BI476: Biostatistics - Case Studies

Lec03: Designing Clinical Trials (临床试验设计)

Maoying,Wu
ricket.woo@gmail.com

Dept. of Bioinformatics & Biostatistics
Shanghai Jiao Tong University

Spring, 2018

Outline

Next Section ...

Pharmaceutical Drug Design: 3 Phases

Phase I trials: Clinical pharmacology and toxicity study

Objective:	The main objective is safety, by providing information on the **pharmacokinetics** and **pharmacodynamics**
Design	Usually single or multiple dose-escalation studies.
Subjects	Normal healthy subjects. Patients may be used, particularly with anti-oncology drugs. The objective in oncology studies is to determine the dose to be used in phase II studies (the maximum tolerated dose, MTD)
Sample size	Approximately 20 to 80 subjects.

Pharmaceutical Drug Design: 3 Phases

Phase I trials: Clinical pharmacology and toxicity study

Objective:	The main objective is safety, by providing information on the **pharmacokinetics** and **pharmacodynamics**
Design	Usually single or multiple dose-escalation studies.
Subjects	Normal healthy subjects. Patients may be used, particularly with anti-oncology drugs. The objective in oncology studies is to determine the dose to be used in phase II studies (the maximum tolerated dose, MTD)
Sample size	Approximately 20 to 80 subjects.

Pharmacokinetics (药代) Process by which a drug is absored, distributed, metabolized, and eliminated by the body.

Pharmacodynamics (药效) Study of action or effects of drugs on living organisms or living systems.

Pharmaceutical Drug Design: 3 Phases

Phase II trials: Initial clinical investigation on treatment effect

Objective: To evaluate the potential effectiveness of a drug based on clinical endpoints for a particular indication or indications, the common short-term side effects, and the risks associated with the drug. Providing data on the doses to be used for III trials.

Design: Often single-arm, to be compared with historical controls or current treatment:

- Randomized dose ranging design
- Randomized titration design
- Two-stage phase II design (oncology)
- Multistage design
- Bayesian design
- Randomized phase II
- Multiple-endpoint design

Subjects: Patients with disease.

Sample size: Often $<$ 100 patients.

Pharmaceutical Drug Design: 3 Phases

Phase III trials: Full-scale evaluation of the effects

Objective:	To compare the efficacy of the new treatment with the standard regimen in a scientifically rigorous manner.
Design:	Randomized trial of the new treatment *versus* the control regimen.
Subjects:	Patients with disease.
Sample size:	Often 100 to 1000+ patients.

We will focus on the phase III study in the following slides.

Phase III Clinical Trials

Randomized controlled trials

- Gold-standard clinical design:
  - **New** intervention compared with a **control**
  - Treatment assignment made randomly.
- Randomization:
  - Remove **bias** in subject allocation to treatments.
  - Tends to produce comparable group (w.r.t. confounders)
  - Allows valid statistical tests to be performed.

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

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

- external/historical control (外部/历史对照)

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

- external/historical control (外部/历史对照)
- no-treatment control (空白对照)

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

- external/historical control (外部/历史对照)
- no-treatment control (空白对照)
- placebo control (安慰剂对照)

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

- external/historical control (外部/历史对照)
- no-treatment control (空白对照)
- placebo control (安慰剂对照)
- active/positive control （阳性对照）

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

- external/historical control (外部/历史对照)
- no-treatment control (空白对照)
- placebo control (安慰剂对照)
- active/positive control （阳性对照）
- dose-response control (剂量-响应对照)

Clinical Trials: Notations and Priciples

A clinical trial is a propspective study comparing the efficacy and safety of an intervention (干预) against a control (对照) in human subjects.

3 Principles

- Randomization (随机)
- Controlled (对照)
- Replication/Sample size (重复/样本量)

The control can be

- external/historical control (外部/历史对照)
- no-treatment control (空白对照)
- placebo control (安慰剂对照)
- active/positive control （阳性对照）
- dose-response control (剂量-响应对照)
- hybrid control (组合对照)

A Randomized Clinical Trial: Example

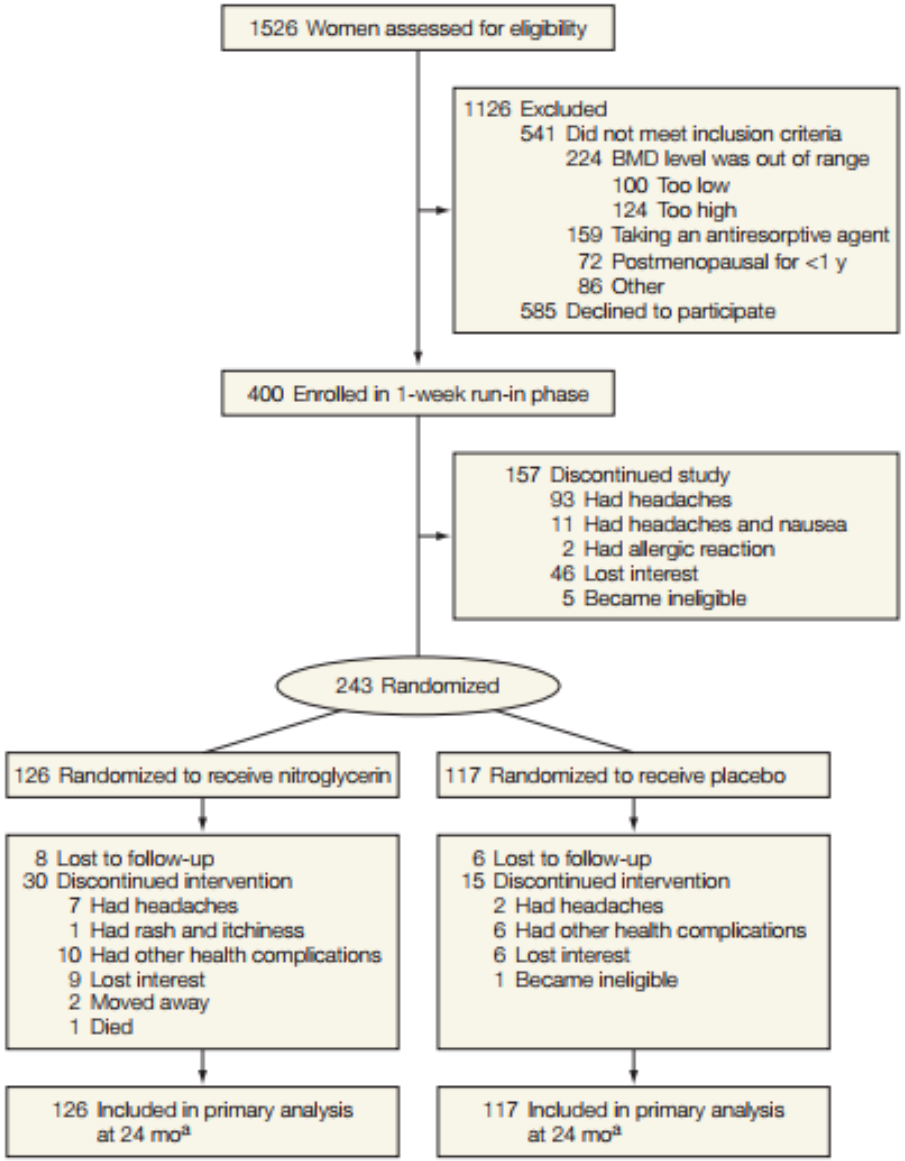
Effect of Nitroglycerin Ointment on Bone Density and Strength in Postmenopausal Women

Sophie A. Jamal, et al. JAMA 2011; 305(8):800-807

Objective	To determine if nitroglycerin increases lumbar spine bone mineral density (BMD)
Design	**Single-center**, **double-blind**, **placebo-controlled** **randomized** trial.
Intervention	Nitroglycerin ointment (15 mg/d) or placebo applied at bedtime for 24 months.
Primary outome	Areal BMD at the lumbar spine, femoral neck, and total hip.
Results	blahblah
Conclusion	Among postmenopausal women, nitroglycerin ointment modestly increased BMD and decreased bone reabsorption.

A Randomized Clinical Trial: Example

Participant flow diagram



Nitroglycerin and BMD: Baseline characteristics

Table : Baseline characteristics of study participants

Characteristics	Nitroglycerin (n=126)	Placebo (n=117)
Age (y)	61.3 (6.6)	61.9 (7.3)
Weight (kg)	70.3 (11.9)	70.9 (13.3)
White race (%)	118 (94)	107 (91)
Years since menopause	11.8 (8.2)	11.8 (8.3)
Walks $\geq 2h$ per wk (%)	104 (89)	109 (87)
Nonsmoker (%)	124 (98)	113 (97)
Vitamine D intake (IU/d)	783.2 (251.2)	753.2 (237.2)
Cacium intake (mg/d)	1548.8 (317.2)	1565.6 (373.6)
T score (Lumbar spine)	-0.9 (0.6)	-1.1 (0.6)
T score (Femoral neck)	-0.9 (0.6)	-0.8 (0.7)
T score (Total hip)	-0.6 (0.7)	-0.6 (0.7)

Results

Table : Absolute BMDs of different sites at baseline, 12 and 24 months

	BMD, Absolute value (95%CI)		
Site and Group	**Baseline**	**12 months**	**24 months**
Lumbar spine			
- Placebo	1.06 (1.05-1.08)	1.06 (1.05-1.08)	1.08 (1.08-1.09)
- Nitroglycerin	1.05 (1.04-1.07)	1.11 (1.10-1.13)	1.14 (1.12-1.15)
Total hip			
- Placebo	0.93 (0.91-0.94)	0.92 (0.91-0.94)	0.92 (0.90-0.94)
- Nitroglycerin	0.92 (0.91-0.94)	0.96 (0.94-0.98)	0.97 (0.96-0.99)
Femoral neck			
- Placebo	0.87 (0.86-0.89)	0.87 (0.85-0.88)	0.86 (0.85-0.88)
- Nitroglycerin	0.88 (0.86-0.90)	0.91 (0.89-0.92)	0.93 (0.92-0.95)

Sample size requirements

In order to compute the sample size required for comparing two means, we need

- $\delta_0 = 0.02$: difference between the nitroglycerin and control group;
- $\sigma = 0.045$: estimated standard deviation for treatment or control group;
- $\alpha = 0.05$: Two-tailed type I error;
- $1 - \beta = 0.90$: Power to detect the difference

Then, sample size for each group is:

$$\begin{aligned} n = & \quad 2 \times \left(\frac{(z_{1-\frac{\alpha}{2}} + z_{1-\beta})\sigma}{\delta_0} \right)^2 \\ = & \quad 106.4 \approx 107 \end{aligned}$$

Statistical test for comparing the differences in changes of BMD

- $H_0 : \mu_1 = \mu_2$;
- $H_a : \mu_1 \neq \mu_2$;
- Two side t-test with $\alpha = 0.05$;
- Statistic: $t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$
- $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_c^2 \frac{n_1 + n_2}{n_1 n_2}}$
- $S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$
- $t = 8.80 \sim t_{df = n_1 + n_2 - 2}$
- $p = 2.69 \times 10^{-16}$
- Reject the null hypothesis.
- 95% confidence interval of difference: $(\bar{x}_1 - \bar{x}_2) \pm t_{0.975, n_1 + n_2 - 2} \times S_{\bar{x}_1 - \bar{x}_2}$

Key issues and the corresponding solution

Table : Issues and Solutions

Issues	**Solutions**
Procedure selection bias	Randomization (随机化)
Assessment bias	Masking/blinding (盲法)
Assessment bias	Objective assessment (客观评价)
Treatment-time confounding	Concurrent controls (同期对照)
Disease remission/progression	Concurrent controls (同期对照)
Variation	Replication (重复，保证足够样本量)

Next Section ...

Clinical Trials: Various Controls

- No control
- Historical control (single-arm)
- Concurrent but non-randomized control
- Randomized

Historical control design

A.k.a single-arm study (单臂试验).

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 population
- 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 new treatment. Control groups taken from the literature are a particularly poor choice. Covariate analysis can be used to adjust for patient selection, but all other

Concurrent control design (同期对照)

- Not randomized
- Patients compared, treated by different strategies, same period

Advantage

- Eliminate time trend
- Data of comparable quality

Disadvantage

- Selection Bias
- Treatment groups not comparable
- Covariance analysis not adequate

Randomization can reduce the selection bias

Table : Clinical trials on the use of anticoagulant therapy on acute MI (1977)

	#studies	#($p < .05$)	Estimated reduction in total mortality
Non-randomized			
- Historical controls	18	15	50%
- Concurrent controls	8	5	50%
Randomized	6	1	20%

Randomization can reduce the selection bias

Table : Clinical trials on the use of anticoagulant therapy on acute MI (1977)

	#studies	#($p < .05$)	Estimated reduction in total mortality
Non-randomized			
- Historical controls	18	15	50%
- Concurrent controls	8	5	50%
Randomized	6	1	20%

- The difference in estimated reduction is probably due to biases in the non-randomized trials.

Randomization can reduce the selection bias

Table : Clinical trials on the use of anticoagulant therapy on acute MI (1977)

	#studies	#($p < .05$)	Estimated reduction in total mortality
Non-randomized			
- Historical controls	18	15	50%
- Concurrent controls	8	5	50%
Randomized	6	1	20%

- The difference in estimated reduction is probably due to biases in the non-randomized trials.
- Selection bias can lead historically controlled studies to inappropriately favor the new intervention.

Randomization can reduce the selection bias

Table : Clinical trials on the use of anticoagulant therapy on acute MI (1977)

	#studies	#($p < .05$)	Estimated reduction in total mortality
Non-randomized			
- Historical controls	18	15	50%
- Concurrent controls	8	5	50%
Randomized	6	1	20%

- The difference in estimated reduction is probably due to biases in the non-randomized trials.
- 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 exists.

Commonly used Phase-III designs

- Parallel (平行)
- Crossover (交叉)
- Factorial (因子)
- Group/Cluster (组)
- Adaptive (适应性)

Parallel design

Patient	Treatment
1	A
2	B
3	B
4	A

- In a parallel study design, each subject is randomized to one and only one treatment.
- Most large clinical trials adopt this approach.
- During the trial, participants in one group receive drug A in parallel to participants in the other group receiving drug B

Crossover design

Patient	Period 1	Period 2
1	A	B
2	B	A

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

Crossover design

Patient	Period 1	Period 2
1	A	B
2	B	A

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

Pros

- Each subject serves as own control $\Rightarrow$ variability reduced.

Crossover design

Patient	Period 1	Period 2
1	A	B
2	B	A

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

Pros

- Each subject serves as own control $\Rightarrow$ variability reduced.

Cons

- Condition must be chronic (e.g., HBP, arthritis), either "cure" or "death" before the second treatment would ruin the design.

Crossover design

Patient	Period 1	Period 2
1	A	B
2	B	A

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

Pros

- Each subject serves as own control $\Rightarrow$ variability reduced.

Cons

- Condition must be chronic (e.g., HBP, 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. Else you need wash-out period to avoid a carry-over effect.

Crossover Trials: Advantage and disadvantages

Advantages

- Reducing the variability since the treatment comparison is only within-subject other than between-subject
- Smaller sample size needed.

Crossover Trials: Advantage and disadvantages

Advantages

- Reducing the variability since the treatment comparison is only within-subject other than between-subject
- Smaller sample size needed.

Disadvantages

- Strict assumption about carry-over effects;
- Only appropriate for chronic diseases;
- Drop-out may occurred before second period;
- Period effect.

High-order Crossover Designs

- Note that in crossover design, the number of periods does not necessarily have to be equal to the number of treatments to be compared.
- Here is an example of 2×3 crossover design for comparing two treatments with three periods.

High-order Crossover Designs

- Note that in crossover design, the number of periods does not necessarily have to be equal to the number of treatments to be compared.
- Here is an example of 2×3 crossover design for comparing two treatments with three periods.

Table : Two-sequence Dual Crossover Design

	Period 1	**Period 2**	**Period 3**
Sequence 1	A	B	B
Sequence 2	B	A	A

High-order Crossover Designs

- Note that in crossover design, the number of periods does not necessarily have to be equal to the number of treatments to be compared.
- Here is an example of 2×3 crossover design for comparing two treatments with three periods.

Table : Two-sequence Dual Crossover Design

	Period 1	**Period 2**	**Period 3**
Sequence 1	A	B	B
Sequence 2	B	A	A

Types of high-order crossover designs

- Balaam's design: AA, BB, AB, BA
- Two-sequence dual design: ABB, BAA
- Double (replicated) design: AABB, BBAA
- Four-sequence design: AABB, BBAA, ABBA, BAAB
- William's design with three treatments: ACB, BAC, CBA, BCA, CAB, ABC
- William's design with four treatments: ADBC, BACD, CBDA, DCAB

Factorial design

Assumed that there are two different treatments: A and B

	Treatment B	Control
Treatment A	$A + B$	A
Control	B	Control

Randomization of subjects to one of 4 possible regimens.

Factorial design

Assumed that there are two different treatments: A and B

	Treatment B	Control
Treatment A	$A + B$	A
Control	B	Control

Randomization of subjects to one of 4 possible regimens.

Pros

- Conduct 2 experiments at once!
- Investigate potential interaction between A and B

Factorial design

Assumed that there are two different treatments: A and B

	Treatment B	Control
Treatment A	$A + B$	A
Control	B	Control

Randomization of subjects to one of 4 possible regimens.

Pros

- Conduct 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 ($\Rightarrow$ reduced power).

Factorial design

Assumed that there are two different treatments: A and B

	Treatment B	Control
Treatment A	$A + B$	A
Control	B	Control

Randomization of subjects to one of 4 possible regimens.

Pros

- Conduct 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 ($\Rightarrow$ reduced power).

But if there are three dffernet treatments?

Factorial design

Assumed that there are two different treatments: A and B

	Treatment B	Control
Treatment A	$A + B$	A
Control	B	Control

Randomization of subjects to one of 4 possible regimens.

Pros

- Conduct 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 ($\Rightarrow$ reduced power).

But if there are three dffernet treatments?
Balanced $2 \times 2 \times 2$ factorial design.

Group allocation design

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.
- Larger sample size required than simple randomized design.

Cluster RCT Example

1. Population settings

Aim of study
To examine the effectiveness of a school intervention for well-being and health risk behaviors.

Study design
Cluster RCT

Source population
Metropolitan Melbourne and rural districts, Australia

Study year
1997

Eligible population
Schools in 12 districts in two education regions in Melbourne and schools in 4 rural districts.

Selected population
26 metropolitan government, independent and catholic schools and country schools.

Age
13-14 years (year 8)

Female
53.2%

Excluded population
Classrooms in government, independent and catholic metropolitan schools and country schools.

Cluster RCT Example

2. Intervention allocation method

Demographic factors.

Intervention/s
Intervention involved institutional and individual-based components based on an understanding of mental health and risk behaviors that derive from social environments.

On a whole school level, intervention involved establishing an "adolescent health team" to identify effective strategies to address risk issues.

The teaching part of the intervention was derived over 10 weeks in 2 school years (years 8 and 9).

Intervention category
School-based

Intervention period
10 weeks during 2 years

Control/s
No intervention.

Sample sizes: Total $n = 26$ schools, 2678 students.
Intervention $n = 12$
Control $n = 14$

Baseline comparisons
The intervention group reported slightly lower-level of parental smoking and parental separation.

Cluster RCT Example

3. Outcome and methods of analysis

Primary outcomes
Smoking Prevalence (any smoking or regular smoker)

Secondary outcomes
None

Follow-up periods
1,2 and 3 years from baseline.

Evaluation
Students completed questionnaires at baseline (beginning of year 8) and were followed-up at 1 (end of year 8), 2 (end of year 9) and 3 years (end of year 10). Absent students were surveyed at a later date or telephoned (along with students who had left the schools).

Analysis method
Multivariate analysis. Stated that analysis was intention-to-treat but it appears that only students that took part in each measurement stage were included in the analysis.

Cluster RCT Example

4. Results

Primary outcomes

Table : Prevalence of smoking (intervention vs. control)

	Year 1	Year 2	Year 3
Any smoking	22.0% vs. 24.9% (OR: 0.89 (0.72-1.12))	25.0% vs. 18.7% (OR: 0.92 (0.63-1.33))	24.9% vs. 28.2% (OR: 0.91 (0.67-1.24))
Regular smoking	4.9% vs. 8.3% (OR: 0.66 (0.46-0.95))	7.7% vs. 11.9% (OR: 0.72 (0.47-1.09))	11.8% vs. 15.6% (OR: 0.79 (0.58-1.07))

Reported ORs were adjusted for baseline measurements and gender, family structure, Australian born and parental structure.

Cluster RCT Example

5. Notes

Limitations identified by author
The small number of schools in the trial limits the effectiveness of the randomization process.

Limitations identified by review team
Although it is implied taht schools were the units of randomization, randomization was primarily by district. It is unclear whether this was taken into account in the analysis.

Evidence gaps and/or recommendations for future research
Research to investigate specific mechanisms that affect change.

Hybrid design

Hybrid design combines historical controls and traditional controls.

These **criteria** must be met:

- Same entry criteria and evaluation factors.
- Participant recruitment by the same clinic or investigator
- Data from historical control participants must be fairly current

Advantages

- Potentially, the need for few participants to be entered into a trial.

Disadvantages

- Bias can be introduced from nonrandomized participants (historical controls).

Adaptive design

Adaptive trial design refers to a clinical trial methodology that allows trial design modifications to be made after patients have been enrolled in a study, without compromising the scientific method.

In order to maintain the integrity of the trial, these modification should be clearly pre-defined in the protocol. When designed well, an adaptive trial empowers sponsors to respond to data collected during the trial.

Adaptive design

Adaptive trial design refers to a clinical trial methodology that allows trial design modifications to be made after patients have been enrolled in a study, without compromising the scientific method.

In order to maintain the integrity of the trial, these modification should be clearly pre-defined in the protocol. When designed well, an adaptive trial empowers sponsors to respond to data collected during the trial.

Types of adaptive trial designs

- Dropping a treatment arm
- Modifying the sample size
- Balancing treatment assignment using an adaptive randomization
- Stopping a study early for success or failure

Next Section ...

Efficacy of the new treatment compared to a control (有效性评估分类)

- The new treatment has superior efficacy to the control (Superiority trial, 优效性试验)
- The new treatment has efficacy equivalent to that of the active control (Equivalence trial，等效性试验)
- The new treatment is not much worse than the active control (Non-inferiority trial, 非劣效性试验)

Superiority Design (优效性试验设计)

Aim

Show that a new drug is **better** than control w.r.t. the efficacy variable of interest.

Statistical Tests

H_0: No difference in effect between the treatment and controls.

Table : Superiority trials

p-value	Indication	Conclusion
$p < .001$	Strong evidence	"is superior"
$p = 0.02$	Some evidence	"Seems superior"
$p = 0.06$	Weak evidence	"Might be superior"
$p = 0.3$	No evidence	"Seems not superior"

Equivalence Design (等效性试验设计)

Purpose

To confirm the absence of a clinically meaningful difference between treatments.

Hypothesis testing Equivalence is inferred when ENTIRE confidence interval falls exclusively within equivalence margin:

$$(-\delta, +\delta)$$

Noninferiority Design (非劣效性试验设计)

A non-inferiority trial aims to demonstrate that the effect of a new treatment is as good as, or better than, that of the standard one.

This is assessed by demonstrating that the new treatment is not worse than the comparator by more than a specified margin (δ_0).

The new treatment might be tested to establish taht it matches the efficacy of standard one, and meanwhile has secondary advantages (e.g., in terms of safety, convenience to the patients, or cost-effectiveness).

Alternatively, it might have potential as a second-line therapy to the standard (in cases when the standard fails or is not tolerated).

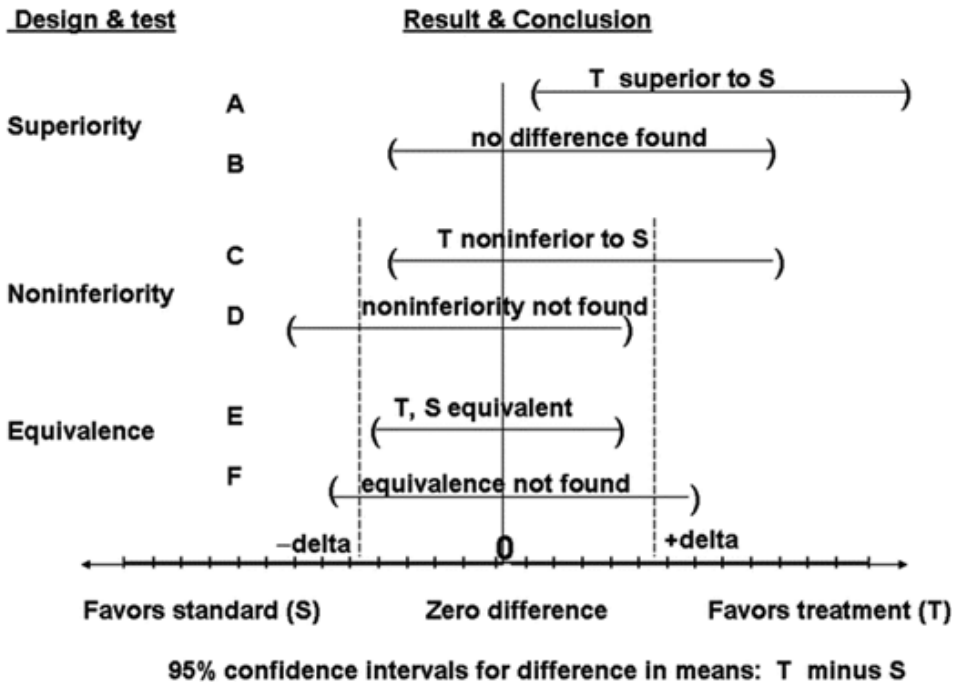
Hypothesis formula

Table : Hypotheses formulation for superiority, non-inferiority, and equivalence trials

Study type	Null hypothesis	Alternative hypothesis	Statistic
Statistical superiority	$H_0 : C - T \geq 0$	$H_a : C - T < 0$	$Z = \delta / s$
Clinical superiority	$H_0 : C - T \geq -\delta_0$	$H_a : C - T < -\delta$	$Z = (\delta - \delta_0)/s$
Non-inferiority	$H_0 : C - T \geq \delta_0$	$H_a : C - T < \delta$	$Z = (\delta + \delta_0)/s$
Equivalence	$H_0 : \lvert C - T \rvert \geq \delta_0$	$H_a : \lvert C - T \rvert < \delta$	$Z_1 = (\delta + \delta_0)/s, Z_2 = (\delta -$

- C: control or standard treatment;
- T: new treatment;
- δ_0: clinically admissible margin of non-inferiority/ equivalence/superiority;
- δ: Observed difference;
- One-sided test is performed in both superiority and non-inferiority trials;
- Two-sided test is performed in equivalence test.

Hypothesis testing: Sample size for continuous outcome



Hypothesis testing: Sample size for continuous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

Hypothesis testing: Sample size for continuous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

- Equivalence design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

Hypothesis testing: Sample size for continuous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

- Equivalence design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

- Statistical superiority design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta} \right)^2 \times s^2$$

Hypothesis testing: Sample size for continuous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

- Equivalence design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times s^2$$

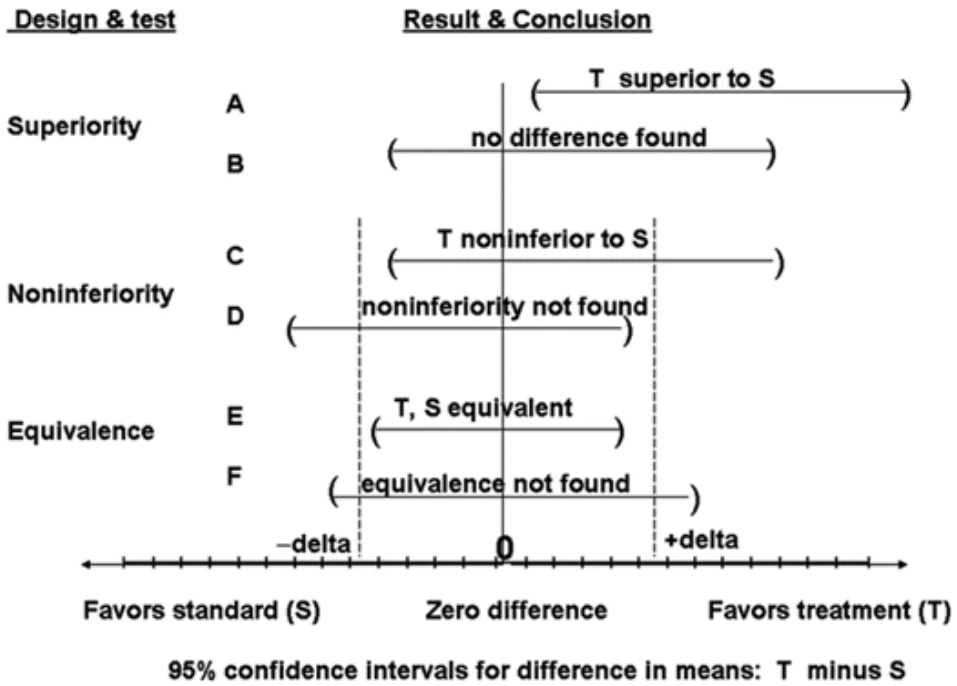
- Statistical superiority design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta} \right)^2 \times s^2$$

- Clinical superiority design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta - \delta_0} \right)^2 \times s^2$$

Hypothesis testing: Sample size for dichotomous outcome



Hypothesis testing: Sample size for dichotomous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1 - p)$$

Hypothesis testing: Sample size for dichotomous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

- Equivalence design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

Hypothesis testing: Sample size for dichotomous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

- Equivalence design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

- Statistical superiority design

$$N = \frac{1}{2} \times \left(\frac{z_{\frac{\alpha}{2}} + z_{1-\beta}}{\arcsin\sqrt{p} - \arcsin\sqrt{p_0}} \right)^2$$

Hypothesis testing: Sample size for dichotomous outcome

- Non-inferiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

- Equivalence design

$$N = 2 \times \left(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\delta_0} \right)^2 \times p \times (1-p)$$

- Statistical superiority design

$$N = \frac{1}{2} \times \left(\frac{z_{\frac{\alpha}{2}} + z_{1-\beta}}{\arcsin\sqrt{p} - \arcsin\sqrt{p_0}} \right)^2$$

- Clinical superiority design

$$N = 2 \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta - \delta_0} \right)^2 \times p \times (1-p)$$

Sample size for dichotomous outcome: Example

Goal To test whether there is a difference in the efficacy of mirtazapine (new drug) and sertraline (standard drug) for the treatment of resistant depression in 6-week treatment duration.

- $p = 0.40$: the response rate of standard treatment group;
- $p_0 = 0.58$: the response rate of new drug treatment group;
- $\delta = p_0 - p = 0.18$: the real difference between the two treatment effects;
- $\delta_0 = 0.10$: clinically admissible margin;
- $1 - \beta = 0.80$: Power;
- $1 - \alpha = 0.95$: Significance level.

Sample size for dichotomous outcome: Example

Goal To test whether there is a difference in the efficacy of mirtazapine (new drug) and sertraline (standard drug) for the treatment of resistant depression in 6-week treatment duration.

- $p = 0.40$: the response rate of standard treatment group;
- $p_0 = 0.58$: the response rate of new drug treatment group;
- $\delta = p_0 - p = 0.18$: the real difference between the two treatment effects;
- $\delta_0 = 0.10$: clinically admissible margin;
- $1 - \beta = 0.80$: Power;
- $1 - \alpha = 0.95$: Significance level.

Sample size calculation

- $N_{NI} = 2 \times \left(\frac{z_{0.95} + z_{0.80}}{0.10}\right)^2 \times 0.40 \times 0.60 = 298$
- $N_{EQ} = 2 \times \left(\frac{z_{0.975} + z_{0.80}}{0.10}\right)^2 \times 0.40 \times 0.60 = 378$
- $N_{SS} = \frac{1}{2} \times \left(\frac{z_{0.025} + z_{0.80}}{\arcsin\sqrt{0.40} - \arcsin\sqrt{0.58}}\right)^2 = 121$
- $N_{CS} = 2 \times \left(\frac{z_{0.95} + z_{0.80}}{0.18 - 0.10}\right)^2 \times 0.40 \times 0.60 = 466$

Next Section ...

Allocation procedures to achieve balance

- Simple randomization (简单随机)

Allocation procedures to achieve balance

- Simple randomization (简单随机)
- Biased coin randomization (有偏投币随机)

Allocation procedures to achieve balance

- Simple randomization (简单随机)
- Biased coin randomization (有偏投币随机)
- Permuted block randomization (随机排列区块随机)

Allocation procedures to achieve balance

- Simple randomization (简单随机)
- Biased coin randomization (有偏投币随机)
- Permuted block randomization (随机排列区块随机)
- Balanced permuted block randomization (平衡随机排列区块随机)

Allocation procedures to achieve balance

- Simple randomization (简单随机)
- Biased coin randomization (有偏投币随机)
- Permuted block randomization (随机排列区块随机)
- Balanced permuted block randomization (平衡随机排列区块随机)
- Stratified randomization (分层随机)

Allocation procedures to achieve balance

- Simple randomization (简单随机)
- Biased coin randomization (有偏投币随机)
- Permuted block randomization (随机排列区块随机)
- Balanced permuted block randomization (平衡随机排列区块随机)
- Stratified randomization (分层随机)
- Minimization method （最小化方法）

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation
  - n known in advance, exactly

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation
  - n known in advance, exactly
  - $n/2$ selected at random and assigned to Trt A, rest to B

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation
  - n known in advance, exactly
  - $n/2$ selected at random and assigned to Trt A, rest to B
- Complete randomization (most common)

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation
  - n known in advance, exactly
  - $n/2$ selected at random and assigned to Trt A, rest to B
- Complete randomization (most common)
  - n not exactly known

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation
  - n known in advance, exactly
  - $n/2$ selected at random and assigned to Trt A, rest to B
- Complete randomization (most common)
  - n not exactly known
  - marginal and conditional prob of assignment = 1/2

Simple Randomization

A specified probability p (usually equal), of patients assigned to each treatment arm, remains constant or may change but not a function of covariates or response.

- Fixed-random allocation
  - n known in advance, exactly
  - $n/2$ selected at random and assigned to Trt A, rest to B
- Complete randomization (most common)
  - n not exactly known
  - marginal and conditional prob of assignment = 1/2
  - analogous to a coin-flip/random-digit

Restricted Randomization

guarantee of procedure balance (过程平衡)

- Biased coin (Efron)
- Urn design (LJ Wei)
- Permuted block

Biased coin design (BCD)

- Allocation probability to Trt A changes to keep balanced in each group nearly equal

Biased coin design (BCD)

- Allocation probability to Trt A changes to keep balanced in each group nearly equal
- BCD(p)

Biased coin design (BCD)

- Allocation probability to Trt A changes to keep balanced in each group nearly equal
- BCD(p)
  (1) Compute the running difference $D = n_A - n_B$ and $n = n_A + n_B$;

Biased coin design (BCD)

- Allocation probability to Trt A changes to keep balanced in each group nearly equal
- BCD(p)
  (1) Compute the running difference $D = n_A - n_B$ and $n = n_A + n_B$;
  (2) Define p to be the probability of assigning Trt $> 1/2$

Biased coin design (BCD)

- Allocation probability to Trt A changes to keep balanced in each group nearly equal
- BCD(p)
  (1) Compute the running difference $D = n_A - n_B$ and $n = n_A + n_B$;
  (2) Define p to be the probability of assigning Trt $> 1/2$
  (3) We can compute p_A:

$$p_A = \begin{cases} 1/2 & D = 0 \\ 1-p & D > 0 \\ p & D < 0 \end{cases}$$

Biased coin design (BCD)

- Allocation probability to Trt A changes to keep balanced in each group nearly equal
- BCD(p)
  (1) Compute the running difference $D = n_A - n_B$ and $n = n_A + n_B$;
  (2) Define p to be the probability of assigning Trt $> 1/2$
  (3) We can compute p_A:

$$p_A = \begin{cases} 1/2 & D = 0 \\ 1-p & D > 0 \\ p & D < 0 \end{cases}$$

- Efron suggests that $p = 2/3$

Urn randomization

- Wei & Lachin: Controlled Clinical Trials, 1988
- A generalization of BCD to correct for the constant correction prob (e.g. $2/3$) regardless of the degree of imbalance
- Urn design modifies p as the function of the degree of imbalance
- $U(n, n)$:
  1. Start with Urn containing n white and n red balls;
  2. Ball is drawn at random and replaced;
  3. If red, assign B; else assign A;
  4. Add 1 ball of opposite color;
  5. Go to 2

Permuted Block Randomization

Basic Idea

(1) Divide potential patients into G groups or blocks of size $2m$

(2) Randomize each block such that m patients are allocated to A, and m to B;

(3) The total sample size is $2 \times m \times G$

(4) For each block, there are $\binom{2m}{m}$ realizations.

(5) Maximum imbalance at any time is m

Permuted Block Randomization: Concerns

If blocking is NOT masked, the sequence become somewhat predictable (e.g., $2m = 4$):

- ABAB BAB?
- AA??

This will lead to selection bias.

Permuted Block Randomization: Concerns

If blocking is NOT masked, the sequence become somewhat predictable (e.g., $2m = 4$):

- ABAB BAB?
- AA??

This will lead to selection bias.

Simple solution

- Do NOT reveal blocking mechanism;
- Use random block sizes;

Permuted Block Randomization: Concerns

If blocking is NOT masked, the sequence become somewhat predictable (e.g., $2m = 4$):

- ABAB BAB?
- AA??

This will lead to selection bias.

Simple solution

- Do NOT reveal blocking mechanism;
- Use random block sizes;

If treatment is double-blinded, no selection bias.

Balancing on Baseline Covariates (基于基线协变量的平衡)

- Stratified Randomization (分层平衡)
  - balanced w.r.t prognostic or risk factors (covariates)
- Covariate Adaptive (协变量自适应平衡)
  - Minimization
  - Pocock & Simon method

Stratififed Randomization (分层随机法)

- For large studies, randomization "tends" to be balanced.
- For small studies, a better guarantee may be needed.
- Divide each risk factor into discrete categories:
  - f: number of risk factors;
  - ℓ_i: number of categories for factor i
- Randomize within each stratum (usually blocked)

Stratififed Randomization (分层随机法)

- For large studies, randomization "tends" to be balanced.
- For small studies, a better guarantee may be needed.
- Divide each risk factor into discrete categories:
  - f: number of risk factors;
  - ℓ_i: number of categories for factor i
- Randomize within each stratum (usually blocked)

Age	Male	Female
40-	ABBA, BAAB, ...	BABA, BAAB, ...
41-60	BBAA, ABAB, ...	ABAB, BBAA, ...
60+	AABB, ABBA, ...	BAAB, ABAB, ...

Minimization

- Balances treatments simultaneously over several prognostic factors (strata)
- Does NOT balance within cross-classified cells; balance over the marginal totals of each stratum separately.
- Used when the number of stratum cells is large relative to the sample size (stratified design will yield sparse cells)
- Can be computerized

Minimization: Method

Three stratification factors:

- Gender (2 levels)
- Age (3 levels)
- Disease stage (3 levels)

Table : Current assignment for 50 patients

		Treatment A	Treatment B
Gender:	Male	16	14
	Female	10	10
Age:	40-	13	12
	41-60	9	6
	61+	4	6
Disease stage:	Stage I	6	4
	Stage II	13	16
	Stage III	7	4
Total		26	24

Minimization

Say the 51st patient enrolled in the study is male, *age* $\geq$ 61, stage III.
Summarize the above table for the corresponding factors:

	Treatment A	Treatment B	Sign of difference
Male	16	14	+
Age: 61+	4	6	-
Stage III	7	4	+
Total	27	24	2+,1-

Minimization

Say the 51st patient enrolled in the study is male, *age* ≥ 61, stage III.
Summarize the above table for the corresponding factors:

	Treatment A	Treatment B	Sign of difference
Male	16	14	+
Age: 61+	4	6	-
Stage III	7	4	+
Total	27	24	2+,1-

Two possible criteria

- Count only the direction (sign) of the difference in each factor (2A vs. 1B) $\Rightarrow$ assign the 51st patient to B.
- Add the total overall categories (27A vs. 24B) $\Rightarrow$ assign the 51st patient to B.

Minimization

Say the 51st patient enrolled in the study is male, *age* ≥ 61, stage III.
Summarize the above table for the corresponding factors:

	Treatment A	Treatment B	Sign of difference
Male	16	14	+
Age: 61+	4	6	-
Stage III	7	4	+
Total	27	24	2+,1-

Two possible criteria

- Count only the direction (sign) of the difference in each factor (2A vs. 1B) $\Rightarrow$ assign the 51st patient to B.
- Add the total overall categories (27A vs. 24B) $\Rightarrow$ assign the 51st patient to B.

The two criteria will usually agree, but not always.

Next Section ...

Blinding

The goal of blinding is to minimize the potential biases resulting from differences in management, treatment, assessment of patients, interpretation of results that could arise as a result of participant or investigator knowledge of assigned treatment.

Blinding

The goal of blinding is to minimize the potential biases resulting from differences in management, treatment, assessment of patients, interpretation of results that could arise as a result of participant or investigator knowledge of assigned treatment.

Types of blinding

Single-blind design: Only participants are blinded.

Double-blind design: Both participants and investigators are blinded.

Triple-blind design: Participants, investigators, and statisticians are blinded.

PROBE: Prospective Randomized Open with Blinded Endpoint Assessment.

Blinding

The goal of blinding is to minimize the potential biases resulting from differences in management, treatment, assessment of patients, interpretation of results that could arise as a result of participant or investigator knowledge of assigned treatment.

Types of blinding

Single-blind design: Only participants are blinded.

Double-blind design: Both participants and investigators are blinded.

Triple-blind design: Participants, investigators, and statisticians are blinded.

PROBE: Prospective Randomized Open with Blinded Endpoint Assessment.

The opposites of a blind trial is open-label trials, which are ethical phase I dose-escalating studies in oncology, pre-marketing, post-marketing surveillance.

Open-label/Unblinded trials

Open-label/Unblinded trials

Advantages

- Simple and fairly inexpensive
- A true reflection of clinical practice.

Open-label/Unblinded trials

Advantages

- Simple and fairly inexpensive
- A true reflection of clinical practice.

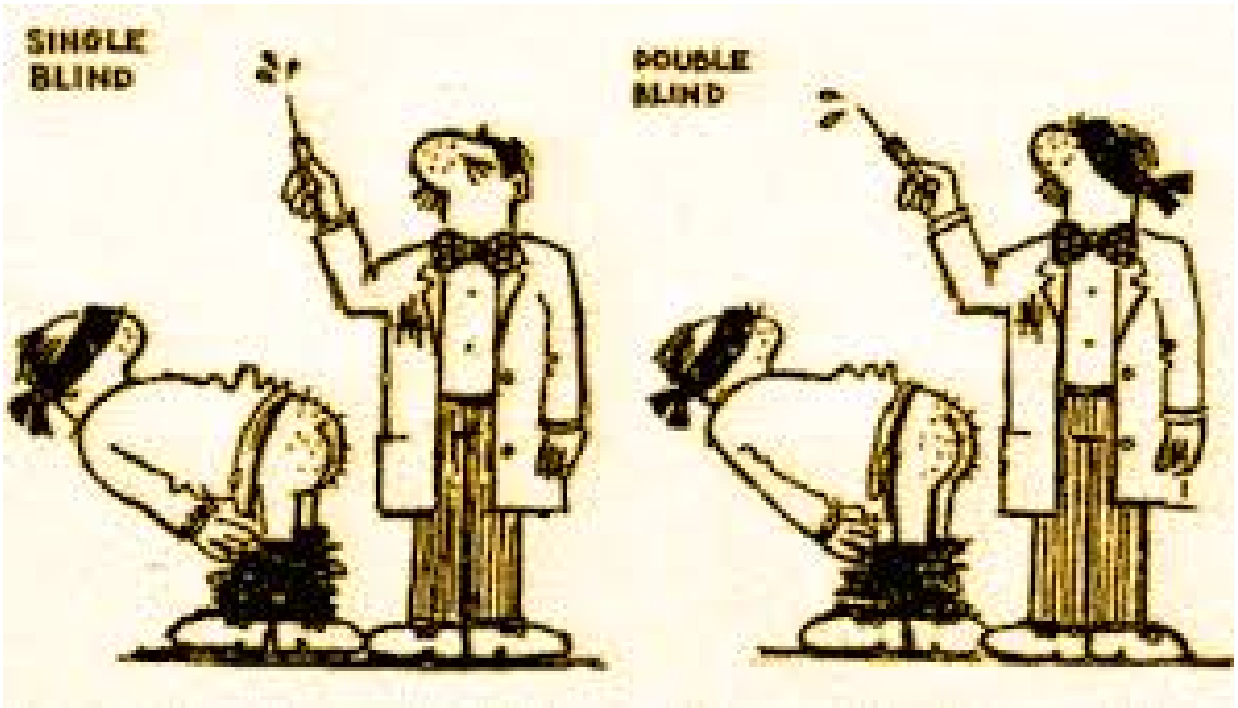
Disadvantages

- Participants may underreport adverse effects of the new treatment.
- Investigators might supply different amounts of concomitant treatments (e.g., only giving analgesics to the surgical group).

Single-blinded (单盲) trials

to counteract expectations and the placebo effect

- The participants should be unware of which treatment they are taking.
- The investigators are aware of whether the treatment is new, standard or placebo.



Single-blinded (单盲) trials

to counteract expectations and the placebo effect

- The participants should be unware of which treatment they are taking.
- The investigators are aware of whether the treatment is new, standard or placebo.

Advantages

The design is simple and allows investigators to exercise their clinical judgement when treating participants.

Single-blinded (单盲) trials

to counteract expectations and the placebo effect

- The participants should be unware of which treatment they are taking.
- The investigators are aware of whether the treatment is new, standard or placebo.

Advantages

The design is simple and allows investigators to exercise their clinical judgement when treating participants.

Disadvantages

- Patients might under- or over-report treatment effects and side-effects, based on some influence or response from the investigators.
- Investigators may give advice or prescribe additional therapy to the control arm if they feel that these patients are disadvantaged in comparison to the active arm - bias.

Double-blinded trials

Double-blinded trials

Advantages

- Reducing the biases incurred by unblindedness.

Double-blinded trials

Advantages

- Reducing the biases incurred by unblindedness.

Disadvantages

- Lessening the ability for investigators to monitor the safety of treatments.

Double-blinded trials

Advantages

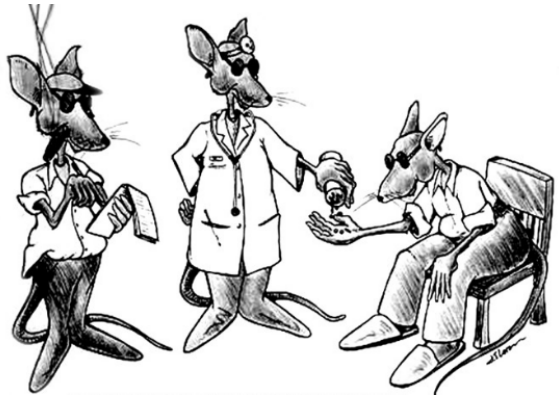
- Reducing the biases incurred by unblindedness.

Disadvantages

- Lessening the ability for investigators to monitor the safety of treatments.

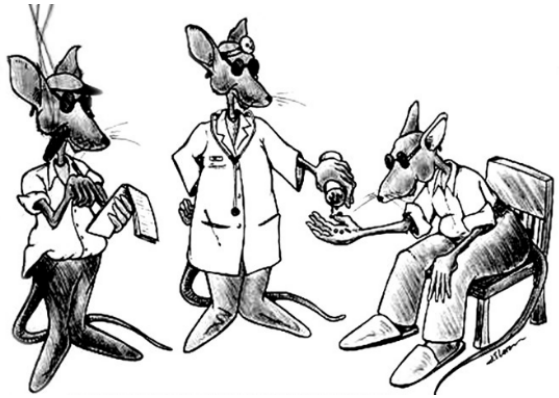
Triple-blinded trials

- Appropriate for studies with low risk of adverse events;
- Not for treatments with critical safety issues.



Triple-blinded trials

- Appropriate for studies with low risk of adverse events;
- Not for treatments with critical safety issues.

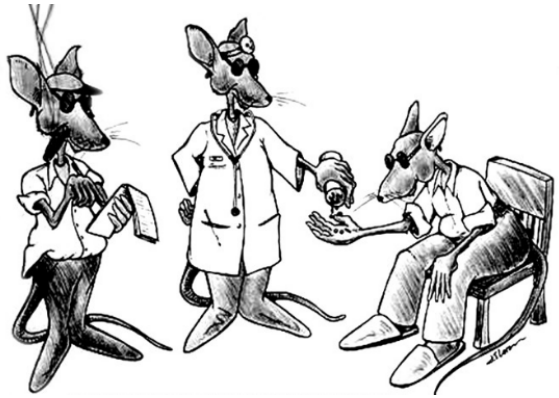


Advantages

-

Triple-blinded trials

- Appropriate for studies with low risk of adverse events;
- Not for treatments with critical safety issues.



Advantages

-

Disadvantages

- Lessening the chance that the trial may stop early to favor either treatment;
- Making the evaluation of results more objective;
- Lessening investigator's ability to monitor safety and efficacy.

PROBE trials

- Prospective, Randomized, Open-label, Blinded Endpoint Design
- Phrase coined by Hannsson et al. (Blood Pressure, 1992)
- Motivation:
  - More similar to clinical practice
  - Easier to enroll trials
  - Better patient compliance
  - Cheaper (?)

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.
- Blinding does not guarantee valid results.

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.
- Blinding does not guarantee valid results.
- Willingness to compromise is essential:

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.
- Blinding does not guarantee valid results.
- Willingness to compromise is essential:
  - Feasibility (cost, time, patient/investigator interest, patient safety)

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.
- Blinding does not guarantee valid results.
- Willingness to compromise is essential:
  - Feasibility (cost, time, patient/investigator interest, patient safety)
  - Necessity and common sense (how much bias)

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.
- Blinding does not guarantee valid results.
- Willingness to compromise is essential:
  - Feasibility (cost, time, patient/investigator interest, patient safety)
  - Necessity and common sense (how much bias)
- In some circumstances, blinding can change the nature of research question.

Blinding: Summary

- Blinding is an effective way to reduce/eliminate bias in clinical trials - do it when you can.
- Blinding does not guarantee valid results.
- Willingness to compromise is essential:
  - Feasibility (cost, time, patient/investigator interest, patient safety)
  - Necessity and common sense (how much bias)
- In some circumstances, blinding can change the nature of research question.
- Consider opportunities for blinding carefully before the trial begins.

Next Section ...

Intention-to-treat (IIT) analysis

Once randomized, always analyzed!

RCTs often suffer from two major complications: noncompliance (违反协议) and missing outcome (数据缺失).

Participants in a trial may experience:

- Completed the treatment (完成治疗);
- Did NOT receive treatment (未接受任何治疗);
- Lost-to-follow-up (失访)
- Discontinued to treatment (中途退出，有部分治疗结果数据);
- Stop treatment early (早期退出，只由很少数据);
- Died (死亡);
- Received incorrect treatment （依从性不佳）

ITT: An example

Patients:
200 patients with cerebrovascular disease

Randomization:

- Group *A*: ASA + Surgery, 100 patients
- Group *B*: ASA only, 100 patients

Follow-up: 1 year.

Result:

- 10 patients in group *A* got stroke before surgery.
- 10 patients in group *B* got stroke within a month.
- 10 more patients in group *A* got stroke after surgery within a year.
- 10 more patients in group *B* got stroke within a year of follow-up.

Any significant difference between the two groups?

Three common approaches

ITT analysis includes every subject who is randomized accordign to randomized treatment assignment by ignoring noncompliance, protocol deviations, withdrawal, and anything that happens after randomization.

Per-protocol (PP) population is defined as a subset of ITT population who complete the study without any major protocol violations.

Modified ITT (mITT) is a subset of ITT population and allows the exclusion of some randomized subjects in a justified way (such as patients who were deemed ineligible after randomization or certain patients who never started treatment.

Three common approaches

ITT analysis includes every subject who is randomized accordign to randomized treatment assignment by ignoring noncompliance, protocol deviations, withdrawal, and anything that happens after randomization.

Per-protocol (PP) population is defined as a subset of ITT population who complete the study without any major protocol violations.

Modified ITT (mITT) is a subset of ITT population and allows the exclusion of some randomized subjects in a justified way (such as patients who were deemed ineligible after randomization or certain patients who never started treatment.

Then, which approach should we choose, for the above example?

Three common approaches

ITT analysis includes every subject who is randomized accordign to randomized treatment assignment by ignoring noncompliance, protocol deviations, withdrawal, and anything that happens after randomization.

Per-protocol (PP) population is defined as a subset of ITT population who complete the study without any major protocol violations.

Modified ITT (mITT) is a subset of ITT population and allows the exclusion of some randomized subjects in a justified way (such as patients who were deemed ineligible after randomization or certain patients who never started treatment.

Then, which approach should we choose, for the above example?

- Ignoring those getting stroke before surgery in group *A*?

Three common approaches

ITT analysis includes every subject who is randomized accordign to randomized treatment assignment by ignoring noncompliance, protocol deviations, withdrawal, and anything that happens after randomization.

Per-protocol (PP) population is defined as a subset of ITT population who complete the study without any major protocol violations.

Modified ITT (mITT) is a subset of ITT population and allows the exclusion of some randomized subjects in a justified way (such as patients who were deemed ineligible after randomization or certain patients who never started treatment.

Then, which approach should we choose, for the above example?

- Ignoring those getting stroke before surgery in group *A*?
- Transfer those getting stroke before surgery in group *A* to group *B*?

Three common approaches

ITT analysis includes every subject who is randomized accordign to randomized treatment assignment by ignoring noncompliance, protocol deviations, withdrawal, and anything that happens after randomization.

Per-protocol (PP) population is defined as a subset of ITT population who complete the study without any major protocol violations.

Modified ITT (mITT) is a subset of ITT population and allows the exclusion of some randomized subjects in a justified way (such as patients who were deemed ineligible after randomization or certain patients who never started treatment.

Then, which approach should we choose, for the above example?

- Ignoring those getting stroke before surgery in group *A*?
- Transfer those getting stroke before surgery in group *A* to group *B*?
- Intention-to-treat?

RCTs Example

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?

Randomized treatments

- L-Pam (orally, 0.15mg/kg body weight, 5 consecutive days every 6 weeks for 2 years, with specified dose modification for hematological toxicity)
- Placebo (physically indistinguishable from L-Pam)

Eligibility

- Radical mastectomy for primary breast cancer (with 4 weeks of starting protocol treatment)
- Historically confirmed axillary node involvement.
- No skin ulceration or peau d'orange
- Age $\leq$ 75 years
- Not pregnant or lactating

Focus is on those patients most likely to benefit.

RCTs Example

A clinical trial for primary breast cancer (1972)

Outcomes

- Primary outcome: disease-free interval = time from mastectomy till fist detection of tumor (local regional, or distance)
- Other outcomes:
  - Survial time = time from mastectomy until death
  - Toxicity = occurrences of hematological toxicity or nausea/vomting

RCTs Example

A clinical trial for primary breast cancer (1972)

The study

- A **written protocol** documented all study procedures and information. Sample size requirements dictated several hundred patients $\Rightarrow$ 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.
- The trial was double-blinded
- **Follow-up** exams were performed every 6 weeks, with test for hematological toxicity every 3 weeks $\Rightarrow$ outcome evaluation was done consistently and objectively.
- Patient **accrual** started Sep 1972 and ended Feb 1975. In total, 370 were accrued.
- **Information 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. #(positive nodes)
- The **central coordinating office** supervised data collection and processing.
- As the data accumulated, **interim analyses** were preformed. There was pressure to release interim results

RCTs Example

A clinical trial for primary breast cancer (1972)

Results

- The final results showed that L-Pam significantly increased the disease-free survival time ($p = 0.009$)
- **Subgroup analysis** revealed that L-Pam had the largest benefit in younger patients ($\leq$ 50 years old).
- **Toxicity**: Hematological toxicity was common (25%), but never life-threatening.
- **Side effect**: Nausea/vomting was experienced by 40% of L-Pam patients and 11% of placebo patients.

Conclusion

L-Pam was adopted as the new "standard" treatment.