

Clinical Trials

Definition: A *clinical trial* is a prospective study comparing the effects and value of intervention(s) against a control in human beings.

- A clinical trial is *prospective*
 - Well defined starting point (baseline or time zero)
 - Follow subjects forward in time
- One or more *interventions*
 - Without intervention a study is descriptive, not an experiment (e.g. follow natural history of a disease)
 - Interventions must be applied in a standard fashion

105

- At least one control group
 - Best current standard therapy (gold standard)
 - No active intervention
 - Placebo
- *Randomized* clinical trials
 - Random assignment of subjects to treatment groups
 - Blocking, matching, or stratification
- **Blindedness**
 - A randomized clinical trial is *double blind* if neither the subjects nor the people administering the treatments and taking measurements know the identity of the treatment assignments

106

Clinical Trial Phases

- **Preclinical trials**
- **Phase I trials:**
 - A few subjects who failed to respond to standard treatments
 - No control group
 - Can the drug be tolerated in humans (after studies in animals)?
 - How large a dose before unacceptable toxicity?

107

- – Find the maximum tolerated dose [MTD]
 - * Start with a low dose (3 to 15 subjects)
 - * Increase dose by steps until unacceptable toxicity is observed
 - Include 3 to 6 more subjects
 - If no further toxicity, continue to next higher dose
 - If additional toxicity, terminate at this MTD
 - * Use smaller steps as dose increases
- Pharmacology studies often done

108

- – References:

- * O'Quigley and Chevret (1991, *Statistics in Medicine*)
- * O'Quigley et al, (1990, *Biometrics*)
- * Storer (1989, *Biometrics*)
- * Gatsonis and Greenhouse (1992, *Statistics in Medicine*)
- * Babbs, Rogatko, and Zack (1998, *Statistics in Medicine*)
- * Freidman, L. M., Furberg, C. D. and Demets, D. L. 1998, *Fundamentals of Clinical Trials, 3rd edition, Springer-Verlag, New York.*

109

- Phase II trials:

- Does the drug have biologic activity or effect?
- What is the rate of adverse events?
- Used to design a Phase III trial
- Commonly used Gehan design
 - * Begin with 14 subjects at dose selected from Phase I study
 - Check for a minimum activity level (20% tumor response)
 - No success in 14 subjects occurs with probability less than .05 if treatment is at least 20% effective
 - * Include 10-20 additional subjects
 - Estimate positive response rate
 - Estimate rate of adverse effects

110

- Some phase II trials explore additional doses

- Status of subjects

- Disease is more advanced in Phase I studies
- More exclusion criteria than in Phase III studies

- Possibly different outcomes

- Phase II cancer study (tumor response)
- Phase III cancer study (survival)

111

- References

- Gehan (1961, *Journal of Chronic Diseases*)
- Fleming (1982, *Biometrics*)
- Sargent and Goldberg (2001, *Statistics in Medicine*)
- Freidman, L. M., Furberg, C. D. and Demets, D. L. 1998, *Fundamentals of Clinical Trials, 3rd edition, Springer-Verlag, New York.*

112

- **Phase III trials:**

- Large randomized clinical trials
- Assess effectiveness of a new intervention relative to at least one control group
- Follow-up time may be short relative to intended use

- **Phase IV trials:**

- Long term surveillance of an intervention
- No control group
- Larger number of subjects
- Efficacy and side effects

113

Ethics of Clinical Trials

- Informed consent
- Active vs inactive control
- Use of finders fees
- Randomization
- Early stopping (Freidman, et al, Chapter 15)
- Patient confidentiality
- Falsification of data
- Human Subjects Review

114

Timing of Clinical Trials

- Stability of the intervention
- Ability to measure efficacy

Well run clinical trials are costly and should be done only when preliminary evidence of efficacy of an intervention justifies the expenses and the potential risks to the subjects.

115

Protocol Document

A scientific planning document for a medical study on human subjects. It provides the study background and justification, objectives, descriptions of the design and organization of the trial, and evaluation criteria.

- Developed before subjects are enrolled
- Should remain essentially unchanged during the trial

116

Outline fo a Typical Protocol

A Background, Objectives, Justification

B Objectives

1. Primary objective and response
2. Secondary objectives and responses
3. Subgroup hypotheses
4. Potential adverse effects

C Study Design

1. Study Population
 - (a) Inclusion criteria
 - (b) Exclusion criteria
2. Sample size considerations

117

3 Subject recruitment and enrollment

- (a) Informed Consent
- (b) Assessment of eligibility
- (c) Baseline examination
- (d) Intervention allocation
 - randomization procedure
 - blinding

4 Intervention

- (a) Description and Schedule
- (b) Measures and Compliance

5 Follow-up schedule

118

6 Data Collection

- (a) Responses that will be measured
- (b) Measurement and Training
- (c) Quality control

7 Data Analysis

- (a) Interim evaluations
- (b) Final evaluation
- (c) Reporting adverse effects
- (d) Assessing Health-Related Quality of Life (HRQL)

8 Termination policy

119

D Organization

1 Participating investigators

- (a) Record treatment assignment
- (b) Prepare treatments
- (c) Reporting forms and software
- (d) Data base management
- (e) Laboratories and special units
- (f) Clinical centers

2 Study Administration

- (a) Review committees
- (b) Funding organization
- (c) Study management

120

Appendices

Definitions of eligibility criteria

Definitions of response variables

Descriptions of measurement procedures

Reference

Freidman, L. M., et al, 1998, *Fundamentals of Clinical Trials, Third edition, Springer-Verlag, New York.*

121

Randomization

Advantages

- Eliminate selection bias
- Averages out potential bias due to unknown factors
- Groups are *alike on average*
- Guarantees that statistical tests will have valid significance levels
- Make causal statements

122

Disadvantages

- Ethical issues
- Interferes with doctor-patient relationship
- Administrative complexity
- *Alike on average* - no guarantee of balanced groups
- Increased variability reduces power of statistical tests

123

Methods of Randomization

- Simple randomization
- Balanced randomization
- Stratified randomization
 - Reduce variation in composition of treatment groups
 - Create strata based on values of prognostic factors measured at or before randomization (e.g. age 30-39 years, 40-49 years, etc)
 - separate random assignment of subjects to treatment groups within each stratum
 - Account for stratification in the analysis of the data

124

Randomization as a Basis for Inference

Example: Effectiveness of Vitamin C as a cold preventative

- 20 subjects available for study
- Randomly divide subjects into two groups, with 10 in each group.
- Randomly select one group to receive Vitamin C (*treatment group*)
- Members of the other group (*control group*) receive a placebo.

125

Double Blind experiment:

Neither the subjects nor the people who administer the treatments and record results know which subjects are receiving the active treatment.

126

		Results		
		No Cold	Cold	
Treated with Vitamin C	$y_{11} = 9$	$y_{12} = 1$	10	
Controls	$y_{21} = 5$	$y_{22} = 5$	10	
		$y_{+1} = 14$ $y_{+2} = 6$		

H_0 : Placebo and Vitamin C are equally effective for preventing colds. (independence)

H_A : Vitamin C is more effective (one-sided alternative)

127

Randomization Argument:

If H_0 is true, then $y_{+1} = 14$ “no colds” and $y_{+2} = 6$ “colds” are features of the 20 subjects used in this study that cannot be changed by random assignment of subjects to groups.

Consequently, all row totals and all column totals are “fixed” quantities when H_0 is true.

		No Cold	Cold	
Vitamin C	Y_{11}			$Y_{1+} = 10$
Placebo				$Y_{2+} = 10$
		$Y_{+1} = 14$	$Y_{+2} = 6$	
		\swarrow \nearrow fixed when H_0 is true		

Other counts in the 2×2 table are determined by the value of Y_{11} .

128

Given the assumption that H_0 is true and all marginal totals are fixed, what is the probability of observing a table of counts at least as inconsistent with H_0 as the observed table of counts?

$$\begin{array}{|c|c|} \hline 9 & 1 \\ \hline 5 & 5 \\ \hline \end{array} \quad Pr(Y_{11} = 9) = \frac{\binom{14}{9}\binom{6}{1}}{\binom{20}{10}}$$

$$= \frac{12,012}{184,756}$$

$$= .0650$$

$$\begin{array}{|c|c|} \hline 10 & 0 \\ \hline 4 & 6 \\ \hline \end{array} \quad Pr(Y_{11} = 10) = \frac{\binom{14}{10}\binom{6}{0}}{\binom{20}{10}}$$

$$= .0054$$

129

$$\begin{aligned} p\text{-value} &= Pr(Y_{11} \geq 9) \\ &= Pr(Y_{11} = 9) + Pr(Y_{11} = 10) \\ &= \underline{.0704} \end{aligned}$$

Conclusion?

Two-sided test:

H_0 : Treatment and placebo
are equally effective

H_A : not H_0 .

$$\begin{aligned} p\text{-value} &= Pr(Y_{11} \geq 9) + Pr(Y_{11} \leq 5) \\ &= .1408 \end{aligned}$$

130

This is referred to as

Fisher's "Exact" Test

It requires an ordering of possible tables of counts with respect to how inconsistent they are with H_0 relative to the specified alternative.

An appropriate ordering is not always obvious or convenient.

131

		Stay	Get	
	Improve	Same	Worse	
Drug A	3	6	11	20
Drug B	8	4	8	20
Drug C	10	5	5	20
	21	15	24	

Is the following table less consistent with H_0 ?

4	7	9
7	3	10
10	5	5

132

Order tables using values of

$$G^2 = 2 \sum_i \sum_j Y_{ij} \log(Y_{ij}/\hat{m}_{ij})$$

or

$$X^2 = \sum_i \sum_j \frac{(Y_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

or use “exact” probabilities (conditional on H_0 is true).

Pr{table of counts}

$$= \frac{\prod_{i=1}^I (Y_{i+}!) \prod_{j=1}^J (Y_{+j}!)}{Y_{++}! (\prod_{i=1}^I \prod_{j=1}^J Y_{ij}!)}$$

These generally produce different orderings and different p -values for testing H_0 .

133

Listing all tables with X^2 values equal to or larger than X^2 for the observed table of counts becomes an overwhelming task as the group sizes and/or the number of response categories and/or the number of groups increase. There are too many possible tables.

- Simulate tables of counts with the appropriate sets of row and column totals and compute the percentage with X^2 values larger than the X^2 value for the observed table
- Use the chi-square approximation for the distribution of X^2 when H_0 is true.

134

Example:

Incidence of Common Colds in a double blind study involving 279 French skiers.

L. Pauling (1971), Proc. Natl. Acad. Sciences, 68 pp 2678-2681
Dykes & Meier (1975, JASA, 231, 1073-1079).

	No Cold	Cold	
Vitamin C	122	17	139
Placebo	109	31	140

H_0 : Vitamin C and the placebo are equally effective in preventing colds

H_A : Not H_0

135

Large sample chi-squared test:

$$\text{Expected count} = \frac{\left[\begin{array}{c} \text{column} \\ \text{total} \end{array} \right] \times \left[\begin{array}{c} \text{row} \\ \text{total} \end{array} \right]}{\left[\begin{array}{c} \text{total for} \\ \text{entire table} \end{array} \right]}$$

115.1	23.9
115.9	24.1

$$X^2 = \sum_i \sum_j \frac{(Y_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

$$= 4.81$$

with $p\text{-value} = 0.028$

136

“Exact” test:

$$p\text{-value} = Pr\{Y_{11} \geq 122\}$$

$$+ Pr\{Y_{11} \leq 109\}$$

$$= \frac{\binom{231}{122} \binom{48}{17}}{\binom{279}{139}} + \frac{\binom{231}{123} \binom{48}{16}}{\binom{279}{139}} + \dots$$

$$= 0.038$$

137

```
/* This program is posted as
   randombin.sas          */
```

```
DATA SET1;
  INPUT ROW COL COUNT;
  CARDS;
1 1 9
1 2 1
2 1 5
2 2 5
run;

PROC FREQ DATA=SET1;
  TABLES ROW*COL / CHISQ EXACT;
  WEIGHT COUNT;
RUN;
```

138

The FREQ Procedure

Table of ROW by COL

	Frequency		
	Percent		
Row Pct			
Col Pct	1	2	Total
1	9	1	10
	45.00	5.00	50.00
	90.00	10.00	
	64.29	16.67	
2	5	5	10
	25.00	25.00	50.00
	50.00	50.00	
	35.71	83.33	
Total	14	6	20
	70.00	30.00	100.00

139

Statistics for Table of ROW by COL

Statistic	DF	Value	Prob
Chi-Square	1	3.8095	0.0510
Likelihood Ratio Chi-Square	1	4.0700	0.0437
Continuity Adj. Chi-Square	1	2.1429	0.1432
Mantel-Haenszel Chi-Square	1	3.6190	0.0571
Phi Coefficient		0.4364	
Contingency Coefficient		0.4000	
Cramer's V		0.4364	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Cell (1,1) Frequency (F)	9
Left-sided Pr <= F	0.9946
Right-sided Pr >= F	0.0704
Table Probability (P)	0.0650
Two-sided Pr <= P	0.1409

140

DATA SET2;

INPUT ROW COL X;

CARDS;

1 1 3

1 2 6

1 3 11

2 1 8

2 2 4

2 3 8

3 1 10

3 2 5

3 3 5

RUN;

PROC FREQ DATA=SET2;

TABLES ROW*COL / EXACT CHISQ;

WEIGHT X; RUN;

141

Frequency

Percent

Row Pct

Col Pct	1	2	3	Total
1	3	6	11	20
	5.00	10.00	18.33	33.33
	15.00	30.00	55.00	
	14.29	40.00	45.83	
2	8	4	8	20
	13.33	6.67	13.33	33.33
	40.00	20.00	40.00	
	38.10	26.67	33.33	
3	10	5	5	20
	16.67	8.33	8.33	33.33
	50.00	25.00	25.00	
	47.62	33.33	20.83	
Total	21	15	24	60
	35.00	25.00	40.00	100.00

142

Statistic	DF	Value	Prob
Chi-Square	4	6.3643	0.1735
Likelihood Ratio Chi-Square	4	6.8949	0.1415
Mantel-Haenszel Chi-Square	1	5.5580	0.0184
Phi Coefficient		0.3257	
Contingency Coefficient		0.3097	
Cramer's V		0.2303	

Fisher's Exact Test

Table Probability (P)	2.040E-04
Pr <= P	0.1575

Sample Size = 60

143

```

# This code is posted as
#     randombin.R

# R and Splus have a built in function
# for the Fisher exact test.
# Computations are based on a C version
# of the FORTRAN subroutine FEXACT which
# implements a procedure developed by
# Mehta and Patel (1986) and improved
# by Clarkson, Fan & Joe (1993). The
# FORTRAN code can be obtained from
# <URL: http://www.netlib.org/toms/643>.
# This fails when the counts are too
# large or there are too many rows or
# columns in the table.
# The p-value is always for a two-sided
# or multi-sided test.

```

144

```

> fisher.test(matrix(c(9,1,5,5),
                    ncol=2,byrow=T))

Fisher's exact test
data:  matrix(c(9, 1, 5, 5),
              ncol = 2, byrow = T)
p-value = 0.1409
alternative hypothesis: two.sided

> fisher.test(matrix(c(3, 6, 11, 8, 4,
                    8, 10, 5, 5), ncol=3,byrow=T))

Fisher's exact test
data:  matrix(c(3, 6, 11, 8, 4, 8,
              10, 5, 5), ncol = 3, byrow = T)

p-value = 0.1575
alternative hypothesis: two.sided

```

145

Randomization with blocking

Example: Gastric pH

- Subjects matched by age, gender, and medical history
- Three dietary interventions

Block	Diet A	Diet B	Diet C
1	1.4	2.8	1.8
2	2.6	1.4	1.6
3	3.5	1.9	2.4
4	3.4	2.2	2.8
5	4.9	3.4	3.6
6	5.9	6.1	4.4
7	2.4	1.3	1.6
8	4.3	2.4	3.7
9	1.5	2.5	1.2
10	4.6	2.2	2.9

H_0 : All three diets are equally effective

H_A : Not H_0

146

- ANOVA

Source of Variation	df	Sum of Squares	Mean Square	F
Blocks	9	38.34	4.26	
Diets	2	4.71	2.35	4.82
Error	18	8.78	0.49	
Total	29	51.83		

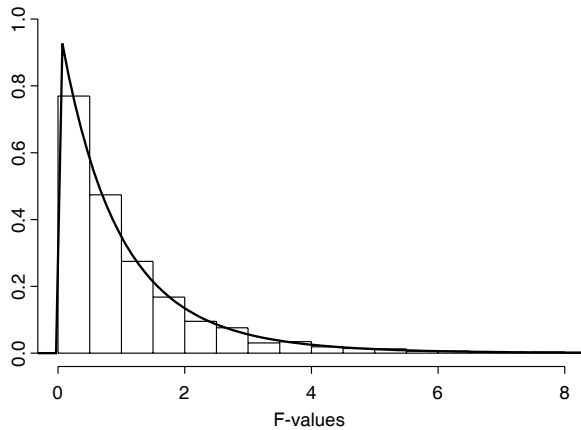
- Sample 5000 possible permutations
p-value = .025

Code is posted as randomblock.R
randomblock.sas

- Comparison with F-distribution
with (2,18) df
p-value = .021

147

F-values for 5000 Permutations



148

- **Adaptive randomization**

Change the allocation probabilities as the study progresses

- Adjust allocation to balance baseline characteristics but do not consider responses

- * Biased coin design (Efron, 1971, *Biometrika*)

- * Urn design (Wei, et al, 1990, *JASA*)

149

- **Adjust allocation using information on response to treatment**

- Play the winner

Stay with a treatment until a failure occurs, then switch

Zelen (1969, *JASA*)

- Two-armed bandit

Adjust allocation probabilities to give currently better treatment a higher probability

- Wei, 1977 *JASA*; 1978 *Annals*

- Yao and Wei, 1996 *Stat. in Med.*

- Begg, 1990 *Biometrika*

- ECMO study

- Ware, 1989 *Statistical Science*

150

Adherence

- **Reasons for nonadherence**

- Failure to understand instructions

- Unwillingness to modify behavior

- Lack of proper support

- Experience adverse side effects

- Length of the study

- **No subjects should be withheld from analysis for lack of adherence**

- Potential bias

- Reduced power \Rightarrow increase sample size

151

- **Percentage mortality in the Coronary Drug Project (1980, *New England Journal of Medicine*)**

	overall	Adherence	
		≥ 80%	< 80%
Clofibrate	18.2	15.0	24.6
Placebo	19.4	15.1	28.2

- **Percentage mortality in the Aspirin Myocardial Infarction Study (Friedman, 1984, *Controlled Clinical Trials*)**

	Overall	Adherence	
		Good	Poor
Aspirin	10.9	6.1	21.9
Placebo	9.7	5.1	22.0

152

- **References**

- Newcombe, 1988, *Statistics in Medicine*
- Detre and Peduzzi, 1982, *Controlled Clinical Trials*
- Oakes, et al, 1993 *JASA*
- **Intention to treat: Daniels, 2001, *Applied Statistics***

153

Surrogate Markers

- **Endpoint measured in lieu of some other so-called true endpoint**
 - tumor shrinkage/mortality from cancer
 - cholesterol level/heart disease
 - CD4 counts or viral load/development (or death) from AIDS

154

- **advantages:**

- determine more quickly the effect of a treatment or intervention
- use in phase I/II trials to help plan a phase III trial
- preliminary approval (accelerated approval) of a drug based on marker effects

- **disadvantages**

- if not a good marker, misleading conclusions

155

- **References:**

- **Definitions and evaluation:** *Statistics in Medicine*, Vol. 8, pp 405-440
- **Application (PTE):** Lin et al (1993, 1997) *Statistics in Medicine*
- **Application (meta-analysis):** Daniels and Hughes (1997) *Statistics in Medicine*
- **Application (AIDS):** Hughes et al. (1998)
- **Buyse et al. (2000)** *Biostatistics*

156

Common Study Designs

- **Completely randomized designs**

- **Single intervention factor**
- **Factorial designs:** look at several treatments/interventions at once

- **Randomized block (matched) designs**

- **Stratify potential subjects (blocking or matching)**
- **Random assignment of subjects to treatment groups within strata**

157

- **Cross-over designs**

- Each participant is his/her own control
 - * Only within subject variability
 - * Use fewer subjects
- Randomize order in which treatments are given
- Effect of intervention in one period may carry over into any subsequent period
- **References**
 - * **Brown, 1980**, *Biometrics*
 - * **Fleiss, 1989**, *Controlled Clinical Trials*
 - * **Koch, et al, 1989**, *Stat. in Med*
 - * **Louis, et al, 1984**, *New Engl J Med*

158

- **Equivalency trials**

- **Is new intervention as good as an established one**
- **Often 'create' an indifference region**
- **References**
 - * **Blackwelder, 1982**, *Cont Clin Trials*
 - * **Westlake, 1979, 1981**, *Biometrics*
 - * **Breslow, 1990**, *Statistical Science*
 - * **Hogan and Daniels, 2001** *Applied Statistics*

159

Statistics 565

What have we covered?

- **Types of studies**
 - **Observational studies**
 - * **Case-control studies**
 - * **Cohort studies**
 - **Randomized clinical trials**
- **Relative risk (odds ratio)**
- **Statistical methods**
 - **matched pairs (blocking)**
 - **independent samples or groups**
 - **randomization tests**

160

Upcoming Topics

- **Analysis of survival data (6 weeks)**
- **Longitudinal Studies (5 weeks)**
- **Missing Data, Meta Analysis (2 weeks)**

161