic technology : not just to assess accuracy but effect on treatment and patient outcome. *Guyatt JCD 1986*

Steps in a trial : how to do a trial?

Step 1 : Define purpose of trial : How to state specific hypotheses?

Step 2 : Design the trial : How to design a trial and write the protocol?

Step 3 : Conduct the trial : How to organise a trial?

Step 4 : Analyse the data : How to analyse trial data?

Step 5 : Draw conclusion : How to interpret data and report a trial?

The following pages provide a step by step account of the design and conduct of clinical trial.

However, please refer Appendix D :

For clinicians, please read Ethical and GCP guidelines too

- Declaration of Helsinki : International guidelines for biomedical research involving human subjects
- CPMP. Good clinical practice (GCP) for trials on medicinal products in the European community. *Toxicology and Pharmacology 1990*

For biostatistician, please read Biostatistical methodology guidelines too.

- CPMP. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products *Statistics in Medicine 1995*

Step 1 : Define purpose of trial : How to state specific research question/hypotheses?

Deming
" until the purpose is stated, there is no right or wrong way of going about the research "

Research question : (*Friedman CT chp 2, 1985*)

1. Primary question : the major question the investigators are most interested in answering.
2. Secondary or subsidiary questions :
   - relate to subsidiary end points, other than primary end point.
   - relate to subgroup hypotheses, other than overall contrast between each arm of trial.
3. Ancillary questions : questions which do not bear directly on the intervention being tested but which are nevertheless of interest; for eg. natural history of disease in the control group , risk factor analysis.

Stating objective/hypotheses :

- Use action verb eg. to determine, to compare, to describe, to establish.
- Use measurable outcomes eg death, QOL measure.
- Use testable proposition; generally no problem for trial, question is clearcut.

Typical examples are :

- to compare the effects of *new treatment A* versus *standard treatment B /placebo* in patients with *disease X* on *outcome measure Z.*
- to compare whether *new treatment A* improves outcome measure Z relative to *standard treatment B /placebo* in patients with *disease X* .
- to establish the equivalence of *new treatment A* with *standard treatment B* in patients with *disease X* with respect to *outcome measure Z.*

Step 2 : Design the trial : How to design a trial and write the protocol?

Issues in trial design : consider the following

a. Study population
b. Sample size calculation
c. Treatment
d. Response or end point evaluation
e. Design techniques
f. Interim analysis
g. Ethical issues

a. Study population : (Friedman CT chp 3 1985)

this is the subset of the general population defined by the eligibility criteria (inclusion and exclusion criteria). The group of patients actually studied in the trial is selected from the study population; this constitutes the sample. Clear definition of study population and description of how patients are actually selected are important for assessing trial generalisability, trial's merit and appropriateness and to allow others to replicate trial if necessary.

	eligible		sampling/selection		consent	
General population	-----→	study population	-----→	sample	----------→	in trial

a. Eligibility criteria : state each and justify

1. Inclusion criteria :
   - definition of the target disease for which the treatment is indicated.
   - patients with target disease who could potentially benefit from the treatment as judged by pathophysiology and mechanism of drug action.

2. Exclusion criteria :
   - patients with target disease who potentially could not benefit from the treatment as judged by pathophysiology and mechanism of drug action, for eg disease too advanced or too mild.
   - patients with target disease who could potentially be harm by the treatment ; eg pregnancy, lactation, hypersensitivity and other contraindications to treatment based on known mechanism of drug action or phase I or II studies.
   - patients with multiple comorbid diseases that affect trial treatment or that complicate ancillary care or co-intervention.
   - patients in whom outcome of interest can be observed with reasonable frequency and in reasonable time. This is purely a design consideration. It reduces sample size and trial duration. This explains why severe disease usually studied first before mild one eg Hypertension trials.
   - patients likely to develop conditions (eg death) which preclude the ascertainment of outcome of interest (competing risk problem). This explains

why complicated disease excluded for patients may die from causes unrelated to disease under trial treatment.

- Patients likely to drop out or not comply with trial treatment.

Decision on above eligibility criteria will impact on patient recruitment for trial as well as generalisability of results. A fundamental consideration is the distinction between explanatory trial (mini) and pragmatic trial (mega); ie strict criteria to obtain homogenous subset vs loose criteria to maximise number and obtain heterogeneous group. (*Schawrtz JCD 1967, Yusuf Stat med 1984 & Topol Br Heart J 1992)*

Patient selection and recruitment :

- Formal sampling is unusual; more typically consecutive eligible patients from participating centres are recruited over a time period.
- consider the source of patient : centre effect(centre of excellence), hospital vs primary care setting.
- if selection on the basis of measurement above some cut-off eg BP>100 diastolic, then consider regression to the mean. May need to elevate cut-off or observe for a period or adjustment (regression dilution bias)
- important to register all potentially eligible patients (log book), state reason for non inclusion (non-eligible), and state reason for non consent if eligible and preferably baseline data too.
- strategies to maximise recruitment and consent : see Trial Conduct.

b. b. Sample size calculation : how many patients are needed?

Essential in planning trial :

1. Sample size needed for adequate power. Underpowered study is unethical. (*Altman BMJ 1980)*
2. Trial feasibility. Required sample size partly determines duration and cost of study, also need for multi-centre study.

Principles underlying sample size calculation： *(Freidman NEJM 1978)*

To understand the basis of sample size calculation, you need to understand how inference is made about treatment effect in a trial. The procedure used is called hypotheses or significance testing.

In a typical trial comparing a new treatment versus placebo, the aim is to estimate the true response μ_t in the treatment group (eg. mean response or proportion responding) and compare it with the estimate of the true response μ_c in the control group. The difference between μ_t and μ_c is the treatment effect. However, patients vary in their characteristics and responses, an apparent difference in response may be due to chance alone. We therefore need to test whether there is any true treatment effect at all (ie $\mu_t - \mu_c > 0$).

We set up the null (H0) and alternative(Ha) hypotheses as follow :

H0 : $\mu_t = \mu_c$ ie no difference.

Ha : $\mu_t > \mu_c$ ie true difference exists.

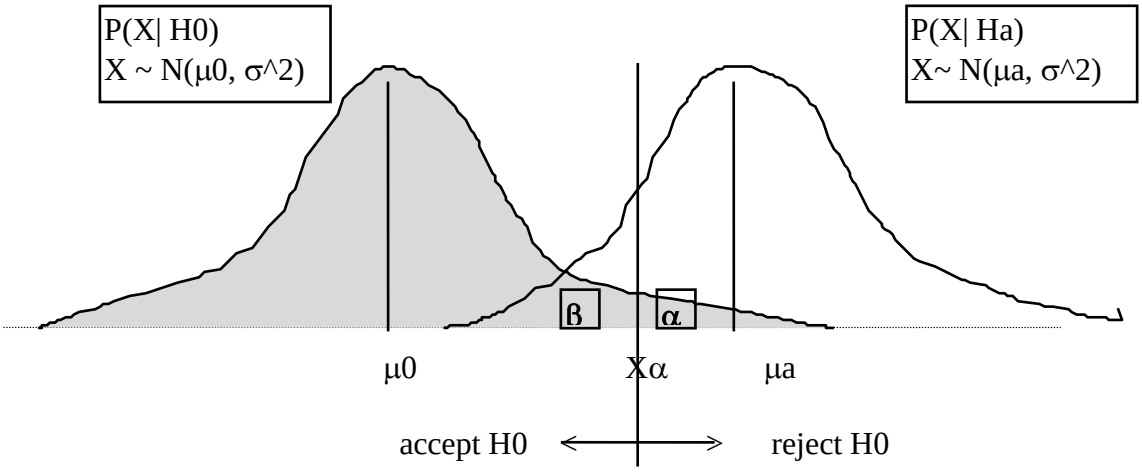
The idea is to assess the weight of evidence on the basis of the observed data. If the evidence favours the null hypotheses, we accept it and conclude that there is no difference in response. On the other hand, if the evidence favours the alternative hypotheses, we reject the null hypotheses and conclude that true difference exists.

The assessment of the weight of evidence is by estimating the probability that the observed difference $\mu_t - \mu_c$ in response could in fact have arisen purely by chance alone. The probability is called the P value. This is the probability that the difference as large or larger than that observed ($\mu_t - \mu_c$) would occur if the H0 were true. If the P value is smaller than some pre-specified level say α , called the significance level, then we reject the H0 and conclude that there is a statistically significant difference in response between the 2 treatment. The α level is simply a rule that determines how small P must be before we are prepared to conclude a difference exists between the 2 treatment. Of course, when we reject the H0, we are running a risk of making the wrong decision. There may in fact be no difference between the 2 treatment yet we concluded that there was. This is called the false-positive error or Type I error (ie error of rejecting H0 when in fact it is true). The probability of Type I error is in fact α the significance level. Thus, by setting α at a low level, we minimise the risk of Type I error.

On the other hand, the estimated P value may be larger than α. In that case, we cannot reject the H0 and we conclude there is no statistically significant difference between the 2 treatment. This may be an error too. There may in fact be a true difference between the 2 treatment and yet we fail to reject the H0. This is called false negative error or Type II error . The probability of Type II error is β.

The above ideas can be summarised graphically and in the following table :

Let the test statistic $X = \mu_t - \mu_c$



$Z\alpha = X\alpha - \mu 0 / \sigma$ under H0 $\qquad$ $Z\beta = X\beta - \mu a / \sigma$ under Ha

$X\alpha - \mu 0 = Z\alpha * \sigma$ $\qquad$ $X\alpha - \mu a = Z\beta\sigma$

$| \mu a - \mu 0 | = | X\alpha - \mu 0 | + | \mu a - X\alpha |$

therefore, $\mu a - \mu 0 = Z\alpha * \sigma + Z\beta * \sigma = (Z\alpha + Z\beta) * s / \sqrt{n}$

from which are derived sample size formulae.

Decision table :

		Truth	
		$\mu_t = \mu_c$	$\mu_t > \mu_c$
Trial conclusion	$\mu_t = \mu_c$	correct	False negative or Type II error (probability is β)
	$\mu_t > \mu_c$	False positive or Type I error (probability is α)	correct

Sample size calculation in practice :

In practice sample size calculation requires the following specifications and considerations. Given the 4 parameters below, standard formulae are available for calculating the sample size, or you could look up tables (eg. Machin and Campbell) or easier still use a computer program (eg POWER).

Refer Appendix A on how to use POWER program for sample size calculation.

Basic parameter specifications required :

1. delta Δ : the clinically worthwhile difference to detect or effect size.. This is a clinical decision and there may be uncertainty concerning what is worthwhile.
2. Alpha α : the probability of Type I error. We want this to be low, typical values are 0.05 or 0.01
3. Beta β : the probability of Type II error. The complement of which is 1-β , referred to as power, is the probability of avoiding type II error. This is the probability of detecting a difference as large or larger than that specified (Δ) if it exists ie probability of rejecting H0 when it is false. We want the power to be high, typical values are 0.8 or 0.9 (corresponding to Type II error of 0.2 or 0.1)
4. A measure of variability of the chosen response measure in the study population. For example , standard deviation for quantitative response measure.

Additional considerations :

1. Types of trial :
   - Equivalence trial vs comparative or superiority trial , *see Jones BMJ 1996*
   - pragmatic/mega trial vs explanatory/mini trial : large heterogenous sample vs smaller but homogenous sample

2. Uncertainty concerning effect size that is worthwhile to detect or that is equivalent:
   - Bayesian approaches : *see Speigelhalter Stat med 1993*
   - patient centred approach : *see Naylor JCE 1994*

3. Research questions :
   - only one overriding end-point
   - multiple end points and subset hypotheses
   - detection of adverse effects if that is of interest; most trial tend to focus on detection of benefit. Adverse effect must be rare or unexpected for trial to proceed to phase III; more demanding on sample size. Hence the need for phase IV post marketing surveillance

4. Number of treatment groups :
   - 2 groups : Rx vs placebo/std. Rx
   - 3 or more groups : how compared?
     a. Rx vs placebo only eg. different doses of same drug vs placebo : equal size for treatment and inflate placebo arm by sqrt(# treatment arms). (*Klimt CCT 1981)*
     b. pairwise comparisons eg. Rx A vs placebo, Rx B vs placebo and Rx A vs Rx B. Then allow for multiplicity problem.

5. Summary statistics for response data type : *Florey BMJ 1993, Day BMJ 1989*
   - proportion
   - mean
   - ordinal data
   - survival function: timing is important. The intended power is realised only when sufficient events had occurred, not number of patients enrolled.
   - rates

For more complex effect estimates, may simply manipulate the trial to a comparison of proportion and use simple formula with minimal assumptions; which is conservative (inflate sample size). (*Pocock, Statistician 1982*)

6. Design considerations :
   - Experimental design : parallel groups with independent observations vs cross over or self-controlled or matched design with dependent or paired observations.
   - For cross-over trial ,using the simplest analysis which is paired t test for continuous response measure, the calculated sample size is that for a single mean :

sample size $n = ((Z\alpha + Z\beta)* \sigma_d) \text{^}2 / \Delta$

where σ_d is std deviation of within pair difference
Δ is clinically worthwhile difference to detect.
$Z\alpha$ is 1.96 for α =0.05 (2 tailed)
$Z\beta$ is 0.84 for 80% power or 1.28 for 90% power.

   - Individual vs group or cluster randomisation
   - Stratified randomisation leads to slight increase power; can be ignored which is conservative.
   - Unequal randomisation or allocation :

for 2 groups design : equal sample size is most powerful. Up to 2:1 allowed to increase experience with use of new drug but may cast doubt on equipoise. Adjust accordingly as follows

a. $m' = (r+1/2r)* m$ (*Campbell BMJ 1995*)

where m = sample size based on equal allocation
m' = size of one group under unequal allocation
and r*m' = size of the other group

b. $n' = 0.25*(1/Q1 + 1/Q2)* n$ (*Lachin CCT 1981*)

where n = total sample size(both combined) under equal allocation
n'= total sample size(both combined) under unequal allocation
and $n= Q1*n + Q2*n$ and $Q1 + Q2 =1$

Both above methods are equivalent.

7. Operational considerations : in particular protocol deviations will lead to power reduction. Therefore need to allow for it esp non-compliance and drop-out as follows :

   adjusted n' = n/(1 - R)^2 (*Lachin CCT 1981*)

   where R is drop out rate eg 0.1 (10% drop out)

8. Analysis :
   - interim analysis : need to inflate size accordingly.
   - final analysis : sample size estimation is based on the method to be used (significance test) to analyse trial data.. These tests require assumptions and sample size calculation must therefore take into account potential departures from these assumptions, typically tends to inflate sample size.
     - sample size based on non-parametric methods.

Finally, note that :

1. sample size calculation is only a rough guide
2. It represents a compromise between available resources and objectives (eg desired effect size). Iterations between sample size required and detectable effect size to arrive at final range of estimates.
3. Sensitivity analysis therefore useful : determine effect on sample size of varying specifications.
4. Assumptions and uncertainty inevitable : Need to check assumption during interim and revise sample size accordingly.
   - use of Internal pilot studies but beware (*see Wittes Stat med 1990, Fayer & Machin Br J cancer 1995)*

c. Treatment and covariates :

1. Treatment :

a. Define treatment procedure including formulation, route, dose, regimen, duration.
b. Dose modification /titration not allowed(fixed dose) or strictly per protocol (standardised) or clinician discretion and judgement (individualised treatment). Distinction between explanatory (mini) trial and pragmatic (mega) trial .(*Schawrtz JCD 1967 , Yusuf Stat med 1984 & Topol Br Heart J 1992*) Individualised treatment also make blinding difficult (see later).
c. Non-drug treatment(eg surgery) need to consider skill (learning curve) and experience of operator; though this apply to lesser degree for drug treatment too. (*Russell BMJ*). Also difficult to blind, and sometimes to randomise if outcome dependent on patients' participation and therefore preference eg education, psychosocial intervention. (*Bradley Diab care 1992*)
d. Treatment compliance : how assessed?; has complex effect on trial results (*Kramer JAMA 1984 , Friedman CT chp 13, Stichele Eur J clin pharm 1991 Stat med Feb issue 1998)*
e. Co-intervention or concurrent treatment and ancillary care : strictly per protocol (complete protocolisation of management) or clinician discretional judgement (flexible protocol). (*Schawrtz JCD 1967 , Yusuf Stat med 1984 & Topol Br Heart J 1992)*

2. Covariates : other variables to include .

a. must include all known prognostic factors with respect to trial outcome. Randomisation does not guarantee baseline balance.
b. Effect modifier : if specified in subset hypotheses.
c. Clinical variables : diagnosis.
d. Other variables : identifier, universal variables, time.

Refer to Annex of GCP guideline (see Appendix D) for a list of items required in data form for use in trial (case report form).

d. Evaluation of patient response (also known as outcome assessment, end-point ascertainment).

a. Outcome measures :
   - Ultimate outcome measures : mortality (may take too long) and quality of life (may be too complex to measure); both often requires much larger sample size esp mortality.
   - Intermediate measures / surrogate measures : endpoint that investigator deems as correlated with an end point of interest but that can perhaps be measured at lower expenses and at earlier time than the endpoint of interest. Examples are morbidity, complication, hospitalisation, clinical measure (BP,

disease severity), lab measure but must correlate with ultimate outcome, acceptable to scientific and medical communities, adequate measurement properties and feasible or acceptable to patients. (*Ellenberg BMJ 1991, Editorial Lancet 1990)*

- Few easy to measure ultimate and relevant end-points eg mortality, other events versus detailed measures of pathophysiologic process that is related to biological basis of treatment response; eg renal death vs reduce intraglom pressure in ACEI trial (mega vs mini trial, *see Topol Br Heart J 1992 Guyatt JCD 1985)*

b. Measurement properties : reliability, validity, discriminatory ability and responsiveness of outcome measures. May need separate validation study.

c. Number of measurements :
   - number of end-points : one only or one primary with other subsidiary ones or multiple equally important ones or multiple related ones that measure different facets of the same quantity (different domains of quality of life (QOL) , different measures of renal function)
   - number of measurement occasions : one only at end of trial or pre and post treatment (pre-post design) or repeated response measurements.

d. Toxicity or side effects : *(Friedman CT chp 11)*
   - Note that trials are designed to detect efficacy, it lacks power to detect side effect at the same time. Therefore need for post-marketing surveillance (phase IV study).
   - Nevertheless, trials should still incorporate adverse effect measures but problematical apart from lack of power :
     - may be difficult to define eg subjective symptoms
     - large number of possible adverse effects
     - sufficient duration of follow-up
     - patients at risk of adverse effect selectively excluded
     - withdrawal or dose reduction because of adverse effect?
     - how measured ? elicit (checklist) vs volunteer (no checklist), frequency within subject, severity
     - attribution : disease itself, co-intervention.

e. Who evaluate?
   - Clinician/treatment team but must be or preferably blinded or use ‘hard’ end point.
   - independent evaluator, central lab, end point reading committee.
   - Training of evaluators esp for unfamiliar end point eg QOL measure.

f. Duration of follow-up : sufficient to observe outcome as well as adverse effects.

e. Design techniques:

Goals are bias avoidance, practicality/feasibility and statistical efficiency. These may conflict with one another.

1. Experimental design

- standard is parallel group design (concurrent controls) : patients are randomised to one of 2 or more arms, one arm is treatment under trial, the others being control treatment such as placebo or an active comparator. (*Lavori NEJM 1983*)
- there are other options :

1. historical controls : not recommended. *(Pocock CT chp 4 1983)*
2. self-controlled (pre-post design) : also not recommended but *see Guyatt JCD 1986*
3. Cross-over trial : each patient is randomised to a sequence of 2 or more treatment, and hence acts as his/her own control for treatment comparisons. (*Louis NEJM 1984, Hills Br J clin pharm 1979*)
4. N of 1 trial : trials in single subject, single patient randomised multiple cross over trials of one treatment versus another. (*Johannessen BMJ 1991, Mahon BMJ 1996, March BMJ 1994, Guyatt NEJM 1986, Mcleod Lancet 1986)*
5. Factorial design : 2 or more treatments are evaluated simultaneously in the same patient population through the use of varying combinations of the treatments. Simplest example is 2X2 design in which patients are randomised to one of 4 possible combinations of 2 treatments; A&B, A alone, B alone, neither. *(Byar Cancer Treat report 1985)*

- Alternative designs to deal with problems in obtaining consent or patient /clinician preferences : *Zelen NEJM 1979, Bradley Diab care 1993, Torgerson BMJ 1998*
  - Zelen's design
  - Brewin & Bradley's design
  - Korn & Baumrind's design

2. Randomisation *(Pocock CT chp 5, 1983)*

This is the process by which trial subjects are assigned to each arm of a trial randomly. Through random allocation, each subject has equal chance of being allocated to each arm of the trial (not necessary each arm of trial has equal chance, unequal randomisation is OK see below). Further, the allocation is unpredictable, for eg, ABABAB..... sequence ensures equal chance of allocation for a 2 arms trial but is predictable.

Purpose of randomisation :

a. Scientific : avoid bias in selecting patients to treatment groups by investigators. Further, randomisation tends to produce study groups

comparable to known as well as unknown prognostic factors; ie avoid confounding bias. However, in any single sample for a trial , no guarantee that randomisation will produce balance baseline covariates. Always need to check and adjust accordingly.

b. Statistical : randomisation provides basis for frequentist inference (Neyman Pearson approach as exemplified by significance test and confidence interval ; but apparently not for Bayesian inference, *Berry Stat med 19##*). Further, it can do so without any assumption of population model. Exact inference (permutation test) bsed on randomisation distribution is theoretically possible and increasingly feasible with cheap computing power and smart algorithm.

Methods of preparing the randomisation list :

- simple randomisation but often unequal number.
- restricted randomisation to ensure equal treatment numbers: methods are replacement randomisation, random permuted block, biased coin method.

- stratified randomisation : random permuted block within strata or minimisation method(only balanced marginals).
  - choice between stratify pre or post randomisation ie stratified randomisation vs stratified analysis/covariate adjustment.
  - Stratify if : a. Multicentre, then by centre; b. Highly influential factor that is also uncommon ;risk of all subjects with factor randomise into one group or severe imbalance, c. Small trial (<100 subjects).

- Individual (patient) randomisation vs group or cluster (clinician or centre) randomisation. Group can avoid contamination bias, important for psychosocial, educational and health service intervention trial

- Equal or unequal randomisation :
  - Equal is consistent with equipoise; no preference for particular treatment and also most powerful.
  - Unequal to gain more experience with use of new drug, to detect uncommon adverse effect. Up to 2:1 allowed to preserve power. (*Pocock CT chp 5, 1984*)

- Problems in randomising patients who have preference or interventions that depends on patient's cooperation that in turn depends on patient's beliefs and preference :
  1. Patient preference trial : use of preference arm, randomised only those without preference. (See alternative designs page 21)
  2. Measure preference and include as covariate (*Torgesson BMJ 1998)*

- Problems in randomising patients when clinicians have preference or unwilling to assign different treatments to similar patients.
  - Composite randomisation : randomise clinician, all patients under the clinician assign to same treatment (*Simon Biometrics 1981)*

* Clinician centred design (*Korn & Baumrind Lancet 1991)*

- Dynamic randomisation or adaptive procedure but problem with equipoise. (*see Simon Stat Med 1991, Begg Biometrics 1980, Armitage JCE 1989*)

Refer Appendix B on how to prepare randomisation list.

Revealing the treatment assignment which should be as late as possible. This is more important than how the randomisation list is prepared. Randomisation can be subverted (*Schultz JAMA 1995*) if one is not careful with this.

4 basic options (*Pocock Statistician 1982*)

1. transfer list to sequence of sealed opaque envelope
2. If trial is double blind, the pharmacist can produce a corresponding sequence of 'drug packages'. The package must be coded and linked to the randomisation list. The code is also written on patient's trial forms.
3. central registration and randomisation : contact by phone esp for multicentre trial,
4. For single centre trial, an independent person can handle allocation.

3. Blinding and placebo: *(Pocock CT chp 6 1983, Friedman CT chp 6)*

one or more of the participants involved in the trial is/are unaware of which treatment the patient is receiving.
Single blind : only patient or evaluator are blinded.
Double blind : both patient and clinician/evaluator are blinded.
Triple blind : all are blinded, including statistician and data monitor.

Justification for blinding :
to avoid bias. This may be due to :

1. Patient : psychological or expectancy effect from being on new treatment; esp in psychiatric illness.
2. Clinician : awareness of treatment assignment may influence clinical management decision eg dose modification, intensity of monitoring, co-intervention or additional treatment.
3. Evaluator : awareness of patient's treatment makes it difficult for evaluator to remain objective in evaluating response to treatment, esp if response evaluation requires clinical judgement (soft end-point).
4. Statistician or data monitor : awareness of treatment assignment can influence analyses esp interim analysis and decision on stopping trial.

To the extent the above biases are important in any particular trial, then blinding should be used.

Feasibility of blinding :
However, blinding is not always feasible .

1. Invasive treatment like surgery, ECT etc: sham operation or procedure ethical? Rare examples in the literature : CABG trial, ECT trial (*Johnstone Lancet 1980*)
2. Treatment that requires active participation by subjects eg behavioural or lifestyle intervention, patient education, health services provision. Participation depends on patient's beliefs and preference which can influence outcome, yet blinding all the more important. (*Buck JCD 1982, Kramer JAMA 1984*)
3. Highly toxic treatment or treatment with distinctive effect are difficult to blind.
4. Treatment that requires complicated dose schedule or dose modification, as opposed to fixed dose treatment.

Alternative strategies are :

- blind independent evaluator only.
- use hard end-point eg mortality, lab measures.
- Omit informed consent (subject therefore unaware being on trial); possible for community trial of behavioural or lifestyle intervention, and health services intervention where group randomisation is used.

Blinding techniques :

1. Use of placebo (treatment that is identical in all respects to the active treatment except the active ingredient is absent):

when no effective treatment available or even if treatment exists, the treatment efficacy may not be proven or probably false positive. *(Pocock CT chp 6 1983, Collier BMJ 1995*). Otherwise, should compare against standard treatment. (*Henry BMJ1995, Rothman NEJM 1994, Rothman BMJ 1996)*

Purpose of placebo :

a. Make blinding possible.

b. Assess non-specific positive treatment response (placebo effect) so that observed effect is real and not attributable to placebo.

2. Invasive treatment like surgery, ECT etc. : sham procedure but practical or ethical ?
3. IV treatment : sham injection but ethical?

4. Randomisation list must not be revealed. Treatment must be coded and linked to the randomisation list but must allow for emergency unblinding (code breaking) in case of severe reaction.

5. Unblinding bias : when codes are broken. Strategy : ask patients at end of trial and compare response. (*Kramer JAMA 1984*)

f. Interim analysis and stopping rules: *Pocock BMJ 1992*

periodic monitoring and analysis of accumulating data during the course of the trial to detect early benefit, unexpected toxicity or lack of efficacy. The trial may then be terminated or otherwise its design altered.

Justification :

1. Ethical : avoid randomising patient to inferior treatment or treatment with serious toxicity. This however needs to balance against the needs to obtain reliable conclusion based on sufficient data. (*Pocock BMJ 1992, Editorial Lancet 1993*)
2. Efficiency : unnecessary prolongation of trial and the cost it entails.

Can be controversial; *see Abrams BMJ 1998, Editorial Lancet 1993*

Feasibility : not always feasible

1. Trial with inadequate sample size to start with.
2. Poorly organise trial with incomplete data.
3. Trial with long term outcome. By the time data ready for interim analysis, all patients would have been randomised.

Statistical stopping rules or guidelines :

This is a technical matter, consult a statistician. Essentially, the problem is that of multiple testing, need to adjust or correct the P value according to the intended stopping rules.

a. Positive stopping rules : stopping trial because the new treatment shows early large unexpected benefit. Many methods are available :
   - continuous sequential method : only for industry
   - group sequential method : usual; many published rules like O'Brien-Fleming, Peto-Haybittle, Pocock, Lan-Demets.
   - stochastic curtailment :
   - repeated confidence interval method
   - two stage design
   - Bayesian method : open trial or continuous assessment (*see Lilford BMJ1995, Berry Stat med 1985, Freedman CCT 1989, Fayers Stat med 1997, Abrams BMJ 1997*), shrinkage estimate (*Pocock CCT 1989)*

b. Negative stopping rules (*Pocock BMJ 1992*) : stopping trial because the new treatment

   - toxic : pre-specified P value for occurrence of toxicity associated with new treatment . Weaker evidence is acceptable for adverse effect, typically $p<0.01$ for negative stopping rule.
   - lacks efficacy :statistically significant effect at trial end improbable.
     Methods : a. Stochastic curtailment
     b. Confidence interval method : CI may overlap but if far from minimally clinically important difference at say 99%.

Interim analysis :

1. must be pre-planned and not ad-hoc.
   - which end-point? : stopping rule should only be based on one major end-point, but check others for consistency.
   - overall treatment effect or sub-group ? : usually just overall but note ISIS2 example.
   - frequency : one or two, maximum 5 (*Pocock CT 1984*)
   - when ? : only when pre-specified sufficient data have accumulated.

2. Data requirement and preparation :
   - must NOT be based on available data since poorer outcome tend to accumulate first while good outcome still on follow-up.
   - based on number of patients who have been on trial from beginning for a specified period. They must all complete interim evaluation then.
   - missing data (incomplete evaluation) if present must assess impact on interim analysis results.

3. Independent data monitoring committee :
   - subject matter experts with experience of RCT and statistician.
   - Not involve in the trial (ie independent)
   - drug company sponsor not allowed
   - confidentiality of interim results crucial

Decision to stop trial :

- should not be based on statistical stopping rules alone (ie corrected P value)

- balance individual ethics and collective ethics ie not wanting to randomise next patient to inferior or toxic treatment versus the need for reliable and credible results that can change medical practice (*Pocock BMJ 1992, Editorial Lancet 1993*).
  - Example : Zidovudine trial in AIDS stopped early based on clinical response because of overriding concern for individual ethics; subsequent European Concorde trial (which wasn't stopped) showed no difference in survival outcome.
  - Example : ISIS2 trial of streptokinase for AMI already shoed definite overall benefit yet randomisation continued because of need for more precise treatment effect estimate , including for subgroups. Concern for collective ethics here is overriding to definitively change medical practice as past results of small trial has failed to do so.

  - Other considerations are :
    - literature review; previous similar trial results.
    - disease pathophysiology and drug action , for eg early benefit may be expected but not sustained based on such biological considerations.

g. Ethical issues :

a. Approval from local technical and ethics committee :

- approval that trial is ethically acceptable
- approval that trial is technically /scientifically sound
- approval for informed consent procedure (omission of informed consent)
- approval for stopping rules

But note that official approval and informed consent do not justify inherently unethical research.

b. Informed consent. *See debate : Doyal BMJ 1997, Tobias BMJ 1997 & BMJ 1993*

1. Not always necessary : (*Pocock CT 1984*)

- community trial with group randomisation
- similar treatment with differing regime only .
- equivalence trial of drug with proven efficacy against placebo
- patient with acute distress; consent from relative?
- cancer : patient may not be informed of diagnosis in the first place.

2. How informed is informed consent?

- fully informed consent, including interim results to mere lip service

3. How done? (*Wager BMJ 1995*)

- patient information leaflet useful
- written consent vs verbal approval
- particularly difficult groups : paediatric, cognitive impaired, acute distress

4. Alternative design :

- randomised consent design (*Zelen NEJM 1979*)
- patient preference trial : only patients with no preference randomised and consent obtained. (see page 21 on alternative designs)

c. Adherence to current guidelines assure trial is ethically acceptable
ie trial provide adequate protection of trial subjects and is scientifically /technically sound.

- see Appendix D for ethical (Heklsinki declaration), GCP and biostatistical guidelines for trial.
- Note the following in GCP guidelines :
  - sponsor's and investigator's responsibilities.
  - need for a system of quality assurance and trial audit to ensure adherence with GCP guidelines.
  - need for a person to monitor trial progress
  - adequate documentation : protocol, operational manual, medical records, data forms, raw data.

Step 3 : Conduct the trial : How to organise a trial?

The actual conduct of a trial may be divided into 3 phases : *Klimt CCT 1981*

1. Planning phase : trial proposal, funding and approval.
2. Preparatory phase : protocol and manual, organisation and logistics, recruitment, data issues, pilot studies, training and instruction.
3. Operational phase : recruitment, registration, informed consent, schedule and termination.

1. Planning phase :

a. Core group of investigators and other support personnel assembled and led by an able leader, the principal investigator.

b. Development of a trial proposal /protocol: proposal in broad outline suffice, only later when trial is to proceed need a detailed protocol and operational manual be drawn up.

Contents of proposal should include :

1. Rationale and motivational background justifying the proposed trial.
2. Literature review
3. Formulate trial objectives / hypotheses (step 1)
4. Trial design :
   - Experimental design : parallel groups etc
   - Study population + sampling and sample size
   - Enrolment of subjects : eligibility screen, consent, baseline examination, randomisation.
   - Treatment : describe, compliance. Other study variables and measurements
   - Follow up visit and schedule
   - Evaluation of patient response : training, data collection, blinding
5. Plan for data monitoring and interim analysis.
6. Plan for data management and analysis.
7. Organisation : investigators, trial administration : committees. Coordinating centre
8. Work plan : time table or time schedule of research project.
9. Pilot study or pre-test planned.
10. Budget : personnel, data handling, travel and accommodation, equipment, materials and supplies.
11. Summary
12. Appendices and references : letters (collaborators, approval), definitions and instruments, description of method (esp unpublished one), references.

c. Funding : must secure necessary funding from various sponsors.

d. Approval : must secure approval from various parties involved.

1. Ethics committee.
2. Collaborating centres.

3. Supporting lab, computer facilities and others.

2. Preparatory phase of a trial :

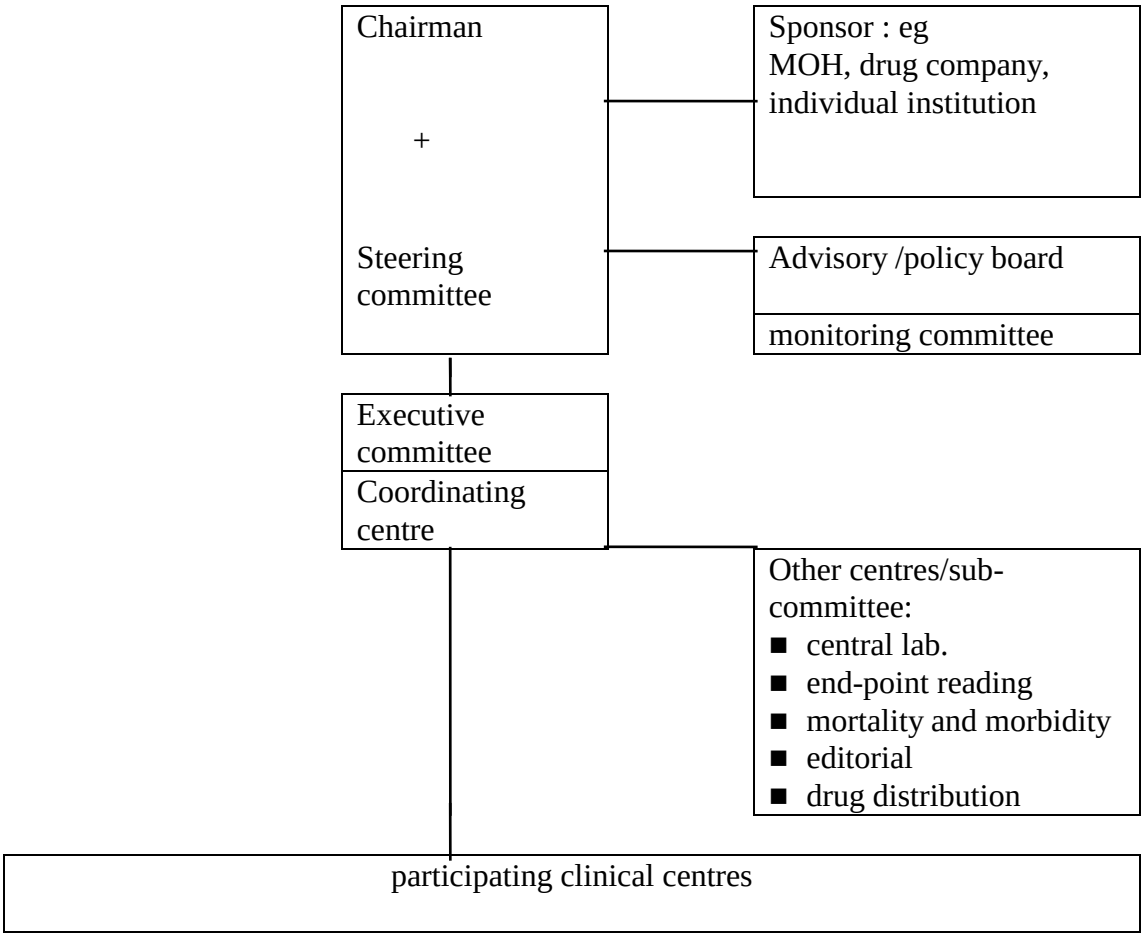
After obtaining funding and approval for the proposed trial, a detailed protocol (document containing information on purpose, design, conduct and analysis of a trial) and operational manual must be drawn up. Protocol and manual may be merged into single document.

Detailed preparation should include the following :

1. Organisation and committee, administrative, participant and staff responsibilities clearly delineated.
2. Recruitment : enlist co-operation/participation by centres and subjects.
3. Data issues : data collection, data management, quality control.
4. Pilot studies/ pre-tests : address methodological and operational problems
5. Training and instruction

1. Organisation : *Meinert CCT 1981, Ederer AJE 1975*

A typical organisational chart of a trial is as follows :



Various committees are required; it would however be a mistake to do a trial by committee. Strong leadership provided by the principal investigator is needed.

Typical membership of the committees are:

1. Clinicians : participating clinical investigators
2. Pharmacologist
3. Epidemiologist, statistician and computer expert.
4. Pathologist
5. End point specialist : reading X ray, ECG, etc
6. Lab. Specialist
7. Nosologist : expert in disease classification
8. Ethicist or medico-legal expert
9. patient representative

Other important staff to recruit :

1. Administrator
2. data coordinator
3. clinic data handler
4. Monitor
5. Data entry clerk
6. Secretariat staff

1. Chairman and the steering committee + executive committee:
   - Clearly defined leadership is important. The principal investigator who is the chairman must provide strong leadership and be ultimately in charge.
   - The steering committee is the leadership body. It typically comprises participating investigators and representatives from coordinating centre, lab, end-point centre and other centres.
   - If the steering committee is too large (can be huge in multi centre trial), then an executive committee is appointed for day-to-day executive functions in directing the operation of the trials, dealing with operational problems and supervise the coordinating centre.

2. Advisory or policy board and monitoring committee :
   - this comprises individuals from outside the trial . It reviews the trial design and operation, and adjudicates controversies.
   - the monitoring committee similarly is comprised of individual from outside the trial but chaired by the steering committee chairman. It may be merged with the advisory board. It has the responsibility for monitoring results of the trial (interim analysis to detect early benefit, toxicity, lack of efficacy) and advises on trial termination accordingly.

3. Co-ordinating centre :
   - this a key unit of any trial. In multi-centre trial, it holds the trial together.
   - it is responsible for overall coordination of the trial and handles all administrative matters. Its day-to-day service functions include patient registration, central randomisation, supplying data forms and other materials, supply drug treatment, dealing with enquiries and providing feedback.
   - the centre also provides methodological, statistical and computer expertise fro trial design, analysis and data management.
   - it is also responsible for data management and data quality control.
   - it monitors trial progress and participant's compliance with trial protocol and guidelines .

4. Other centres and sub-committees :
   - central lab to handle uniform lab tests
   - end point reading committee : eg for reading HPE, ECG, X ray etc
   - disease, mortality and morbidity classification
   - editorial : report writing and publication
   - drug supply centre

2. Recruitment : enlist co-operation/participation by centres and subjects.
This is crucial. The 2 most important participants are :

1. Participating centres : they recruit/enlist study subjects, institute intervention and collect data.

   Methods for promoting trial acceptance and enlisting cooperation of centres :

   a. Leadership influence, professional body or peer influence, social network.
   b. Timing is important widely accepted procedure not accepted for trial.
   c. Address ethical concerns : lack of individual equipoise
   d. Education : awareness of importance of research, contribution to science
   e. Participation : sense of ownership of data.
   f. Recognition : publication
   g. Feedback : regular meeting to update on progress, feedback analysis, demonstrate utility of data.

2. Study subjects :

identifying potentially eligible subjects and obtaining their cooperation. Ideally we wish to ascertain all eligible subjects from a geographically defined area ie population based study.

Strategies and methods for recruiting subjects:

a. local effort at individual participating centre.
b. Influence of key people : doctor, community leader.
c. Directly contacting patients: from existing register, mass mailing, health fair, community screening.
d. Contact local medical community or organisation (non-participating centre) to send patient
e. Publicity via media; must brief local community first
f. Appeal to sense of contribution to science
g. Incentive : payment.
h. employ recruitment agency
i. Active case finding : search register of ward, clinic, lab, X ray, admin database Death register etc.

ref :
Friedman CT chp 9 1985
Prout CCT 1981
McIntyre BMJ 1991
Baum Lancet 1993

3. Data issues : data collection, data management, quality control.

No trial is better than the quality of its data (this apply for all types of research). This is a crucially important issue.

1. Data collection : *Friedman CT chp 10, 1985*

this entails the process of :

a. Obtaining measurements from data study subjects .
b. converting the raw data into categories
c. recording the data on data form or data sheet .
d. transferring data into computer (unless data captured electronically).

Preparations for data collection must include :

1. Obtaining measurement :
   - instrument prepared : equipment, observational method, questionnaire.

2. Data conversion :
   - defined rules or criteria for conversion.
   - for numeric data expression, no need conversion.
   - for verbal data expression, allow only simple conversion into categories. Always capture component data needed for logical or judgmental conversion.

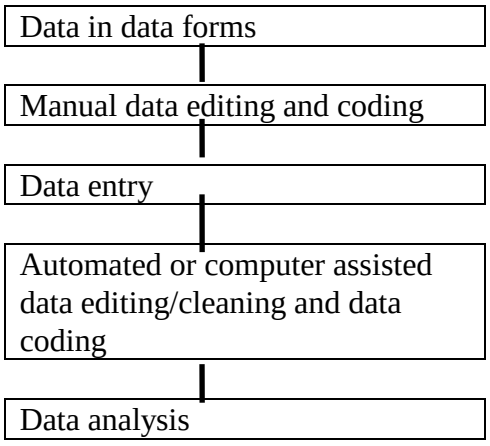
3. Data recording :
   - how many data forms : baseline, follow-up evaluation, flow sheet for serial measurements, clinical encounter form, outcome evaluation.
   - data form or case report form design (**see Appendix C : Data form design**)
   - Note that data form has other uses apart from recording data. They are : as integral part of instrument(eg observational guide & questionnaire), as instruction sheet, and as coding key or frame and code sheet.

4. Data transfer into computer : computer, programs, DE clerk.

5. All of above :
   - pre-test and training
   - time or examination schedule
   - logistical prep : supplies, field personnel/data collectors, supervisor.
   - communication and coordination

2. Data management :
   how data are managed after they are collected and returned to coordination centre.

It is concerned with :
a. Data editing
b. Data coding
c. Data entry
d. Data preparation for analysis.



1. On receipt of a data form, data coordinator manually checks form for incomplete data (missing data) and egregious error. The form is then passed on for data entry.
2. Data entry using data entry screen with in-built validity and range checks. If necessary double enter all data.
3. Post data entry computer assisted data editing :
   - data editing consists of error detection and data corrections.
   - error detection consists of checking each field for missing data and non-valid entry, and checking certain predetermined combinations of fields for inconsistency. The computer system should provide a listing and documentation of the errors detected in a format amenable to subsequent review and correction by data collection personnel. Informal editing should be avoided.
   - data correction consists of updating the database with the corrected values. Ideally, the system should leave an 'audit trail' so that the existence of data at any point in time can be accurately recovered. More typically however, system only generates the latest most error-free data; and data 'freezing' prior to each analysis.

References :
1. *Karrison CCT 1981*
2. *Naus JI. Data quality control and editing. NY Marcel Dekker 1975*
3. *Armstrong Principles of exposure measurement OUP 1992*

3. Quality control procedures :

assuring that the data collected are of good quality, and that these data properly managed, processed and analysed.

a. General :

- it is not possible to have quality control procedure for all measurements. Be selective, concentrate on important measures eg main outcome and study variables.
- do not collect unnecessary data.

b. Preparation :

- well designed instrument and data form minimise error
- all study procedures (instrument, data collection forms) must be pre-tested.
- explicit and specific description of measurement and data collection procedures in the operational manual; in excruciating details, illustrate with photo if necessary.
- establish formal training and certification procedure; either centrally or on the job.

c. Data collection :

- responsibility for seeing that staff is familiar with procedure and protocol rests with centre investigator, and aided by designated participating centre data coordinator/clinic monitor. Monthly phone to discuss problem.
- on site editing of completed data form by data collector as well as an editor.
- address problem identified through monitoring immediately, correct error by call back to subjects or check back to records.
- periodic meetings with data collection personnel to review procedures and discuss problems, retraining.
- periodic site visits to monitor adherence with procedure and protocol : on site observation of the study procedure, on site replication of procedure by independent observer, random checks on data recording and equipment calibration and accuracy.
- replicate some proportion of ata collection (eg 10%) to identify problems.
- for lab test, internal quality control (eg control chart to check drift) , external quality control program; send unknown standards at periodic intervals to evaluate the precision and variability of each labs.
- coordinating centre monitor each participating centre performance : number recruited, # missed visits, # missing data/data error, # protocol deviations.
- data analysis to check error : for repeated measures eg BP readings, variability less than expected means ‘chartology’, compare distribution of study variables among data collectors, trends in variable over time.
- maintain editor’s log of all problems in recording, coding, specimen handling, analysis etc
- enter data contemporaneously with data collection.

d. Data coordinating centre :

- monitoring the coordinating centre itself : external audit or review (‘bank examiners’) to check documentation of operational procedure (data coding,editing entry etc), review data processing procedure, file maintenance, data analysis.
- Sometimes independent centre replicate parts of data entry, processing and analysis to estimate error in data management.

references :
1. Knatterud CCT 1981
2. *Armstrong Principles of exposure measurement OUP*
3. *Hilsenbeck, Glaefke, Feigl, Lane et al. Quality control for cancer registries. US Department of Health and Human sercices Washington DC 1985*
4. *Statistical analyssi and Quality control center. Quality control for cancer registries 1985 Seattle WA*
5. *Horwitz, Yu. Assessing reliability of epidemiologic data obtained from medical records JCD 1984;37:825-31*
6. *Stellman. The case of the missing eights. AJE 1989;129:857-60*
7. *Kaplan (ed) Clinical chemistry : theory, analysis and correlation Mosby 1989*
8. *Tietz (ed) Textbook of clinical chemistry Saunders 1986*

4. Pilot studies/ pre-tests

this is important to consider . It is difficult to plan a trial without some ideas of how things might turn out in practice. Do not assume people will bother to read operational manual.

Pre-test : small trial to test isolated problem of the design.
Pilot study : a small scale replica of the main trial to identify and eliminate problems in the large scale trial proper.

Potential problems to address are :

1. Methodological problems :
   a. Study population :
      - adequacy of recruitment rate
      - accessibility and non-response/consent rate to expect.
      - variability of response measure : needed for sample size calculation

   b. Measurement and data collection :
      - validation study for new or non-standard or subjective measures.
      - trial of alternative methods : eg face to face vs mail vs phone interview
      - trial data collection to identify problems : trial interview, trial examination schedule and observational method.
      - adequacy of data forms.

2. Operational and organisational problems :
   - efficiency of study organisation in the clinics, centre etc
   - efficiency of logistics and communication
   - adequacy of supervision and motivation of clinic staff or data collectors or participating investigators.
   - estimate cost and duration of study.

5. Training and instruction

- training of participants and research assistants on patient recruitment, informed consent procedure, use of instruments and data forms, and other trial operational procedures (do not assume people will bother to read operational manual).
- training of staff on organisational and operational procedure at office, data management, quality control procedure.

3. Operational phase :

Operational manual drawn up, organisation developed, recruitment procedure ready, data issues settled, pilot studies done and problems ironed out, all necessary training and instruction completed, the research is ready to be implemented.

Consider the following sequence of events in the operation of a trial and how they are to be carried out.

Recruitment : All potentially eligible patients over a period should be invited to participate.

Eligibility screen.

eligible

Patient registration (log book) and baseline evaluation: all eligible patients regardless of subsequent consent to participate in trial.

Informed consent.(May randomised first prior to consent)

Run in or qualification period . : optional

Revealing randomisation and blinding procedures.

Follow-up :

- Follow up visit and evaluation schedule
- Protocol deviations
- Trial monitoring. : check compliance with protocol and trial guidelines, monitor adverse effect, organise data processing, feedback info to participants, stopping rules.

Trial termination and close out.

1. Recruitment : ideally we wish to ascertain all patients from a geographically defined area. In practice rare for clinical trial.
   - local effort at individual participating centre.
   - directly contacting patients and enlist
   - briefing local medical community (non-participating centre) to send patient
   - National recruitment drive via public media
   - employ recruitment agency

2. Screening :
   - confirm diagnosis
   - satisfy eligibility criteria
   - ideally also obtain baseline data of non-eligible patients

3. Patient registration (log book) and baseline evaluation: *Friedman CT chp 8, 1985*
   - all eligible patients regardless of subsequent consent to participate in trial must be registered and baseline evaluation completed.

4. Informed consent procedure :
   - who approach patient?
   - patient information : leaflet, video
   - consent form
   - difficult group : paediatric, acute distress, cognitive impaired.

5. Run in period or qualification period + placebo run in (all patients given palcebo): (*see Knipschild JCE 1991, Senn BMJ 1997*)
   - check eligibility criteria and complete baseline exam
   - washout previuos traetment
   - stabilise condition and baseline response parameter.
   - cooperation with end point evaluation eg self completed questionnaire, diary etc
   - weed out non-compliant patients and placebo responder.
   - test and adjust therapeutic response
   - familiarise participants with trial procedure.

6. Revealing randomisation :
   4 choices ; see randomisation page 23

7. Follow-up period :
   - Follow-up schedule : non-annual and annual follow-up and examination schedule
   - adherence with protocol : both patients and participating clinicians. *Pocock CT chp 12 1983*
   - overcome boredom : regular meetings, feedback
   - monitoring committee active

8. Trial termination and close out : *(Friedman CT chp 17*)

Step 4 : Analyse the data : How to analyse clinical trial data?

You will certainly consult a statistician for this. All the issues here are highly technical and difficult to teach non-statistician. Instead, I shall show numerous illustrative examples from trial reports and emphasise data interpretation.

Analysis of clinical trial data entails :

1. **Preliminary clarifications**
2. **Initial data analysis**
3. **Definitive analysis**
   - Summarising treatment response and effect.
   - Dealing with complications : further issues in data analysis
   - Statistical inference

1. Preliminary clarifications :

Before looking at the data, clarify the following with the researcher:

a. Clarify research purpose :

For clinical trial, this is usually obvious ie an estimate of the effect of the treatment under trial with respect to the primary end point is required, as pre-specified in the primary research question.. However, there are usually subsidiary questions involving other endpoints or subset hypotheses, and further there may be ancillary questions that may not necessary be pre-specified. All these require careful handling (see later).

b. Pre-specified methods of analysis :

The methods of analysis are pre-specified in advance in trial protocol . However unanticipated problems or features of importance may be found and deviations are allowed but must be justified.

c. Trial and measurement design, and data collection must be considered:
   - Experimental design, randomisation and blinding.
   - measurement quality : reliability, validity

d. Context : prior or background knowledge available :
   - what is known about the treatment, and its effect

2. Initial data analysis (IDA):

an informal exploratory look at data to get a feel for them. It is a systematic way of assessing, exploring and describing a given dataset prior to definitive analysis. It consists of :

1. Data scrutiny :
   - assess data quality : errors, outliers, missing data , biases in measurements
   - assess data structure : sample size, number and nature of variables
2. Data summary : descriptive statistics, graphs and tables to summarise dataset
3. Data modification necessary : transformation, categorising variable, etc
4. Highlighting interesting or important features in data and suggest direction for definitive analysis or deviation from pre specified analysis.

3. Definitive analysis

- Summarising treatment response and effect.
- Dealing with complications : further issues in data analysis
- Statistical inference

Summarising treatment response and treatment effect : *see page 44 for details*
Choice depends on :

- measurement scales : continuous, ordinal or categorical(event) response data.
- number of measurements : one, pre-post, repeated measurements.

Further issues in data analysis

a. Complex end point :

- Quality of life measure: implication for analysis and difficulty in interpretation

b. Analysis of equivalence trial : *see Jones BMJ 1996*

c. Handling protocol deviations : *Rabeneck Arch int med 1992*

- Modification of intended treatment : protocol error, poor compliance, cross-over or contamination.
- Intention to treat analysis versus per protocol/on treatment analysis. *(Lewis Br J Cancer 1993 Lee Stat med 1991)*
- adjustment for non compliance and contamination *(Stat med Feb issue 1998, Cuzick Stat med 1997)*

- Lack of end point (response/outcome) measure :
- competing event : competing risk model
- withdrawal : informative vs non-informative; always obtain information on reasons for withdrawal. May be possible to adjust *(Murray Stat med 1988, Heyting Stat med 1992)*
- missing data : *Stat med March/April issue 1998*

d. Baseline or prognostic imbalance .

- testing for baseline imbalance? (*Altman Lancet 1990, Armitage CCT 1981, Senn Stat med 1994)*
- adjustment for imbalance vs stratified analysis with stratified randomisation.

e. Subset or sub-group analysis : *Oxman Ann Int med 1992, Bulpitt Lancet 1988, Simon Br J clin pharm 1982,*

- Interaction between treatment effect and patient subset
- significance test within subset.

f. Omitted covariate : linear vs non-linear model *(Gail Biometrics 1984)*

g. Sensitivity analysis and correction for other biases :

- Selection bias : centre effect, non-response(non-consent bias); compare consenting and non-consenting eligible patients
- Regression to the mean : adjustment for regression dilution bias
- Unblinding bias : compare those who guess correctly their allocation and those not.

Statistical inference : Drawing conclusion with the aid of probability statement from a sample to a larger universe. *See Data interpretation page 57*

- frequentist inference : P value, confidence interval
- Bayesian inference
- Multiple comparisons : *see Perneger BMJ 1998*
  - define one end point as primary, the rest subsidiary.
  - multiple end points of equal importance.
  - multiple related end points : single multivariate outcome

Summarising treatment response and treatment effect :

In a trial, subjects are randomly assigned to each treatment arm and their responses to treatment are then measured. Subsequent analysis of the response data requires appropriate measures to summarise and compare treatment responses. Treatment effect refers to the comparison of treatment responses between the 2 arms or among 3 or more arms of a trial. These summary measures of treatment response and treatment effect are called summary or descriptive statistics.

Choice of summary statistics depends on :

- types of response data (measurement scale) : continuous, ordinal or categorical(event) response data.
- number of measurements : one post-treatment , 2 at pre and post treatment, 3 or more repeated measurements.

Wide variety of summary statistics are used to describe treatment response and effect. Most are technical and difficult to explain. Instead I shall emphasise interpretation of the statistics through illustrative examples from trial reports. I categorised them on the basis of types of response data and number of measurements as follows :

	Types of response data and number of measurements	Examples	Summary statistics : treatment response	treatment effect
1. a	Single terminating event	death, success/failure	event risk, event rate	Risk difference or ratio; odds ratio. Rate difference or ratio
1.b.	Recurrent events	number of epileptic fits, recurrent peritonitis, recurrent bladder tumour	event risk, event rate	Risk difference or ratio. Rate difference or ratio
2.a	Single post treatment ordinal response	tumour response(remission, partial response, progress), severity (mild,mod,severe).	distribution of ordered response categories	rank sum estimator, generalised odds ratio, cumulative odds ratio.
2.b.	Repeated ordinal response data	hynotic response, tumour response,	distribution of ordered response categories	cumulative odds ratio, continuation odds ratio, adjacent cat. odds ratio
3.a.	Single post treatment quantitative response	BP in mmHg, renal function GFR in ml/min, lung function	mean, median	difference in mean or median
3.b.	Two measurements at baseline and post treatment	grip strength in arthritis, lung function, Hamilton score in depression.	Post-treatment mean, mean change, mean percentage change.	difference in mean
3.c.	Repeated quantitative measurements	creatinine or GFR in renal failure progression, grip strength in arthritis, lung function, Hamilton score in depression.	summary measures of each patient's responses, single multivariate end-point	difference in mean of summary measures

1. **Event data (categorical data) :**

These are occurrence of events like death, complication, success/failure, respond/non-respond. These are the most commonly encountered response data in trial.

a. Single terminating event : (Sinclair JCE 1994, Laupacis NEJM 1988, Cook BMJ 1995)

Examples are death or other terminating event data defined for a trial like success/failure(eg. transplant graft failure), respond/non-respond.

Summary statistics for single terminating event :

1.Risk :

this is the probability (denoted P) of an individual having an event over a specified period of time (conditional on the individual not dying from any causes during that period).As a probability measure, it can vary between 0 and 1, and it is dimensionless. As a probability measure, risk is straightforward to interpret. The higher the risk, the more likely the event will occur. A risk of 1 means the event will certainly occur during the specified period, while a risk of 0 means the event certainly will not occur. Another way of expressing probability is in terms of odds (denoted O) . This is the probability of having an event divided by the probability of not having it; O = P / 1-P

2. Rate :

Rate is the instantaneous potential for occurrence of event per unit change in time, relative to the size of the population at risk at time t. Rate is a measure of the rapidity of change. Conceptually , it expresses the pressure or force of event occurrence in the population at an instant. Rate is expressed in 'per unit time', eg per week, per year. It can range from 0 to infinity; there is no upper bound.

Unlike probability, rate has no natural interpretation. A rate say of 0.05 per week is not meaningful in itself. Hence, risk is generally the preferred summary measure of event occurrence in clinical trial.

Summary statistics for treatment effect :

1. Absolute effect measures; which are differences in risk or rate measures between groups.
2. Relative effect measures ;which are ratio of risk (or odds) or rates between groups.

1. Absolute effect measures :

Risk difference :

Risk in treated group = P1

Risk in control group = P2

.. Risk difference d = P1-P2

Interpretation :

a. Risk difference measure the absolute difference in risk of event occurrence between treated and control groups.

b. Another useful interpretation is in term of number needed to treat (NNT):

NNT = 1 / d

This expresses the number of patients who must be treated in order to reduce one adverse event; and therefore provides a measure of therapeutic effort in relation to therapeutic yield.

Rate difference :
same as above , except we have difference of rates . NNT interpretation however does not apply unless for small rate which approximates risk.

2. Relative effect measures :

Risk ratio or relative risk :
P1 = risk of treatment
P2 = risk of control
Risk ratio or relative risk RR = P1 / P2

interpretation :
a. RR expresses the relative probability that an event will occur when 2 groups are compared.

Eg RR = 0.6 < 1 means treatment reduces risk of event.

b. It is often useful to express in terms of relative risk reduction.

RRD = 1 - RR
eg. RRD = 1-0.6 = 0.4 means treatment reduces occurrence of event by 40%.

c. However, without knowledge of baseline risk (ie control group risk), it can be difficult to interpret RR. The magnitude of RR depends on the magnitude of the baseline risk. The same RR could correspond to greatly differing absolute effect.

Eg. RR = 2 = 0.002/0.001 ; with d = 0.001 which is trivial

or = 0.8/0.4 ;with d= 0.4 which is a very impressive absolute risk reduction.

Rate ratio :
same as above except we have ratio of 2 rates.

Odds ratio :
since probability can alternatively be expressed as odds O = P/1-P; we can similarly have ratio of 2 odds.

Odds ratio OR = O1 / O2 = $\frac{P1/1-P1}{P2/1-P2}$

Interpretation :
a. Odds ratio is hard to interpret. Most people except gambler are more comfortable with risk than with odds. Odds ratio can easily be converted to risk ratio

RR = OR / (1+Ic*(OR-1)) where RR is risk ratio, OR is odds ratio and Ic is event risk in control group. Therefore I can't see any reason for using odds ratio.
b. It is useful in certain specific situation. In certain case control study; OR can be used to estimate RR but could hardly be justified for clinical trial.
c. Odds ratio however has useful statistical and mathematical properties which is why it is popular with statistician but we are clinicians.

Example:

In a trial comparing effect of Cyclosporine A(CsA) versus conventional imunosuppression on kidney graft failure, the following data are obtained :

	CsA	No CsA
Number of subjects	100	100
Number of graft failures at 1 year post-transplant	10	15
Risk of graft failure at 1 year post-transplant	0.1	0.15

- no patient died.

Using no CsA group as reference,
1. What is the risk difference?
2. How many patients need to be treated with CsA in order to save one graft at 1 year?
3. What is the risk ratio and odds ratio?
4. If the odds ratio is 0.5, what is the risk ratio? Which do you prefer?

b. Recurrent event data :

Examples are recurrent peritonitis on CAPD(a type of dialysis), occurrence of epileptic fits on anti-convulsant, acute exarcebation of asthma on steroid, recurrence of bladder tumour on thiopeta etc. Very common medical response data but unfortunately difficult to handle. You are advised to consult a statistician on this.

Summary statistics for recurrent event :

1. Risk : when number of events per subject is small and risk of event varies over time, use multivariate failure time methods; eg. Prentice WP model, Wei LW model and Andersen-Gill model (Clayton SMMR 1994, Stat med issue 1997 April)
2. Rate : when events are frequent and the hazard varies little over time, use methods based on event count over defined time interval. (Clayton SMMR 1994)

Summary statistics for treatment effect :

1. Absolute effect measure : risk or rate difference
2. Relative effect measure : risk or rate ratio

Example:
In a trial comparing effect of disconnect CAPD system versus conventional UVXD system on occurrence of peritonitis, the following data are obtained :

	Disconnect	UVXD
Number of patient-months of observation in trial	1200	1200
Total number of peritonitis during trial	66	78
Peritonitis rate in number of episode of peritonitis per patient-month (pt-mo/epi)	0.055 (18)	0.065 (15)

- no patient died.

Using UVXD group as reference,

1. What is rate difference?
2. What is the rate ratio?

2. Ordinal response data :

This refers to response which is ordered. Examples are tumour response (complete remission, partial response, progression), anti-TB treatment response, Lanza score used in ulcer treatment trial, etc.
This is a difficult subject, you are advised to consult a statistician.

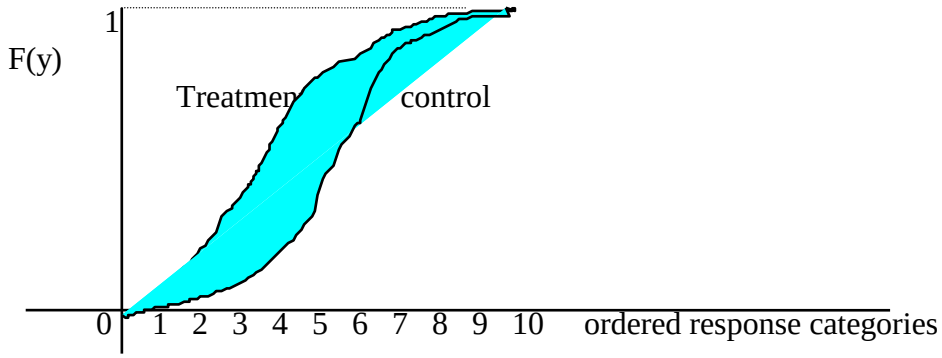
a. Single post treatment ordinal response :

Summary statistics for ordinal response :

1. mean : inappropriate
2. median : also inappropriate but commonly seen
3. display distribution of ordered response categories : recommended

Summary statistics for treatment effect :

1. rank sum estimator, delta δ : the probability of a treatment group response is better than a control group is p1 and that of control being better than treatment is p2, then δ = p1-p2.
2. generalised odds ratio, alpha α : α = p1/p2 (Agresti Biometrics 1980)
3. Continuation odds ratio
4. Adjacent category odds ratio
5. Cumulative odds ratio, theta θ : currently the preferred choice. It means the odds of larger values of response measure is θ times for treatment versus placebo; the odds remains the same regardless of choice of cut points to define larger values of response. Graphically, this indicates cumulative distribution function F(y) of response is stochastically higher for treatment than for placebo as shown below:.



Example :
In the classic MRC trial in 1948 comparing Streptomycin treatment for pulmonary TB versus no treatment(control), the response data are as follows :

category	Response	treatment	Control	Total
1	Death	4	14	18
2	Marked deterioration	6	6	12
3	Mod. Deterioration	5	12	17
4	No change	2	3	5
5	Mod. Improvement	10	13	23
6	Marked improvement	28	4	32
Total		55	52	107

1. The rank sum estimator, delta is 0.498. Interpret.
2. The generalised odds ratio, alpha is 3.75. Interpret
3. The cumulative odds ratio, theta is 5.43 Interpret.

b. Pre-post or repeated ordinal response :

This is an active area of current statistical research, consult a statistician. I shall only illustrate with an example :

Example :
In a trial comparing a new hypnotic versus placebo in patients with insomnia, the outcome is patient response to the question ' How quickly did you fall asleep after going to bed?' using categories (<20, 20-30, 30-60, >60) minutes. Responses were obtained at baseline and at 2 weeks after treatment. The results are as follows :

Table # : Proportion distribution of time to falling asleep by treatment and measurement occasion.

		Responses			
Treatment	Occasion	<20 min	20-30	30-60	>60 min
Hypnotic	Baseline	10%	17%	34%	40%
	Follow-up	34%	41%	16%	9%
Placebo	Baseline	12%	17%	29%	42%
	Follow-up	26%	24%	29%	21%

1. The response distribution shifts downwards on follow-up for both treatments indicating placebo effect or spontaneous resolution. However the degree of shift seems greater for hypnotic treatment.
   The cumulative odds ratio at follow up was 2. Interpret.

3. Continuous or quantitative response data :

These are familiar measurements to all doctors. Examples are blood pressure in mmHg, lung function test, renal function test eg, GFR in ml/min, etc.

a. Single post treatment quantitative response :

This is generally straightforward to describe.

Summary statistics for quantitative response :
1. mean : if symmetrically distributed
2. median : if skewed

Summary statistics for treatment effect :
1. difference in mean or median

b. Two measurements : pre and post treatment quantitative measures :

This is also quite common. Many trials include a baseline measure obtained before treatment and a response obtained after treatment. Examples include : exercise tolerance time in CHF trial, grip strength in arthritis trial, lung function in asthma trial, Hamilton scores in depression trial etc.

Many options are available for summarising response and treatment effect:

1. Post treatment mean : ignore the baseline measure, only summarise and compare the post-treatment response measure with or without covariance adjustment.
2. Mean change : compare change in measures ie (post-Rx response - baseline)
3. Mean percentage change : compare mean change relative to baseline ie (100* change/baseline).

Erroneous results are obtained if inappropriate methods are used. Therefore wise to consult statistician.

Example :

In an anti-asthma treatment trial comparing 2 treatment (Rx 1 and 2), the baseline is the last pre-treatment measurement of percentage of predicted normal FEV1 and the response is the maximum percentage of predicted FEV1 in the first 4 hours post-treatment. The results by steroid dependence status are as follows :

	Treatment 1		Treatment 2	
	not steroid dependent n=141	steroid dependent n=107	not steroid dependent n=143	steroid dependent n=104
Baseline FEV1	60.1	52.6	59.8	55.4
Response :				
■ percentage change relative to baseline	39.8	48.2	43.0	45.7
■ change from baseline	24.9	24.1	25.2	25.3

1. One interpretation of the above results was that steroid dependent patients respond better to treatment (especially for treatment 1) than non-steroid dependent patients. Do you agree?

b. Repeated pre and post treatment quantitative measures :

This is a difficult subject for non-statistician. Therefore wise to consult one.
Some trials include baseline measurements as well as repeated measurements of patients' responses at several pre-defined time after treatment. Examples include : sr. Creatinine or GFR measures in renal failure progression trial, exercise tolerance time in CHF trial, grip strength in arthritis trial, lung function in asthma trial, Hamilton scores in depression trial etc.

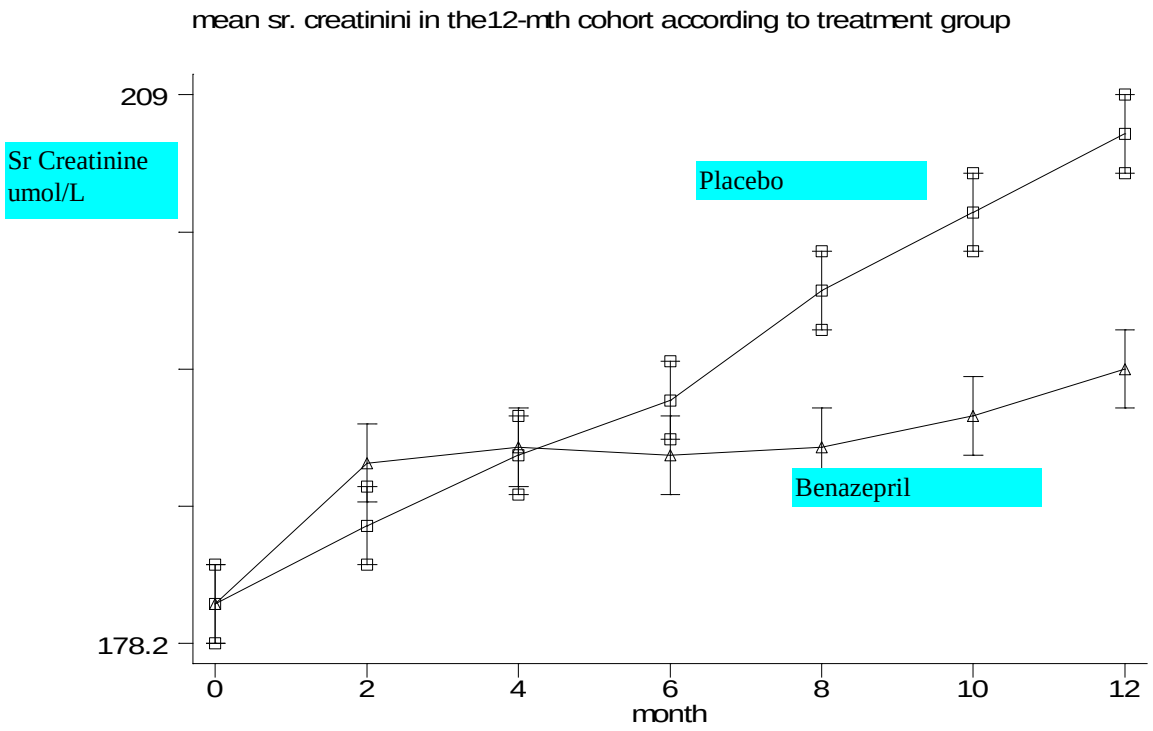
Many options are available for summarising response and treatment effect:

1. Mean response at every time points. Frequently seen in the past but obviously incorrect method.
2. Repeated measures as multiple end-points and use appropriate multiple testing procedure but difficult to interpret. (Pocock Biometrics 1987, O'Brien Biometrics 1984)
3. Summary measures of each patient's responses, then use standard 2 or multiple group comparison methods to assess differences in mean of summary measures :
   a. Post treatment mean : ignore the baseline measure, only summarise and compare the post-treatment response measure with or without covariance adjustment.
   b. Mean change : compare change in measures ie (post-Rx response - baseline)
   c. Mean percentage change : compare mean change relative to baseline ie (100* change/baseline).
   d. Rate of change (slope)
   e. Time average measure (AUC).

4. Complex models : repeated measures ANOVA, MANOVA, Random effect model/conditional model, GEE/marginal model, Markov model etc.

Example :

The APRI trial (NEJM 1996) compared the effect of Benazepril versus placebo in patients with mild to moderate renal impairment (creat.cl. 30-60 ml/min) on progression of renal failure. Primary end-point was doubling of baseline sr. Creatinine or need for dialysis. One of the secondary end-point was repeated sr. Creatinine on treatment. Results are as follows :



How would you summarise the treatment effect on serial sr, creatinine?

Step 5 : Draw conclusion : How to interpret and report a trial?

- Data interpretation.
- Writing report and preparing manuscript for publication

Drawing conclusion : interpreting research results

1. What conclusion is made?
2. Is the conclusion valid?

1. **What conclusion is made ? : interpretation given to the data**
   - Anticipated result : treatment under trial found to be effective or equivalent to standard treatment
   - "Negative" result : result inconclusive or absent of difference compared to placebo, this means either truly no difference or trial lack power (see below). If comparator is an active treatment, absent of difference of does not mean no difference or equivalent. (*see Jones BMJ 1996)*

2. **Is the conclusion valid ? interpretation justified by the data.**

Threats to study validity : the degree to which the inference drawn (conclusion made) from a study are warranted when account is taken of the study methods and data analysis.
Consider the following :
1. Measurement validity.
2. Statistical validity.
3. Internal validity
4. External validity (generalizability)
5. Substantive considerations
6. Causality inference

1. Measurement validity : relevance and soundness of measurement method.
   - relevant : appropriate to study purpose. Example: OKT3 reduces CD4 cell count, CD4 count is a relevant outcome measure in basic research on OKT3 property, but not in clinical trial of its efficacy.
   - valid : what is measured must approximate the truth.

2. Statistical validity : soundness of method used to summarise data and make statistical inference.
   Study results are basically descriptive or summary statistics (treatment effect measure) calculated from the data, accompanied by a measure of statistical uncertainty like P value or confidence interval.
   Measure of uncertainty of inference based on sample data; measure of whether findings could arise by chance. Typically by significance testing (P value) or 95% confidence estimate.

- formal assumptions required are satisfied eg homocedasticity required for t test, linear regression. Large sample assumption in many approx test.
- multiple testing : ? need for correction to preserve overall type I error. (*Savitz AJE 1995, Jones IJE 1982, Perneger BMJ 1998)*
- non pre-specified hypotheses : subgroup analysis
- Negative results interpreted as no difference : may lack power (post hoc power analysis not recommended); check 95% CI indicate range of compatible results in likelihood or Bayesian inference (*Goodman Ann Int med 1994)*
- Positive results may still be false positive : 5% alpha error doesn't mean false positive error cannot occur; more likely in published research.

3. Internal validity :

absence of alternative explanation for the findings (apart from chance or other statistical data torturing). The most important alternative explanation is :

- bias : in selection, confounding, measurement and information. Consider all sources of bias in the trial, whether effort has been taken to avoid them, otherwise, due allowance for them in analysis. *See Schulz JAMA 1995*

4. External validity (generalizability) : results that is valid in one study population may not be in another.

- assess study population eligibility criteria
- assess sample characteristics : comparable to study population of interest, intended target population.
- mechanistic generalisation : same pathological mechanism apply in other segment of population. (*Davis CCT 1994*)
- absence of important interaction with intrinsic explanatory variable like age, sex etc(*Cox JRSS A 1992*). But there is a limit, see *Thompson JCE 1994*.

5. Substantive considerations :

Even if results seems valid by above considerations (statistical, internal +/- external), results must be interpreted in the context subject matter.

- each study result is just one piece of a puzzle which when assembled will help either to confirm or refute a hypotheses. Are results consistent with other studies?
- effect size is important for assessing quantitative or clinical importance.
- biological plausibility : disease pathophysiology and mechanism of drug action or treatment makes sense .

Writing report and preparing manuscript for publication :

The object of publishing a scientific paper is you provide a report which contains sufficient information to enable readers to :

- assess the observations you made
- repeat the study if they wish
- determine whether the conclusions drawn are justified by the data (see Data interpretation above).

The goal of writing is simple and clear communication of information.

Consider the following rules :

1. Follow the instruction or guidelines given in 'Instructions to author' page which every journal publishes. *Eg BMJ 1996;312:41*. Journals often have their own house style, though most now followed the Vancouver style (*International committee of medical journal editors. Uniform requirements for manuscript submitted to biomedical journals BMJ 1982;284:1766-80; 1991;302:338).*
2. Follow the guidelines on how to write a report (medical writing) below.
3. Avoid redundant or overlapping publication: reporting the same study in 2 journals or splitting the same study into 2 or more parts. (*Kassirer NEJM 1995*)
4. Avoid hype, the temptation to overstate the case, exaggerating importance, suppressed biases etc. Better to come clean and reveal flaws, recognise limitation of work, add caveats where necessary. (*Editorial Lancet 1994)*
5. For RCT, specific guidelines available on reporting trial results. (*CONSORT guideline JAMA 1996 , see Appendix D)*

How to write a report : guide to medical writing

1. Structure and content
2. Revision
3. Style

1. Structure and content :

   The basic structure of a scientific paper has the following sections :

   - Introduction : what question was asked?
   - Methods : how was it studied?
   - Results : what was found?
   - Discussion : what do the findings mean?
   - + title, abstract, references.

   Introduction : what question was asked?

   - state clearly the research question, objective or hypotheses.
   - review briefly the relevant literature
   - justify the proposed research

Methods : how was it studied?

- describe, and if necessary defend, the study design, study population selected and sampling design, measurement methods used.
- for standard measurement method, just cite reference. Otherwise (new or modify existing method), give complete details and assess validity and reliability.
- cite references for statistical methods employed.

Results : what was found?

- describe study sample first.
- give only overall description or summary of major findings (both positive and negative findings), avoid presenting every scrap of data.
- present findings in logical sequence, proceed form simple descriptive to more complex analysis.
- presentation of data using either figure (visual impact), table (large amount of info yet concise) or simple text summary. Avoid unnecessary repetition.
- always include sample size (n=###)
- do not comment on the findings in the result section, only present a clear description of findings.

Discussion : what do the findings mean?

- interpretation : interpret results based on actual evidence of your data.
- limitations : discuss limitations of study; especially biases, problems with methods.
- disputation : how do your results fit in with the general literature on the subject? Explicitly refer to the literature (avoid preferential citing), compare findings and account for differences if any.
- disquisition : discuss implications of your findings. You may speculate, state opinion.
- conclusion : produce a succinct and strong conclusion. Don't summarise (leave it to the abstract), use signal words (finally, in conclusion)return to beginning by answering the question posed in the Introduction, if unable to do so suggest further work.

References :

- include only relevant ones
- avoid preferential citing, it will be obvious to reviewer.

Abstract :

- write this section last.
- summarise all parts of articles, give important results and conclusions, be specific and avoid statements like " the findings are discussed".
- great care needed as abstract used for indexing, cross referencing and most readers only read abstract (and title).

Title :

- be snappy, eye catching and accurate to entice readers.

2. Revision :
   - the first draft is always too long, trim to no more than 2500-3000 words.
   - put aside paper for 1 to 2 weeks, then revise. Often get a better picture of things.

- get other people to revise the paper too, to detect ambiguity (not critical appraisal, this need to be sorted before paper even written). One cannot be objective with one's own writing.
- Check order of presentation : any sections can be rearranged to improve logical flow of paper.
- Check accuracy of figures, tables and results.
- finally, check style and grammar.

3. Style : manner of expression in language
   some elements of style (see Chp 15 How to write a paper, Thorne's medical writing).
   - keep it short : short word, phrase, sentence and article.
   - avoid figure of speech or idiom
   - use passive voice sparingly
   - avoid foreign, technical word or jargon
   - use abbreviations with acre, spell it out the first time it is used.
   - avoid tautology : saying the same thing in different words eg red in colour
   - avoid circumlocution : saying the same thing in a more complicated manner.
   - use verbs and nouns rather than adverbs and adjectives
   - preposition rather than string of nouns (noun salad) eg quality improvement programme cf programme to improve quality, longer but easier to understand.
   - keep verb close to noun (avoid dangler)

References :

1. *Archibald C. A busy physician's guide to medical writing.*
2. *Lock. Thorne's better medical writing . Pittman's medical 1977*
3. *Hall (ed). How to write a paper. BMJ publication1994*

Appendix :

A. How to calculate sample size using POWER program ?

B. How to prepare randomisation list ?

C. How to design data form (case report form)?

D. Some important guidelines for clinical trials :

1. Declaration of Helsinki : International guidelines for biomedical research involving human subjects

2. CPMP. Good clinical practice for trials on medicinal products in the European community. *Toxicology and Pharmacology 1990*

3. CPMP. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products *Statistics in Medicine 1995*

4. Begg C, Cho M, Eastwood S, Horton R et al. Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA 1996*

Appendix A : How to calculate sample size using POWER program ?

POWER is a public domain DOS based program for calculating sample size. Refer *Dupont WD, Plummer WD, Power and sample size calculations : a review and a computer program. Controlled clinical trials 1990;11:116-128* for details.

POWER program : overview
It does not calculate sample size required for one sample problem. You don't need this for sample size planning in RCT.
It does for the following situations :
1. Comparison of means for independent or dependent samples.
2. Comparison of proportions for independent or dependent samples.
3. Comparison of survival times for independent samples using the log rank test.
4. It handles other study designs like case-control study apart from cohort study/RCT.
5. It can also handle matched pair design like cross over trial.

Given any two of the following parameters, POWER can calculate the third :
1. power
2. sample size
3. detectable difference.

POWER program is not case sensitive, you can use either upper or lower case characters.

Installing POWER program :
simply copy the file a:\power.exe to a sub-directory of your choice in your hard disk

Starting, logging output and exiting POWER :
to start : cd\power <enter>
power <enter>

After starting POWER, the program asks for the name of a log file to save the output of your session. If none is specified, the default name is "power.log". This is an ASCII file, you can read it in DOS EDITOR or import it into any word processor and print thereof.

to exit anytime : type Ctrl-C, output however is not saved.
to exit and save output : type Ctrl-Z <enter> enough times.

Notations and commands :

General :

ALPHA	type I error probability, example 0.01 or 0.05. POWER always assume a two-sided test.
POWER	the desired power (1- type II error), example 0.8 or 0.9

Comparison of means for independent or dependent samples.

DELTA	clinically worthwhile difference in means to detect
S	standard deviation of response measure in the population, assume the same in each group.
M	the number of controls per treatment subject (case). For equal sized groups M is equal to one
N	the number of subjects in the treatment group, ie the required sample size

Comparison of proportions for independent or dependent samples.

P0	the proportion of subjects in the control group having outcome event
P1	the proportion of subjects in the treatment group (cases) having outcome event
R	relative risk of events between control and treatment groups
PSI	odds ratio of exposure between control and treatment groups, for case control study design
M	number of controls per treatment subject (case)
N	number of cases, the required sample size

Comparison of survival times for independent samples using the log rank test.

M1	median survival time in control group
M2	median survival time in treatment group
R	relative risk (hazard ratio), the risk of outcome event in control group relative to treatment group.
A	accrual time, time period over which subjects are enrolled into trial from the first to last subject.
F	follow up time, follow up time from end of accrual period to termination of trial.
N	number of subjects per group, the required sample size

Other commands :

E	to edit previously entered values
R	to obtain references
?	to get context sensitive help message

Appendix B : How to prepare randomisation list ?

The most common method used is Random permuted block method. This is shown here. For other methods, refer *Pocock. Methods of randomisation Chapter 5 in Clinical Trials : a practical approach Wiley 1985*.

There are computer programs to generate randomisation list which I don't have. It is easy to write your own program to do this so long you have access to any software that can generate random number (most statistical software can). However, for small sample size (less than 1000), it is easy to do manually as shown below :

1. Decide on a block size (n), typically 4 or 6 or 8.

2. Calculate the number of treatment arm (T) 'combinations' (strictly distinguishable permutations) for a block size of n :

   For example : for a 2 arms trial (treatment vs control) T= 2
   say we choose block size n=4

   then the number of 'combinations' are ${}_nC_T = 4C2 = 4x3 / 2x1 = 6$

3. Write down all 'combinations' :

   Example : for T=2 labelled A and B, n=4 and 4C2 = 6

   the 6 'combinations' are : AABB, ABAB, ABBA, BBAA, BABA, BAAB.

4. Assign a digit number 0 to 9 to each 'combination'.

   Example :

for digit	1	assign	AABB
"	2	"	ABAB
"	3	"	ABBA
"	4	"	BBAA
"	5	"	BABA
"	6	"	BAAB

5. Use a random number table (provided, see next page) to obtain a list of random numbers . Allocate the treatment according to the list.

Note :
With small block size like n=4 or 6 and non-double blind trial, an astute investigator who knows something about random permuted block method (like you do now) can always predict the next treatment assignment at the end of each block.

One could use large block size like n=12 or 20 but more tedious to do manually and careful if you need to do interim analysis, or use random block size like say 4, 8, 2,4,6,8,6,6,8, etc

Permutations of A and B in block size = n for random permuted block randomisation for trials with 2 treatment arms:

Number of permutations for block size n and for trial with x treatment arms are :
= n C n/x

for example : n=4, x=2, n/x=2 therefore 4C2 = 6
n=6, x=2, n/x=3 6C3 = 20
n=8 x=2, n/x=4 8C4 = 70

n = 4
# of permutations = 6

number				
1	B	A	B	A
2	A	B	A	B
3	A	A	B	B
4	A	B	B	A
5	B	B	A	A
6	B	A	A	B

n = 6
# of permutations = 20

number						
1	A	A	A	B	B	B
2	A	A	B	A	B	B
3	A	A	B	B	A	B
4	A	A	B	B	B	A
5	A	B	A	A	B	B
6	A	B	A	B	A	B
7	A	B	A	B	B	A
8	A	B	B	A	A	B
9	A	B	B	A	B	A
10	A	B	B	B	A	A
11	B	A	A	A	B	B
12	B	A	A	B	A	B
13	B	A	A	B	B	A
14	B	A	B	A	A	B
15	B	A	B	A	B	A
16	B	A	B	B	A	A
17	B	B	A	A	A	B
18	B	B	A	A	B	A
19	B	B	A	B	A	A
20	B	B	B	A	A	A

n = 8
# of permutations = 70

number								
1	A	A	A	A	B	B	B	B
2	A	A	A	B	A	B	B	B
3	A	A	A	B	B	A	B	B
4	A	A	A	B	B	B	A	B
5	A	A	A	B	B	B	B	A
6	A	A	B	A	A	B	B	B
7	A	A	B	A	B	A	B	B
8	A	A	B	A	B	B	A	B
9	A	A	B	A	B	B	B	A
10	A	A	B	B	A	A	B	B
11	A	A	B	B	A	B	A	B
12	A	A	B	B	A	B	B	A
13	A	A	B	B	B	A	A	B
14	A	A	B	B	B	A	B	A
15	A	A	B	B	B	B	A	A
16	A	B	A	A	A	B	B	B
17	A	B	A	A	B	A	B	B
18	A	B	A	A	B	B	A	B
19	A	B	A	A	B	B	B	A
20	A	B	A	B	A	A	B	B
21	A	B	A	B	A	B	A	B
22	A	B	A	B	A	B	B	A
23	A	B	A	B	B	A	A	B
24	A	B	A	B	B	A	B	A
25	A	B	A	B	B	B	A	A
26	A	B	B	A	A	A	B	B
27	A	B	B	A	A	B	A	B
28	A	B	B	A	A	B	B	A
29	A	B	B	A	B	A	A	B
30	A	B	B	A	B	A	B	A
31	A	B	B	A	B	B	A	A
32	A	B	B	B	A	A	A	B
33	A	B	B	B	A	A	B	A
34	A	B	B	B	A	B	A	A
35	A	B	B	B	B	A	A	A
36	B	A	A	A	A	B	B	B
37	B	A	A	A	B	A	B	B
38	B	A	A	A	B	B	A	B
39	B	A	A	A	B	B	B	A
40	B	A	A	B	A	A	B	B
41	B	A	A	B	A	B	A	B
42	B	A	A	B	A	B	B	A
43	B	A	A	B	B	A	A	B
44	B	A	A	B	B	A	B	A
45	B	A	A	B	B	B	A	A

46	B	A	B	A	A	A	B	B
47	B	A	B	A	A	B	A	B
48	B	A	B	A	A	B	B	A
49	B	A	B	A	B	A	A	B
50	B	A	B	A	B	A	B	A
51	B	A	B	A	B	B	A	A
52	B	A	B	B	A	A	A	B
53	B	A	B	B	A	A	B	A
54	B	A	B	B	A	B	A	A
55	B	A	B	B	B	A	A	A
56	B	B	A	A	A	A	B	B
57	B	B	A	A	A	B	A	B
58	B	B	A	A	A	B	B	A
59	B	B	A	A	B	A	A	B
60	B	B	A	A	B	A	B	A
61	B	B	A	A	B	B	A	A
62	B	B	A	B	A	A	A	B
63	B	B	A	B	A	A	B	A
64	B	B	A	B	A	B	A	A
65	B	B	A	B	B	A	A	A
66	B	B	B	A	A	A	A	B
67	B	B	B	A	A	A	B	A
68	B	B	B	A	A	B	A	A
69	B	B	B	A	B	A	A	A
70	B	B	B	B	A	A	A	A

Appendix C : How to design data form (case report form)?

Definition and uses of form
Planning and designing the data form

Data form or case report form : the record use for recording the data collected in a research though it has other uses.

1. as record for recording data collected
2. as integral part of the instrument : questionnaire, structured observation guide
3. as instruction sheet : instruction on questionnaire administration, performing observation or data abstraction; instruction on data conversion into categories
4. as coding key or frame and as code sheet : data form is pre-coded , then as coding key or data form has space for post coding , then code sheet

Planning and designing the data form :
General considerations :

1. How many forms ?
   - consider multiple forms if data collected at different places, different sources (primary vs secondary), diff collectors, diff time, diff categories of patients .

2. Consider the different users and uses of the form :
   a. The form filler :
      - subject himself : clear instruction, language and wording, etc
      - data collector : instruction on use, definition of terms or jargon used, judgmental conversion allowed only for expert data collector (eg clinician)

   b. The cataloguer : the person who does something to the form after it is completed.
      - coding : preferably pre-coded, otherwise leave space for post coding,
      - data entry clerk : ease of data entry, consistency in use of code.
      - editing

3. Always pre-test form and train user on use of form

Form design :

a. Content and individual items in the from.
b. Form construction : putting it all together.

a. Content and individual items in the from.
   - content : list all variables(each represented by one or more items) required, its definition, data expression, complexity (simple or composite).
   - **Refer to Annex of GCP guideline (see Appendix D) for a list of items required in data form for use in trial (case report form).**
   - Each item : an item comprises a stem statement or question, instruction on how to complete response and the response format.

- Stem : clear and unambiguous if necessary defined explicitly, wording or language preferably short, list better than prose, use 2 separate questions rather one complex conjunction, phrase such as to minimise amount of reading required, lower case easier to read.
- Instruction : always include, use different typeface eg *italic*
- Response : most critical part in determining ease of use of from.
- avoid judgmental conversion of data.
- for numeric response, record as is.
- for verbal response :
  - simple conversion into categories if possible (structured response format).
  - simple ticking is easiest
  - tick (boxes) what apply rather than delete what doesn't
  - use vertical response format if space allowed
  - pre-code all close ended question and use consistent coding
  - in the sequence of response, list common ones first or present some gradation
  - list all possible responses or allow for indefinite response (don't know) or contingency field (other,specify...)
  - response options may be mutually exclusive or multiple responses allowed
  - in wording response options use positive comparative term (eg better, bigger, more etc)
  - avoid reliance on small function word like preposition with critical effect on meaning (eg reduction *by* less than 50% cf smaller, 50-99% of original size.)
  - for verbal data that cannot be categorised eg name, address, too many categories, subtle nuances or much details needed, then verbatim recording (free format) of raw expression , provide enough space, avoid character segmentation of space, post code later.

b. Form construction : putting it all together

1. Overall layout :

Title
Preliminary : introduction , instruction
Identification : name, adddress, socio demographic, ID# : centre, who record and when data recorded
Body : sections or parts and sub-sections : each item : stem, instruction and response format
Ending

2. Items order :
- proceed from general to particular or specific about a topic
- within any topic, items should follow logical sequence.
- items or topics sequence may follow temporal sequence of data availability, order of observation made, structure of records for data abstraction.
- items on the same topic grouped together

- items may also be grouped based on uniform response format, then allow use of single instruction, matrix response format (row item stem and column response option)
- items grouped on who is the data collector
- items grouped on data abstraction from common place in the record for abstraction
- always number item (1,2,3...) and grouping of topic (section A, B, etc)

3. Form structure :
   - Introduction : include research organisation, sponsor.
   - Instruction on completing form
   - Linking phrases between topics, to introduce topic, to break monotony.
   - Skips or branches (where succeeding items are not applicable): skip instruction must be placed immediately after answer that leads to the branch point; or avoid by including 'not applicable' category in all items.
   - Allow space for comments
   - clear demarcation between sections/groups; use of line, box, white space

4. Form presentation :
   - Title , names of organisation
   - Numbering, print, layout, colour.

References :

Wright and Haybittle. Design of forms BMJ 1979

Appendix D : Some important guidelines for clinical trials :

1. Declaration of Helsinki : International guidelines for biomedical research involving human subjects

2. CPMP. Good clinical practice for trials on medicinal products in the European community. *Toxicology and Pharmacology 1990*

3. CPMP. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products *Statistics in Medicine 1995*

4. Begg C, Cho M, Eastwood S, Horton R et al. Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA 1996*