Society for Clinical Trials 31st Annual Meeting

Workshop P1
Essentials of Randomized Clinical Trials

Sunday, May 16, 2010
8:00 AM - 5:00 PM
Essex AB
WORKSHOP 1 - Essentials of Randomized Clinical Trials

1. Overall, did the subject context of this workshop meet your expectations and needs?  
   Yes ( )          No ( )  
   If yes, in what way? If no, why not?  ________________________________  
   ___________________________________________________________________

2. Was the content of this workshop of value to you personally or on the Job?  
   Yes ( )          No ( )  

3. Was the content of the workshop:  
   New ( )  New/Review ( )  Review ( )

4. The level and complexity of this workshop was:  
   Too elementary ( )  Correct ( )  Too advanced ( )

5. Rate the extent to which this workshop:  
   a. Presented content clearly  1  2  3  4  5  
   b. Allowed sufficient time for discussion and audience participation  1  2  3  4  5  
   c. Provided useful information  1  2  3  4  5  
   d. Utilized appropriate teaching methods, i.e., audiovisual, handouts, lectures  1  2  3  4  5

6. Please rate each workshop faculty member:

<table>
<thead>
<tr>
<th>Name</th>
<th>Knowledge of Subject</th>
<th>Organization/Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christopher S. Coffey</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Dixie Ecklund</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Marta M. Gilson</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Laura Lovato</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Michele Melia</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Yves Rosenberg</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>
1. Are you currently working in a clinical trial? (Yes) (No)

2. What is your job title? __________________________________________________________

3. Do you have any suggested topics for workshops at future meetings? If so, please list below:
   ____________________________________________________________________________
   ____________________________________________________________________________

4. What aspect of the workshop did you like best?
   ____________________________________________________________________________
   ____________________________________________________________________________

5. What aspect of the workshop would you change if this workshop were offered again?
   ____________________________________________________________________________
   ____________________________________________________________________________

6. Additional Comments: _________________________________________________________
   ____________________________________________________________________________
Part I: Introduction

Yves Rosenberg, M.D, M.P.H.
Program Director, Acting Branch Chief
Atherothrombosis and Coronary Artery Disease Branch
Division of Cardiovascular Sciences
National Heart, Lung and Blood Institute, NIH
6701 Rockledge Dr., Rm. 8148
Bethesda, MD 20892-7956
Tel.: 301-435-1292
Fax: 301-480-3667
E-mail:rosenbey@nih.gov

SCT Pre-Conference Workshop - Baltimore May 16, 2010
Essentials of Randomized Clinical Trials

Introduction to Randomized Clinical Trials

Outline I

• Historical perspective
• Rationale for randomized clinical trials
  – Rationale for randomization
  – The equipoise issue
  – To blind or not to blind?
• Key issues in the design of a RCT:
  – What is the study question? Defining hypothesis, objectives and end-points
  – Defining selection criteria: generalizability vs. homogeneity
  – Selecting the control group: the placebo vs. “usual care” issue

Outline II

• The different phases of a RCT
• Basic RCT Designs
  – Parallel, cross-over, factorial and cluster designs
  – Large Simple Trials
  – Comparative Effectiveness trials
  – Superiority, Equivalence and Non-Inferiority trials
• Key elements of a RCT Protocol
• Some ethical considerations
  – Informed Consent Process
  – Patient safety issues
Historical perspective

Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink. Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the King's meat; and as thou seest, deal with thy servants. So he consented to them in this matter, and proved them ten days. And at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the King's meat.

Book of Daniel, Chapter 1, Verses 12 -15

Historical perspective

I raised myself very early to visit them when beyond my hope I found those to whom I had applied the digestive medicament, feeling but little pain, their wounds neither swollen nor inflamed, and having slept through the night. The others to whom I had applied the boiling oil were feverish with much pain and swelling about their wounds. Then I determined never again to burn thus so cruelly the poor wounded by arquebuses.

Ambroise Paré (1510 – 1590)

Historical perspective

Lind's Scurvy Study

Nb of Patients: 12
Test Treatments:
- Cyder, 1qt/day
- Elixir vitriol, 25 gutts, 3 times/day
- Vinegar, 2 tsp, 3 times/day
- Bigness of nutmeg 3 times/day
- orange (2) ; lemon (1) /day
Control Treatment
- Sea-water, ½ pt/day
Follow-up: 6 days
Outcome: fit for duty

Lind's Treaty on Scurvy, 1753
Historical perspective
Key Dates in the History of RCT

• 1747 Lind’s Scurvy experiment
• 1800 Waterhouse’s smallpox experiments
• 1863 Gull’s use of Placebo Treatment
• 1923 Fisher’s 1st application of randomization
• 1931 1st use of randomization (and blindness) in a clinical trial
• 1946 Nuremberg Code for Human Experimentation
• 1962 Hill AB Statistical Methods of Clinical and Preventive Medicine
• 1979 Society for Clinical Trials
• 2006 Clinical and Translational Science Awards (CTSA) program
• 2009: The Recovery Act (ARRA) provides $1.1 billion for Comparative Effectiveness Research.

Introduction to Randomized Clinical Trials
Outline I

• Historical perspective
• Rationale for randomized clinical trials
  – Rationale for randomization
  – The equipoise issue
  – To blind or not to blind?
• Key issues in the design of a RCT:
  – What is the study question? Defining hypothesis, objectives and end-points
  – Defining selection criteria: generalizability vs. homogeneity
  – Selecting the control group: the placebo vs. “usual care” issue

Randomized Clinical Trials
Some Terminology

• Clinical Trial:
  – An experiment testing medical (e.g. drug, surgical procedure, device or diagnostic test) treatments on human subjects
    • Experiment: a series of observations made under conditions controlled by the scientist
    • Prospective (≠ case-control study)
    • Comparative (≠ cohort study)
    • Involves human subjects
  – A research activity that involves administration of a “test treatment” to some “experimental unit” in order to evaluate that treatment
**Randomized Clinical Trials**

Some More Terminology

- **Randomization**: the process of assigning patients to treatment using a random process (such as a table of random numbers)

- **Randomized controlled clinical trial (or randomized clinical trial-RCT)**:
  - Clinical trial with at least one control treatment and one test treatment
  - In which the treatment administered are selected by a random process

---

**Randomized Clinical Trials**

Why Randomize?

“The goal of randomization is to produce comparable groups in terms of general participant characteristics, such as age or gender, and other key factors that affect the probable course the disease would take. In this way, the two groups are as similar as possible at the start of the study. At the end of the study, if one group has a better outcome than the other, the investigators will be able to conclude with some confidence that one intervention is better than the other.”

*Friedman et al. Fundamental of Clinical Trials, Mosby Press*

---

**Randomized Clinical Trials**

Why Randomize?

- To find out which (if any) of two or more interventions is more effective
- Produce comparable groups
  - Protect against both known and unknown/unmeasured confounders (prognostic factors)
  - Eliminate treatment selection bias
- Best to establish causality
- Can define “Time zero”

---
Randomized Clinical Trials
Why Randomize?

- Necessary to detect small but clinically important treatment differences
- Protect against possible time trends in:
  - Patient population and disease characteristics
  - Diagnostic methods and supportive care
- Provides a valid basis for statistical tests of significance

Randomized Clinical Trials
Why Randomize: The Hormone Replacement Therapy Story

Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the nurses' health study

METHODS. We followed 48,470 postmenopausal women, 30 to 63 years old... During up to 10 years of follow-up (337,854 person-years), we documented 224 strokes, 405 cases of major coronary disease (nonfatal myocardial infarctions or deaths from coronary causes), and 1263 deaths from all causes.

RESULTS. After adjustment for age and other risk factors, the overall relative risk of major coronary disease in women currently taking estrogen was 0.56 (95 percent confidence interval, 0.40 to 0.80...)

CONCLUSIONS. Current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease, but it is not associated with any change in the risk of stroke.

Randomized Clinical Trials
ESTROGEN FOREVER?
The prevailing medical view is that most should stay on estrogen for the long haul... At the turn of the century, women died soon after their ovaries quit. Now they live to face heart disease, osteoporosis, increased fractures... There may be other risks and other advantages of HRT, but what doctors know is limited by the type of research that has been done. Instead of setting up a group of women on HRT and a carefully matched control group that does not take hormones, studies like the Nurses trial simply look at populations of women who made their own choice whether to take estrogen. The problem with this... is that women who take hormones go to doctors more, eat well, exercise and are in better health generally than women who don't take hormones. Thus it is hard to tell whether their lower rates of heart disease or colon cancer or fractures reflect HRT or these other healthy habits.
Design. Estrogen plus progestin component of the Women's Health Initiative, a randomized controlled primary prevention trial (planned duration, 8.5 years) in which 16608 postmenopausal women aged 50-79 years with an intact uterus at baseline were recruited by 40 US clinical centers in 1993-1998.

Interventions. Participants received conjugated equine estrogens, 0.625 mg/d, plus medroxyprogesterone acetate, 2.5 mg/d, in 1 tablet (n = 8506) or placebo (n = 8102).

Main Outcomes Measures. The primary outcome was coronary heart disease (CHD) (nonfatal myocardial infarction and CHD death), with invasive breast cancer as the primary adverse outcome. A global index summarizing the balance of risks and benefits included the 2 primary outcomes plus stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.

Conclusions. Overall health risks exceeded benefits from use of combined estrogen plus progestin for an average 5.2-year follow-up among healthy postmenopausal US women. All-cause mortality was not affected during the trial. The risk-benefit profile found in this trial is not consistent with the requirements for a viable intervention for primary prevention of chronic diseases, and the results indicate that this regimen should not be initiated or continued for primary prevention of CHD.
Randomized Clinical Trials
When Randomize?

• Finding “window of opportunity”
  – Too early
    • Not enough “preliminary” evidence: biological plausibility, epidemiologic studies
    • Intervention not “mature” enough (e.g., surgical technique)
  – Too late: intervention already established in clinical practice
• Clinical Equipoise
• Changing Clinical Practice Guidelines

Randomized Clinical Trials
To Blind or not to Blind?

• Definition: concealment (masking) to the patient (single blind), investigator (double) and the monitors (triple) of the identity of the intervention.
  (Opposite = unblinded or open trial)
• Goal: avoid bias (systematic error= anything that does not occur by chance!)
The more subjective the intervention, the more important the blinding!
Bias can occur at any stage of the study: patient assignment, data collection, event ascertainment...

Randomized Clinical Trials
To Blind or not to Blind?

• Unblinded trial
  – May be the only option: strategies of treatment (drug vs. surgery) behavioral interventions...
  – “True” blinding may be hard: expected biological effect of intervention
  – Easier to carry out and less expensive but...
    Risk of bias generally outweigh benefits!
• Alternative to blinding intervention (if not possible): blind outcome assessment
Introduction to Randomized Clinical Trials

Outline I

• Historical perspective
• Rationale for randomized clinical trials
  – Rationale for randomization
  – The equipoise issue
  – To blind or not to blind?
• Key issues in the design of a RCT:
  – What is the study question? Defining hypothesis, objectives and end-points
  – Defining selection criteria: generalizability vs. homogeneity
  – Selecting the control group: the placebo vs. “usual care” issue

Elements of a RCT

What is the Study Question (Who-What-When)?

• Primary question tests the hypothesis
• Hypothesis must include:
  – Population studied
  – Primary outcome of interest
  – Intervention studied
  – Period of observation
• Objective: phrase the research question in concise, quantitative terms

Elements of a RCT

Primary and Secondary Objectives

• Primary objective needs to be defined (determine sample size)
• Secondary objective needs to be:
  – Defined a priori (avoid post hoc “fishing expedition”)
  – Chosen parsimoniously (avoid false positive)
• Primary vs. secondary:
  – Question of greatest interest/relevance
  – Consider feasibility (e.g. mortality vs. morbidity)
Elements of a RCT

The Endpoints

• Quantitative measurement required by the objectives (= outcome, response variable)
• Event/condition the trial is designed to ameliorate, delay, prevent…
• Primary endpoint: need to be clearly and rigorously defined (what is survival?)
• Endpoints defined by type of measurement used:
  – Discrete, dichotomous (dead or alive?), count
  – Continuous (BP change), ordered (toxicity)

Elements of a RCT

The Endpoints: what makes a good Primary Endpoint?

• Must answer the primary question (Co-primary?)
• Frequency of occurrence must be known in control (determine sample size)
• Must be able to estimate treatment effect: clinical relevance (minimum desired effect to change practice?)
• Must be assessed/evaluable in all participants
• Can be measured accurately/reliably/objectively
  – Blinded randomization
  – Blinded assessment (soft end point?)
• All patients must be evaluated (no post randomization exclusion/no lost to follow up)

Elements of a RCT

Other Types of Endpoints

• Intermediate and surrogate
• Combined
• See Part V
Elements of a RCT
Defining the Study Population

- Subset of population with disease/condition of interest
- Patients enrolled = subset of study population defined by the eligibility criteria
- Inclusion criteria: Define “at risk” population
  - Less inclusive (= more homogeneous population): potential for benefit increase
    - but need to understand mechanism of action of intervention
    - Cannot generalize to other “subgroups”
  - More inclusive (= more heterogeneous population):
    - Increase generalizability
    - But may dilute effect of intervention (increase sample size)
  - Select group more likely to benefit from intervention
    - Higher risk: increase number of events, decrease sample size
    - But: are results applicable to lower risk?

- Exclusion criteria:
  - Insure patient safety (risk/benefit in specific subgroups)
  - Assess competitive risk
  - Assess likelihood of adherence to protocol and intervention

Eligibility criteria will be defined by goal of trial: efficacy vs. effectiveness trial?

---

Homogeneity
- Divergent subgroup of patients (i.e., “atypical” patients) can distort findings for the majority
- Restriction of population reduces “noise” and allows study to be done in a smaller sample size
- Restrict population to homogenous group

Generalizability
- At the end of the study, it will be important to apply findings to the broad population of patients with the disease
- It is questionable to generalize the findings to those excluded from the study
- Have broad inclusion criteria “welcoming” all

From: Virginia Howard
### Elements of a RCT

#### Defining the Study Population: Efficacy vs. Effectiveness trial

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Efficacy Trial</th>
<th>Effectiveness Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>Test biological question</td>
<td>Assess &quot;real life&quot; effect of intervention</td>
</tr>
<tr>
<td>No participants</td>
<td>&lt; 1,000</td>
<td>&gt; 10,000</td>
</tr>
<tr>
<td>Cohort</td>
<td>Homogeneous</td>
<td>Heterogeneous</td>
</tr>
<tr>
<td>Data collection</td>
<td>Extensive</td>
<td>Limited</td>
</tr>
<tr>
<td>Focus of inference</td>
<td>Internal validity</td>
<td>Generalizability</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>Strict</td>
<td>Broad</td>
</tr>
</tbody>
</table>


### Elements of a RCT

#### Choosing an Effectiveness Design

- **Define the question**: What is the purpose of the trial?  
  - Does the intervention work when applied in usual practice?
- **Define the setting**: under which conditions will the trial results be applicable?  
  - Ideal setting vs. normal practice?
- **How are participants selected?**  
  - Eligibility criteria mostly defined by the condition of interest
- **Outcomes of interest?**  
  - Direct relevance to practice  
  - Will influence clinical decisions and health policy decisions

### Randomized Clinical Trials

#### Nature of “Intervention”

- Drug (or drug regimen)
- Surgical procedure
- Medical device
- Therapeutic modality (radiation, biologic therapy, etc)
- Diet
- Behavioral intervention (education)
- Clinical approach to diagnosis, treatment, symptom management, palliative care, etc. (e.g. strategy)

The common denominator: there is a choice between two alternative approaches; uncertain which is preferable (e.g. equipoise)
Randomized Clinical Trials
Selecting the Control Group

- Four different types:
  - Placebo
  - No Treatment
  - Different doses or regimens of the treatment
  - Different active treatment (including usual care)

- Control group will be classified based on:
  - Type of treatment used
  - Method of assignment in control group
  - May be more than one control!

Randomized Clinical Trials
Selecting the Control Group: The Placebo Issue

- Definitions
  1. Clinical: "A substance having no pharmacological effect but given merely to satisfy a patient who supposes it to be a medicine"
     Goal: to distinguish pharmacological effects from the effects of suggestion
  2. Research: "A substance having no pharmacological effect but administered as a control in testing experimentally or clinically the efficacy of a biologically active preparation."
     Goal: to obtain an unbiased assessment of the result of an experiment

Randomized Clinical Trials
Placebo Control: Scientific Justification

- Minimize subject and investigator bias (when used with randomization and blinding)
- Maximize likelihood of establishing efficacy: encourage optimal conduct of the trial: decrease "incentive" for poor trial conduct (drop-outs, cross-overs, etc)
- Enable distinction between adverse effects of drug/intervention and disease
- Allow for measurement of true effect size: account for the "placebo effect"
Randomized Clinical Trials
The Active Control

- Positive control: new therapy compared to known active therapy (randomized, can be blinded)
  - Goal: effectiveness or non-inferiority
  - Based on assumption that previous treatment shown to be effective! (external validation needed)
- Challenges:
  - Effect size and safety assessment more difficult
  - Larger sample size
  - Many possible bias: non adherence, concomitant therapies, randomization of inappropriate patients

Randomized Clinical Trials
Usual Medical Care as Control Group

- State of equipoise: is there a “standard of care”?
- Potential advantages:
  - Increase relevance
  - Increase external validity
  - Increase practicality
- Interpretation of evidence:
  - Is usual care validated by research? Is there a consensus on what is “usual care”?
  - Adherence to guidelines/evidence-based care?

Introduction to Randomized Clinical Trials
Outline II

- The different phases of a RCT
- Basic RCT Designs
  - Parallel, factorial, cluster and cross-over designs
  - Large Simple Trials
  - Comparative effectiveness trials
  - Superiority, Equivalence and Non-Inferiority trials
- Key elements of a RCT Protocol
- Some ethical considerations
  - Informed Consent Process
  - Patient safety issues
### The Different Clinical Trial Phases

#### Phase I
- First in humans
- Small, uncontrolled
- Healthy volunteers/failed conventional therapy
- Dose-escalation protocols
- Tolerability/toxicity study: Maximum Tolerated Dose (MTD)
- Dose-response models

#### Phase II
- Test biologic activity/effect
- Estimate rates of adverse events
- Performed in patients with disease/condition of interest
- With or without comparison group
- Strict eligibility criteria

#### Phase III
- Determine the effectiveness (overall benefit/risk-cost assessment) of new therapies relative to standard therapy
- Large sample size
- Multicenter
- Superiority, equality, equivalence or non-inferiority
The Different Clinical Trial Phases
Phase IV
• Long term surveillance studies ("post marketing") for safety
• New dosing regimens/indications
• Look for rare side effects
• Often non randomized

Introduction to Randomized Clinical Trials
Outline II
• The different phases of a RCT
• Basic RCT Designs
  – Parallel, cross-over, factorial and cluster designs
  – Large Simple Trials
  – Comparative Effectiveness Trials
  – Superiority, Equivalence and Non-Inferiority trials
• Key elements of a RCT Protocol
• Some ethical considerations
  – Informed Consent Process
  – Patient safety issues

Basic RCT Designs
Parallel Design
FREEDOM Design
Future REvascularization Evaluation in patients with Diabetes mellitus: Optimal management of Multivessel disease
Eligibility: DM patients with MV-CAD eligible for stent or surgery
Exclude: Patients with acute STEMI, cardiogenic shock
Randomized 1:1
MV-stenting
With Drug-eluting stents
And abciximab
CABG
With or without CPB
Basic RCT Designs

Cross Over Design

- Participant = own control
- Randomize: order of treatment for each patient (e.g. AB vs. BA)
- Advantages
  - Reduce variability
  - Detect difference in response in individual patient
- Disadvantages
  - Order of treatment should not matter
  - No carry over of effect

Advantages
- Reduce variability
- Detect difference in response in individual patient

Disadvantages
- Order of treatment should not matter
- No carry over of effect

Analysis of a 2 x 2 factorial RCT

Effect of A: ac vs. bd *
Effect of B: ab vs. cd *

*If no treatment interaction

Basic RCT Designs

Factorial Design

Intervention A

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Intervention B

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cells

a= Active A + Active B
b= Control A + Active B
c= Active A + Control B
d= Control A + Control B

Analysis of a 2 x 2 factorial RCT

Effect of A: ac vs. bd *
Effect of B: ab vs. cd *

*If no treatment interaction

Basic RCT Designs

Factorial Design

BP

<table>
<thead>
<tr>
<th></th>
<th>Intensive (SBP=120)</th>
<th>Standard (SBP=140)</th>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensive Glycemic Treatment (A1C=6%)</td>
<td>1178</td>
<td>1193</td>
<td>1383</td>
<td>1374</td>
</tr>
<tr>
<td>Standard Glycemic Treatment (A1C 7-7.9%)</td>
<td>1184</td>
<td>1178</td>
<td>1370</td>
<td>1391</td>
</tr>
</tbody>
</table>

Lipid

<table>
<thead>
<tr>
<th></th>
<th>Intensive (SBP=120)</th>
<th>Standard (SBP=140)</th>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensive (Cholesterol)</td>
<td>2362*</td>
<td>2371*</td>
<td>2765*</td>
<td>2765*</td>
</tr>
</tbody>
</table>

*Primary analyses compare marginals for main effects

ACCORD (Action to Control Cardiovascular Risk in Diabetes)

(ACCORD Study Group, Am J Cardiol 2007;99[suppl]:21i-33i)
Basic RCT Designs
Factorial Design

- Advantage:
  - Two trials for (almost) the price of one
  - Design is best if two intervention have different mechanisms of actions or different outcomes (e.g. cancer for A and CV disease for B)

- Disadvantages:
  - Need to test for possibility of interaction (e.g. A differs based on the presence or absence of B)
  - Test for interaction not very powerful
  - Need to consider gain in cost vs. increased complexity, recruitment and adherence issues and potential for adverse events

Basic RCT Designs
Cluster Design

- Cluster design = group randomization
- Group = schools, clinics, villages...
- Sample size: based on number of groups (not individuals)
  - Need to be adjusted by factor Nm (where N = number of cluster each of size m)
  - Need to take into account within-cluster correlation of response (correlation = loss of efficiency)

- Analysis:
  - Cannot use classic statistical methods (correlation)
  - Random effect model
  - Use sensitivity analyses

Basic RCT Designs
Cluster Design: The Public Access Defibrillation (PAD) Trial

RESUSCITATION. 2003 Feb;56(2):135-47
Basic RCT Designs
Large Simple Trials

- Provide a more reliable estimate of the effect of intervention
- Needed to uncover smaller treatment effects
  *That are important in common conditions*
- Increase generalizability
  *But limit data collection/subgroups and secondary analyses*
- Decrease cost by simplifying design and management
  *But need strong randomization procedures and reliable outcomes assessment*

Basic RCT Designs
Large Simple Trials

Comparative Effectiveness Trials

- A type of health care research that compares the results of one approach for managing a disease to the results of other approaches.
- Comparative effectiveness usually compares two or more types of treatment, such as different drugs, for the same disease. Comparative effectiveness also can compare types of surgery or other kinds of medical procedures and tests.
- Comparative effectiveness research is designed to inform health care decisions by providing evidence on the effectiveness, benefits, and harms of different treatment options. The evidence is generated from research studies that compare drugs, medical devices, tests, surgeries, or ways to deliver health care.
Basic RCT Designs

Comparative Effectiveness Trials

Superiority, Non-Inferiority and Equivalence Trials

- Superiority trial
  - Is new intervention better than no (placebo) intervention or standard intervention?
  - Goal: Demonstrate a difference!

- Non-inferiority trial
  - New intervention not worse than standard? (not less effective, but safer, cheaper, etc.)
  - Goal: Demonstrate that new intervention is not worse than the standard by a prespecified \( \Delta \) (minimum clinically significant difference)

- Equivalence trial
  - Are the effects of the two interventions very similar?
  - Goal: Demonstrate that the two interventions are not different by more than the prespecified \( \Delta \)

Introduction to Randomized Clinical Trials

Outline II

- The different phases of a RCT
- Basic RCT Designs
  - Parallel, cross-over, factorial and cluster designs
  - Large Simple Trials
  - Comparative Effectiveness Trials
  - Superiority, Equivalence and Non-Inferiority trials
- Key elements of a RCT Protocol
- Some ethical considerations
  - Informed Consent Process
  - Patient safety issues
Key elements of a RCT Protocol
Study Design: Preliminary Considerations

- Demonstrate need for trial
- Establish study objectives
- Choose best approach to problem/question
  - Small vs. large?
  - Less is more!
- Objectives ≠ study goals
  - Objectives: statement about question to answer
  - Goals: what you need to achieve to answer the question

Key elements of a RCT Protocol
Study Design: Framing the Question

- Toxicity? Efficacy? Effectiveness?
- Feasibility
- Proof of concept
- Pilot study

Why? How? Importance?

Outcome

Key elements of a RCT Protocol
Study Design: Key Steps to Follow

1. Establish study objectives
2. Choose basic study design
3. Determine primary and secondary outcomes
4. Choose type of control
5. Determine need/feasibility of blinding
6. Choose randomization procedure
7. Sample size and power
8. Determine screening, baseline, treatment and follow-up periods
9. Choose patient population
10. Establish treatment modalities
Elements of a RCT
Protocol: Table of contents (I/IV)

- Abstract
- I. STUDY HYPOTHESIS
- II. INTRODUCTION AND BACKGROUND
- III. OBJECTIVES OF THE STUDY
  - A. Primary objective
  - B. Secondary objective
- IV. STUDY ENDPOINTS
  - A. Primary Endpoint
  - B. Secondary Endpoints
- V. STUDY DESIGN
- VI. PATIENT SELECTION
  - A. Inclusion criteria
  - B. Exclusion Criteria
- VII. INFORMED CONSENT PROCEDURE

Elements of a RCT
Protocol: Table of contents (II/IV)

- VIII. RANDOMIZATION PROCEDURE
- IX. ADMINISTRATION OF STUDY DRUG
- X. DATA MANAGEMENT, QUALITY ASSURANCE & MONITORING PROCEDURES
  - A. Data collection and management
  - B. Monitoring reports
    - 1. Executive Committee
    - 2. Steering Committee
    - 3. Data and Safety Monitoring Board
  - C. Quality Assurance
- XI. STATISTICAL ANALYSES
  - A. Primary endpoint
  - B. Sample size and power
  - C. Subgroup and secondary analyses
  - D. Interim analyses

Elements of a RCT
Protocol: Table of contents (III/IV)

- XII. STUDY ORGANIZATION
  - A. National Heart, Lung, and Blood Institute
  - B. Steering Committee
  - C. Clinical Trial Center
  - D. Data and Safety Monitoring Board
- XIII. SUBSTUDIES AND ANCILLARY STUDIES
  - A. Introduction
  - B. Ancillary studies
  - C. Databank studies
  - D. Application review process
  - E. Data storage and analysis
Elements of a RCT
Protocol: Table of contents (IV/IV)

XIV. PUBLICATION POLICY
  A. Data analysis and release of results
  B. Review process
  C. Primary outcome papers, abstracts and presentations

XV. CLOSEOUT PROCEDURES
  A. Interim
  B. Reporting of Study Results

XVI. REFERENCES

Appendices
- Mode Informed Consent
- Conflict of Interest Policies

Introduction to Randomized Clinical Trials
Outline II

The different phases of a RCT
Basic RCT Designs
  Parallel, cross-over, factorial and cluster designs
  Large Simple Trials
  Comparative Effectiveness Trials
  Superiority, Equivalence and Non-Inferiority trials
Key elements of a RCT Protocol
Some ethical considerations
  Informed Consent Process
  Patient safety issues

Ethical Issues
Specific to Clinical Trials

Special ethical concerns because treatment is determined by chance
The arms of the clinical trial must be in clinical equipoise
Principle of non maleficence, withholding proven treatment from control group
Periodic analysis of interim data by independent Data and Safety Monitoring Board
Some Ethical Considerations

**Informed Consent Process**

- Purpose of the trial
- Nature of the trial
- Procedures of the trial
- Risks and potential benefits and alternatives to participating
- Procedures to maintain confidentiality
- Assurances and contact information

**Informed Consent Issues**

- Withdrawal
  - Participant is free to withdraw at any time
- New findings
  - Obligation to tell participant of any significant new findings that may affect his/her willingness to continue
- Potential for coercion

**Health Information Portability and Accountability Act (HIPAA)**

- Research subjects must sign an authorization form that describes the use and disclosure of their protected health information (PHI) for research purposes
- HIPAA authorization wording may be part of informed consent document or a separate form
- Subject must be given signed copy of form with HIPAA disclosure information
Some Ethical Considerations
Where to Go for More Info

- Human Subjects Research Protection
  - http://www.hhs.gov/ohrp/
- Registry of clinical Trials and Background:
  - http://clinicaltrials.gov/
- Regulations and Ethical Guidelines:
  - 45 CFR 46 Protection Of Human Subjects
  - Guidelines for Conduct of Research Involving Human Subjects at NIH (Gray Booklet) (pdf file)
  - The Belmont Report Ethical Principles and Guidelines for the Protection of Human Subjects of Research
  - Nuremberg Code Directives for Human Experimentation
  - World Medical Association Declaration Of Helsinki

Randomized Clinical Trials
Some Key Points

- Important
  - in evaluating interventions for the prevention, diagnosis, and treatment of disease
  - Important to obtain unbiased comparisons of interventions
- Ethical
  - in the presence of uncertainty (equipoise)
  - present the best choice of therapeutic options to the patients
- Robust
  - large trials recommended to increase reliability
- Applicable to studies of efficacy and of effectiveness
- Can answer more than one question at a time (factorial trials and other designs)
- In some situations, can randomize entire groups (e.g., communities, medical practices)

Randomized Clinical Trials
Some Key References

- Fundamental of Clinical Trials. Lawrence M Friedman, Curt D Furberg, David L DeMets. Springer Verlag Editors
Project Management in Clinical Trials

- Requirements for Clinical Trials vary widely which drives the Project Management model

- Big Pharma clinical trials
  - Initiated, developed, and managed by Industry Sponsor (e.g., Merck)
  - Data Coordinating Center (DCC)
  - Clinical Coordinating Center (CCC)
  - Statistical Coordinating Center (SCC)
  - Participating Clinical Centers (PCC)
  - Sponsor

- Industry or Federally-funded clinical trials
  - Initiated and developed by Industry Sponsor or NIH-funded Principal Investigators
  - Managed by Contract Research Organization (CRO) or Academic CRO
  - Data Coordinating Center (DCC)
  - Clinical Coordinating Center (CCC)
  - Statistical Coordinating Center (SCC)
  - Participating Clinical Centers (PCC)
  - Sponsor
Project Management in Clinical Trials

- Requirements for Clinical Trials vary widely which drives the Project Management model
- Federally-funded clinical trials
  - R01 Grants
  - Managed by Academic CRO or Traditional Data Coordinating Center (DCC)
- Data Coordinating Center
- Statistical Coordinating Center
- Interact with CCC and Sponsor
- Important to define who is doing what

Project Management in Clinical Trials

Which model of Data Coordinating Center?
DCC Requirements to Successfully Manage Multi-Site Clinical Trials

- Build Good Teams
- Phase I: Grant/Protocol Development
- Phase II: Implementation
- Phase III: Up and Running
  - Study Start-Up Activities
- Phase IV: Ongoing Activities
  - Study Continuation
- Phase V: Study Close-Out

Bring Together a Good Team

- Data Coordinating Center (DCC) Teams
  - Biostatistics
    - Protocol Development
    - Statistical Analysis Plans
    - Report Generation
    - Interim Analysis
    - Final Analysis

Bring Together a Good Team

- Data Coordinating Center (DCC) Teams
  - Protocol Coordinators
    - Clinical Coordinators
    - Manage sites
    - Manage and resolve data queries
    - Develop study materials
    - Maintain study supplies
Bring Together a Good Team

- Data Coordinating Center (DCC) Teams
  - **Data Managers**
    - Technical coordinators
    - Develop user’s specifications for data entry systems
    - Develop testing plans for data entry systems
    - Validate data entry systems
    - Documentation of validation

Bring Together a Good Team

- Data Coordinating Center (DCC) Teams
  - **Information Technology (IT) Developers**
    - Develop Web Applications
    - Develop Data Entry Applications
    - Data Storage
    - Data Back-up and Recovery
    - 21 CFR Part 11 Compliance

Bring Together a Good Team

- Data Coordinating Center (DCC) Teams
  - **Regulatory**
    - Responsible for Trial Master File
    - Monitor Site Regulatory Binders
    - IND Safety Reports
    - MedDRA Coding
    - FDA Submissions
Bring Together a Good Team

• Data Coordinating Center (DCC) Teams
  • **Fiscal/Administrative**
    • Develop grant budgets
    • Monitor expenditures
    • Human Resource functions
    • Coordinate meeting and travel arrangements

• Data Coordinating Center (DCC) Teams
  • **Medical Monitors**
    • MedDRA coding
    • Medical writing
    • Aggregate review of Adverse Events
    • Individual review of Serious Adverse Events

• Data Coordinating Center (DCC) Teams
  • **Quality Management**
    • Backbone of all processes
    • Develop and monitor SOPs
    • Standardize training/education
    • Develop center-wide metrics to monitor quality
    • Develop study-specific metrics to monitor quality
Bring Together a Good Team

- Clinical Coordinating Center (CCC) Team
  - Lead Principal Investigator (PI)
  - Lead Study Coordinator
  - Support Staff
- Participating Clinical Center (PCC) Teams
  - Site PI
  - Site Study Coordinator
  - Support Staff
- Sponsor
  - NIH
  - Foundations
  - Industry

Bring Together a Good Team

- Is there a need for additional subcontractors?
  - Drug distributor
  - Specimen kit distributor
  - Central laboratories
  - Onsite monitoring
  - Recruitment support
  - DSMB Support

Glue the teams together

- Written Standard Operating Procedures (SOPs)
- Written Study-Specific Project Work Instructions (PWIs)
- Training and education programs
  - Cross-train whenever possible
- Quality Management initiatives
Develop a Study Plan

• Outline areas of oversight
• Delineate areas of responsibility
• Form Steering Committee
  • Voting and non-voting members
  • Unblinded biostatistician
• Identify and empanel the DSMB
• Identify need for FDA meetings

Develop a Study Plan

• Structure Committees

Develop a Study Plan

• Structure Committees
  • Steering Committee
  • Publications
  • Ancillary studies
    • Mechanistic
    • Quality of Life
  • Clinical Events/Adjudication
Phase I: Development

- **Protocol Development**
  - PIs for scientific and medical input
  - Biostatisticians for design and analysis input
  - Study Coordinators for practical input
  - Medical Writer to help with readability
Phase I: Development

- **Study Materials Development**
  - Investigator Brochure (IB) for IND/IDE studies
  - Manual of Procedures (MOP)
  - Laboratory Manuals
  - Informed Consent Templates
  - Web and Data System User’s Guides
  - Adverse Event System User’s Guides

- **Select Qualified Investigators**
  - Search FDA warning letters for debarred investigators
  - Appropriate clinical expertise
  - Adequate staff to perform studies
  - Adequate facilities to perform studies
  - Pool of eligible subjects
  - Conflict of Interest Disclosures

- **Select Qualified Subcontractors**
  - **On-site Monitoring**
    - Performed by DCC or Contract Research Organization (CRO)
    - Qualified Data Auditors
    - Qualified CRAs
    - Adequate personnel to meet the requirements of the monitoring plan
**Phase I: Development**

- **Select Qualified Subcontractors**
  - *Specimen kit assembly and distribution*
    - Configure specimen kits
    - Supplies on-hand to manufacture kits
    - Ability to meet deadlines to manufacture specimen kits
    - Ability to distribute specimen kits and shipping supplies
    - Ability to collaborate with Specimen Tracking System (if in place)

- **Central laboratories**
  - Appropriate certifications
  - Ability to handle throughput
  - Specialty laboratory specific to research question
  - Able to provide results to Clinical Centers
  - Able to provide results via data transfer to DCC
  - QC processes in place

- **Drug distributor**
  - Receive Investigational Products from Manufacturers
  - Receive approved drugs obtained through Clinical Trial Agreements or purchase
  - Label Investigational Products
  - Appropriate storage facilities
  - Appropriate inventory support
  - Appropriate distribution processes
  - Ability to ship out of country if needed
  - Ability to accept returned products
  - Ability to destroy expired or returned products
Phase I: Development

- FDA submission for IND/IDE approval
  - Work with Regulatory Team and Sponsor
  - Assistance from CTSA staff may be available
- Investigator Meeting
  - Protocol finalization
  - Procedural discussions
  - Recruitment goals
- Coordinator Training
  - May be held in conjunction with Investigator Meeting

Phase I: Development

- Develop Source Documents
- Develop Case Report Forms
- Finalize Protocol
- Finalize Case Report Forms
  - And then finalize them again

Phase II: Implementation

- Develop and validate data entry system
  - Data Management Plan
  - User’s specifications
  - Testing plans
  - Validation documentation
Phase II: Implementation

- Configure specimen kits
  - Must be user friendly
  - Consider specimen tracking system

- Package and label investigational products

Phase II: Implementation

- Work with sites on IRB approvals

Phase II: Implementation

- Establish and maintain Trial Master File
  - May be held by Sponsor

- Develop Site Regulatory Binders
  - Prepare tabs and binders
Phase II: Implementation

• Develop recruitment plan and materials
  • Identification of Investigators
  • Public website (clinical trials.gov)
  • Call Centers
  • Brochures
  • Google ad campaign
  • Television/radio spots
  • Newspaper advertisements

Phase II: Implementation

• Develop On-site Monitoring Plan
  • 100% informed consents
  • 100% inclusion/exclusion criteria
  • Random selection of % of subjects enrolled
  • Site regulatory files

• Execute subcontracts
  • This can be a very lengthy process
  • Legal talking to Legal…

• Prepare for initial DSMB meeting for protocol approval

Phase II: Implementation

• Develop Safety Monitoring Plan
  • Identify Medical Monitor(s)
  • Determine level of reporting required
  • Adjudication of events between Medical Monitors
**Phase II: Implementation**

- **Site Initiation Visits**
  - An opportunity to begin a study on the right path
  - May be done in person or through teleconference or webinars
  - Important agenda items
    - Protocol training
    - Good Clinical Practices
    - Study Coordinator Training on Procedures
    - Data Entry Training and Certification
    - Review of Facilities (if not previously done)

**Phase III: Up and Running**

- **Collect, QA and Monitor Site Regulatory Documents**
  - 1572s
  - Delegation of Responsibility Log
  - Investigator CVs
  - Investigator Licenses
  - Laboratory Certifications
  - Laboratory Normal Ranges
  - IRB approvals

- **Monitor Site IRB approvals**
  - Activate sites when approvals received
Phase III: Up and Running

- Develop User Access Policies for the Web and Data Entry Systems
  - Set up user accounts
  - Verify users through Delegation of Responsibilities Log

- Develop ongoing study training materials
  - Webinars
  - ppt. presentations
  - Revisions to MOP

Phase III: Up and Running

- Develop Statistical Analysis Plan (SAP)
  - Submit to FDA

- Distribute study supplies after site activation
  - Study drug
    - Investigational and approved products
  - Specimen collection kits
  - Kits and shipping supplies
  - Study supplies and equipment
    - Study-specific (e.g. Blood Pressure monitors, EKG machines, Glucometers, etc.)

Phase III: Up and Running

- Develop communication plan
  - Identify areas of responsibility
  - Who is the first level of communication in each area?
  - Who is the next level of communication if an item needs escalation?

- Establish relationships with site staff
  - Coordinator teleconferences
Phase III: Up and Running

• Study-specific QA activities
  • Replicate samples
  • Equipment calibrations
  • Data entry QA
  • Project Work Instructions

Phase IV: Ongoing Activities

• Protocol amendments
  • Submit to FDA
  • Submit to IRBs

• Monitor IRB approvals and renewals

• Monitor recruitment, retention and adherence
  • Site performance tracking tools

• Monitor data entry for timeliness
  • Query resolution
  • Missing data
    • Reports to monitor missing data
    • Work closely with study coordinators to receive all obtainable data

• Ongoing collection and QA of regulatory documents
Phase IV: Ongoing Activities

- Develop report shells
  - Enrollment report
  - Ineligibility report
  - Adverse Events/Serious Adverse Events
  - Protocol deviations
  - Missing data
  - Study-specific reports

Phase IV: Ongoing Activities

- On-site monitoring according to monitoring plan
  - Provide CRA with data listings
  - Source document verification
  - Tools for resolving data discrepancies
  - Monitor drug accountability logs
  - Monitor site regulatory documents
  - Monitor Adverse Event/Serious Adverse Event reporting
  - Monitor protocol deviation reporting
Phase IV: Ongoing Activities

• Distribute study drugs and supplies
  • Monitor site utilization
  • Monitor expiration dates
  • Establish trigger points for re-order

• Monitor drug and supply accountability logs
  • Internal DCC monitoring to ensure sites don’t run out of drug or supplies

Phase IV: Ongoing Activities

• Site retraining on protocol and procedures
  • Study coordinator turnover
  • One on one webinars
  • On-site training
  • Training for cause

Phase IV: Ongoing Activities

• Data entry system enhancements
  • Initial version released at study start-up
  • Don’t get caught up in never-ending tweaking of the data entry system
  • Develop enhancements in batches and release preferably no more often than quarterly
  • Hot fixes only for bugs that prohibit data entry
  • Change management software is very useful
Phase IV: Ongoing Activities

- Safety Review
  - Definitions for AEs
    - NCI Common Toxicity Criteria
    - Revised study-specific criteria
  - MedDRA coding
    - Training of coders
    - Agreement of coders
  - Monitor follow-up until resolution
    - Resolution with sequelae
  - Aggregate review of AEs for trends
  - Write safety narratives for Annual Report
  - Sponsor review
  - Submit to FDA

- Safety Review
  - Definitions for SAEs
    - NCI Common Toxicity Criteria
    - Revised study-specific criteria
  - Who determines relatedness?
  - Who determines expectedness?
  - Who determines need to expedite?
  - Who does the expedited safety reporting?
  - Monitor follow-up until resolution
  - MedDRA coding
  - Write safety narratives for Annual Report
  - Sponsor review
  - Submit to FDA
Phase IV: Ongoing Activities

- Import central laboratory data into database
  - Establish standardized data sets
  - Establish timeline for data transfers
  - Upload comma-delimited files
  - Email vs. encrypted transfer

- Interim analyses as described in SAP
  - Monitor for stopping rules
  - Futility analysis

Phase IV: Ongoing Activities

- FDA Annual Report
  - Determine due date and data lock date
  - Describe protocol activity
  - Describe safety profile
  - Describe new findings
**Phase IV: Ongoing Activities**

- **Lost to Follow-up**
  - Important to make all efforts to obtain endpoint data
  - May need IRB approval to make final contact

- **Investigator Payments**
  - Based on recruitment activities
  - Based on data entry completion

- **Newsletters**
  - Keep sites informed year-round of study activities
  - Relevant "hot" topics
  - Updates from study PI
  - Updates from DCC
  - Recruitment tips

- **Update Investigator’s Brochure**
  - Review safety profile
  - Submit to FDA
  - Submit to IRBs

- **Annual Investigator/Study coordinator meeting**
  - Opportunity to kick-start recruitment
  - Opportunity for training
  - Protocol
  - Procedures
  - Data entry
  - Collaborate on publications
Phase IV: Ongoing Activities

- **DSMB Meetings**
  - Follow the charter
  - Teleconference or face-to-face?
  - Establish data lock dates
    - Give coordinators plenty of notice to complete data entry
  - Establish timeline after data lock to:
    - Run reports
    - QA reports
    - Print reports
    - Distribute reports

Phase IV: Ongoing Activities

- **Steering Committee Meetings**
  - Establish frequency of meetings
  - Teleconference or face-to-face?
  - Steering Committee Chairperson
  - Who are the voting members and what constitutes a quorum
  - Who has responsibility for:
    - Agenda items
    - Minutes
    - Action items

Phase IV: Ongoing Activities

- **Subcommittee Meetings**
  - Establish frequency of meetings
  - Teleconference or face-to-face?
  - Committee Chairperson
  - Who has responsibility for:
    - Agenda items
    - Minutes
    - Action items
Phase V: Study Close-Out

- Site Close-Out Visits
  - Final On-Site Monitoring Visit
  - Reconcile Files
  - Final Drug and Supply Accountability
  - Close-Out Letter

- Resolve all queries and data issues
- Data lock
Phase V: Study Close-Out

- Return all unused study drug (if applicable)
  - Will there be ongoing study drug treatment?
  - When will subjects/sites be unblinded?

- How (if) are subjects informed of results?
  - Keep IRB open at sites if recontact is anticipated

- Analysis programs are developed and debugged

Phase V: Study Close-Out

- Final analysis done per SAP
- Reports are written, reviewed, and accepted by Steering Committee
- Sponsor and FDA receive final reports
- Publications
  - Lead authors determined through Publication Policy
  - DCC assists with additional analyses as requested
- Submit data sets to clinical trials.gov

Overall Project Management Tips

- Develop good teams and working relationships
  - Identify the Project Champion
  - Provide the teams with the tools and training to successfully accomplish their goals
  - Monitor for meetings that have served their purpose and should be discontinued
  - Monitor for redundancy in meetings
  - Acknowledge and reward exceptional behavior
  - Find the strengths in each team member
Overall Project Management Tips:

- Ensure that all team members are aware of areas of responsibilities
- Never walk out of a meeting without a clear understanding of the deliverables
- Or who is responsible for the deliverable
- Or what the expected timeline is for the deliverable

- Require documentation for all proceedings
- Don’t rely on memory for previous decisions
- Distribute minutes and action items after all meetings
- Post minutes and action items in a shared drive or on the web

- Be flexible when needed
- Good communication will reveal problem areas
- Must always be willing to re-examine and reprioritize
- Be willing to look at things from a different viewpoint
- Solicit input from the staff regularly
- Disagreement can be healthy if handled well
Overall Project Management Tips

• Follow-up on progress
  • Hold team members accountable for timelines
  • Expect progress reports on regular intervals
  • Look for ways to improve efficiencies
  • Look for ways to maintain staff satisfaction
  • Have some fun along the way!

Conclusion

• There are many components to juggle in clinical trials research
• Good project management makes clinical trials research doable
• Clinical trials work can be very rewarding
Part III: Data Collection, Reporting, and Quality Control Issues

Laura Lovato, MS
Senior Biostatistician
Department of Biostatistical Sciences
School of Medicine
Wake Forest University
SCT Pre-Conference Workshop
Essentials of Randomized Clinical Trials

Outline

• Introduction (GCPs, QC, QA, SOPs)
• Primary sources of error
• Steps in Data Collection
  • Design of data collection forms
  • Standardization of procedures
  • Types of data entry/management systems
• Quality control methods and reporting

Introduction

“No study is better than the quality of its data.”
-Friedman, Furberg and DeMets

“To err is human.”
Introduction: Guidelines for Good Clinical Practice

• Unified standard
• For design, conduct, analyses and reporting of clinical trials that involve human subjects
• To ensure that patients’ rights, safety and confidentiality are protected
• To promote scientific validity and data integrity

Introduction: Specific Principles of GCP Applicable to Data Collection

• Confidentiality of records should be protected
• All clinical trial data should be handled in a way to ensure accurate reporting, interpretation and verification
• An audit trail should be maintained for changes/corrections to forms and electronic data

Introduction: Useful web sites – GCPs and SOPs

From the U.S. FDA:
http://www.fda.gov/oc/gcp/

From Wake Forest University:
http://www2.wfubmc.edu/NR/rdonlyres/0C87FF91-78E1-48D0-8E13-8DB9244BF577/0/GoodClinicalPracticePresentation.ppt
Introduction: Data Collection and Quality Control

"Any procedure, method, philosophy … that is aimed at maintaining or improving the reliability or validity of the data and the associated procedures used to generate them."

- Curtis Meinert

Introduction: Quality Control (QC) vs Quality Assurance (QA)

QC involves all process controls and monitoring performed by local staff on a day-to-day basis to maintain data quality

QA involves independent review or auditing of key processes to uncover and remedy problems

Primary sources of error in data collection process

• Missing data – incomplete or irretrievable
• Incorrect data – more difficult to recognize
• Excess variability – can reduce the opportunity to detect real change
Steps in Data Collection

- Define key variables
- Standardize & train on procedures (MOP)
- Data Collection
  - Acquisition
  - Recording
  - Entry
  - Study Closeout
  - Preparation for analysis

Define Key Variables

Define key variables

- Depends on trial type and outcomes
- At Baseline: characteristics of enrolled/non-enrolled participants related to major eligibility requirements
- Primary/Secondary outcome measures
- Variables that might confound/mediate/modify association
- Monitoring adherence to the protocol
Focus on key variables

Only important data should be collected

- As the volume of noncritical data increases, forms become burdensome and complicated leading to confusion
- Clinical care data often not needed as part of trial database

Steps in Data Collection

Standardization and Training

Pre-trial Quality Control Activities:
- Obtain adequate resources
- Design of case report forms
- Pre-testing
- Design of data management system
- Manual of Procedures (MOP)
- Hiring qualified personnel
- Training and certification
**Standardization & Training**

Manual of Procedures
(prior to and during the study)
- Standardized procedures
- Clearly written, detailed instructions
- Timely updates and clarifications
- Accessibility is essential

**Training and Certification**

- Central, regional, or local
- "Train the trainer" model
- Audio-visuals
- Certification/recertification to maintain skill set

**Design of data management system**

- Security features/protection of human subjects' rights (privacy and confidentiality)
- Controlled Access
- Identification and authentication
Standardization & Training

Design of data management system
• Data entry/editing capability
• Desirable features:
  • Ease of screen set up and use
  • Range, field type, skip pattern checks
  • Query system
  • Ability to accommodate double data entry
  • Word processing or spreadsheet software not advocated

Standardization & Training

Design of data management system
• Web-based systems also have administrative functions
  • Communications hub,
  • Information/Resource Center,
  • Coordination of publications process,
  • Management of Adjudication System

Steps in Data Collection

Data Acquisition
Design of Case Report Forms

• Purpose:
  • To collect complete and accurate data
  • To ensure standardization and consistency
  • In some cases, to reinforce the protocol

Design of Case Report Forms

• Clean, concise, consistent
• Well-organized with logical flow
• Few “write-in” or “text” answers
• No essay questions!

Design of Case Report Forms

• Selection of items to be collected
• Timing of visit schedule
• Ordering of Procedures
Steps in Forms Development

- Examination of Existing Forms (not necessary to "reinvent the wheel")
- Data Collection forms in Clinical Trials (Spilker B, Shoenfelder J, Raven Press, New York, 1991)
- The Annotated Bibliography of Epidemiologic Methods for Cardiovascular Research
- Use the web – similar studies may have examples on the public side of their web sites

Pre-Testing

- Mock visits/procedures conducted
- Simulation with practice participants
- Debriefing is essential to improve procedures
- Procedures/forms revised accordingly
Changes to Study Forms

- Often done early on to improve data collection
- Can be problematic when done repeatedly throughout the trial
  - Results in multiple versions of data sets
  - Can increase risk of errors (clinic, data entry, analysis)

Changes to Study Forms

Initial Version
Troponin results
1 At least 5x upper limit of normal
2 At least 2x upper limit of normal but less than 5x
3 Greater than upper limit of normal but less than 2x
4 Within normal limits

New Version
Troponin results
1 At least 5x upper limit of normal
2 At least 3x upper limit of normal but less than 5x
3 At least 2x upper limit of normal but less than 3x
4 Greater than upper limit of normal but less than 2x
5 Within normal limits
Data Recording

- Traditionally, refers to transcribing information onto case report forms
- Trend toward direct computer entry (computer-assisted data collection, e.g., on PDAs) with no prior hard copy
- Both approaches depend on well-designed forms/data entry screens

Types of Case Report Forms

- Paper forms
- Scannable forms (NCR)
- FAX-based forms (Teleform)
- Direct web-based entry
Steps in Data Collection

Data Entry

Modes of Entering information into central database

- Direct computer entry
  - Sometimes includes hand-held devices
  - Data entry screens resemble forms
  - Built-in logic and range checks
- Optical mark reading (scanning)
- Optical character recognition (DATAFAX)

A Note on Direct Data entry with no source forms

- Convenient
- Can allow for direct participant data entry
- No source form for audit/data errors
- Sample with both hard copy and electronic
- Built-in real time error checks
Data Entry
Types of data entry systems

• Local
  • Data keyed onsite by clinic personnel
  • Potential for quick resolution of data omissions, errors, and inconsistencies

• Central
  • Forms mailed/faxed to sponsor or data coordinating center
  • Data entered by experienced keyers
  • Forms stored centrally.

Data Entry
Web-based data entry systems

• Provides flexibility
  • Data entry can be local or mix local/central
  • No specific hardware requirements
  • No specific software requirements for internet browser
  • Secure link provided
  • Data from multiple sources are consolidated on a central server

Data Entry
Web-based data entry systems

• Security features/protection of human subjects’ rights (privacy and confidentiality)
• Controlled Access
• Identification and authentication
  • Requires valid user id and password
  • Password expire every 90 days
  • Specific access rights based on study function
Data Entry
Web-based data entry systems

• Audit trail
• Each and every access into the system is documented
• Every page that is accessed is documented
• All versions of any record entered are kept and date/time stamped (with user id)

Data Entry
Web-based data entry systems

• Virus protection/scanning strategies to monitor and eliminate security threats
• Database server behind firewall
• Disaster recovery plan
• Regular backup for all data

Example of a Multi-center Study web-site
Web site as a communication hub

Steps in Data Collection
Closeout

Special notes on study closeout

- Continuous monitoring throughout the trial reduces the clean-up job at the end of the study
- Lost-to-Follow-up (National Death Index, web-based searches, paid search firm)
- "Freezing" data at various points of cleanliness
- Data dictionaries created
- Responsibilities to sponsor (i.e., public use datasets, storing study materials)
Steps in Data Collection
Preparation for Analysis

Data Preparation for analysis

- Cleaning/editing
  - Inconsistencies
  - Omissions/discrepancies
- Merging records
- Documenting analysis files
  - Definition of variables/cut points
  - Validation of calculated variables
  - Verification of statistical outliers/distribution of data

Site Visits

Quality assurance visit of a clinical trial unit (e.g., clinical centers, coordinating center, central lab, etc.) by a team of experts to observe operations and assess performance
Scientific Misconduct in Clinical Trials

Data Fraud:
• reported in a small number of clinical trials
• refers to:
  • Fabrication (making up data)
  • Falsification (changing or removing data values)

High Quality Data

• Good Clinical Practice Guidelines
• Good clinical research practice
• SOPs
• Ethical/scientific integrity
• "GIGO"
  • Inaccurate data are worse than no data
  • Garbage in, garbage out

Quality Control Monitoring Reports
Basic Monitoring Reports

- Data Monitoring
- Quality Control reports

Data Monitoring Reports

Examples of the following:
- Recruitment
- Baseline and Follow-up data collection (includes lab, ecg, drug distribution, etc.)
- Adherence to protocol (clinicians and participants)
- Lost to follow-up, Refusals

Recruitment Monitoring Example
Recruitment Monitoring Example

![Graph](image)

Monitoring Baseline Assessments

Are the study groups comparable at the time of randomization?

- Risk or prognostic factors, important demographic characteristics, medical history
- Randomization on average produces balance between groups – no guarantee!
- Correcting an imbalance: adjust in randomization or in analysis

Monitoring Baseline Assessments

Easiest way: compare each variable by treatment assignment using means, medians, ranges

Note that the groups will never be identical: 5% of the comparisons will show differences at the 0.05 significance level
Monitoring Follow-up assessments

1. Number of Visits completed as planned: %

2. Completeness of data: missing forms, missing data on forms

3. Quality of data received: data queries on each field (at data entry and/or retrospective data queries)
Monitoring data quality

Monitoring Adherence

- Come at adherence from many different angles:
  - Participant adherence
  - Clinical site staff adherence to the protocol
- Long-term trials, look at changes over time
- Separate by calendar time, clinic visit, by clinic if a multi-center trial
- Tables and/or graphs

<table>
<thead>
<tr>
<th>ID #</th>
<th>Trial Status</th>
<th>Date of Last Form</th>
<th>Days</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ppt#1</td>
<td>Non-adherent</td>
<td>15 JAN 2008</td>
<td>216</td>
<td>10/20/2007 by Jill Jones (CCN): Elevated CK &gt; 5X ULN on 2 occasions. Does patient have symptoms?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>01/15/2008 by Joe Smith (CS): Yes patient has symptoms.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>04/17/2008 by Jill Jones (CCN): Looks like both blinded lipid med and statin were stopped. Last CK is 120. Consider checking CK next visit off all lipid meds, check in 6-8 weeks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>04/28/2008 by Joe Smith (CS): Participant rechallenged on low-dose blinded med, will check in 6-8 weeks.</td>
</tr>
<tr>
<td>Ppt#2</td>
<td>LTF</td>
<td>12 DEC 2008</td>
<td>89</td>
<td>01/09/2009 by Joe Smith (CS): This patient has moved to Papua New Guinea for his work and couldn't come for his interval visit in December. Not forwarding address.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>01/12/2009 by Jill Jones (CCN): Per our phone conversation, try to get in contact with patient to make sure he can get phone contact info for an events assessment at minimum.</td>
</tr>
</tbody>
</table>

Enter a comment
Monitoring Lost to Follow-up, Refused

- Separate groups: Lost to Follow-up versus Participant refusals (withdrawn consent)
- Investigators will want to know why participants are lost (e.g., moved out of range) and refused (e.g., withdrawn consent due to problems with protocol)
- Anticipate participants prone to becoming lost: monitor missed visit patterns and what happened to them
- Second tier: participants not officially LOST or REFUSED, but are no longer coming to the clinic or taking study medications

Summary

- Learning Objectives
  - GCPs, QC, QA, SOPs
  - Primary sources of error
  - Steps in Data Collection
    - Design of data collection forms
    - Standardization of procedures
    - Types of data entry/management systems
    - Quality control methods and reporting
Part IV: Treatment Allocation

Michele Melia, Sc.M.
Senior Statistician
Jaeb Center for Health Research
Tampa, FL
SCT Pre-Conference Workshop
Essentials of Randomized Clinical Trials

Outline
- What randomization is and why it is used
- Truly random versus not random allocation
- Simple, block, and stratified randomization and when to use them
- Adaptive randomization and some of its pros and cons
- How to administer randomization in a trial

What is randomization?
A process by which subjects are randomly assigned to a treatment in a clinical trial
- Neither the participant nor the investigator knows what treatment the participant will receive
**Why is randomization used?**

- Problems arising with treatment assignment in clinical practice:
  - Individuals with certain disease characteristics are generally more likely to receive certain treatments (confounding by indication)
  - Inability to characterize why individuals were assigned to a particular treatment, leading to non-homogeneous groups with different (and unquantifiable) underlying risk
  - Wide variation in outcomes relative to the magnitude of differences due to treatments; treatment differences easily obscured by bias

**How does randomization work?**

- Randomization does:
  - Reduce bias in assigning patients to treatments
  - Ensure valid statistical tests
  - Reduce unwanted variation resulting in improved power for statistical tests (more about this later)
- Randomization does not:
  - Guarantee equal distribution of prognostic factors among treatment groups
  - For large studies, the chance of imbalances is small
  - For small studies, the chance of imbalances is larger

**When is randomization used?**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Frequency</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Rarely</td>
<td>Not generally necessary to achieve phase I goals of establishing toxicity/maximum tolerated dose/dose response.</td>
</tr>
<tr>
<td>II</td>
<td>Sometimes</td>
<td>When comparison group is helpful in defining possible biologic and adverse effects, e.g. for highly subjective endpoints. When required by FDA.</td>
</tr>
<tr>
<td>III</td>
<td>Almost always</td>
<td>&quot;Gold standard&quot; for reducing bias in assignment of patients to treatment and estimation of treatment effects</td>
</tr>
</tbody>
</table>

Other methods of (non-random) treatment allocation are also sometimes used in CTs:

- Single group with or without historical controls
- Non-random allocation of 2 or more groups
Non-random methods of treatment allocation

- Alternating treatments (1st patient gets A, 2nd gets B, 3rd gets A, etc.)
- Alternating assignment by date or day of week (patient gets A if enrolled on even date, B if odd date)
- Using patient initials to determine assignment
  - A-K → treatment 1
  - M-Z → treatment 2

Problems with non-random treatment allocation

- Treatment assignment of next patient can be predicted in advance; therefore,
  - Not truly random
  - Open to manipulation
  - Goal of bias reduction can be subverted

Basic types of randomization

- Simple
- Block
- Stratified / stratified block
Simple Randomization

A sequence from a random number table or generator is used to assign sequential patients to a study treatment using a pre-defined rule. E.g. Even number→A and Odd number→B.

<table>
<thead>
<tr>
<th>Sequence from random number table</th>
<th>Treatment assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
</tr>
</tbody>
</table>

Simple Randomization

- **Advantages**
  - Simple
  - Each new assignment made without regard to previous assignments
- **Disadvantages**
  - No guarantee of equal or approximately equal sample size in each treatment group at any stage of the trial (including at the end)
    - Imbalance reduces statistical power
    - Estimates of treatment effect are not affected; only precision
  - No protection against long runs of one treatment

Block randomization

- Block size that is an integer multiple of the number of treatments is chosen (integer≥2)
- Equal numbers of patients are assigned to each treatment within a block
  - Numbers are proportional rather than equal in the case of unequal allocation
- Overcomes some disadvantages of simple randomization
Example: Block Randomization for 2 Treatments

- Possible block sizes are 4, 6, 8, etc.
- For block size of 4, there are 6 treatment-balanced permutations
  - ABAB, AABB, ABBA, BABA, BBAA, BAAB
- These may be chosen at random with replacement

<table>
<thead>
<tr>
<th>Sequence from random number table</th>
<th>Treatment assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>--</td>
</tr>
<tr>
<td>7</td>
<td>--</td>
</tr>
<tr>
<td>9</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>AABB</td>
</tr>
<tr>
<td>1</td>
<td>ABAB</td>
</tr>
<tr>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>6</td>
<td>BAAB</td>
</tr>
</tbody>
</table>

Block randomization – cont’d

- Large block size does not protect as well against long runs as small block size
- Small block size makes it easier to guess next treatment
- To make it harder to guess the next allocation when small block sizes are used, block size can be chosen at random from a pre-defined list of block sizes, e.g. 4, 6, 8
- Simple and block randomization do not guarantee balance of treatment groups on important prognostic factors

Stratification

- With stratification, a separate, independent randomization sequence is used for each prognostic group (or strata)
- To guarantee treatment balance within strata at all stages of the trial, stratification is combined with blocking
  - Use of simple randomization within strata will not guarantee treatment balance within strata
  - Consequence of imbalance on a prognostic factor is bias in the estimated treatment effect
Example – Blocked and stratified randomization

- A randomized trial comparing near versus distance activities while patching for amblyopia (lazy eye) in children 3 to <7 years old
  - Pilot study data suggested that near activities might be less effective in moderate as compared to severe amblyopia
  - Randomization was stratified by amblyopia severity; random block sizes of 4 and 6 also were used

Example - continued

- If even, use block size=4; otherwise block size=6
- Use a random shuffle of the block elements

<table>
<thead>
<tr>
<th>Moderate amblyopia</th>
<th>Random No.</th>
<th>Block size</th>
<th>Random sequence</th>
<th>Treatment assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>6</td>
<td>7 9 2 1 6 6</td>
<td>A A A B B B B B</td>
<td>0 1 2 6 7 9</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>8 5 1 3 9 7</td>
<td>A A A B B B B B</td>
<td>0 1 3 5 7 8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severe amblyopia</th>
<th>Random No.</th>
<th>Block size</th>
<th>Random sequence</th>
<th>Treatment assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6 3 1 2</td>
<td>A A B B</td>
<td>1 2 3 6</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0 9 5 7 3 4</td>
<td>A A A B B B B B</td>
<td>0 2 3 5 7 9</td>
</tr>
</tbody>
</table>

Stratified randomization – cont’d

- Chance of imbalance on prognostic factors is small with large sample size
  - Stratification is more important when sample size is small
- As number of stratification factors increases, the number of strata grows very fast, and efficacy with respect to achieving desired balance may decrease
  - Think of case where # strata = sample size
- Be judicious in choice of stratification factors
Stratified randomization – cont’d

- If many prognostic factors must be controlled:
  - Consider combining them into an overall index and stratifying on index
  - Consider minimization (more on this in a few moments)
- When analyzing data, it is important to account for stratification
  - If ignored, variability due to the stratification factor is included with error variance
  - If included, variability due to stratification factor is removed from error term, increasing precision

Unequal Treatment Allocation

- With unequal treatment allocation, the study is designed to have unequal numbers of patients on the treatments
- Treatment groups of equal size are desirable from a statistical perspective for making treatment group comparisons
  - Maximizes power for a given sample size
  - However, loss of power may not be too severe as long as imbalance is not severe, e.g. 2:2:1

Unequal Treatment Allocation – cont’d

- Some reasons to consider unequal allocation:
  - More information is needed on effect of a new treatment (e.g. adverse effects, effect of dose)
  - Patients may be unwilling to be randomized if probability of assignment to control or placebo is high
  - To reduce study cost when one treatment is a lot more expensive than the other
- Principles of basic randomization regarding use of blocking and stratification still apply
Cluster Randomization

- Clusters of patients are randomized rather than the individual patients
  - Example: In trial of vitamin A supplementation for prevention of mortality in preschool children in Nepal, administrative wards were randomized to supplement or placebo (West KP, Lancet 1991)
- Cluster randomization reduces statistical efficiency (i.e. it requires more patients)
- Usually used when it is not feasible to randomize individual patients

Adaptive Allocation (aka Adaptive Randomization)

- Information on previously enrolled patients is used to modify (or adapt) the allocation ratio, i.e. the probability of being assigned to each treatment
- Information used typically is one of:
  - Treatment
  - Covariates (prognostic factors)
  - Response (outcome)
- Other terms:
  - Biased-coin design
  - Urn design
  - Play-the-winner design

Treatment Adaptive Randomization

- Allocation ratio is adjusted using the number of patients previously assigned to each treatment
- Basic idea (for trial with 1:1 allocation):
  - If current proportion of patients randomized to A is less than $\frac{1}{2}$, assign current patient to A with probability greater than $\frac{1}{2}$. 
Treatment Adaptive Randomization

- Advantages
  - Balance on # of patients in each treatment group is achieved at all stages of the trial
  - Harder to guess next assignment than for randomized block design with small block size
- Disadvantages
  - Increased administrative complexity
  - Analysis is more complicated – probability for each assignment is needed

Covariate Adaptive Randomization

- Also known as minimization
- Basic idea:
  - If number of previous patients with covariate profile matching the current patient is higher in group A than B, then probability the current patient is randomized to B is increased to greater than \( \frac{1}{2} \).

Covariate Adaptive Randomization – cont’d

- Advantages
  - Achieves balance among treatments on important covariates
- Disadvantages
  - Intensive administrative effort may be needed (especially if number of covariates is large)
  - Increased risk of breaking masking
  - Unnecessary matching
    - Large sample size alone is likely to result in good balance on covariates
    - Randomization and analysis have been complicated unnecessarily
Response Adaptive Randomization

- Also known as ‘Play-the-winner’ designs
- Basic idea:
  - If current trial results favor treatment A, probability that the patient is randomized to A is increased to greater than ½
- Famous example: ECMO Study (Bartlett, Pediatrics1985)
  - Start with 2 balls in an urn marked E(cmo) and C(control)
  - If treatment is successful, add a ball marked with that treatment into the urn (along with the selected ball)
  - If not successful, add a ball marked with the opposite treatment (along with the selected ball)

Response adaptive allocation - ECMO Study

- Trial ends when 10 balls of 1 type are added with that type declared the winner
- Assuming one treatment has substantially greater chances of survival, this design has high probability of selecting the better treatment as the winner

ECMO Study Results

- E(cmo) selected
- Patient lives
- C(control) selected
- Patient dies
- E selected
- Patient lives
- 4th-10th balls: E selected
- Patients all live
ECMO Study Results

- 10 E balls were added, so ECMO declared the winner
- 2 more patients given E; both lived
- Final counts:
  - 0/1 control patients lived
  - 11/11 emco patients lived
- Might be tempted to analyze using Fisher’s Exact Test, but cannot, as marginal totals are random variables that contain information about the outcome

Response Adaptive Allocation – cont’d

- Advantages
  - Increases chances that patients will get the better treatment
  - Ethically appealing
- Disadvantages
  - Increased administrative complexity
  - Not always possible (e.g. long-term response)
  - Analysis is more complicated; appropriate statistical tests may not exist
  - Ethical difficulties if allocation ratio becomes highly skewed to one treatment

Summary – Adaptive Allocation

- Simple randomization or stratified block randomization are generally perfectly adequate when sample size is large
- Consider complex alternatives only if sample size is small
Administration of randomization codes

- When the study protocol is finalized, but before the study begins patient enrollment:
  - The randomization schedule is generated (for a non-adaptive randomization scheme)
  - Procedures for obtaining a randomization code for a study patient are defined
  - Procedures for unmasking are defined
  - System for tracking randomizations issued, errors and deviations from schedule, and unmasking is in place

Generating the randomization schedule

- A Standard operating procedure (SOP) for generating randomization schedules is desirable. Elements of the SOP should include:
  - Who may generate a schedule (preferably this is done by a statistician not involved in day-to-day study operations)
    - Statistician ensures that the schedule adheres to the study design
  - Procedures for schedule/code checking

Generating the schedule - continued

- Documentation of how the schedule was generated
  - Programs & pseudonumber generator used
  - How to use them
  - Seed(s) used to obtain the schedule in question
- For studies being submitted to FDA, the programs must be validated (and periodically re-validated) and results of validation must be documented
Procedures for obtaining a randomization code

- There are many procedures that are commonly used including:
  - Centrally administered
    - Telephone call to coordinating center or its surrogate (e.g. answering service)
    - Web-based system
  - Locally administered
    - Sequential drug kits
    - Envelope system
    - Computer program installed on local PC

Procedures for obtaining a randomization

- Procedures should take into account:
  - Allowable time between request for randomization and issuance of randomization
  - Times of day and days of week that patients will be randomized and attendant staffing needs
    - Coverage for all time zones
  - Ease and convenience for investigators and patients

Procedures for obtaining randomization – cont’d

- Procedures should take into account:
  - Vulnerability to manipulation or tampering
    - Centrally-administered systems generally easier to secure
    - Secure local systems are possible with proper safeguards
  - Need for fall back procedure in event that primary procedure isn’t working (e.g. web site outage)
# Procedures for unmasking

- Under what circumstances is unmasking permitted?
- Who may be unmasked?
- How will unmasking be performed?

---

# Summary

- Randomization is the primary means for controlling bias in allocation of patients to treatment in a clinical trial
- Randomization helps to generate comparable groups of patients on each treatment
- Randomization enables valid statistical tests for the evaluation of the treatments
- Judicious use of stratification with appropriate analysis can improve statistical power

---

# Selected References

- Controlled Clin Trials 1988; Volume 9, issue 4 has a series of articles on randomization in clinical trials by John Lachin
Software

nQuery Advisor can be used to generate randomization lists

For links to randomization software (free) and services (not free) developed and maintained by Martin Bland at University of York see:

http://www-users.york.ac.uk/~mb55/guide/randsery.htm

Disclaimer: endorsement of software and services on this website is not implied
Part V: Choice of Endpoints

Marta Marsh Gilson, Ph.D.
Assistant Professor
Center for Surgery Trials & Outcomes Research (CSTOR)
Department of Surgery, Johns Hopkins University
Baltimore, Maryland

SCT Pre-Conference Workshop
Essentials of Randomized Clinical Trials

Learning Objectives

By the end of the course, attendees should be able to:
• Identify possible endpoints for their study
• Assess the pros and cons for possible endpoints
• Be able to ‘better’ choose endpoints that meet study needs

• Missing Data
• Intention to Treat

Primary vs. Secondary Endpoints

Endpoint (outcome)
Determined in each study subject / participant / unit

Primary outcome variable
“… designated or regarded as key in the design or analysis of the results of a trial.” – Meinert, CL

Secondary outcome variable
“ any other outcome variable used for treatment evaluation” – Meinert, CL
Questions vs. Endpoints

- Research question(s) – What we want to show
  - hypothesis
- Endpoint(s) – How to show it
  - single primary outcome
  - limited number of secondary outcomes
- Endpoint(s) are much more specific than question(s)

Endpoint Considerations

Choice of endpoint will affect:

Personnel
Equipment
Facilities
Study duration
Sample size calculations

Resources will affect choice of endpoint

Personnel

Who (skill level)
  - HS educ vs. special training vs. machine
What
  - Examination vs. photos vs. lab values
Where
  - Local clinic vs. home visit vs. central facility
When and how often
  - One point in time vs. repeated measures
Personnel turnover
**Equipment**

- Specialized vs. standard
- Specific make model vs. approved subset vs. any
  - If more than one type – can you switch
- Move equipment to people or people to equipment
- Technology stable vs. changing/improving
- Any information comparing equipment (endpoint)

**Facilities - Local**

- Size of room
  - conduct visit, store files**, forms
  - measure outcome (distance vision)
- Location of room
  - elderly population – stairs, long walk
- Privacy (quality of life, or personal interview)
  - shared space vs. dedicated trial space

**Facilities - Central**

- Reading Center (photographs, ultrasound, X-rays, etc)
- Pathology Center (tissue/ slides)
- Radiation Physics Center (dose curves)
- Space – specimens, gradings, storage **
- Ancillary study use of materials
  - Committee to approve use
  - Archiving committee

****Know the rules for how long you must keep data forms / specimen
your institution / study sponsor

---

**Facilities - Local**

- Size of room
  - conduct visit, store files**, forms
  - measure outcome (distance vision)
- Location of room
  - elderly population – stairs, long walk
- Privacy (quality of life, or personal interview)
  - shared space vs. dedicated trial space

**Facilities - Central**

- Reading Center (photographs, ultrasound, X-rays, etc)
- Pathology Center (tissue/ slides)
- Radiation Physics Center (dose curves)
- Space – specimens, gradings, storage **
- Ancillary study use of materials
  - Committee to approve use
  - Archiving committee

****Know the rules for how long you must keep data forms / specimen
your institution / study sponsor
Study Duration

When is endpoint assessed
• Day 1 vs. 8 weeks vs. all cause mortality
• Length of follow-up beyond primary outcome
• Frequency of assessment

Rate of occurrence
rare event vs. common event

Single vs. multicenter

Sample Size

To be covered next (Part VI) in much more detail

Choice of endpoint determines sample size
Different endpoints – different sample size

Balance and Adjustments

ENDPOINT

Practical considerations

Scientific considerations
Essential Scientific Features

• Rigorously defined
• Rigorous assessment methods
• Relevant to study goals
• Reproducible in research study
• Assessable in all groups to be evaluated or compared
• Unbiased (minimize bias)
• Chosen before data collection
• Anticipate data analysis methods/needs

Fictitious Example A

Research Question:
Does wearing near correction for reading decrease the incidence of myopia in school age children?
Study design: Prospective

Rigorously define:
• By how much?
• Any / XX or more
• Since when?
• Baseline / last visit
• Relative or absolute difference

Fictitious Example A

Research Question:
Does wearing near correction for reading decrease the incidence of myopia in school age children?
Study design: Prospective

Rigorously define:
• When:
  • Ever / by age A / by grade G / per year
  • Single observation / confirmed observation
  • How much time between observations
Fictitious Example A

Research Question:
Does wearing near correction for reading decrease the incidence of myopia in school age children?

Study design: Prospective

Rigorously define:

- Myopia:
  - Any / Threshold –XX or more / Need glasses
  - How are you measuring myopia?
- Uncorrected distance vision / Refractive correction / Autorefraction device / Axial length of eyeball

Percent of children who are myopic* within 3 years of baseline

* Myopic = refractive error** -1.00 or more
** as determined by subjective refraction per study protocol

Rigorous Assessment Methods

Study protocol should specify….

- Equipment needed (camera, charts…)
- Environment (lighting levels, test distance)
- Time of evaluation
- Who determines endpoint
- Proxy or substitute endpoints allowed?
Relevant to Study Goals

<table>
<thead>
<tr>
<th>Study Goal</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Large simple trial</td>
<td>• Easy to obtain</td>
</tr>
<tr>
<td>• Screening program for population</td>
<td>• Cheap/ fast/ easy</td>
</tr>
<tr>
<td>• Select between treatments</td>
<td>• Relevant clinical measure</td>
</tr>
<tr>
<td>• Estimation</td>
<td>• Reliable measure</td>
</tr>
</tbody>
</table>

Reproducible in Research Study

Internal Data
Duplicate measures
Sample / Total study population
Same / different assessors
Same / different methods
Same / different days

External Data
Similar method
Similar personnel
Similar training

Reproducible in Research Study


• The HemoCue machines were checked daily against a standard.

• If participant refused venous blood draw, finger prick blood draw

• In a study by Sari et al., mean Hb measurements using venous or capillary blood was identical when using the Hemocue machine

• Capillary blood lower sensitivity (70.6%) but comparable specificity (95.2%) relative to venous blood (82.4 and 94.2%, respectively) for identifying anemic subjects (Hb 110 g/L)
### Assessable in All Groups

- Same methods for all
- Documentation of methods (protocol)
- Same time points for all

### Unbiased - Minimizing Bias

- Objective vs. Subjective
- Masked (blinded) examiner or assessor
- Masked study participant
- Assessed same way for all (per protocol)

### Anticipate Data Analysis Needs

- **Different endpoints**
  - Median survival
  - 5-year survival
  - Death rate during first year post-treatment

- **Different methods**
- **Different data collection forms**
- **Different analysis plan**
Data Analyses

- Descriptive Statistics
  - location (mean, median)
  - variability
  - frequency
- Hypothesis Testing
- Estimation and Confidence Intervals
- Exploratory Data Analysis
- Adjusted Analyses
  - summary measure from stratified analysis
  - regression model (linear / logistic)
- Overviews (Meta analyses)

Options

ENDPOINT

Practical considerations

Scientific considerations

Endpoint Options

Endpoint units of measurement

- Quantity (continuous or numerical scale)
  - Quality of Life Scores; BP; reading speed; CD4 count
- Likert scale (ordered categorical)
  - None / Very mild / Mild / Moderate / Severe / Very severe
- Dichotomous (Binary=2 choices)
  - [None / Some] [Disease / Disease-free] [Success / Failure]
- Time-to-event
  - Disease-free-survival / Overall Survival
- Person-years
Mortality Options

- All cause
- Disease specific
- Proxy (next of kin)
- MD (doctor)
- Death certificate
- Mortality Coding Committee (MCC) confirmed

Peri-operative or Post-operative Endpoints

Nausea / vomiting / pain / infection / adverse events

- Window of Time
  - Within – possibly attributed to operation
  - Outside – not attributed to operation

- List of possible diagnoses/complications
- Specific methods/tests

Surrogate Endpoints - Definition

Surrogate outcome variable

“A test, measurement, score, or some other similar variable that is used in place of a clinical event in the design of a trial, or in summarizing results from it.”

- Believed to be correlated with clinical event
- Perceived utility in yielding detectable treatment difference

– Meinert, CL
Surrogate Endpoints – Fictitious Example

Research Question: Does vitamin supplementation of pregnant women reduce infant mortality.

Ideal endpoint: 6-month survival of infants

Suppose very mobile population, women busy, not likely to come back at 6 months, hard to find out vital status of baby once they leave hospital

Possible surrogate endpoint: birth weight of baby (% low birth weight)

Surrogate Endpoints – Not Universal

Surrogate endpoints are not universal; what works for one study may not work for other studies

Nepal Nutrition Intervention Project – Sarlahi (NNIPS)

• Randomized community trial
• Combinations of micronutrients
• 6-month infant mortality
• Birth weight
• Higher mortality with larger birth weight

Criteria for Good Surrogate Endpoint

• Strong statistical association with primary endpt.
• Change in surrogate strongly correlated with change in primary endpoint (but: correlation ≠ causality)
• Surrogate is in the biological pathway of the disease (there may be > 1 pathway)
• Short latency (↑ surrogate followed by rapid onset of disease)
• Responsive to treatment (effect on surrogate may not equal effect on disease)
Surrogate Endpoints – Example 1
Cardiac Arrhythmia Suppression Trial (CAST)

• Prior evidence of association between arrhythmia and sudden death.
• Wide use of medication to suppress arrhythmia
• Enrolled: patients with asymptomatic or mildly symptomatic ventricular arrhythmia (six or more ventricular premature beats per hour) after myocardial infarction.
• Treatment(s): antiarrhythmic therapy (encainide, flecainide, or moricizine)
• Endpoint(s): death from arrhythmia / initial suppression of their arrhythmia (as assessed by Holter recording)

Surrogate Endpoints – Ex. 1 Continued
Cardiac Arrhythmia Suppression Trial (CAST)

• March 30, 1989 Results:
  • 75% had initial suppression of their arrhythmia (surrogate)
  • higher rate of death (primary) from arrhythmia in patients assigned to active drug than the patients assigned to placebo
  • “We conclude that neither encainide nor flecainide should be used in the treatment of patients with asymptomatic or minimally symptomatic ventricular arrhythmia after myocardial infarction, even though these drugs may be effective initially in suppressing ventricular arrhythmia.”

• Evidence that effect on possible surrogate outcome may differ from effect on clinical outcome

Surrogate Endpoints – Example 2
Prospective Randomized Milrinone Survival Evaluation (PROMISE) trial

• Prior evidence: Milrinone, a phosphodiesterase inhibitor, enhances cardiac contractility by increasing intracellular levels of cyclic AMP
• Enrolled: Severe chronic heart failure (New York Heart Association class III or IV) and advanced left ventricular dysfunction
• Treatment(s): 40 mg of oral milrinone daily or placebo (all patients received conventional therapy w/ digoxin, diuretics, and converting-enzyme inhibitor)
• Endpoint: mortality from all causes
**Surrogate Endpoints – Ex. 2 Continued**

Prospective Randomized Milrinone Survival Evaluation (PROMISE) trial

- Results: Milrinone therapy was associated with
  - 28 percent increase in mortality from all causes
  - 34 percent increase in cardiovascular mortality.
- "Our findings indicate that despite its beneficial hemodynamic actions, long-term therapy with oral milrinone increases the morbidity and mortality of patients with severe chronic heart failure. The mechanism by which the drug exerts its deleterious effects is unknown."

- Evidence that effect on possible surrogate outcome differs from effect on clinical outcome

---

**Surrogate Endpoints**

AIDS trials (systematic review)

- Use of CD4 counts
- HIV Viral load

- Differences between the statistical significance of surrogate and clinical endpoints

---

**Surrogate Endpoints**

<table>
<thead>
<tr>
<th>Disease</th>
<th>Definitive Endpoint</th>
<th>Surrogate Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVD</td>
<td>MI</td>
<td>Cholesterol level</td>
</tr>
<tr>
<td></td>
<td>CHD</td>
<td>Carotid IMT</td>
</tr>
<tr>
<td></td>
<td>Heart Failure</td>
<td>BNP</td>
</tr>
<tr>
<td></td>
<td>Stroke</td>
<td>Blood pressure</td>
</tr>
<tr>
<td>Cancer</td>
<td>Mortality</td>
<td>Tumor size reduction</td>
</tr>
<tr>
<td>Prostate Cancer</td>
<td>Disease progression</td>
<td>PSA</td>
</tr>
<tr>
<td>HIV Infection</td>
<td>AIDS/Death</td>
<td>CD4+ count</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>Vision Loss</td>
<td>Intraocular pressure</td>
</tr>
</tbody>
</table>
Surrogate Endpoints - Advantages

Possible Advantages
- Smaller sample size
- Endpoint earlier than ideal endpoint
- Easier
- Less costly

Surrogate Endpoints - Disadvantages

Possible Disadvantages
- Not well correlated to ideal endpoint
- Mechanism of action unclear
- Less acceptable
- Less clinical relevance
- NO SURROGATE for Safety

Composite Endpoints

Composite event
…considered to have occurred if any one of several different outcomes are observed
- e.g. angina pectoris, transient ischemic attack, or myocardial infarction = composite vascular event
**Composite Endpoints - Advantages**

Possible Advantages
- Increases expected event rate
- Increases power
- Reduces sample size
- Shorter study duration
- Combine benefits and risks
- Reduce bias
- Allow multiple important outcomes

**Composite Endpoints - Disadvantages**

Possible Disadvantages
- Confusion in interpreting results
- Additional ‘noise’ may hide differences
- Correlated events - smaller advantage
- Sample size – “minimum clinically important difference”

**Composite Endpoints – Example 1**

Women’s Health Initiative (WHI)
- **Enrolled:** Postmenopausal women aged 50-79 years with an intact uterus
- **Treatment:** Estrogens plus medroxyprogesterone acetate or placebo
- **Endpoint:** CHD (nonfatal MI and CHD death), with invasive breast cancer as the primary adverse outcome. A global index summarizing the balance of risks and benefits included the 2 primary outcomes plus stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.
- **Results:** Invasive breast cancer exceeded the stopping boundary and global index statistic supported risks exceeding benefits.
Composite Endpoints – Example 2

Term Breech Trial (TBT)

- *General consensus:* planned caesarean section is better than planned vaginal birth for the delivery of the fetus in the breech presentation at term
- *Enrolled:* women with a singleton fetus in a frank or complete breech presentation
- *Treatment:* planned caesarean section or planned vaginal birth.
- *Primary endpoint:* perinatal or neonatal mortality, or serious neonatal morbidity;
- *Secondary endpoint:* maternal mortality or serious maternal morbidity
- *Results:* Composite and individual outcomes significant

Adjudicated Endpoints

- *Subjective Endpoints*
- *Multiple assessments/assessors – then adjudicate*
- *Committee – Equal experience or Experts and non-experts*
- *Where are people located?*
- *Adjudicate in person / e-mail*
- *How often does adjudication happen?*
- *What materials does committee need?*
- *Grade independently or all together?*

Missing Patients (Endpoints)

- *Exclusions (never randomized)*
  - No bias in randomized comparison
  - Does influence interpretation and generalization
- *Withdrawals (deliberately omitted from analysis)*
  - Severe bias may arise
  - Withdrawals may be acceptable if based on eligibility criteria determined at baseline and not affected by events subsequent to randomization
- *Losses to follow-up (missing outcome data)*
  - Bias may arise if the loss is related to the intervention and the outcome
Missing Data

• Treatment dropouts do not necessarily have missing outcome data
  – we should design trials (& informed consent processes) so that treatment modifications and/or dropout do not lead to "off-study"
  – such patients should still be followed for outcome

• Patients who need (or want) to modify their therapy may be prognostically different from those who are maintained on the therapy initially assigned (and this may vary by treatment group)

Intention to Treat (ITT) Analyses

• Include all individuals randomized
• Include in the group to which they were randomized
• Regardless of what treatment they received or what occurs subsequently
• First analysis of any randomized trial
• Supported by the randomization
• Maintains comparability (expectation)

Intention to Treat - 2

• Provides a test of the "policy" ("strategy", "intention")
• Estimate of effectiveness (real world)
  – Efficacy – analyse as treated (ideal world)
• May need to adjust sample size for non-compliance
What is the goal?

- "pragmatic efficacy" [intent-to-treat] or
- "biologic efficacy" [full compliance] *

* may not be attainable (intolerance or toxicity)
  danger of false optimism

* may not be straightforward: danger of bias

---

In equivalence trials, excessive noncompliance may lead to apparent equivalence which does not reflect reality

- here, intent-to-treat analysis does not have the usual advantage of "conservatism"

---

Severe bias may arise if deliberately omitted from analysis

comparing compliers in both groups may be biased;

"as treated" analysis may even be worse

> lose the comparability provided by randomization
Learning Objectives

By the end of the course, attendees should be able to:

• Identify possible endpoints for their study
• Assess the pros and cons for possible endpoints
• Be able to ‘better’ choose endpoints that meet study needs

• Missing Data
• Intention to Treat

References: Surrogate Outcomes


References: Surrogate Outcomes


References: Composite Outcomes


Tugwell Pet al. Powering our way to the elusive side effect: A composite outcome ‘basket’ of predefined designated endpoints in each organ system should be included in all controlled trials. Journal of Clinical Epidemiology 2005;58: 785–790.

Writing Group for the Women’s Health Initiative Investigators. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women. Principal Results From the Women’s Health Initiative Randomized Controlled Trial. JAMA. 2002;288:321-333.
A primary objective of most clinical trials is to demonstrate the effectiveness and safety of a treatment under investigation.

The purpose of such trials is to:
- Find out which (if any) of the treatments are more effective
- Convince others of the results
In designing such trials, we need to keep in mind two issues related to participant (patient) heterogeneity:
- The effect of chance
- The effect of bias (whether conscious or unconscious)

These are addressed by:
- Using randomization for treatment assignment
- Having adequate numbers of participants in study

Hypothesis testing involves:
- Collecting a sample and using the sample to estimate unknown population parameters.
- Comparing the sample estimate(s) to some hypothesized population value to see if the sample came from the specified population.

Hypothesis: Statement about a population parameter
Null Hypothesis ($H_0$): A hypothesis of no difference or status quo; often what we would like to disprove
$$H_0: \mu = 0$$

Alternative Hypothesis ($H_A$): A statement which contradicts the null hypothesis
$$H_A: \mu \neq 0$$

The goal of hypothesis testing is to collect a sample and determine which hypothesis is 'more likely' to have generated the observed sample.
Testing

Test Statistic: A statistic computed from the sample upon which we will base our decision.

Acceptance Region: The range of values for which \( H_0 \) is not rejected.

Rejection Region: The range of values for which \( H_0 \) is rejected.

The test statistic must fall into one of these regions.

The test statistic must fall into one of these regions:

- If the test statistic falls into the rejection region, the test is said to be statistically significant.
- If we don’t reject \( H_0 \), we can’t claim to ‘accept \( H_0 \)’
  - Suppose one makes a statement ‘all swans are white’
  - To examine this statement, a sample of swans is drawn
  - Two things can happen:
    a) All swans in the sample are white
    b) At least one swan in the sample is not white
  - The event (b) establishes the falsehood of statement
  - However, the event (a) does not prove the statement!

Type I Error: Rejecting null hypothesis when true (i.e., conclude benefit when none actually exists)

\[ \alpha = \Pr\{ \text{Type I error} \} = \Pr\{ \text{Reject } H_0 \text{ when true} \} \]

Type II Error: Not rejecting null hypothesis when false (i.e., fail to conclude benefit when actual benefit exists)

\[ \beta = \Pr\{ \text{Type II error} \} = \Pr\{ \text{Fail to reject } H_0 \text{ when false} \} \]
The decision may be summarized as follows:

<table>
<thead>
<tr>
<th>Evidence of Treatment Effect</th>
<th>No Treatment Benefit</th>
<th>Treatment Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error (False Positive)</td>
<td>Correct Result (True Positive)</td>
<td></td>
</tr>
<tr>
<td>Correct Result (True Negative)</td>
<td>Type II error (False Negative)</td>
<td></td>
</tr>
</tbody>
</table>

Statistical tests quantify the probability of a type I error (false positive result).

For example, an observed difference with $p \leq 0.01$ implies that the probability of obtaining a difference this extreme (or more so) by chance alone is less than or equal to 1%.

There is a tradeoff between the probability of a type I error and a type II error. Traditionally, type I errors are of greater concern. Hence, we often fix $\alpha$ at 0.05 and try to take a large enough sample to ensure $\beta$ is at a reasonable level ($<0.20??$). Should this always be the case?
Example (from Rosner, p. 193-194):

It has been suggested that a certain hospital has lower birthweight babies than the national average. To see if a special care nursery is needed, a sample of birthweights from the hospital are collected and used to test:

\[ H_0: \mu \geq \text{national average} \quad \text{vs.} \quad H_\alpha: \mu < \text{national average} \]

If \( H_0 \) is rejected, the hospital will add a special care nursery.

- If a type I error is made, the extra cost of adding a special care nursery will be recommended when it is not needed
- If a type II error is made, a needed special care nursery will not be funded.
  - As a result, some low-birthweight babies may not receive the special attention that they need

A confidence interval quantifies the uncertainty around the estimated intervention effect. CI's also indicate the range of values within which we think the true intervention effect lies.

Relationship between CI's and hypothesis tests:

- A \((1-\alpha)\%\) confidence interval for \(\mu\) consists of all values for which \(H_0\) could not be rejected at the \(\alpha\) level.
TESTING

<table>
<thead>
<tr>
<th>Jury Trial (criminal law)</th>
<th>Clinical Trial (statistical testing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Presume innocent</td>
<td>• Assume the null hypothesis</td>
</tr>
<tr>
<td>• Goal: Convict the guilty</td>
<td>• Goal: Detect a true difference (Reject the null hypothesis)</td>
</tr>
<tr>
<td>• “Beyond reasonable doubt”</td>
<td>• “Level of significance”</td>
</tr>
<tr>
<td>• Requires evidence: Convincing testimony</td>
<td>• Requires evidence: Adequate sample size</td>
</tr>
<tr>
<td>• Mistake: Convict an innocent person</td>
<td>• Mistake: False positive (Type I error)</td>
</tr>
</tbody>
</table>

Acknowledgement to Susan Hilsenbeck and Sylvan Green

POWER

A primary objective of most clinical trials is to demonstrate the effectiveness and safety of a treatment under investigation.

Hence, sample size calculation plays an important role at the planning stage to ensure sufficient subjects for answering the question of interest.

If sample size is too large, study will waste resources

If sample size is too small, study underpowered and a potentially useful treatment may be discarded.

POWER

Sample size calculation is usually performed based on some statistical criteria controlling Type I and/or Type II errors.

With a fixed sample size:

- $\alpha$ increases as $\beta$ decreases
- $\alpha$ decreases as $\beta$ increases

The only approach to decrease both $\alpha$ and $\beta$ is to increase the sample size.
Two common approaches to choosing sample size:

- **Precision Analysis**: Sample size chosen such that there is a desired precision at a fixed confidence level (i.e., fixed Type I error)
- **Power Analysis**: Sample size chosen to achieve desired power for detecting clinically/scientifically meaningful difference at a fixed Type I error rate.

In this course, we focus on sample size calculation based on a power analysis for various situations in clinical trials.

---

Power of the test is defined as the probability of correctly rejecting the null hypothesis when false.

- Power = 1 - β
  
  = Pr{ Reject H₀ when false }
A valid sample size calculation MUST be based on appropriate tests for hypotheses that reflect study objectives under a valid study design. Hence, it is important that the following are aligned:

- Study Objective (Hypothesis)
- Study Design
- Statistical Analysis (Test Statistic)
- Sample Size Calculation

Any discrepancies between these items can distort the validity and integrity of the trial.

What must be known to compute sample size?

1. Type of data
2. Type of test
3. Measure of precision or variability
4. The magnitude of treatment difference that the study should be able to detect (δ)
5. Specified Type I error (α)
6. Target Power [or specified Type II (β) error]

Type of Data:

- Dichotomous responses (success or failure; presence or absence)
- Continuous responses (blood pressure; length of hospitalization)
- Time to event responses (time to occurrence of a clinical event)

Sample size estimates for response variables that do not fall into these categories can usually be approximated by one of them!
POWER

Type of Test:

- **Test for Equality**: Show one treatment is more effective than another
  
  \[ H_0: \delta = 0 \text{ vs. } H_A: \delta \neq 0 \]

- **Test for Superiority**: Show test drug is more effective than an active agent or standard therapy
  
  \[ H_0: \delta \leq \varepsilon \text{ vs. } H_A: \delta > \varepsilon \]
  
  where \( \varepsilon \) is the *superiority* margin.

- **Test for Non-inferiority**: Show test drug is as effective as an active agent or standard therapy
  
  \[ H_0: \delta \leq -\varepsilon \text{ vs. } H_A: \delta > -\varepsilon \]
  
  where \( \varepsilon \) is the *non-inferiority* margin.

- **Test for Equivalence**: Show no difference of clinical importance between two treatments
  
  \[ H_0: |\delta| \geq \varepsilon \text{ vs. } H_A: |\delta| < \varepsilon \]
  
  where \( \varepsilon \) is the *equivalence* margin.

Important to ensure that the sample size calculation parallels the planned primary analysis.

The hypothesis of interest should be clearly stated when performing a sample size calculation.

Each of the above hypotheses has a different sample size requirement in order to achieve a desired power for the corresponding test.

For this course, we will primarily focus on tests of equality between two treatments.
POWER

Precision and Variance:

- A more precise method of measurement (i.e. small $\sigma$) will permit detection of any given $\delta$ with a smaller sample size.
- The importance of precision increases as the desired size of the effect becomes smaller.
- A study with a small sample size will have more uncertainty and will only show statistically significant differences if there is a large difference between the two groups.

POWER

Treatment Effect:

- The choice of $\delta$ is critical for study planning
- Different choices of $\delta$ have major effects on the sample size requirements.
- If $\delta$ is small, a large sample size will be required
- Important to ensure the treatment effects have both clinical and statistical meaning
- Possible to design study to detect reduction of onset time of local anesthesia from 60 to 59 seconds, but likely not of clinical importance.

POWER

Type I Error (Significance level):

- Pre-set by researchers early in study planning
- Common $\alpha$ values are 0.01, 0.05, and 0.10.
- Often, choose $\alpha = .05$ more by convention than design
- This implies that we would expect to reject the null hypothesis 5% of the time when it is true (there is no effect).
- May need to adjust for multiple testing
POWER

Power:
- Typically set at 80% or 90% for planning purposes
- Power curves are useful since study planning often involves a trade-off between desired sample size, cost, and patient resources.

Power (cont.):
- Power curves typically have a sigmoidal shape, with increasing power as n or \( \delta \) increases.
- Impact of small changes in design parameters depends on shape of power curve.
- If trial design lies near shoulder, small changes in design parameters can seriously affect power.
- Typically, trials designed with 80% power are more susceptible to inaccuracies in design parameters than trials designed with 90% power.

To determine power, we need to specify:
- The sample size - \( N \)
- The significance level - \( \alpha \)
- A clinically important difference that we wish to detect - \( \delta \)
- Any additional nuisance parameters
To determine sample size, we need to specify

- Target power - $P_t = 1 - \beta$
- The significance level - $\alpha$
- A clinically important difference that we wish to detect - $\delta$
- Any additional nuisance parameters

Sample size estimates are approximate:

- Equations often based on approximations to the exact statistical distributions.
- Parameters used in calculations are guesses and have an element of uncertainty

Researchers hope that any errors are small and that the computed sample size is close to the actual number truly needed.

Be conservative (but realistic – always round up!) when estimating sample size!

Small changes in design parameters may yield large changes in the required sample size.

Required sample size increases with:

- Variance of the treatment difference
- Decreasing type I error
- Increasing desired target level of power
- Smaller treatment effects of interest
Note that we cannot separate power from either size of study or magnitude of treatment effect. Hence, the following statement is ambiguous:

"The trial has 90% power."

All three values must be discussed simultaneously:

"With 500 subjects per group, the trial has 90% power to detect a decrease of 10 mmHg in blood pressure due to the new treatment at the 5% significance level."

Sample size calculation provides the number of *evaluable* subjects required for achieving a desired level of power.

If drop-outs are expected, the sample size should be adjusted upward to ensure a sufficient number of evaluable subjects.

If the response variable can be partially explained by other covariates, the required sample size may be reduced.

Suppose that there are two groups of observations:

\[ x_i, i = 1, \ldots, n_1 \text{ (treatment)} \]
\[ y_i, i = 1, \ldots, n_2 \text{ (control)} \]

Assume that \( x_i \) and \( y_i \) are independent and normally distributed with means \( \mu_1 \) and \( \mu_2 \), respectively, and a common variance, \( \sigma^2 \).
Suppose the hypothesis of interest is:

\[ H_0: \mu_1 = \mu_2 \text{ vs. } H_A: \mu_1 \neq \mu_2 \]

Assuming equal variance and equal sample sizes in the two groups, use the test statistic:

\[ Z = \frac{\bar{x} - \bar{y}}{\sigma \sqrt{2/n}} \]

Under the null hypothesis of no treatment effect:

\[ Z \sim N(0,1) \]

Hence, we reject the null hypothesis when:

\[ |Z| > z_{\alpha/2} \]

Under alternative hypothesis that \( \mu_1 = \mu_2 + \delta \) (where \( \delta \) is a clinically meaningful difference), the distribution is centered away from 0. Power is the area under the alternative distribution that lies in the rejection region.
CONTINUOUS RESPONSE

For given $\alpha$, $\beta$, $\delta$, and $\sigma$, the total required sample size is given by:

$$2N = \frac{4(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2} \sigma^2$$

NOTE: This formula is based on a normal (not a t) distribution and assumes either $\sigma$ is known or $N$ is large enough to make this assumption valid.

VI-43

CONTINUOUS RESPONSE

Example:

- In a study of a new diet to reduce cholesterol, a 10 mg/dl difference would be clinically significant.
  \[ \delta = 10 \]
- From other data, $\sigma$ is estimated to be 50 mg/dl.
  \[ \sigma = 50 \]
- We want a two-sided test with equal sample sizes, $\alpha = 0.05$, and we desire 90% power.
  \[ Z_{\alpha/2} = 1.96, \ Z_{\beta} = 1.28 \]

VI-44

CONTINUOUS RESPONSE

Substituting those values into the formula gives:

$$2N = \frac{4(1.96 + 1.28)^2 (50)^2}{(10)^2} = 1049.8$$

Rounding up yields a required sample size of $2N = 1050$, or $N = 525$ in each group.

VI-45
CONTINUOUS RESPONSE

How different would the required sample size be if \( \sigma \) were actually 60:

\[
2N = \frac{4(1.96 + 1.28)^2 (60)^2}{(10)^2} = 1511.7
\]

Rounding up yields a required sample size of \( 2N = 1512 \), or \( N = 756 \) in each group.

This is a big difference in the required sample size considering the relatively small increase in \( \sigma \).

Be conservative in estimates of \( \sigma \)!!

---

DICHOTOMOUS RESPONSE

Compare Drug A (standard) vs. Drug B (new)

\( p_A = \) Proportion of failures expected on drug A

\( p_B = \) Proportion of failures on drug B which one would want to detect as being different

We want to test

\[ H_0: p_A = p_B \text{ vs. } H_A: p_A \neq p_B \]

With significance level \( \alpha \) and power = 1 - \( \beta \) to detect a difference of \( \delta = p_A - p_B \).

---

DICHOTOMOUS RESPONSE

The estimates of \( p_A \) and \( p_B \) are:

\[
\hat{p}_A = \frac{r_A}{N} \quad \text{and} \quad \hat{p}_B = \frac{r_B}{N}
\]

With \( r_A \) and \( r_B \) the number of events in the two groups and \( N \) the number of subjects in each group.

The usual asymptotic test statistic is:

\[
Z = \frac{(\hat{p}_A - \hat{p}_B)}{\sqrt{\frac{1}{N} \left( \hat{p}(1 - \hat{p}) \right)}}
\]

where \( \hat{p} = (\hat{p}_A + \hat{p}_B)/2 \)
The total sample size required (N in each group) is:

\[ N = \frac{Z_{\alpha/2} \sqrt{2 \bar{p}(1-\bar{p})} + Z_{\beta} \sqrt{p_s(1-p_s) + p_b(1-p_b)}}{(p_A - p_B)^2} \]

where \( \bar{p} = (p_A + p_B)/2 \) and \( Z_{\alpha/2} \) and \( Z_{\beta} \) are critical values of the standard normal distribution.

In general, the variance is largest when \( p = 0.5 \) and smallest when \( p \) is near 0 or 1.

Hence, larger sample sizes are required to detect a change in \( p_A - p_B \) when \( p_A \) and \( p_B \) are near 0.5.

Example:
- In a clinical trial, the cure rate for the active control agent is assumed to be 65%.
  \( p_A = 0.65 \)
- We want to detect an increase of 20% in cure rate.
  \( p_B = 0.85 \rightarrow \delta = (0.85 - 0.65) = 0.20 \)
- We want a two-sided test with equal sample sizes, \( \alpha = 0.05 \), and 80% power.
  \( Z_{\alpha/2} = 1.96, \ Z_{\beta} = 0.84 \)
DICHOTOMOUS RESPONSE

Substituting those values into the formula gives:

\[ n = \frac{z_{0.025}(1-\beta) + z_{0.025}(1-\beta)}{\beta^2} \]

\[ = \left[ 1.96\sqrt{2(0.75)(1-0.75)} + 0.84\sqrt{0.65(1-0.65)} + 0.85(1-0.85) \right]^2 \]

\[ = 73 \]

Hence, we require a total sample size of 73 in each group (146 total).

TIME TO EVENT RESPONSE

Proportional Hazards Model (Two Groups)

\[ h(t, x_i) = h_0(t) \cdot \exp(\beta x_i) \]

\( x_i = 1 \), if new treatment
\( = 0 \), if standard treatment

Hazard for person \( i \) at time \( t \) is a function of:

• \( h_0(t) \): the hazard for those on the standard treatment, \( i.e. \ x_i = 0 \)
• A linear function of group membership (\( x_i \))

From this model, the hazards for subjects in the two treatment groups are:

Standard Treatment (\( x_i = 0 \)): \( h(t, 0) = h_0(t) \)

New Treatment (\( x_i = 1 \)): \( h(t, 1) = h_0(t) \cdot \exp(\beta) \)

Hence, to compare the hazards for an individual on the new treatment vs. one on the standard treatment:

\[ HR = \frac{h(t, 1)}{h(t, 0)} = \frac{h_0(t) \cdot \exp(\beta)}{h_0(t)} = \exp(\beta) \]
Hence, a unit increase in \( x \) multiplies the hazard by an amount that is constant over time:

\[
HR = \exp(\beta)
\]

Hence, the log-hazard ratio (\( \beta \)) is an unknown coefficient that describes the way survival time is affected by the covariate:

- \( \beta = 0 \): no effect
- \( \beta > 0 \): survival is worse with new treatment
- \( \beta < 0 \): survival is better with new treatment

Thus, a test of difference in survival times for the two groups corresponds to a test of:

\[
H_0: \beta = 0.
\]

We will compute the required sample size based on the log-rank test. However, the log-rank test is equivalent to the score test from a Cox regression model with a single dichotomous covariate.

In order to compare the groups we need to have a reasonable number of events, NOT total observations. Hence, sample size calculations for comparing two survival curves consists of a two step process:

1) Calculating the Required Number of Events
2) Calculating the Required Number of Patients

Furthermore, the required sample size depends on the accrual and follow-up time for the study.
To determine the required number of events, we need to specify:

- $\beta$ = Effect (log HR) we wish to detect
- $\alpha$ = Significance level used for test
- $P$ = Target power
- $\pi_1$ = Proportion of observations in group 1

The required number of events for a given study is then given by:

\[
\text{required # of events} = \frac{\left(\frac{z_{\alpha/2} + z_P}{\pi_1(1 - \pi_1)}\right)^2}{\beta^2}
\]

To calculate the required number of patients to be enrolled, we need to consider the probability of the event over the course of the study. Once probability of the event has been determined, the required number of subjects can be found from:

\[
\text{required sample size} = \frac{\text{required # of events}}{\text{Pr}\{\text{event}\}}
\]

Subgroup analyses refer to analyses using a subset of participants in a trial. For example, in a trial comparing treatment to placebo, we may be interested in assessing the treatment effect in men and women separately.

Subgroup analyses are important for several reasons:

- Clinical decision making
- Regulatory requirements
- Hypothesis generating
SUBGROUP ANALYSES
Potential Problems with Subgroup Analyses:
- Insufficient power
  - Trials powered to detect an overall treatment effect will be underpowered to detect similar effects in subgroups
- Multiple comparisons
  - “If you torture the data long enough, eventually it will confess to anything”
Whenever possible, important subgroup analyses should be defined in the protocol a priori.

SOFTWARE
Software for power calculations (among many):
- Commercial packages:
  - SAS (PROC POWER)
  - NCSS PASS
  - NQuery
- Free packages:
  - Dr. Russell Lenth’s website: http://www.stat.uiowa.edu/~rlenth/Power/index.html
  - PS: Power and sample size calculation http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize

INTERIM MONITORING
Data and Safety Monitoring Boards (DSMBs) are often given the responsibility of monitoring the accumulating data.
The DSMB is responsible for assuring that study participants are not exposed to unnecessary or unreasonable risks.
The DSMB is also responsible for assuring that the study is being conducted according to high scientific and ethical standards.
INTERIM MONITORING

Why have DSMBs?

- Protect safety of trial participants
- Investigators are in a natural conflict of interest
  - Vested in the study
  - They, and their staff, are paid by the study
- Having the DSMB externally review efficacy and safety data protects:
  - The credibility of the study
  - The validity of study results

INTERIM MONITORING

Principle 1 – Composition. The DSMB should have multidisciplinary representation, including topic experts from relevant medical specialties and biostatisticians.

Principle 2 - Conflicts. Individuals with important conflicts of interest (financial, intellectual, professional, or regulatory) should not serve on a DSMB.

Principle 3 – Confidentiality Issues. Trial integrity requires DSMB members not to discuss details of meetings elsewhere.

INTERIM MONITORING

DSMB’s should periodically review study data.

The study protocol should include a section describing proposed plan for interim data monitoring. This plan should detail:

- What data will be monitored?
- The timing of all interim analyses?
- The frequency of data reviews.
- Criteria that will guide early termination

INTERIM MONITORING

- Protect safety of trial participants
- Investigators are in a natural conflict of interest
  - Vested in the study
  - They, and their staff, are paid by the study
- Having the DSMB externally review efficacy and safety data protects:
  - The credibility of the study
  - The validity of study results

INTERIM MONITORING

Principle 1 – Composition. The DSMB should have multidisciplinary representation, including topic experts from relevant medical specialties and biostatisticians.

Principle 2 - Conflicts. Individuals with important conflicts of interest (financial, intellectual, professional, or regulatory) should not serve on a DSMB.

Principle 3 – Confidentiality Issues. Trial integrity requires DSMB members not to discuss details of meetings elsewhere.

INTERIM MONITORING

DSMB’s should periodically review study data.

The study protocol should include a section describing proposed plan for interim data monitoring. This plan should detail:

- What data will be monitored?
- The timing of all interim analyses?
- The frequency of data reviews.
- Criteria that will guide early termination

INTERIM MONITORING

Principle 1 – Composition. The DSMB should have multidisciplinary representation, including topic experts from relevant medical specialties and biostatisticians.

Principle 2 - Conflicts. Individuals with important conflicts of interest (financial, intellectual, professional, or regulatory) should not serve on a DSMB.

Principle 3 – Confidentiality Issues. Trial integrity requires DSMB members not to discuss details of meetings elsewhere.

INTERIM MONITORING

DSMB’s should periodically review study data.

The study protocol should include a section describing proposed plan for interim data monitoring. This plan should detail:

- What data will be monitored?
- The timing of all interim analyses?
- The frequency of data reviews.
- Criteria that will guide early termination
Typical agenda of initial DSMB meeting:
- Develop and agree on charter
- Approve the protocol

Frequency of additional DSMB meetings depends on disease topic and specific intervention.
For trials funded by NIH, most DSMB’s meet twice per year – once by conference call and one in-person meeting.

Early DSMB meetings almost exclusively focus on:
- Quality of conduct (recruitment, timeliness of data entry, etc.)
- Trial integrity (protocol adherence, etc.)

As more data accrue, DSMB meetings will focus on safety issues as well.
Later DSMB meetings may include formal efficacy or futility analyses.

A typical agenda for a DSMB meeting:
- Closed executive session
  - Review of agenda, additions to agenda
- Open session with investigators
  - Review current status and conduct of study
  - Accrual update
- Closed session with unblinded investigators
  - Review safety data
  - Review interim analysis (if appropriate)
- Closed executive session
- Open session with investigators
  - Discussion/Recommendations
INTERIM MONITORING

Typically, at end of each meeting the DSMB makes one of the following recommendations to the sponsor:

- Study should continue without modification
- Study should continue with the following modifications
- Study should be temporarily stopped until a specific issue is addressed
- Study should be stopped for safety/efficacy/utility

INTERIM MONITORING

At end of each meeting, DSMB also summarizes any areas of concern regarding performance and/or patient safety.

Soon thereafter, the DSMB chair will provide a written summary of the board’s recommendations.

These letters are extremely important for IRB submissions at each individual site.

INTERIM MONITORING

Ethical principles mandate that clinical trials begin with uncertainty as to which treatment is better. (clinical equipoise)

This uncertainty should be maintained during study.

If interim data become sufficiently compelling, ethics would demand that the trial stop and the results made public.

Hence, interim monitoring of safety and efficacy data has become an integral part of modern clinical trials.
**INTERIM MONITORING**

Early termination of a trial should be considered if:

- Interim data indicate intervention is harmful
- Interim data demonstrate a clear benefit
- Significant difference by end of study is probable
- No significant difference by end of study probable
- Severe logistical or data quality problems exist

---

**INTERIM MONITORING**

The decision to stop a trial early is complex, requiring a combination of statistical and clinical judgment.

Stopping a trial too late means needlessly delaying some participants from receiving the better treatment.

Stopping a trial too early may fail to persuade others to change practice.

Statistical methods have been developed for interim monitoring of clinical trials to minimize the role of subjective judgment.

---

**EFFICACY MONITORING**

Consider a clinical trial to compare two normally distributed groups with $K$ interim analyses.

The objective of the trial is to test the null hypothesis of no treatment effect at each interim analysis:

$$H_0: \delta = 0 \text{ vs. } H_A: \delta \neq 0$$

where $\delta$ equals difference between treatment means.

At each interim analysis, the null hypothesis is tested using the test statistics $Z_1, \ldots, Z_K$ ($Z$-statistic for all data observed up to time of $k$th interim analysis)
EFFICACY MONITORING

Under $H_0$ (no difference between groups), repeated testing at level $\alpha$ inflates the probability of making at least one type I error.

Even 5-10 tests can lead to serious misinterpretation of trial results.

<table>
<thead>
<tr>
<th># of tests</th>
<th>True type I error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td>0.14</td>
</tr>
<tr>
<td>10</td>
<td>0.19</td>
</tr>
<tr>
<td>20</td>
<td>0.25</td>
</tr>
<tr>
<td>1000</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Solution is to adjust stopping boundaries in such a way to ensure that overall type I error is equal to $\alpha$:

- Pocock (1977): Same critical value at each interim look
- O’Brien & Fleming (1979): Nominal significance levels needed to reject $H_0$ increase as study progresses.
- Haybittle (1971) & Peto et al. (1976): Reject $H_0$ if $|Z_k| \geq 3$ for all interim tests ($k < K$)

A comparison of the critical values for the Pocock, O’Brien-Fleming, and Haybittle-Peto methods for $k = 5$ looks and $\alpha = 0.05$ is given below:
EFFICACY MONITORING

There is a slight loss of power with multiple testing.
To account for this, sample size calculations must adjust the sample size upward.
This is accomplished by the following process:
- Compute the required sample size under a fixed sample design.
- Multiply this sample size by an appropriate ratio to account for the multiple testing.

EFFICACY MONITORING

The original methodology for group sequential boundaries required that the number and timing of interim analyses be specified in advance.
DSMB’s sometimes may require more flexibility as beneficial or harmful trends emerge.
Lan & DeMets (1983, 1989) proposed an ‘alpha spending function’ which provides more flexible group sequential boundaries.
The approach lends itself well to the accommodation of irregular, unpredictable, and unplanned interim analyses.

FUTILITY MONITORING

Power tells whether a clinical trial is likely to have high probability to detect a pre-defined treatment effect of interest.
Very low power implies that a trial is unlikely to reach statistical significance even if there is a true effect.
One should never begin a trial with low power.
However, sometimes low power becomes apparent only after a trial is well under way.
FUTILITY MONITORING

Stochastic curtailment uses the concept of conditional power:

\[ P_k(\theta) = Pr\{ \text{reject } H_0 \mid \theta \text{ and observed data so far} \} \]

Initially, when \( k = 0 \), this is the usual power function.

At the planned termination of the study (stage \( K \)), this probability is either 0 or 1.

At interim stage \( k \), conditional power depends on \( \theta \).

May want to stop trial for futility if the conditional power drops below some specified level (i.e., 20%).

FUTILITY MONITORING

If early results show:

- Intervention better than expected
  \( \rightarrow \) conditional power high
- Intervention worse than expected
  \( \rightarrow \) conditional power low
  (unless sample size increased)

Group sequential methods focus on existing data.
Stochastic curtailment methods consider future data.

FUTILITY MONITORING

Clearly, the futility rule is heavily influenced by the assumed value of the treatment difference, \( \theta \).

Making an overly optimistic assumption about \( \theta \) delays decision to terminate the trial.

Several options for the value of \( \theta \) have been proposed:

- Lan, Simon, & Halperin (1982): Evaluated at value of \( \theta \) corresponding to alternative hypothesis.
- Evaluated under the null hypothesis.
- Evaluated at the observed treatment effect.
FUTILITY MONITORING
One limitation of conditional power is that no adjustment is made to account for associated prediction error if observed treatment effect is used. Interim futility monitoring may also be conducted using other approaches:
- Predictive Power: Mixed Bayesian-Frequentist approach
- Predictive Probability: Bayesian approach

SOFTWARE
Software packages for group sequential methods:
- S+SeqTrial (Insightful Corporation)
- EaST (Cytel)
- PEST 4 (University of Reading)
- LanDeM (University of Wisconsin)
- SAS (through the use of Macros)

ADAPTIVE DESIGNS
There may be limited information to guide initial choices for study planning. Since more knowledge will accrue as the study progresses, adaptive designs allow these elements to be reviewed during the trial. An adaptive design allows for changing or modifying the characteristics of a trial based on cumulative information.
ADAPTIVE DESIGNS

Adaptive designs are NOT new.

The broad definition includes topics such as group sequential designs and covariate adaptive randomization techniques.

However, because this is a rapidly expanding area of research, more practical experience is needed.

Both Bayesian and Frequentist approaches should be considered.

SUMMARY

The size of a study should be considered early in the planning phase.

**Fundamental Principle:** Clinical trials should have sufficient statistical power to detect differences between groups considered to be of clinical interest.

Therefore, calculation of sample size with provision for adequate levels of significance and power is an essential part of planning.

SUMMARY

There are a variety of approaches for interim monitoring of clinical trial data.

The relationship between clinical trials and practice is very complex, and this complexity is evident in the data monitoring process.

The appropriate monitoring plan depends on the goals of the trial.
Because of the repercussions of stopping a trial early, the decision to stop a trial is complex and requires both statistical and clinical judgment. Hence, these methods should not be used as a sole basis in the decision to stop or continue a trial. Other considerations that play an important role in decision making process cannot be fully addressed within the statistical sequential testing framework.