INTRODUCTION OF CLINICAL TRIALS

Jie Yang, PhD

Outline

- What is clinical trial
- Types of clinical trial
- Phase I trial: purpose, typical design
- Phase II trial: purpose, Simon's two stage design
- Phase III trial: types of design, issues in trial design
- Statistical issues involved
- Other types of trials

Two broad areas in the Study of Disease

1. **Epidemiology**

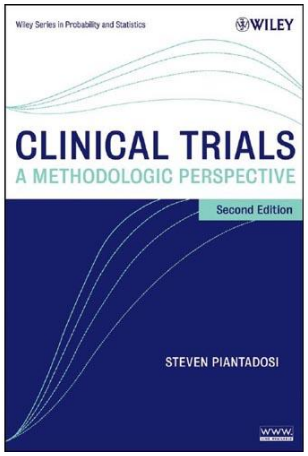
Systematic study of causes and origins of disease using observational data

- Second hand smoking and lung cancer
- Air pollution and respiratory illness
- Diet and risk of heart disease

(Examples: Cross sectional, longitudinal, case-control studies)

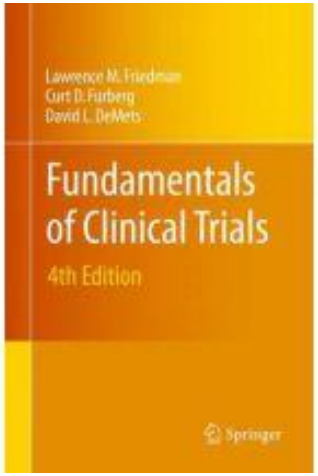
2. **Clinical Trials**

Definition of a Clinical Trial



An experiment testing medical treatments on human subjects

Piantadosi – *Clinical Trials: A Methodologic Perspective*

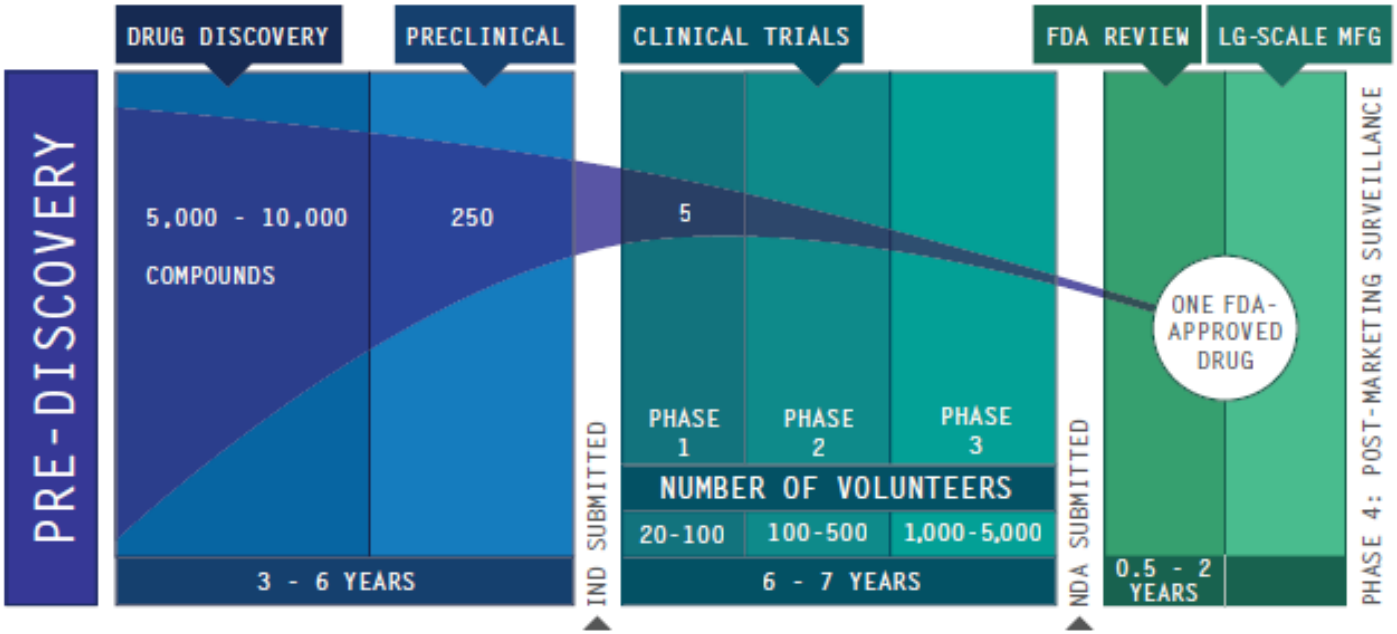


*A **prospective** study comparing the effect and value of intervention(s) **against a control** in human beings*

Friedman et al. – *Fundamentals of Clinical trials*

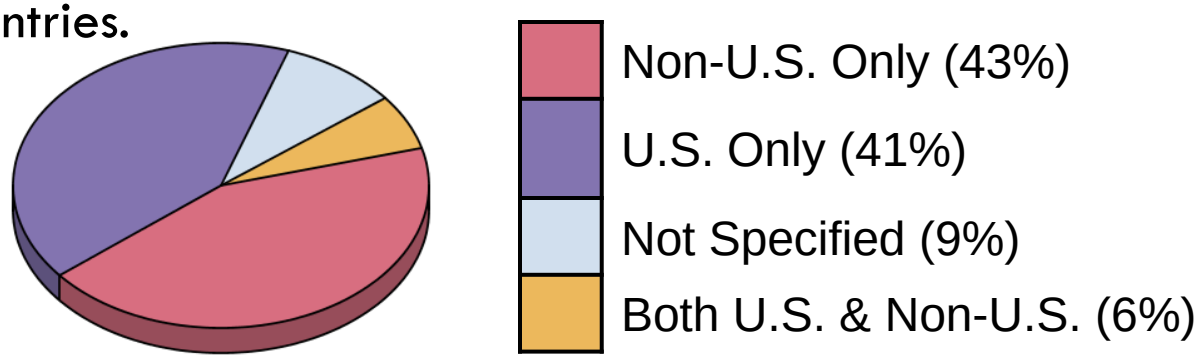
New Drug Development

- Discovery/Preclinical testing
- Phase I:How the drug works (PK/PD) and whether it is safe
- Phase II: Effectiveness and side effects
- Phase III: To confirm effectiveness and monitor
- Phase IV: Additional post-marketing testing



ClinicalTrials.gov

- A service of US National Institute of Health
- *A registry and results database of publicly and privately supported clinical studies of human participants conducted around the world.*
- currently lists 144,308 studies with locations in all 50 states and in 185 countries.



- In Sep 2005, International Committee of Medical Journal Editors (ICMJE) requires trial registration as a condition of publication under the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (URM).
- In Dec 2007, the expanded registration requirements of FDAAA began and were implemented.

Phase I clinical trials

- Perform initial human testing in a small group of healthy volunteers
- Primary purpose is to establish a safe dose and maximum tolerated dose (MTD) and schedule of administration
- Explore drug toxicities
- Other secondary outcomes: PK/PD, efficacy
- Ethical concerns require patients to be treated at lower doses before administering higher doses
- Starting dose is chosen based on pre-clinical data
- Remaining doses to be examined are specified in the protocol

Traditional Phase I Clinical Trial Design

- Dose escalation usually follows a '3+3' design

- Treat patients in groups of 3
  - If 0/3 experience a DLT -> escalate
  - If 1/3 experience a DLT -> enroll 3 more patients
  - If 1/6 experience a DLT -> escalate
  - If ≥ 2/3 experience a DLT -> reduce dose to previous level and add 3 patients (if only 3 studied at that dose level)

Traditional Phase I Clinical Trial Design

- **Maximum Tolerated Dose (MTD)**
  - Dose level immediately below the level at which ≥2 patients in a cohort of 3 to 6 patients experienced a DLT
- Usually go for "safe dose"
  - MTD or a maximum dosage that is pre-specified in the protocol

Phase I '3+3' trials – statistical operating characteristics

True probability of DLT for dose level	.05	.1	.2	.3	.4	.5	.6	.7
Probability of halting dose escalation after accruing either 3 or 6 patients (≥2 DLT) (1)	.03	.09	.29	.51	.69	.83	.92	.97
Probability of continuing escalation after only 3 patients (0 DLT) (2)	.86	.73	.51	.34	.22	.13	.06	.03
Probability of halting escalation after only 3 patients (≥2 DLT) (2)	.01	.03	.10	.22	.35	.50	.65	.78

(1) This row gives probabilities of halting dose escalation, at a given dose, if the true probability of DLT for that dose level is as indicated.

(2) These rows gives probabilities of continuing or halting dose escalation after accruing only 3 patients, at a given dose, if the true probability of DLT for that dose level is as indicated. We see that, in all cases, the cohort will be limited to 3 patients with at least 50% probability, and for the more extreme DLT probabilities (.05 or .7), the cohort will be expanded to 6 patients with less than 20% probability.

Rubinstein, Lawrence V., and Richard M. Simon. "Phase I clinical trial design."*Handbook of Anticancer Drug Development, Budman DR, Calvert AH, Rowinsky EK (eds). Elsevier: Amsterdam* (2003).

Phase I trials

- Dose limiting toxicity (DLT) must be defined
- Decide a few dose levels (e.g. 4)
- At least three patients will be treated on each dose level (cohort)
- Not a power or sample size calculation issue
- Entry of patients to a new dose level does not occur until all patients in the previous level are beyond a certain time frame where you look for DLT
- We are using the dose determined on a small number of patients here for subsequent Phase II/III trials – which may partially contribute to the failure of these trials downstream

Phase II trials

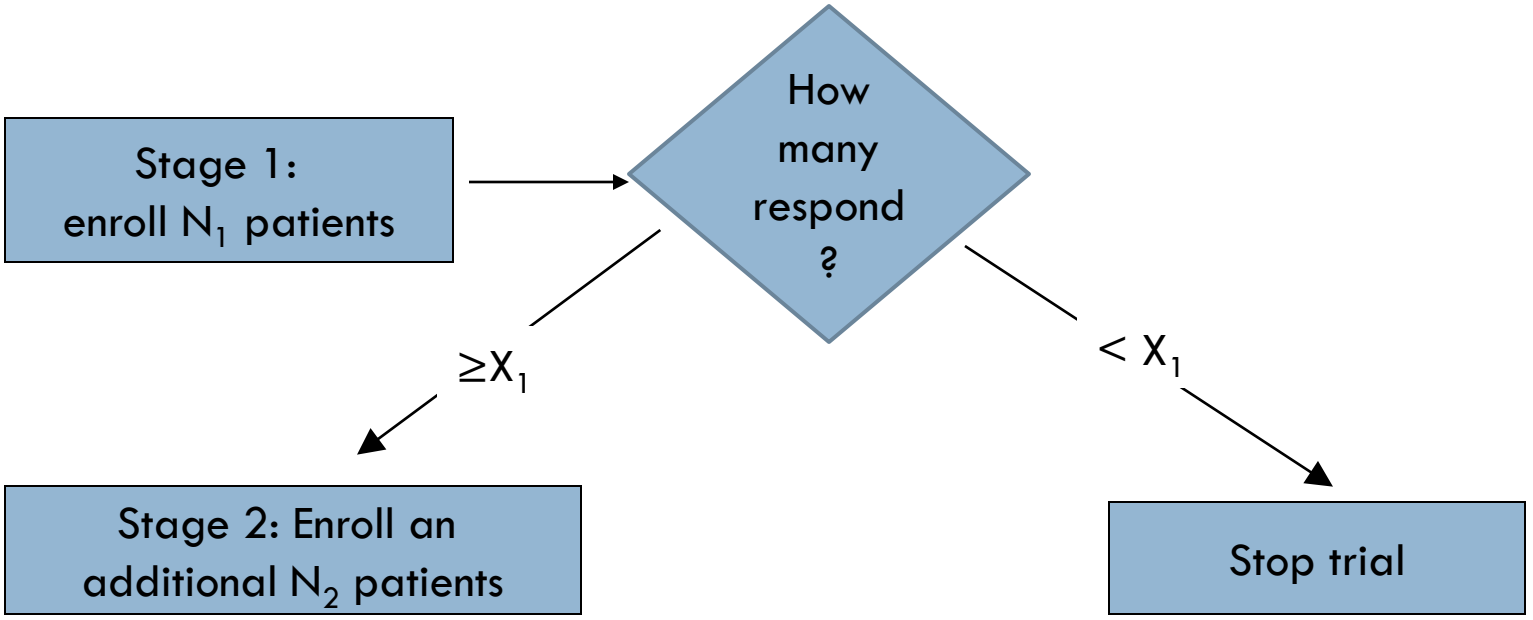
- Primary goal is to identify therapies that warrant further investigation
- Examines whether there is sufficient activity to be tested in a randomized Phase III study
- Safety assessment is still vital
- Goal is usually to estimate a clinical endpoint with a specified precision
- Single-arm trials, or multiple arms (parallel design)

Phase II trials – single stage, one arm

- Early-stage studies of efficacy
- All enrolled patients receive same therapy
- Patient selection – homogeneous, restricted to those most likely to show response (strict inclusion/exclusion criteria)
- Shorter-term endpoint, sometimes a surrogate

Phase II trials – Simon's two stage design

- Single arm, two stage, using an optimal design & predefined response
- Optimal – higher probability of early termination if treatment is not effective



Phase II Example: Two-Stage Optimal Design

- Rule out response probability of 20% (H_0: $p \leq 0.20$)
- Level that demonstrates useful activity is 40% (H_1:$p \geq 0.20$)
- Let $\alpha = 0.1$ (10% probability of accepting a poor agent)
- Let $\beta = 0.1$ (10% probability of rejecting a good agent)
- Charts in Simon (1989) paper with different $p_1 - p_0$ amounts and varying α and β values

Table from Simon (1989)

Table 1 Designs for $p_1 - p_0 = 0.20$[a]

		Optimal Design				Minimax Design			
		Reject Drug if Response Rate				Reject Drug if Response Rate			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)
0.05	0.25	0/9	2/24	14.5	0.63	0/13	2/20	16.4	0.51
		0/9	2/17	12.0	0.63	0/12	2/16	13.8	0.54
		0/9	3/30	16.8	0.63	0/15	3/25	20.4	0.46
0.10	0.30	1/12	5/35	19.8	0.65	1/16	4/25	20.4	0.51
		1/10	5/29	15.0	0.74	1/15	5/25	19.5	0.55
		2/18	6/35	22.5	0.71	2/22	6/33	26.2	0.62
0.20	0.40	3/17	10/37	26.0	0.55	3/19	10/36	28.3	0.46
		3/13	12/43	20.6	0.75	4/18	10/33	22.3	0.50
		4/19	15/54	30.4	0.67	5/24	13/45	31.2	0.66
0.30	0.50	7/22	17/46	29.9	0.67	7/28	15/39	35.0	0.36
		5/15	18/46	23.6	0.72	6/19	16/39	25.7	0.48
		8/24	24/63	34.7	0.73	7/24	21/53	36.6	0.56
0.40	0.60	7/18	22/46	30.2	0.56	11/28	20/41	33.8	0.55
		7/16	23/46	24.5	0.72	17/34	20/39	34.4	0.91
		11/25	32/66	36.0	0.73	12/29	27/54	38.1	0.64
0.50	0.70	11/21	26/45	29.0	0.67	11/23	23/39	31.0	0.50
		8/15	26/43	23.5	0.70	12/23	23/37	27.7	0.66
		13/24	36/61	34.0	0.73	14/27	32/53	36.1	0.65
0.60	0.80	6/11	26/38	25.4	0.47	18/27	24/35	28.5	0.82
		7/11	30/43	20.5	0.70	8/13	25/35	20.8	0.65
		12/19	37/53	29.5	0.69	15/26	32/45	35.9	0.48
0.70	0.90	6/9	22/28	17.8	0.54	11/16	20/25	20.1	0.55
		4/6	22/27	14.8	0.58	19/23	21/26	23.2	0.95
		11/15	29/36	21.2	0.70	13/18	26/32	22.7	0.67

[a]For each value of (p_0, p_1), designs are given for three sets of error probabilities (α, β). The first, second and third rows correspond to error probability limits (0.10, 0.10), (0.05, 0.20), and (0.05, 0.10) respectively. For each design, EN(p_0) and PET(p_0) denote the expected sample size and the probability of early termination when the true response probability is p_0.

Blow up: Simon (1989) Table

Table 1 Designs for $p_1 - p_0 = 0.20^a$

		Optimal Design			
		Reject Drug if Response Rate			
p_0	p_1	$\leq r_1/n_1$	$\leq r/n$	EN(p_0)	PET(p_0)
0.05	0.25	0/9	2/24	14.5	0.63
		0/9	2/17	12.0	0.63
		0/9	3/30	16.8	0.63
0.10	0.30	1/12	5/35	19.8	0.65
		1/10	5/29	15.0	0.74
		2/18	6/35	22.5	0.71
0.20	0.40	3/17	10/37	26.0	0.55
		3/13	12/43	20.6	0.75
		4/19	15/54	30.4	0.67

Phase II Example

- Initially enroll 17 patients.
  - 0-3 of the 17 have a clinical response then stop accrual and assume not an active agent
- If $\geq 4/17$ respond, then accrual will continue to 37 patients.
- If 4-10 of the 37 respond this is insufficient activity to continue
- If $\geq 11/37$ respond then the agent will be considered active.
- Under this design if the null hypothesis were true (20% response probability) there is a 55% probability of early termination

Phase II trials

- Designed to show activity
- Designed to not expose too many patients to toxic or inactive therapy
- Not considered definitive, however **we use the results of Phase II trials as a basis for moving towards larger, randomized trials in Phase III testing.** This does not always work

Phase II trial with historical control

- The control group received a precisely defined treatment in previous study.
- Eligibility criteria, procedures, and evaluation must be the same.
- Important prognostic variables must be known and similar for both groups.
- There is no reason to believe there is an external factor that may lead to different results

(Gehan, 1984; Pocock, 1976)

Randomized Phase II studies

- Need a control arm - limited or inapplicable historical data
- Designed to reinforce potential efficacy and support further study in a large Phase III trial
- Typically based on a shorter-term endpoint

Limitations of Randomized Phase II Trials

- Increased sample size
- Reduced number of new drugs and regimens that can be screened
- Difficulty in accruing to subsequent phase III trials
- Small randomized trials are often no more effective than small single arm trials for identifying treatments with moderate effects

Other Randomized Phase II Studies

- Most common – randomized between multiple treatment groups *(Pick the winner)*
- All arms treated as single Phase II studies using Simon's two-stage design
- If one arm is stopped for futility or toxicity, subsequent patients are enrolled in the remaining arms

Phase II trials

Use of biomarkers (or surrogate endpoints) instead of clinically relevant outcomes

Annals of Internal Medicine

Surrogate End Points in Clinical Trials: Are We Being Misled?

Thomas R. Fleming, PhD, and David L. DeMets, PhD

Why Use Surrogate Endpoints?

- Reduction in sample size, duration of trial and cost
  - Easier to show benefits: weeks/months vs. years
- Assess benefits of drug where measurement of clinical outcomes would be unethical/invasive (Alzheimer's)
- Because death and other "harder" outcomes are uncommon or delayed well into future

Surrogate markers

Example: Osteoporosis clinical trials

- Use of bone mineral density (BMD) as a surrogate for new fracture risk (clinical outcome of interest) is standard

- Does increase in BMD correspond to a decrease in bone fracture risk (clinical outcome)?

- Is BMD in the causal pathway of the disease process?

Most patients don't care about bone density. What they care about is having to endure a broken bone

Biomarker vs. surrogate endpoint

- Biomarker - a characteristic that is objectively measured and evaluated as an indication of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention

- **Surrogate – a biomarker intended to substitute for a clinical endpoint**

- Biomarkers often cheaper and easier to measure than 'true' endpoints

Surrogate markers/endpoints

- 10 biomarkers currently approved by the FDA for use as surrogate endpoints.
- Some of these markers include LDL cholesterol, blood pressure, hemoglobin A (1C), glucose, insulin, HIV-1 and CD4+cells
- Drugs for hypertension are approved entirely on the basis of the blood pressure endpoint, statins are approved on their ability to the lower LDL cholesterol

Phase II trials

True and validated surrogates (strictly defined - see Prentice, 1989), are rare.

'*a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint*'

Often an outcome that is *correlated* with the true outcome of interest is used as it *is likely* that it will predict benefit

Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989; 8:431-40.

Phase II Surrogate endpoints

- Example 1: **Serum cholesterol concentration**
  - Only about 10% of those who are going to have a stroke or heart attack have a serum cholesterol concentration above the reference range.
  - However, is used as a marker of therapeutic response to cholesterol lowering drugs.
- Example 2: **prostate specific antigen**
  - Good diagnostic marker (PSA > 4.0, PSA velocity, doubling time,…) and helps in disease management
  - PSA changes in advanced disease not strongly correlated with overall survival

Note about surrogate endpoints

- Correlation of a measure with clinical progression does not automatically confer "surrogate" status on that measure. **Necessary condition, but not sufficient.**

- It is critical to recognize that the correlation between the surrogate and the clinical outcome of interest **may not persist under drug treatment**

- Surrogates depend on class of drug under study

Example 3: Failure of a surrogate

- **Ventricular arrhythmias** cause sudden death
- It was therefore expected that anti-arrhythmic drugs would prevent sudden death.
- Cardiac Arrhythmia Suppression Trial - Class I anti-arrhythmic drugs increased sudden death significantly in patients with asymptomatic ventricular arrhythmias after a myocardial infarction
- Trial was stopped prematurely

	Sudden Death	Total Death
Drug	43	63
Placebo	9	23

The Cardiac Arrhythmia Suppression Trial-II Investigators. Effect of antiarrhythmic agent moricizine on survival after myocardial infarction: the Cardiac Arrhythmia Suppression Trial-II. *N Engl J Med.* 1992;327:227-233.

Accelerated Drug Approval

- Found at Subpart H of 21 CFR 314.500 (Code of Federal Regulations), and known as the **Accelerated Approval provisions,** the FDA for the provided an explicit standard **for drug approval based on the effect of a drug on a surrogate marker, and not a clinical outcome**

- If the confirmatory trial shows that the drug actually provides a clinical benefit, then the FDA grants traditional approval for the drug. If the confirmatory trial does not show that the drug provides clinical benefit for patients, FDA has regulatory procedures in place that could lead to removing the drug from the market.

Accelerated Drug Approval

- In a 2009 report, 90 drugs were given accelerated approval from 1992-2008. Of these, 79 were to treat cancer, HIV/AIDS, and anthrax
- As of 12/2008, 64% were considered closed by the FDA, meaning sufficient evidence in confirmatory trials
- "*Weaknesses in FDA's monitoring and enforcement process hamper its ability to effectively oversee post marketing studies*"

New Drug Approval: FDA Needs to Enhance Its Oversight of Drugs Approved on the Basis of Surrogate Endpoints GAO-09-866 September 23, 2009

Gleevec

- CML – malignant cancer of the bone marrow
- Median survival – 3-5 years from DX
- Philadelphia chromosome + CML - An enzyme (TKL) is produced that causes the marrow to make too many white blood cells
- Three stages – chronic, accelerated, blast phase, according to proliferation of (immature) blast cells
- Gleevec = *a TKL inhibitor*

Gleevec

- In 2001, Gleevec was granted accelerated approval for the treatment of CML

- Approval was granted on the basis of results from three single-arm studies conducted in patients with Philadelphia chromosome–positive CML. A total of 1027 patients were enrolled on the studies, which evaluated cytogenetic response rate (chronic phase) and hematologic response rate (accelerated phase and blast crisis) as primary endpoints.

- Novartis Pharmaceutical Corporation committed to complete a randomized phase III study

Druker, B. J., *et al*. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *New England Journal of Medicine* **344**, 1031–1037 (2001

Gleevec

- Novartis Pharmaceutical Corporation, readily admitted:

"Accelerated approval of Gleevec was based on (cytogenic, hematogenic response) time to progression as the primary surrogate endpoint. Approval under these regulations requires further adequate and well-controlled studies to verify and describe clinical benefit. ***There are no controlled trials demonstrating a clinical benefit, such as improvement in disease-related symptoms or increased survival."***

Gleevec

- In 2003, accelerated approval was converted to regular approval on the basis of the submission of follow-up Phase III clinical trial data.

- One of the most successful drugs of this decade
- In 2006, NEJM published the results of a trial showing an **89% five-year survival rate in patients**
- Brought its maker, Novartis, nearly $1.9 billion in sales in the first six months of 2009.
- By 2011, FDA-approved to treat 10 different cancers

Phase 0, 2A and 2B trials

- **Phase 0 trial:** The FDA has recently endorsed "microdosing" or the "Phase 0 trials", which allows researchers to test a small drug dose in fewer human volunteers to quickly weed out drug candidates that are metabolically or biologically ineffective.

- **Phase 2A trial:** Sometimes combined with a Phase I trial, a Phase 2A trial is aimed not only at understanding the safety of a potential drug, but also getting an early read on efficacy and dosage in a small group of patients.

- **Phase 2B trial:** Resulting from Phase 2A trial, a Phase 2B trial is designed to build on these results in a larger group of patients for the sake of designing a rigorous and focused Phase III trial.

Phase III trial

- Test in a large group of patients to show safety and efficacy
- Provides the basis for labeling instructions to help ensure proper use of the drug
- Both the costliest and longest trials.
- Hundreds of sites around the United States and the world participate in the study to get a large and diverse group of patients
- Mostly: Randomized and controlled clinical trials (RCT)
- RCTs – considered the **gold standard** for evidence of treatment effect
- Required by FDA to market a new drug

Placebo-controlled RCT

- Placebo-controlled: Some subjects will receive the new drug candidate and others will receive a placebo.
- Provide the highest level of confidence in the validity of the results. Hence many drug trials are placebo-controlled, randomized and double-blinded.
- This method of testing provides the best evidence of any direct relationship between the test compound and its effect on disease because it minimizes human error.
- Sometimes it is unethical to give out placebo to patients and hence the drug candidate need to be tested against another treatment rather than a placebo.

Common Elements of a Protocol

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

Common Elements of a Protocol

VI. STUDY POPULATION AND PATIENT SELECTION

IX. RANDOMIZATION PROCEDURES

X ADMINISTRATION OF STUDY DRUG

XI DATA MANAGEMENT, QUALITY ASSURANCE & MONITORING

XII. STATISTICAL CONSIDERATION

A. Historical data
B. Sample size and power
C. Statistical analysis plan for primary objective
D. Subgroup and secondary analyses
E. Interim analyses and study monitoring plan

Common Elements of a Protocol

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

Primary and Secondary Objectives

- 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
- A good primary objective = clear power computations and clear analysis

Primary vs. Secondary Questions

- **Primary**
  - Most important, feasible, and central question
  - Ideally, only one
  - Stated in advance
  - **Basis for design and sample size**
- **Secondary**
  - Related to primary
  - Defined *a priori* (avoid post hoc "fishing expedition")
  - Chosen parsimoniously (avoid false positive)

Common Elements of a Protocol

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

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

- The form of the response variable (outcome) may differ
  - Binary outcome (Relapse event vs. no relapse event in one year)
  - Time to event variable (Time to relapse event from randomization)
  - Continuous
  - Longitudinal variables

A good primary endpoint

- Must answer the primary question
- Frequency of occurrence/average(SD)/survival must be known for control (determine sample size)
- Must be able to estimate treatment effect: clinical relevance
- Must be assessed/evaluable in all participants
- Can be measured accurately/reliably/objectively
- All patients must be evaluated (no post randomization exclusion)

Surrogate endpoints

- Endpoint that is measured *in lieu* of a more definitive, meaningful (clinical) endpoint
- Used when definitive endpoint trial would be too long or costly. Surrogate endpoint trial generally smaller because:
  - Continuous measurements
  - More frequent events
  - Measure subclinical disease
- And if:
  - Accurate and reliable measurement
  - Acceptable to the participants (*invasive?*) and investigators (*cost, ease of use?*)

Surrogate endpoints

Disease	Definitive Endpoint	Surrogate Endpoint
Cardiovascular disease	Myocardial infarction	Cholesterol level
	CHD	Carotid IMT
	Heart Failure	BNP
	Stroke	Blood pressure
Cancer	Mortality	Tumor size reduction
Prostate Cancer	Disease progression	PSA
HIV Infection	AIDS/Death	CD4+ count
Glaucoma	Vision Loss	Intraocular pressure

Common Elements of a Protocol

VI. STUDY POPULATION AND PATIENT SELECTION

IX. RANDOMIZATION PROCEDURES

X ADMINISTRATION OF STUDY DRUG

XI DATA MANAGEMENT, QUALITY ASSURANCE & MONITORING

XII. STATISTICAL CONSIDERATION

A. Historical data
B. Sample size and power
C. Statistical analysis plan for primary objective
D. Subgroup and secondary analyses
E. Interim analyses and study monitoring plan

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)

Study Population

Exclusion criteria:

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

Study Population

Homogeneity

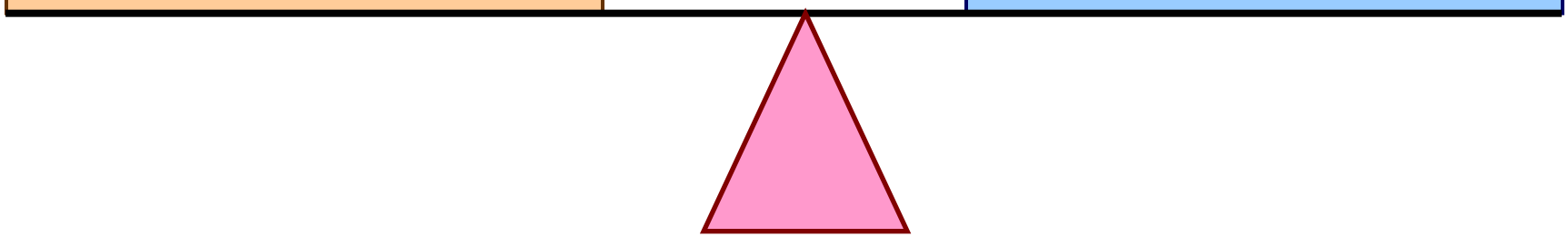
- **Divergent subgroup of patients (i.e., "atypical " patients) can distort findings for the majority**
- **Restriction of population reduces "noise" and allows study to be done in a smaller sample size**

→ 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



Common Elements of a Protocol

VI. STUDY POPULATION AND PATIENT SELECTION

IX. RANDOMIZATION PROCEDURES

X ADMINISTRATION OF STUDY DRUG

XI DATA MANAGEMENT, QUALITY ASSURANCE & MONITORING

XII. STATISTICAL CONSIDERATION

A. Historical data
B. Sample size and power
C. Statistical analysis plan for primary objective
D. Subgroup and secondary analyses
E. Interim analyses and study monitoring plan

RCT: Randomization

- Random assignment to treatment versus control/standard therapy
- In doing this, we technically divide patients into similar groups with respect to all known and unmeasured factors
- Patients are similar **except for treatment assigned**
- Removes **allocation bias**

Ethics of Randomization

- Must sell benefits of randomization
- Ethics $\Rightarrow$ MD should do what he thinks is best for his patient
- If MD "knows" best treatment, should not participate in trial (lack of equipoise)
- If in doubt, randomization gives each patient equal chance to receive one of therapies

Confounding

- One of the biggest issues with observational studies, and will always cast doubts on observational studies (especially if different results vs. RCT)
- The treatment effect on outcome is mixed in with the effect of the confounder
- As a result, the treatment effect is biased

Confounding Bias & Randomization

- Randomization ensures that potential confounding factors, known or unknown, are evenly distributed among the study groups
- Randomization takes care of bias due to factors *before* treatment

Types of Randomization

- Simple
- Blocked Randomization
- Stratified Randomization

Simple Randomization

- Randomize each patient to a treatment with a known probability
  - Corresponds to flipping a coin
- Could have imbalance in # / group or trends in group assignment
- Could have different distributions of a trait like gender in the two arms
- TTTTTTTTTTTTTTCCCCTTTCTCTCTTCCCCTTCC

Long string of patients being assigned treatment

Block Randomization

- To reduce confounding
- Insure the # of patients assigned to each treatment is not far out of balance at any point in time
- Common: **Permuted Block Designs**
  - Block of size 4 – random assignment of 2 treated, 2 control – 6 permutations possible
  - Variable block size (2,4,6,8)
  - An additional layer of blindness

Stratified Randomization

- A priori certain factors likely important (e.g. Age, Gender)
- Randomize so different levels of the factor are balanced between treatment groups
- Stratified block randomization –permuted block randomization within each stratum
- Common strata
  - Clinical center, Age, Gender

Stratification MUST be taken into account in the data analysis

Masking or blinding

Keeping treatment assignments masked for:

1. Subject (blinded)
2. Investigator (double blinded)
3. Committee monitoring response variables (triple blinded)

- Purpose of masking: bias reduction
  - Each group masked eliminates a different source of bias
- Masking is most useful when there is a *subjective* component to treatment or evaluation

Masking or blinding

- No Blind
  - All patients know treatment
- Single Blind
  - Patient does not know treatment
- Double Blind
  - Neither patient nor health care provider know treatment
- Triple Blind
  - Patient, physician and statistician/monitors do not know treatment
- Double blind recommended when possible

Masking or blinding

- Assures that subjects are similar with regard to post-treatment variables that could affect outcomes
- Minimizes the potential biases resulting from differences in management, treatment, or assessment of patients, or interpretation of results
- Avoids subjective assessment and decisions by knowing treatment assignment

Common Elements of a Protocol

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

Phase III trials

Superiority trial

- Ho: *Treatment is no better than control*
- Ha: *Treatment is better than control*

- Survival-based endpoints (oncology), or surrogate endpoints (*for example, viral load, blood pressure, HbA1c reduction*)
- Treatment effect based on estimate of the hazard ratio, odds ratio, or mean difference

Phase III Trials

- Superiority trials are appropriate when comparing to placebo or control
- There are situations where placebo-control is unethical
- There are also situations where the new treatment being introduced is not expected to be significantly better than the current therapies, but offers some advantage (cost, safety, ease)

Superiority, Non-Inferiority and Equivalence trials

Superiority Trial: Show that new treatment (cheaper, safer, easier) is better than the control or standard therapy

Non-inferiority Trial: Show that the new treatment is not worse that the standard by more than some margin (M) with 95% certainty. Can also show superiority with this design

Equivalence Trial: Show that the differences between control and study treatments are not large in either direction

Superiority, Non-Inferiority and Equivalence trials

- In a superiority trial, we would compare the event rate and reject Ho (treatment effect) and conclude treatment is superior (95% CI does not contain 0).
- In a equivalence trial, we would reject Ho and conclude treatment is equivalent if the upper bound and lower bound of the 95% CI is between ± 2
- In a non-inferiority trial, we would reject Ho and conclude treatment is non-inferior if the upper bound of the 95% CI is less than 2

Non-inferiority trials

- Goal of a Non-Inferiority Study: "With 95% certainty can I say that T(reatment) is no worse than a *margin M of equivalence* than C(ontrol)"?

H_0 : T-C > M (Treatment inferior)

vs.

H_A : T-C ≤ M (Treatment is at least equivalent)

Rejection of Ho –> conclude non-inferiority

Non-inferiority Trials

Example:

The yearly rate of hip fractures in women under standard therapy is 4%

For non-inferiority, we expect the rate under our new therapy to be no more than 6% (margin)

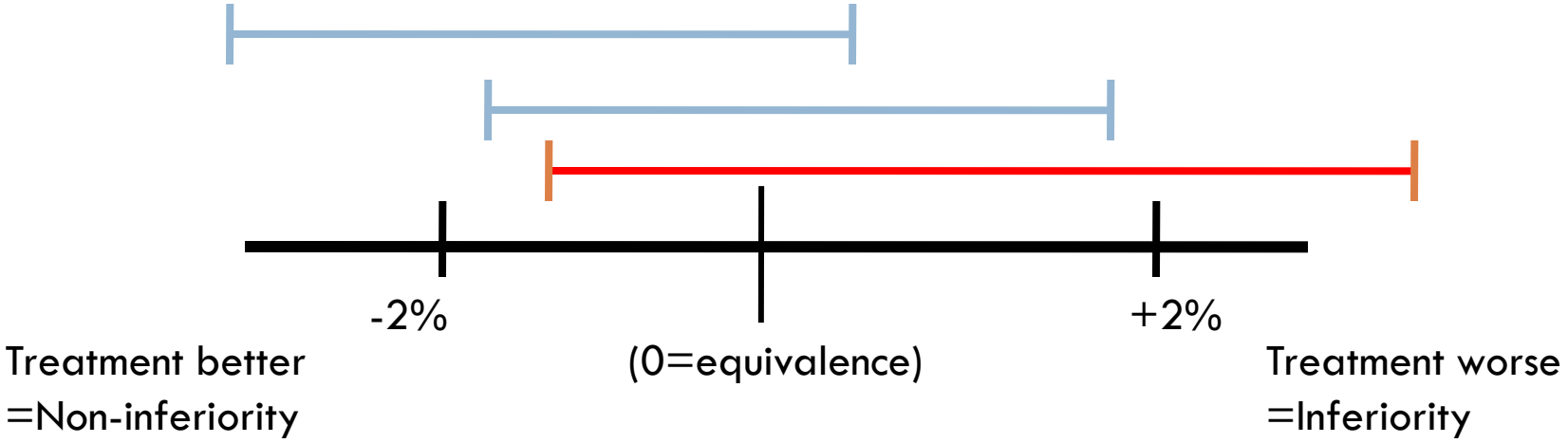
H_0 : T-C > 2% (treatment inferior)

H_A : T-C ≤ 2% (treatment equivalent or non-inferior to control)

We can compute the 95% CI for the treatment effect to see plausible values for the true treatment effect

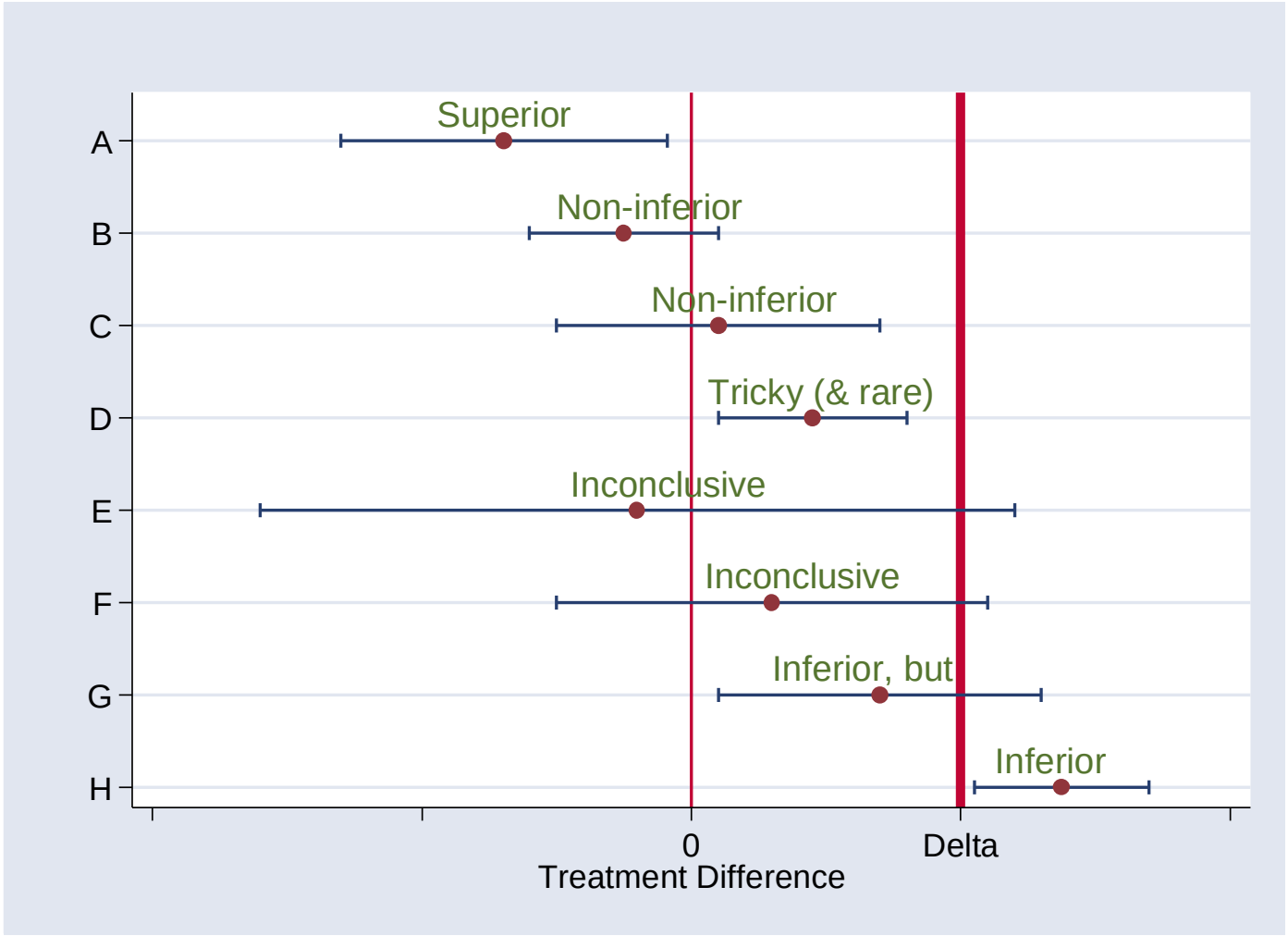
Non-inferiority Trials

95% CI for Trt effect (non-inferiority vs. inferiority):



T-C

Possible outcomes in a non-inferiority trial (Observed difference & 95% CI)



← New Treatment Better

New Treatment Worse →

Non-inferiority Trials

- Not the same as an analysis for superiority - small sample size, leading to low power and subsequently lack of significant difference, does not imply "equivalence"

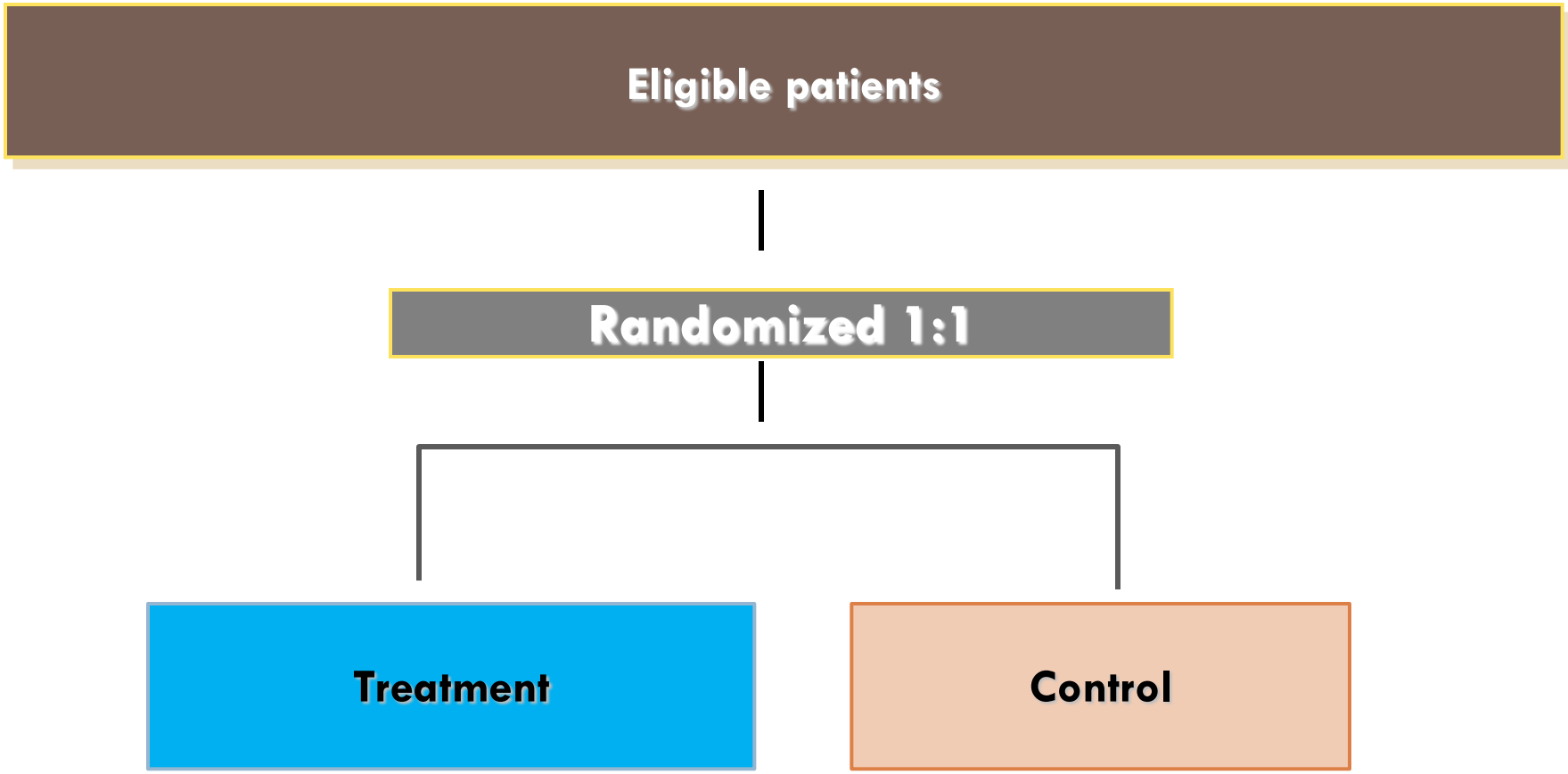
- The choice of M has important practical consequences. The smaller the margin, the smaller the upper bound of the 95% two-sided confidence interval must be, and the larger the sample size that will be needed.

Commonly Used Phase III Designs

- **Parallel** (78%*)
- **Cross Over** (16%*)
- **Group/Cluster** (<5%*)
- **Factorial** (<5%*)

**Based on survey of 2006 PubMed articles*

Phase III: Parallel Designs



Phase III: Crossover Designs

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

Example: Crossover study

- A phase III randomized double-blind placebo-controlled 2-period crossover study to assess the effect of Indacaterol on exercise endurance in patient with moderate to severe COPD

- Arm 1: Indacaterol -> 3 week washout -> placebo
- Arm 2: placebo -> 3 week washout -> Indacaterol

- Primary outcome: Exercise duration time at the end of treatment period
- Secondary outcome: Inspiratory capacity assessed at rest

Common Elements of a Protocol

VI. STUDY POPULATION AND PATIENT SELECTION

IX. RANDOMIZATION PROCEDURES

X ADMINISTRATION OF STUDY DRUG

XI DATA MANAGEMENT, QUALITY ASSURANCE & MONITORING

XII. STATISTICAL CONSIDERATION

A. Historical data

B. Sample size and power

C. Statistical analysis plan for primary objective

D. Subgroup and secondary analyses

E. Interim analyses and study monitoring plan

Analysis Follows Design

Questions → Hypotheses →

Experimental Design → Samples →

Data → Analyses →Conclusions

- Sample size and power calculation should follow the study design.
- Analysis for the primary endpoint should be reflect the study design.

Intent-To-Treat (ITT) analysis

- All patients on a randomized clinical trial should be analyzed as part of the treatment group to which they were assigned, regardless of adherence, withdrawal, treatment received
- Should be the primary approach to test the null of *Ho: No treatment effect*
- Modified ITT (mITT) – all patients who were randomized and received at least one dose
- Other (secondary) approaches: Treatment received (TR) and Adherers only

Types of analysis

- In a parallel study, we are comparing the HbA1c levels of two independent groups
- In the cross-over study of COPD, we are comparing the exercise duration times of the **same person** (treated and on placebo)
- In the latter design, we must analyze the data taking the correlated nature of the outcome data into account

Cross-over trials = Correlated data

- We can expect that the two duration times of individual A are more alike than a time from individual A and a time from an individual B
- Why? Unmeasured or measured patient-level characteristics
- The power of the cross over design is that we get to hold all of these patient level characteristics fixed in comparing treatment

Correlated data

- An appropriate test for the COPD example is the paired t-test (unadjusted for any factors)
- In survival data, we use frailty models
- In binary data, we may McNemar's test, LR with GEE, or mixed models (adjust for period, baseline exercise endurance rate)
- What happens if we do not account for the correlation between patient outcomes?

Interim analysis

- Analysis of the data before the study is ended with the intention of possibly terminating the study early
- Appropriate monitoring involves more . . .
  - Slow accrual
  - Poor data quality
  - Poor adherence
  - Resource deficiencies
  - Unacceptable adverse effects
  - Emerging information makes trial irrelevant
  - Fraud

Interim Analysis

- Focus on statistical stopping rules
  - Adjustment for multiple looks at data
- If traditional tests are used at both the middle and the end of the study, Type I error get inflated

# of interim analysis	0	1	4	9
Type I error	0.05	0.08	0.14	0.2

- To maintain the whole Type I error, α needs to be adjusted at each interim analysis

Group sequential design

- Purpose: controlling the overall type I error
- Pocock boundaries:
  - Same critical value at every monitoring point
  - Relatively early stopping for declaring a treatment effect
  - Critical value at the maximum sample size can be very different from the fixed sample size

Stopping rules for interim analysis

- O'Brien-Fleming boundaries
  - Stop early only if the treatment effect is very (very) large
  - Critical value at the maximum trial size is close to fixed sample value

Pocock and O'Brien-Fleming boundaries

# of looks	Pocock	O'Brien-Fleming	
		First	Last
1	1.96	1.96	
2	2.18	2.96	1.96
5	2.41	4.88	2.03
10	2.54	7.00	2.08

Early stopping for futility

- Stop the study if there is no hope to establish treatment effect
- Purpose:
  - Avoid exposing subjects to ineffective intervention
  - Save resources
- Consequences:
  - It increases type II error (thus power reduced)
  - No effect on type I error

Alpha spending function

- Flexible design: neither the number nor the time of the interim analyses needs to be specified in advance
- The frequency of the interim analyses can be changed during the trial and still preserve the pre-specified α value
- No change of the spending function itself is permitted during the trial

Adaptive Design

- When overall lower event rates or increased variability, or when emerging trends are smaller than planned for but yet of clinical interest
- Sample size increased by extending the recruitment phase, extending following up time, increasing the enrollment rate
- Adjustment should be made as early in the trial as possible or as part of a planned adaptive design strategy.

Note about interim analysis

- Interim analyses results are generally highly confidential
- Results from any statistical method for monitoring can only be reviewed as a piece of evidence, not an exclusive rule for decision making
- Data and safety monitoring is an active process -- additional tabulation and analysis are suggested

Missing data in clinical trials

- It should be the aim of those conducting clinical trials to achieve complete capture of all data from all patients, including those who discontinue from treatment
- Primary analysis is commonly performed on the full analysis set as this analysis is consistent with the intention to treat (ITT) principle
- If data for some subjects are missing for the primary endpoint it is necessary to specify how all randomized patients can be included in the statistical analysis.
- However, there is no universally applicable method that adjusts the analysis to take into account that some values are missing, and different approaches may lead to different conclusions.

Missing Data in clinical trials

- Essential to pre-specify the selected methods in the statistical section of the study protocol
- ALL approaches to analysis rely on assumptions that cannot be fully verified
- The strategy employed to handle missing values might in itself be a source of bias

Methods for Missing data

- Completers only
  - Reduces power
  - Modified-ITT
  - May introduce bias
- Imputation
  - LOCF, BOCF
  - Average
  - Best/worst case
    - May introduce bias
    - Ignores uncertainty in imputed values
- Multiple Imputation

Conduct a Sensitivity Analysis

Subgroup analysis

- Listed in secondary aims
- Stratified randomization allows adequate sample sizes for these aims
- Inflation of type I error if many hypotheses tested
- Main argument against - heterogeneity of relative treatment effect (i.e., benefit in one subgroup and harm in another) is very rare

Subgroup analysis

- Test enough subgroups – probably a false positive result by chance alone

- Worse yet, investigators might just report the statistically significant results

Subgroup analysis

- Pre-specify
- Report all subgroup analyses done
- Use tests of interaction
- If exploratory analyses, report as such and report all exploratory analyses done

Bonferroni-adjustment

- Simple and conservative
- Instead of alpha=0.05 for each of our 5 tests, use an alpha of 0.05/5=0.01
- Maintains an overall significance level of 0.05
- Use when number of tests is low, and correlation between outcomes is small
- Always, method for controlling type I error should be stated in the protocol

Example: Phase III study of PC

Phase III Study of Immunotherapy to treat advanced prostate cancer (clinicaltrials.gov NCT01057810)

- Arm I (Treatment): Ipilimumab,induction+maintenance
- Arm II (Placebo): IV solution,induction+maintenance
- Primary: *To compare OS of subjects, defined as the time from the date of randomization until the date of death or last follow-up time.*
- Patients assessed at each study visit while on treatment and every 12 weeks during f/u

Example: Phase III study of PC

- Inclusion criteria:
  - Metastatic PC
  - Asymptomatic, minimally symptomatic
  - ECOG PS 0-1
  - Progression during HT
  - 18 years or older male

- Exclusion criteria:
  - Lung, brain, liver mets
  - Prior immunotherapy or chemotherapy for PC
  - Autoimmune disease
  - HIV, Hepatitis B,C infection

Example: Phase III study of PC

- Secondary outcomes:
  - PFS
  - Time to pain progression
  - Time to non-hormonal systemic therapy
  - Characterization of safety profile
- Allocation: Randomized
- Sample size to be enrolled: n=600
- Masking: Double blind
  (Subject, Caregiver, Investigator, Outcomes Assessor)

Observational Studies vs. RCTs

- Observational studies (e.g. case-studies, cohort studies)
- Reflects "real world" conditions, broader group of participants
- Often done out of necessity, hypothesis generating
- Use of statistical techniques can adjust for known confounders
- Example: Hormone replacement therapy (HRT) and the Women's Health initiative (WHI)

Women and HRT use: Observational studies vs. RCTs

Observational studies (meta-analysis of 21 studies) showed that current HRT users had a reduction in:

- **CHD incidence** (RR=0.80; (0.68,0.95))
- **Mortality** (RR=0.62; (0.40,0.90))

- Further control for SES, alcohol consumption, physical activity, CHD risk factors) – *no statistically significant association between HRT and CHD*

Women and HRT use: Observational studies vs. RCTs

- WHI – Women's Health Initiative
- Large NIH-funded RCT of estrogen plus progestin vs. placebo (one of four trials)
- 16,608 postmenopausal women aged 50-79 years with an intact uterus at baseline were recruited by 40 US clinical centers in 1993-1998.
- During planning of WHI, the view that HRT prevents CHD was so entrenched that some argued unethical to deny women HRT

Women and HRT use: Observational studies vs. RCTs

- On 5/31/02, (mean of 5.2 years of f/u), the DSMB recommended stopping the trial for safety
- Found that Prempro (estrogen+progestin) increases risk of breast cancer, CHD, stroke, and pulmonary embolism
- Est. hazard ratio (HR) for CHD: 1.29, with 95% CI (1.02-1.63)
- "*…results indicate that this regimen should not be initiated or continued for primary prevention of CHD…*"

Women and HRT use

- What could cause observational studies to differ from RCTs?
  - **Selection bias** – why women selected to be on HRT
  - **Confounding bias** – what is associated with being on HRT as well as CHD?
  - **Non-generalizability** – women who take part in WHI RCT may be very different that general population of menopause-aged women

Women and HRT use

- What could cause observational studies to differ from RCTs?
  - Residual confounding in observational study
  - "Healthy-user" bias – current users of HRT are a self-selected group with favorable characteristics
  - Women recruited for randomization had to be willing to start taking HRT or placebo— blinded—for several years at 'the flip of a coin'.

Disadvantages to RCTs

- Large, expensive, logistically difficult
- High variance with respect to estimated treatment effect across multiple trials, contradictory results
- Different management of subjects, lack of uniform use of therapeutics
- Experimental protocol may not be representative of clinical practice

Seamless phase II/III Trials

- Due to the total length of time needed for Phase II and then Phase III testing, a newer approach is to use a seamless design

- Data collected in the phase II stage are combined with those data obtained at the confirmatory (phase III) stage for final analysis

An example of a seamless phase II/III study

Phase II stage:

- Patients accrued until time t1.
- At t1 accrual will be suspended and patients will be followed for a minimum time f1.
- A comparison of the treated versus control groups based on progression-free survival (PFS)
- If significant, accrual will resume until a total of M patients are accrued.

Phase III stage:

- follow-up will continue for an additional time. At the end of the study OS will be evaluated on all M patients. The total sample size M is that of the phase III study.

Comparative Effectiveness Research (CER)

- RCTs measure *efficacy* under ideal, well-controlled clinical conditions instead of *effectiveness* in a "real world" environment

- The core question of comparative effectiveness research is which treatment works best, for whom, and under what circumstances.

- Large variations observed in physician practices, tests conducted, and outcomes

CER - example

Example: High blood pressure

- Over 20 million Americans are treated with prescription drugs for high blood pressure
- An older set of drugs known as diuretics had been preferred, but have possible side-effects
- Shift to ACE inhibitors and calcium channel blockers. (significantly more expensive than the diuretics.)

CER- example

- NIH study (ALLHAT) in 2002 – randomized controlled CER
- Randomized ~33k people to three different anti-hypertensive therapies (ACE, CCB, diuretics)
- Primary outcome: fatal CHD or non-fatal MI

CER- example

- Analysis concluded there was no difference between the three interventions
- Showed that hypertension control to prevent heart failure can be achieved in most patients regardless of treatment used
- Also showed that diuretics (older, less expensive) can be an effective first-step of the antihypertensive treatment regimen.

(ALLHAT) JAMA 2002;288:2981-2997.

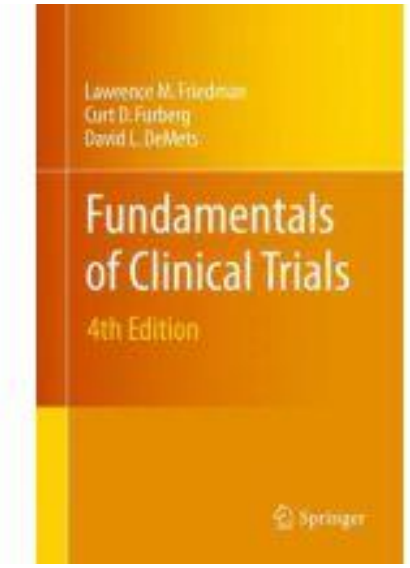
CER-example

- The design of a CER trial is typically based on a hierarchical structure

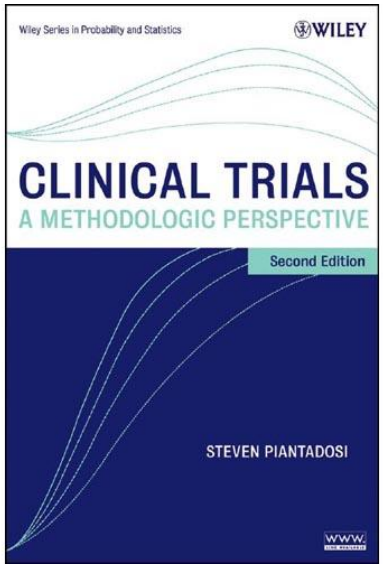
  (region -> clinics -> MD -> patient)

- Due to the likely correlation between patient outcomes (region, clinic, practice) and the potential for contamination bias, cluster randomization and analysis is performed

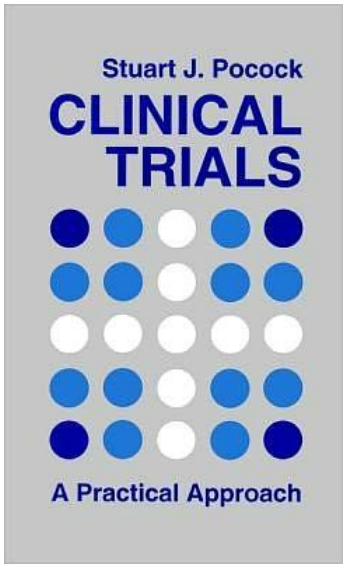
Books recommend



L. M. Friedman
C. D. Furberg
D. L. Demets
(2010)



S. Piantadosi
(2005)



S. J. Pocock
(1984)