COURSE 4: STUDY DESIGN II: PHASE III TRIALS & ETHICAL ISSUES

PHASE III TRIAL DESIGNS.................................................................44
  Randomized Controlled Clinical Trials (Parallel Design) ..........44
  Uncontrolled Trials (Single Treatment).................................47
  Trials with Historical Controls .............................................48
  Nonrandomized Concurrent Trials ........................................49
  Crossover Designs ..................................................................49
  Withdrawal Studies ..................................................................52
  Factorial Designs ...................................................................52
  Group Allocation Designs .....................................................53
  Studies of Equivalency ..........................................................54

ETHICAL ISSUES ...........................................................................55
  Basic Principles .......................................................................55
  Medical Research Combined with Professional Care ..............56
  Implementing Ethical Guidelines ............................................56
  Informed Patient Consent (as of 1983) .....................................56

BASIC STUDY DESIGN REVIEW ....................................................57
PHASE III TRIAL DESIGNS

Randomized Controlled Clinical Trials (Parallel Design)

Best design:
- **New** intervention compared with a **control**
- Treatment assignments made **randomly**

Randomization:
- Removes **bias** in subject allocation to treatments
- Tends to produce comparable groups (w.r.t. risk factors)
- Allows valid statistical tests to be performed

Randomized clinical trials are the standard to which other designs are compared.

**Example of bias reduction with randomization:**
Clinical trials on the use of anticoagulant therapy in patients with acute MI (Chalmers et al, 1977)

<table>
<thead>
<tr>
<th></th>
<th># of studies</th>
<th># significant at p ≤ 0.05</th>
<th>Estimated reduction in total mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-randomized</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical controls</td>
<td>18</td>
<td>15</td>
<td>50%*</td>
</tr>
<tr>
<td>Concurrent controls</td>
<td>8</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td>6</td>
<td>1</td>
<td>20%*</td>
</tr>
</tbody>
</table>

* This difference is probably due to biases in the non-randomized studies.

**Selection bias** can lead historically controlled studies to inappropriately favor the new intervention.

However, small sample sizes in randomized trials lead to missing benefits of new treatments that truly exist.

**Example of a randomized controlled trial**
A clinical trial for primary breast cancer (1972)

**Study question**
Does L-Pam prolong the disease-free interval of primary breast cancer patients after radical mastectomy? (This is a roughly-stated research question. Specifics are given below.)
Randomized treatments

→ L-Pam (orally, 0.15 mg/kg body weight, 5 consecutive days every 6 weeks for 2 years, with specific dose modifications for hematological toxicity)

Mastectomy →

→ Placebo (physically indistinguishable from L-Pam)

Eligibility

- Radical mastectomy for primary breast cancer (within 4 weeks of starting protocol treatment)
- Histologically confirmed axillary node involvement
- No skin ulceration or peau d’orange
- Age ≤ 75 years
- Not pregnant or lactating

(Focus is on those patients most likely to benefit.)

Endpoints

- Primary endpoint: disease-free interval = time from mastectomy until first detection of tumor (local, regional, or distant)
- Other endpoints:
  - Survival time = time from mastectomy until death
  - Toxicity = occurrences of hematological toxicity or nausea/vomiting

The study

- A written protocol documented all study procedures and information. Sample size requirements dictated several hundred patients ⇒ a multicenter trial (37 hospitals) was undertaken.

- Randomization was performed by phoning the central office. Patients were stratified by age (< 50, ≥ 50), nodal status (1 to 3 vs. 4+ positive axillary nodes), and institution. (Stratification assures that groups are balanced w.r.t. treatment for each of the stratification factors.)

- This trial was double blind (neither patient nor physician knew which treatment was being given; drug containers were anonymous).

- Follow-up exams were performed every 6 weeks, with test for hematological toxicity every 3 weeks. ⇒ endpoint evaluation was done consistently and objectively.
• Patient accrual started September 1972 and ended February 1975. In total, 370 were accrued.

• **Informed consent** was obtained from all patients.

• The **trial committee** reviewed the study regularly. After the first few months they decided to relax eligibility w.r.t. # of positive nodes.

• The **central coordinating office** (NSABP in Pittsburgh) supervised data collection and processing.

• As the data accumulated, **interim analyses** were performed. There was pressure to release interim results.

**Results**

• The final results showed that L-Pam significantly increased the disease-free survival time \( p = 0.009 \).

• **Subgroup analyses** revealed that L-Pam had the largest benefit in young patients \( \leq 50 \) years old.

• **Toxicity**: Hematological toxicity was common \( (> 25\%) \), but never life-threatening.

• Nausea/vomiting was experienced by 40% of L-Pam patients and 11% of placebo patients.

**Conclusion**

L-Pam was adopted as the new “standard” treatment.

**Run-in periods**

Before patients enter a clinical trial, a run-in (or lead-in) period of placebo with no active treatment is usually employed prior to randomization. The inclusion of a run-in period prior to the active treatment has the following advantages and disadvantages.

**Pros**

• It acts as a washout period to remove effects of previous therapy.
• It can be used to obtain baseline data and to evaluate if patient fulfills study entry criteria.
• It can be used as a training period for patients, investigators, and their staff.
• It helps in identifying placebo responders.
• It provides useful information regarding patient compliance.
• It can be used to estimate and compare the magnitude of possible placebo effects between groups.

In many clinical trials, it is not uncommon to observe the placebo effect for many drug products. For example, for antidepressant agents, an intensive care period may significantly improve the patients’ depression without any drug treatment. At the end of the active treatment period, it is important to determine whether the observed significant effect is due to the placebo or treatment. To eliminate the possible placebo effect, it is suggested that a run-in period be included to establish patient comparability between treatment groups at baseline, and this helps to remove placebo effect from the comparison at the endpoint evaluation.

Cons
• A run-in period may not be suitable for patients whose conditions are acute and require immediate treatment or for cancer patients.

Uncontrolled Trials (Single Treatment)

In uncontrolled trials, huge differences in response rates may be found.

For example, in 6 trials of 40 to 150 patients each using the drug 5-FU to treat advanced bowel cancer, the response rates varied from 11% to 55% (Moertel and Reitmeier, 1969).

Reasons for differences:
• Different patient characteristics
• Random variation

Uncontrolled studies are more likely than properly controlled studies to lead to enthusiastic recommendations of the treatment.

For example, of 52 published uncontrolled trials in psychiatry, 85% reported a treatment success. Of 20 published trials with a control group, only 25% reported a treatment success (Fouls, 1958).
Trials with Historical Controls

In trials with historical controls, a new treatment is used in a series of subjects; the outcome is compared with previous series of comparable subjects.

Pros
- Rapid, inexpensive, good for initial testing of new treatment

Cons
- Vulnerable to bias. Changes in outcome over time may come from:
  - Change in underlying patient populations
  - Change in criteria for selecting patients
  - Change in patient care and management peripheral to treatment
  - Change in diagnostic or evaluating criteria
  - Change in quality of data available

Studies with historical controls tend to exaggerate the value of a new treatment.

Control groups taken from the literature are a particularly poor choice.

Even historical controls from a previous trial in the same institution or organization may be problematic. Pocock (1977) looked at 19 cases where the same treatment was used in two consecutive multicenter trials. Differences in survival for comparable groups ranged from -46% to +24%. Four differences were statistically significant.

Covariate analysis can be used to adjust for patient selection, but all other biases will remain.
Nonrandomized Concurrent Trials

If treatment assignment is made by date of birth, date of presentation, SSN, or alternate assignment, such that the investigator knows in advance which treatment the patient will receive, s/he may choose not to enroll a patient.

⇒ selection bias

Such bias will not occur if all patients are automatically enrolled in the study.

If assignment is made by institution, e.g.,

Institution #1 ⇒ treatment A
Institution #2 ⇒ treatment B

or, similarly, assignment by hospital ward or other unit, bias may arise from other differences in the two environments (personnel, procedures, equipment, etc.).

Crossover Designs

<table>
<thead>
<tr>
<th>Patient</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Each patient is randomly assigned to “A, then B” or “B, then A.”

Pros
• Each subject serves as own control ⇒ variability reduced.

Cons
• Condition must be chronic (e.g., high b.p., arthritis) (either “cure” or “death” before the second treatment would ruin the design).

This design must assume no carryover (residual) effect of the first treatment. The statistical test for carryover has low power.

Higher-order crossover designs

Note that in the crossover design, the number of treatments to be compared does not necessarily have to be equal to the number of periods. One example is a 2x3 crossover design for comparing two treatments, as illustrated in the
In this design there are two treatments, but three periods. Patients in each sequence receive one of the treatments twice.

### Two-sequence Dual Crossover Design

<table>
<thead>
<tr>
<th>Period</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>→</td>
<td>Sequence 1:</td>
<td>A</td>
<td>washout</td>
</tr>
<tr>
<td>Patients → Randomization →</td>
<td>→ Sequence 2:</td>
<td>B</td>
<td>washout</td>
</tr>
</tbody>
</table>

### Types of higher-order crossover designs

- Balaam’s design
  - AA
  - BB
  - AB
  - BA

- Two-sequence dual design
  - ABB
  - BAA

- Doubled (replicated) design
  - AABB
  - BBAA

- Four-sequence design
  - AABB
  - BBAA
  - ABBA
  - BAAB

- Williams’ design with three treatments
  - ACB
  - BAC
  - CBA
  - BCA
  - CAB
  - ABC

- Williams’ design with four treatments
  - ADBC
  - BACD
  - CBDA
  - DCAB
Example of joint application of parallel-group and crossover design

761 men screened

255 with serum cholesterol \( \geq 232 \) mg/dL

188 willing to continue

120 eligible

Placebo run-in period (4 to 6 weeks)

60 randomized

**Explanation:** As little or no information regarding the interaction between diet and statins (3-hydroxy-3-methylgluaryl coenzyme A [HMG-CoA] reductase inhibitors) was available in the literature, Jula et al. (2002) reported a randomized, controlled crossover trial to evaluate the separate and combined effects of diet and simvastatin therapy on serum levels of lipids, lipoproteins, antioxidants, and insulin.

This study employed a two-group parallel design and the standard 2 x 2 crossover design to investigate the joint effects of statins and diet, as seen
above. The two-group parallel design in this study is to compare two diets, whereas the effects of statins versus placebo were investigated through the standard 2 x 2 crossover design. This is also an example of factorial trials, which will be discussed later.

**Withdrawal Studies**

To evaluate the duration of benefit of an intervention already known to be useful.

(Study only those benefiting from treatment?)

**Pros**
- Used to assess treatments that have not been shown to be beneficial, although in common use.

**Cons**
- Evaluating withdrawal in a selective sample can overestimate benefit, underestimate toxicity.

**Factorial Designs**

**Two treatments**

<table>
<thead>
<tr>
<th></th>
<th>Treatment B</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment A</td>
<td>A+B</td>
<td>A</td>
</tr>
<tr>
<td>Control</td>
<td>B</td>
<td>Control</td>
</tr>
</tbody>
</table>

Randomize subjects to one of 4 possible regimens.

**Pros**
- Do 2 experiments at once!
- Investigate potential interaction between A and B.

**Cons**
If interaction exists, each treatment must be tested separately within each level of the other (⇒ reduced power).
Three treatments

Eight treatment groups in a balanced 2 x 2 x 2 factorial design:

<table>
<thead>
<tr>
<th>Group</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>n</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>n</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>n</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>n</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>n</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>n</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>n</td>
</tr>
</tbody>
</table>

Group Allocation Designs

Groups (clinics, communities) are randomized to treatment or control (e.g. trials on fluoridated water).

Pros
- Sometimes logistically more feasible
- Avoids individual consent problem

Cons
- Many units must participate to overcome unit-to-unit variation ⇒ requires larger sample size than simple randomized design.

Example of group allocation design

Hutchinson Smoking Prevention Project (HSPP)

The number one cause of preventable death in the United States is cigarette smoking. In addition, more than 90% of adult smokers started smoking by or before the age of 20. To counter this serious problem, in 1983 the U.S. National Cancer Institute initiated a request for applications (RFA) for school-based intervention studies to evaluate the long-term effectiveness of interventions on the prevention of habitual cigarette smoking among youth.
Select and recruit 40 Washington school districts

Randomize

20 experimental school districts
$N = 4,177$ 3$^{rd}$ graders

20 control school districts
$N = 4,211$ 3$^{rd}$ graders

Intervention grades 3 to 12

Endpoint data collection (12$^{th}$ grade)

Endpoint data collection (2 years post-high school)

**Studies of Equivalency**

Question: Is new (easier or cheaper) treatment as good as the current treatment?

Must specify: How close is “equivalent”? Can’t statistically “prove” equivalence, only show that difference is $< x$ with probability $p$.

Sample size issues are crucial. Small sample size $\Rightarrow$ low power $\Rightarrow$ no significant difference. This does not imply “equivalent.”

Course 13 will discuss equivalency in detail.
**ETHICAL ISSUES**


**Basic Principles**

(Six out of eighteen principles from the Declaration of Helsinki are listed below.)

1. The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol which should be transmitted to a specially appointed independent committee for consideration, comment, and guidance.

2. Biomedical research involving human subjects cannot legitimately be carried out unless the importance of the objective is in proportion to the inherent risk to the subject.

3. Every biomedical research project involving human subjects should be preceded by careful assessment of predictable risks in comparison with foreseeable benefits to the subject or to others.

   **Concern for the interests of the subject must always prevail over the interests of science and society.**

4. In publication of the results of his or her research, the doctor is obliged to preserve the accuracy of the results. Reports of experimentation not in accordance with the principles laid down in this Declaration should not be accepted for publication.

5. In any research on human beings, each potential subject must be adequately informed of the aims, methods, anticipated benefits and potential hazards of the study, and the discomfort it may entail. He or she should be informed that he or she is at liberty to abstain from participation in the study and that he or she is free to withdraw his or her consent to participation at any time. The doctor should then obtain the subject’s freely-given informed consent, preferably in writing.

6. The research protocol should always contain a statement of the ethical considerations involved in the study and should indicate that the principles enunciated in the present Declaration are complied with.
Medical Research Combined with Professional Care
(Two out of six principles from the Declaration of Helsinki are listed below.)

1. In any medical study, every patient—including those of a control group, if any—should be assured of the best proven diagnostic and therapeutic method.

2. The refusal of the patient to participate in a study must never interfere with the doctor-patient relationship.

Implementing Ethical Guidelines

Institutional review board (IRB)

Every medical institution that participates in research must have an IRB, a committee that includes both clinical and lay people. The IRB reviews the protocol for any proposed research. IRB approval is required before any research program is begun. (Such decisions may not be easy, especially if prior data on efficacy or toxicity of a treatment are ambiguous.)

Informed Patient Consent (as of 1983)

U.S.  Written informed consent is legally required.

U.K.  Depends on the trial; consent may be written, verbal, or not required; decided by local ethics committees.

France  “It is customary not to obtain informed consent, particularly in cancer clinical trials” (Pocock, 1983). (Cancer patients are not usually told that they have cancer; it is considered “not in their best interests” to know. (!))

Germany  Every trial protocol must be examined by a lawyer who decides on the trial’s legal and ethical status. (Delicate issue; earlier attempts to make clinical trials illegal.)

Hill (1963)  “If the patient cannot really grasp the whole situation [the concept of randomized trials], or without upsetting his faith in your judgment cannot be made to grasp it, then in my opinion the ethical decision still lies with the doctor...He cannot divest himself of it simply by means of an illusory consent.”
BASIC STUDY DESIGN REVIEW

- Randomized controlled design
  - (Standard)

- Crossover design
  - (R.C. design)

- Factorial design
  - (R.C. design)

- Historically controlled design
  - (NO → randomized)
  - (OK → phase II trial)

- Group allocation design
  - (Cluster randomization design)
  - (Study unit → group, not patient)

- Withdrawal study

- Nonrandomized concurrent control design

- Hybrid design