Biostatistics for Health Care Researchers: A Short Course

Issues in Clinical Trials

Chandan Saha, PhD

Division of Biostatistics

Indiana University School of Medicine

Objectives

- Become familiar with the features of a well-designed clinical trial and common clinical trials designs
- Understand the steps in data analysis and presentation of results

Outline

- Definition and types of clinical trials
- Ethics
- Design
- Randomization
- Sample Size and Power
- Data monitoring
- Analysis
- Presentation/publication of results

Definition

▲ Word “clinical” is derived from the Greek *kline* which means bed.

▲Thus, the term seems to be related to the “bedside” aspect of the patient/physician relationship.

▲In modern terms, clinical trial is every study that has human beings as its *experimental subjects*.

▲A clinical trial is an experiment testing an intervention on human subjects.

Types of clinical trials

▲ **Phase I:** Primarily dose-finding trials.
Primary concern: safety.
A great deal of the pharmacology and biological effects *in vivo* is learned in these trials.

▲ **Phase II:** Primarily feasibility and estimating treatment efficacy trials.
A fixed dose of a drug is used on a small number of patients.
A balance between efficacy and feasibility (side effects & toxicity) is considered for a large-scale study.

▲**Phase III:** Large comparative efficacy clinical trials.
A new treatment vs. standard or placebo (if no standard therapy exists).
Such trials will be favorable to the new treatment only when they demonstrate large benefits that can not be explained by chance alone.

▲**Phase IV:** Post-marketing observational studies.
Attempts to “catch” rare adverse reactions to the medication or procedure.

Ethics

Three Basic ethical principles

- **Autonomy** (right of patients for self-governance)

  To exercise this right, a patient must be informed on the benefits and potential risks of the new treatment/procedure (related to the requirement of obtaining *informed consent* from patients).

- **Beneficence** is the patient's right to be benefited from therapy, and the physician's duty not to harm the patient.

- **Justice** or fairness of distribution of the burdens and benefits of the research. For example, testing on poor people or minorities, and then distributing to the privileged would be in direct violation of this principle.

Design of clinical trials

- Proper design and analysis allows the investigator to attribute observed differences reliably either to the factors under control or to random error.

- Easier to control many sources of variation through design for the laboratory or industrial based experiments, but not for the clinical trials.

- This is one of the reasons that requires clinical trials relatively larger number of subjects to provide control over random variation.

- Skillful analyses can almost never correct design faults. For example, selection bias.

Features of well-designed clinical trials

1. Clearly stated objectives

2. Well-defined endpoints or quantifiable measures derived from these objectives

3. A priori stated decision rules for success or failure of the experimental treatment based on statistical tests involving these endpoints

4. When necessary, a clearly presented calculation of the sample size and its associated power

5. Well-described patient inclusion and exclusion criteria, and patient screening and randomization

Features of well-designed clinical trials (continued)

6. A system of data monitoring; this includes:
   i. Safety and efficacy monitoring, possibly by an external body (e.g., a Data Safety Monitoring Board or DSMB), with possibly explicit rules for early study termination
   ii . Data quality monitoring and error correction

7. All examinations, test, and evaluations described in detail along with a schedule of when they are to be performed

8. A data collection system which is based on data collection instruments called Case Report Forms (CRFs) as well as a system of digital data entry

Some Common Designs

Phase I Design (Dose Finding Study):

- A commonly used method is
  - ♦ Assign a group of 3 subjects to the first dose level (1 by 1)

  - ♦ Follow the following rule for dose selection
    --- no toxicity ➔ escalate the dose by one level

    --- two toxicities ➔ terminate and declare the previous level the maximum tolerated dose (MTD)

    --- one toxicity ➔ add 3 more subjects and
      ►stops trial and declare previous dose the MTD if there are ≥ 2/6 toxicities
      ►escalate the dose by one level if 1/6 experiences toxicity and continue the trial.

Some Common Designs (continued)

Comparative Study (Phase III):

- **Two groups parallel study**

  Randomly assign patients to one of the two treatment groups

  Patients and/or physicians can be blinded

- **Cross-over trials** (for example, 2 treatments with 2 periods)

  Suppose A and B are two treatments

  Assign the treatment sequence AB to the randomly selected half of the patients

  Assign the treatment sequence BA to the other half

  Use a certain washout period in between the two treatment assignments

Some Common Designs (continued)

Comparative Study (Phase III):

♦ Factorial designs (for example a 2 X 2 design)

Consider a study for blood pressure control with two treatments: amiloride and spirinolactone

Spirinolactone	Amiloride		Total
	Yes	No	
Yes	*a*	*b*	*a+b*
No	*c*	*d*	*c+d*
Total	*a+c*	*b+d*	*N*

Can test if there in an interaction effect between the two drugs.

Randomization and Stratification

Randomization is used to

- Control the variability of clinical outcome
- Combat selection bias (where a certain “type” of patient maybe more likely to be selected for one treatment versus the other)
- Creating homogeneous risk strata

Stratification is used to

- Balance *known* risk factors between the treatments under comparison

Randomization schemes

- The simplest way to randomize patients in two treatments is by "flipping a coin" (the first treatment takes the "heads" and the second treatment the "tails")

- However, there is no guarantee of equal number of subjects assigned to the two treatments

- For example, for N=100, the probability of equal treatment allocation is about 8% only.

Randomization using *permuted blocks*

- To improve balance between treatment assignments we can use permuted blocks.

- Within each block, the two treatments are balanced, and the order of treatment allocation is changed (permuted) from block to block. At the end of each block both treatments are balanced.

Within-block assignment	Permutation number					
	1	2	3	4	5	6
1	A	A	A	B	B	B
2	A	B	B	B	A	A
3	B	A	B	A	B	A
4	B	B	A	A	A	B

Stratification by important prognostic factors

- Every prognostic factor or prognostic factor combination (in the case of multiple prognostic factors) is made into a separate stratum.
- Treatment assignment is then balanced in each stratum yielding balanced representation of each stratum in each treatment.
- Then the blocks shown in the previous table, can be allocated at random into *each* stratum (thus balancing treatment allocation into each stratum). For example,

Stratum	Treatment assignment w/ a block size of 6					
	1	2	3	4	5	6
Old male	A	A	A	B	B	B
Young male	A	B	B	B	A	A
Old female	B	A	B	A	B	A
Young female	B	B	A	A	A	B

Determination of the sample size

Determining the size of the patient sample is a crucial proposition.

On the one hand, ethics dictate that no more subjects than necessary must be submitted to the experimental treatment.

On the other hand, a sufficient number of subjects must be experimented upon so that the question of interest is unequivocally answered.

Note that experimentation with more subjects than necessary violate individual patient rights, while experimenting on a less than adequate number of subjects violates communal patient rights.

Sample Size and Power

- **Power** : Probability of rejecting H_0 (no difference between two treatments) when H_A (treatments are different) is true.

- **Example** : "Using an alpha of 0.05 and a two sample t-test, a sample size of 21 in each of olanzapine and placebo groups will have 80% power to detect a minimum difference of 80 calories in the change in basal metabolic rate (BMR)."

  Interpretation : Suppose the minimum difference in the change in BMR between the olanzapine and placebo group is 80 cal.

  If we repeat a study a large number of times with 21 subjects in each group, at least 80% of the times we will reject the null hypothesis that there is no difference between the placebo and olanzapine groups.

Sample size for comparing two proportions

- Suppose p_1 and p_2 are the proportions of success for treatment A and B respectively.

- Question: What sample size is required for 80% power if $p_1 = 0.65$ and $p_2 = 0.85$?

- The sample size *n* is given from the following formula (w/o continuity correction):

$$n = \frac{\left[\left(z_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})}\right)+\left(z_{\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)}\right)\right]^2}{\Delta^2}$$

  Where $\bar{p} = \frac{p_1+p_2}{2}$ and $\Delta = p_1 - p_2$

- Answer: 73 per group.

Monitoring a clinical trial

Continuous monitoring of a clinical trial is required, as there are several reasons that a clinical trial may stopped before completion.

- Based on the opinion of experts
- Based on statistical tests, because of compelling evidence of efficacy or compelling evidence of non-efficacy
- Poor accrual
- Practical reasons making completion of the trial impossible
- Evidence from another study or discovery
- Obsolescence of the trial objective
- Unexpected adverse events that minimize any other benefits that may be procured by the trial

Technical challenges in clinical trial monitoring

- The most significant problem when undertaking repeated analyses of study data is that the Type I error is adversely affected.
- Type I error rate is greatly inflated when several interim analyses are performed.
- It should be clear that if more than one analyses are carried out at the pre-specified α level (say 5%), the overall (Type I) error rate will be higher than α.
- In the extreme, if the study is analyzed a very large number of times, the probability that one of the analyses will produce a significant result (even if there is no difference between the treatments) is virtually 100%!

Group-sequential methods

We need to develop procedures that “allocate” the total alpha level so that, at the end of the day, the overall chance of a false positive (i.e., a significant result where none is warranted) is still α.

Statistical rules used in monitoring a clinical trial are called *group-sequential* methods, distinguished from purely sequential methods by the fact that each interim analysis follows accrual of a *group* of subjects. (see Souhami & Whitehead (1994) and Cornfield (1966).

Generic description of group sequential methods

In its simplest form, there are R-1 interim analyses and 1 final analysis, producing each time a test statistic $Z_1, Z_2, \cdots, Z_R$ and at each point a concomitant *boundary* value $B_1, B_2, \cdots, B_R$

The trial will be stopped if

$$|Z_i| \geq B_i \text{ for } 1 \leq i \leq R$$

The Pocock and O'Brien-Fleming group sequential designs

The Pocock method (Pocock 1998): All analyses are performed at the same α level, determined so that the overall α is the same as the fixed-trial one. Stopping early is not difficult but the final test is carried out at a much smaller α level than the desired one

The O'Brien-Fleming procedure (O'Brien & Fleming, 1979):

This procedure has the advantage that the final test is carried out almost at the fixed trial alpha level, but early stoppage of the trial is much more difficult than the Pocock procedure

The O'Brien-Fleming design: An example

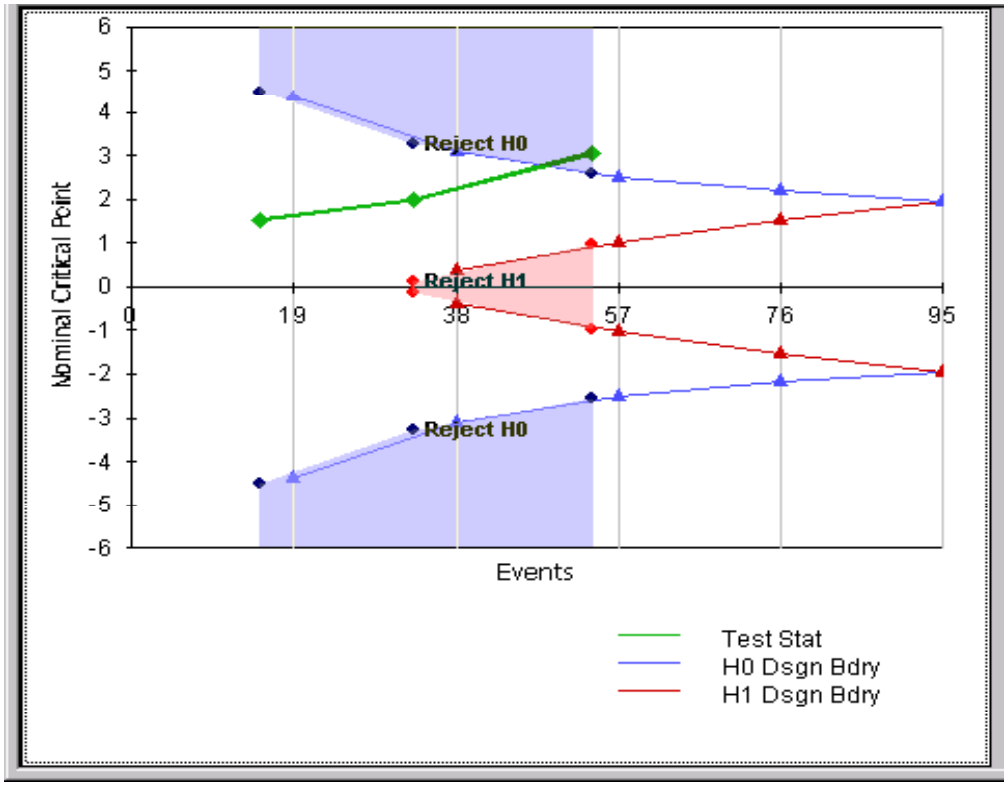

Figure Boundaries of an O'Brien-Fleming sequential design involving 4 interim and a final analysis. The study was interrupted in favor of the alternative hypothesis at the third analysis (from http://icssc.org/ReportLevel3_Jul2003.doc).

Pocock method: An example

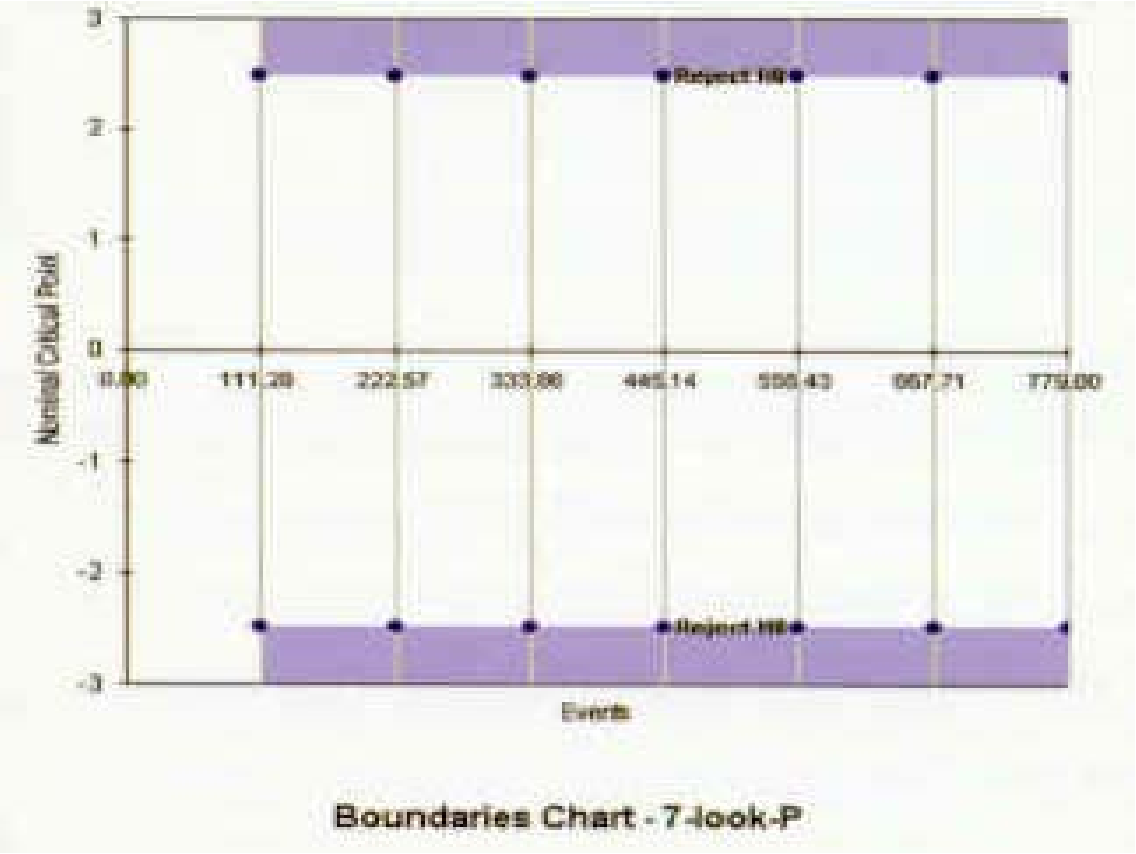

Figure Boundaries of a Pocock sequential design inolving 6 interim and a final analysis. Notice that the final analysis will be require a *z* score well above 2.5 (i.e., a *p* value less than 0.006) to reject the null hypothesis.

Data analysis and presentation of results

Types of Data

- Qualitative response
  - "Success"/"Failure"
  - Remission/relapse
- Quantitative response
  - Blood pressure
  - HIV RNA concentration
  - Percent change of neuropathic pain
- "Time to" analyses and survival data
  - Time to relapse or death
  - Time to development of toxicity

Steps in data analysis and presentation of results

After completion of part or all of the study, the following preparatory steps must be followed:

- Data cleaning
- Description of data
- Statistical Inference
  - Testing
  - Estimation
- Reporting and presenting the results

Statistical analyses

Descriptive statistics

- Qualitative data
  - Frequency tables
  - Frequency histograms

- Quantitative data
  - Descriptive statistical tables
  - Histograms (preferably with error bars)
  - Line graphs (especially if longitudinal trials)

- Survival data
  - Frequency tables (number of subjects that failed)
  - Kaplan-Meier (KM) plots

Statistical inference

- Qualitative data
  - Tests on proportions (binomial, normal)
  - Chi-square or Mantel-Haenszel test
  - Logistic regression and Generalized Linear Models (GLM)

- Quantitative data
  - T-tests, analyses of variance and covariance, GLM
  - Non-linear models (particularly in pharmacokinetic studies)

- Survival data
  - Log-rank test (nonparametric)
  - Cox Proportional Hazards (semiparametric), accelerated failure-time (AFT) models
  - Parametric models (e.g., Weibull, Gompertz, exponential)

Report and presentation of results

Study results

- Are reported in scientific journals and meetings
- Form the basis for submission to regulatory agencies
- Comprise detailed reports to colleagues and interested parties
- Form the basis for advertising by pharmaceutical companies

Components of a study report (Pocock, 1998)

- Title
- Summary or abstract
- Introduction ← Why did you start?
- Methods ← What did you do?
- Results ← What did you find?
- Discussion ← What does it mean?

Pitfalls and critical evaluation of scientific reports

Deficiencies in trial reporting

- Inadequate definition of eligible patients, treatment schedules, methods of evaluation
- Lack of an appropriate control group
- Failure to randomize patients to alternative treatments
- Lack of objective patient evaluation
- Failure to use proper blinding techniques

Pitfalls and critical evaluation of scientific reports

Deficiencies in reporting of results

- Too few patients ← Power
- Failure to account for all patients ← Bias
- Inappropriate statistical methods ← Validity

Deficiencies in editorial standards

- Inadequate reports produced and allowed to be published
- Journals favor positive findings
- Editors are reluctant to publish confirmatory (replicating) studies

Bibliography

- Cornfield, J. (1966). Sequential trials, sequential analysis, and the likelihood principle. *Am Statistician* 20: 18-23.
- O'Brien, PC, and Fleming, TR (1979) A multiple testing procedure for clinical trials. *Biometrics* 35:549-556.
- Piantadosi S. (1997) *Clinical Trials. A Methodologic Perspective*. John Wiley and Sons, Inc. New York, NY
- Pocock S J (1998) *Clinical Trials. A practical approach*. John Wiley and Sons, New York.
- Souhami, RL and Whitehead, J. (Eds.) (1994). Workshop on early stopping rules in cancer clinical trials, Robinson College, Cambridge, U.K., 13-15 April, 1993. *Stat Med* 13:1289-1500