Part II: Basic Statistical Concepts

Michele Melia, Sc.M.

Senior Statistician
Jaeb Center for Health Research
Tampa, FL

SCT Pre-Conference Workshop
Essentials of Randomized Clinical Trials

II-1

Outline

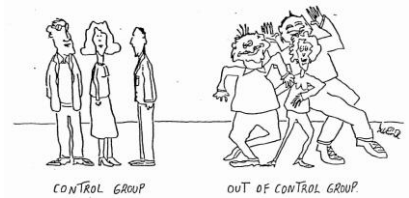
- Randomization
- Blinding (masking)
- Hypothesis testing
- Confidence intervals
- Sample size
- Intention to treat

II-2

What is randomization?

A process by which subjects are randomly assigned to a treatment in a clinical trial

- Neither the participant nor the investigator knows what treatment the participant will receive



skip

II-3

Why is randomization used?

- Problems arising with treatment assignment in clinical practice:
  - Individuals with certain disease characteristics are generally more likely to receive certain treatments (confounding by indication)
  - Inability to characterize why individuals were assigned to a particular treatment, leading to non-homogeneous groups with different (and unquantifiable) underlying risk
  - Wide variation in outcomes relative to the magnitude of differences due to treatments; treatment differences easily obscured by bias

skip

II-4

How does randomization work?

- Randomization does:
  - Reduce bias in assigning patients to treatments
  - Ensure valid statistical tests
  - Reduce unwanted variation resulting in improved power for statistical tests
- Randomization does not:
  - *Guarantee* equal distribution of prognostic factors among treatment groups
    - For large studies, the chance of imbalances is small
    - For small studies, the chance of imbalances is larger

II-5

When is randomization used?

Phase I	Rarely	Not generally necessary to achieve phase I goals of establishing toxicity/maximum tolerated dose/dose response
Phase II	Sometimes	When comparison group is helpful in defining possible biologic and adverse effects, e.g. for highly subjective endpoints. When required by FDA.
Phase III	Almost always	"Gold standard " for reducing bias in assignment of patients to treatment and estimation of treatment effects

Other methods of (non-random) treatment allocation are also sometimes used in CTs:

- Single group with or without historical controls
- Non-random allocation of 2 or more groups

II-6

Non-random methods of treatment allocation

- Alternating treatments (1st patient gets A, 2nd gets B, 3rd gets A, etc.)
- Alternating assignment by date or day of week (patient gets A if enrolled on even date, B if odd date)
- Using patient initials to determine assignment
  - A-K → treatment 1
  - M-Z → treatment 2

II-7

Basic types of randomization

- Simple
- Block
- Stratified / stratified block

II-8

Simple Randomization

A sequence from a random number table or generator is used to assign sequential patients to a study treatment using a pre-defined rule. E.g. Even number→A and Odd number→B.

Sequence from random number table	Treatment assignment
7	B
7	B
9	B
2	A
1	B
0	A
6	A

II-9

Simple Randomization

- Advantages
  - Simple
  - Each new assignment made without regard to previous assignments
- Disadvantages
  - No guarantee of equal or approximately equal sample size in each treatment group at any stage of the trial (including at the end)
    - Imbalance reduces statistical power
    - Estimates of treatment effect are not affected; only precision
  - No protection against long runs of one treatment

II-10

Block randomization

- Block size that is an integer multiple of the number of treatments is chosen (integer≥2)
- Equal numbers of patients are assigned to each treatment within a block
  - Numbers are proportional rather than equal in the case of unequal allocation
- Overcomes some disadvantages of simple randomization

II-11

Example: Block Randomization for 2 Treatments

- Possible block sizes are 4, 6, 8, etc.
- For block size of 4, there are 6 treatment-balanced permutations
  - ABAB, AABB, ABBA, BABA, BBAA, BAAB
- These may be chosen at random with replacement

Sequence from random number table	Treatment assignment
7	--
7	--
9	--
2	AABB
1	ABAB
0	--
6	BAAB

II-12

Block randomization – cont'd

- Large block size does not protect as well against long runs as small block size
- Small block size makes it easier to guess next treatment
- To make it harder to guess the next allocation when small block sizes are used, block size can be chosen at random from a pre-defined list of block sizes, e.g. 4, 6, 8
- Simple and block randomization do not guarantee balance of treatment groups on important prognostic factors

II-13

Stratification

- With stratification, a separate, independent randomization sequence is used for each prognostic group (or strata)
- To guarantee treatment balance within strata at all stages of the trial, stratification is combined with blocking
  - Use of simple randomization within strata will not guarantee treatment balance within strata
  - Consequence of imbalance on a prognostic factor is bias in the estimated treatment effect (unless analysis is adjusted for the factor)

II-14

Example – Blocked and stratified randomization

- A randomized trial comparing near versus distance activities while patching for amblyopia (lazy eye) in children 3 to <7 years old
  - Pilot study data suggested that near activities might be less effective in moderate as compared to severe amblyopia
  - Randomization was stratified by amblyopia severity; random block sizes of 4 and 6 also were used

skip

II-15

Example - continued

- If even, use block size=4; otherwise block size=6
- Use a random shuffle of the block elements

Moderate amblyopia

Random No.	Block size	Random sequence	Treatment assignments
7	6	792106 AAABBB	012679 BBABAA
1	6	851307 AAABBB	013578 BABABA

Severe amblyopia

2	4	6312 AABB	1236 BBAA
3	6	095734 AAABBB	023579 ABBABA

skip

II-16

Stratified randomization – cont'd

- Chance of imbalance on prognostic factors is small with large sample size
  - Stratification is more important when sample size is small
- As number of stratification factors increases, the number of strata grows very fast, and efficacy with respect to achieving desired balance may decrease
  - Think of case where # strata = sample size
- Be judicious in choice of stratification factors

II-17

Stratified randomization – cont'd

- If many prognostic factors must be controlled:
  - Consider combining them into an overall index and stratifying on index
  - Consider minimization
- When analyzing data, it is important to account for stratification
  - If ignored, variability due to the stratification factor is included with error variance
  - If included, variability due to stratification factor is removed from error term, increasing precision

II-18

Unequal Treatment Allocation

- With unequal treatment allocation, the study is designed to have unequal numbers of patients on the treatments
- Treatment groups of equal size are desirable from a statistical perspective for making treatment group comparisons
  - Maximizes power for a given sample size
  - However, loss of power may not be too severe as long as imbalance is not severe, e.g. 2:2:1

II-19

Unequal Treatment Allocation – cont'd

- Some reasons to consider unequal allocation:
  - More information is needed on effect of a new treatment (e.g. adverse effects, effect of dose)
  - Patients may be unwilling to be randomized if probability of assignment to control or placebo is high
  - To reduce study cost when one treatment is a lot more expensive than the other
- Principles of basic randomization regarding use of blocking and stratification still apply

II-20

Cluster Randomization

- Clusters of patients are randomized rather than the individual patients
  - Example: In trial of vitamin A supplementation for prevention of mortality in preschool children in Nepal, administrative wards were randomized to supplement or placebo (West KP, Lancet 1991)
- Cluster randomization reduces statistical efficiency (i.e., it requires more patients)
- Usually used when it is not feasible to randomize individual patients, or treatment contamination is a concern

II-21

Crossover Randomization

- Each patient is randomized to a sequence of treatments
  - Example: Effect of low-fat versus high-fat meal on AUC180 in type I diabetes (AUC180 = area under the 180 mg/dl line of the blood glucose over time curve)
  - Patients randomized to low-fat then high-fat meal or vice-versa
- Crossover randomization increases statistical efficiency (reduces sample size); each patient acts as his/her own comparator providing control for all patient-level factors
- Crossover randomization is suitable only in certain circumstances: chronic disease with stable course; treatments with rapid and short-acting effects

II-22

Adaptive Allocation (aka Adaptive Randomization)

- Information on previously enrolled patients is used to modify (or adapt) the allocation ratio, i.e. the probability of being assigned to each treatment
- Information used typically is one of:
  - Treatment
  - Covariates (prognostic factors)
  - Response (outcome)
- Other terms:
  - Biased-coin design
  - Urn design
  - Play-the-winner design

II-23

Summary

- Randomization is the primary means for controlling bias in allocation of patients to treatment in a clinical trial
- Randomization helps to generate comparable groups of patients on each treatment
- Randomization enables valid statistical tests for the evaluation of the treatments
- Judicious use of stratification with appropriate analysis can improve statistical power

II-24

Blinding (Masking)

- What is it?
  - Concealment of the treatment assignment
- Purpose? Avoid bias in:
  - Delivery of treatment / other care
  - Compliance
  - Outcome assessment
- Who can/should be masked?
  - Trial dependent
  - Patient, treating clinician, possibly other clinical personnel, outcome assessor
- More subjective outcome➔more important to mask
- Perception of quality/integrity is important

II-25

Examples of Masking in Practice

- Collaborative Ocular Melanoma Study (COMS) trial for medium-sized tumors – no masking
  - Eyes with melanoma randomized to enucleation (removal of the eye) or radiation (eye is retained)
  - Impossible to mask patient or clinical staff
  - Primary outcome: all-cause mortality (objective)
  - Secondary outcome: melanoma-related mortality (subjective; cause of death assignment made by masked mortality coding committee)

II-26

Examples of Masking in Practice

- Patching versus Atropine for Amblyopia (lazy eye)
  - Patients randomized to patching or atropine drops in good eye
  - Not possible to mask patient, parent, or treating clinician
  - Primary outcome: visual acuity at 17 weeks (subjective)
  - Visual acuity examiner is masked (by patching good eye)
- Levodopa for Residual Amblyopia (lazy eye that did not fully respond to conventional treatment)
  - Patients randomized to levodopa or placebo (2:1)
  - All patients patch good eye 2 hours/day
  - Primary outcome: visual acuity at 18 weeks (subjective)
  - Patients and clinical personnel are masked by using placebo

II-27

Hypothesis Testing

A primary objective of most clinical trials is to demonstrate the effectiveness and safety of a treatment under investigation.

The purpose of such trials is to:

- Find out which (if any) of the treatments are more effective
- Convince others of the results

II-28

Hypothesis Testing

Hypothesis testing involves:

- Collecting a sample and using the sample to estimate unknown population parameters.
- Comparing the sample estimate(s) to some hypothesized population value to see if the sample came from the specified population.

II-29

Null and Alternative Hypotheses

Hypothesis: Statement about a population parameter

Null Hypothesis (H_0): Usually, hypothesis of no difference or status quo; or, what we would like to disprove

$$H_0: \mu = 0$$

Alternative Hypothesis (H_A): A statement which contradicts the null hypothesis

$$H_A: \mu \neq 0$$

The goal of hypothesis testing is to collect a sample and determine which hypothesis is 'more likely' to have generated the observed sample.

II-30

Test Statistic

Test Statistic: A quantity computed from the data used to measure the plausibility of the null hypothesis

- Appropriate test statistic depends on the type of outcome, e.g. continuous, categorical, time to event

II-31

Example

Does performing near activities while patching for amblyopia result in more improvement in visual acuity (VA) among children age 3 to < 7 yrs as compared to distance activities?

H_o: treatment group difference in mean change in VA = 0

Results: Mean VA change in near group = 2.5 lines
Mean VA change in distance group = 2.6 lines
Difference = -0.03 lines

Test statistic (from t-test)

= (sample mean – hypothesized population mean) / std(mean)

= (-0.03 – 0) / 0.16 ≈ -0.19

II-32

P-values

- P-value gives the probability of obtaining a test statistic as extreme as the observed test statistic (or more so) given H_o is true

  Example: T-test statistic = -0.19

  P-value ≅ 0.85

- The p-value indicates that getting a test statistic value of -0.19 (or value more extreme) was very likely (85% chance) under H_o

  H_o is not rejected

II-33

Rejecting the Null Hypothesis

- If we don't reject H_o, we can't claim to 'accept H_o'
  - Suppose one makes a statement 'all swans are white'
  - To examine this statement, a sample of swans is drawn
  - Two things can happen:
    a) All swans in the sample are white
    b) At least one swan in the sample is not white
  - The event (b) establishes the falsehood of statement
  - However, the event (a) does not prove the statement!

Acknowledgment to Chris Coffey

II-34

Defining type I and II errors

Study conclusion	Truth	
	Treatments not different (Δ=0)	Treatments differ (Δ≠0)
Treatments not different (Δ=0)	True negative	False negative Type II error (β)
Treatments differ (Δ≠0)	False positive Type I error (α)	True positive

- There always is a chance of drawing a non-representative sample that leads to a wrong conclusion
- There are 2 types of wrong conclusions, referred to as type I or type II error

II-35

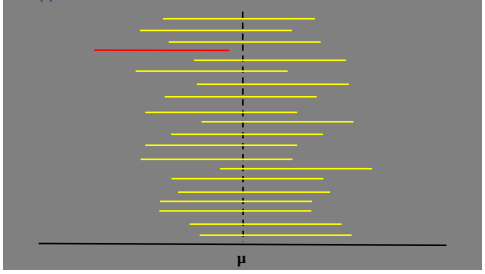
Confidence Intervals

- Provide information about uncertainty in the estimate of the population parameter (Ex: Mean difference in VA) by including lower/upper bounds around the sample estimate
  - COMET2: 0.28 D (95% CI: 0.01 – 0.55 D)
- Express how certain (confident) we are that the procedure used to generate the interval includes the population parameter
  - Higher confidence level corresponds to wider interval and improved likelihood of capturing the population parameter

II-36

95% Confidence Interval

A 95% CI is generated by a statistical procedure that captures the population parameter (μ) in 95% of its applications



Relationship with Hypothesis Testing

What is the relationship between confidence intervals and hypothesis testing?

- Decision to reject/fail to reject the null hypothesis depends on whether the confidence interval includes value(s) consistent with the null hypothesis
  - If CI includes null hypothesis → Fail to reject
  - If CI excludes null hypothesis → Reject

Example:

- COMET2: 0.28 D (95% CI: 0.01 – 0.55 D)
- Interval does not include H_o value of 0, so H_o is rejected

II-38

Sample Size

- Sample size should be assessed as early as possible during the design phase
  - Avoid waste of time designing a study with a sample size that is not feasible

II-39

Basis of Sample Size Determination

- In all clinical trials, we are selecting a sample from a target population
- The possibility exists that the sample we select will not be representative of the outcome rate or treatment effect
- Goal is to select sample size to:
  - ensure high chances of getting the correct answer
    - AND
  - enroll as few subjects as possible

II-40

What information is needed?

- Basic trial design features:
  - Comparative type (equality, equivalence, non-inferiority)
  - Randomization design (parallel group, crossover, cluster)
- Number of treatment groups
- Primary outcome, outcome rate or variance in control group, analysis method for primary outcome
- Size of treatment effect to be detected
- Risk we are willing to take that study will "miss" a true treatment difference (β=type II error; 1-β is the study power)
- Risk we are willing to take that study will erroneously conclude treatments are different (α=type I error)

II-41

Effect of basic trial design elements on sample size

Randomization Design	Relative Effect on Sample Size
Parallel Group	(Reference)
Crossover	Smaller
Cluster-randomized	Larger
Comparative Type	
Efficacy (rx difference≠0)	(Reference)
Equivalence	Larger
Non-inferiority	Smaller or larger

II-42

Outcome variable

- What is the most important outcome variable?
- What is the expected proportion with the outcome (or variance of the outcome if continuous) in the control and treatment groups?
  - Continuous outcomes usually have smaller SS than a proportion using the same measurement, but may be less clinically interpretable
  - E.g. % with systolic BP<120 and diastolic<80 versus mean reduction in mean arterial blood pressure
- Accurate estimate of control group outcome is key
- Analysis method for the primary outcome
  - Method for sample size calculation parallels the method for analysis

II-43

Outcome Variable

The smaller the treatment effect to be detected, the larger the required sample size.

- Sample size increases exponentially as the effect size decreases, regardless of the type of outcome

II-44

How to estimate outcome in the control group?

- Previous randomized trials
- Previous prospective studies
- Retrospective studies
- Pilot study

II-45

How to determine alpha (α) and beta (β)

Although α often is set at 0.05 and β at either 0.10 or 0.20, they should be study-specific

- Seriousness of disease and impact of treatment
- Public health importance of disease and treatment
- Availability of other treatments
- Cost of treatment

Sample size increases as α and β decrease

II-46

Choosing α and β

Example*: It has been suggested that a certain hospital has lower birth weight babies than the national average.

- To see if a special care nursery is needed, a sample of birth weights from the hospital are collected and used to test:
  - H_0: μ ≥ national average

    vs.
  - H_A: μ < national average

*From Rosner, Fundamentals of Biostatistics. Duxbury Press, 2010. II-47

Choosing α and β

- If H_0 is rejected, the hospital will add a special care nursery.
- If a type I error is made, the extra cost of adding a special care nursery will be recommended when it is not needed
- If a type II error is made, a needed special care nursery will not be funded.
  - As a result, some low-birth weight babies may not receive the special attention that they need

II-48

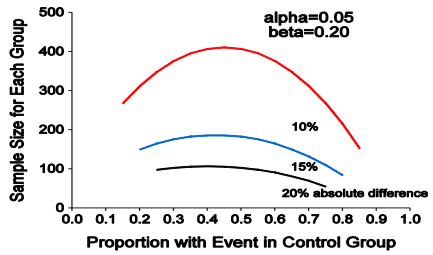
Effect of Variance of Outcome on Sample Size

Larger variance➔larger sample size

- For a proportion, variance is function of the outcome rate in the treatment groups
  - Variance ∝ P*(1-P), which is largest for P=0.5
- P is the average of the control and treated group outcome proportions

II-49

Sample size according to absolute difference in proportion of events between treatments



II-50

Adjustments to sample size estimate

- Losses to follow up
- Treatment group crossovers
- Poor treatment adherence
- Ineligible patients enrolled
- Misclassification of outcome

Presence of any of these increases sample size.

Your statistician can adjust the sample size accordingly, but this does **NOT** address possible bias that may be introduced

II-51

Ways to decrease sample size (for dichotomous outcome)

- Increase the number of outcomes in control group (assumes events per group are proportional, e.g. 2:1)
  - Lengthen follow up
  - Change primary outcome or widen outcome criteria
  - Switch to a surrogate outcome
  - Limit enrollment to higher risk patients
- Increase magnitude of treatment effect to be detected
- Minor: increase alpha or beta; change to one-sided

II-52

Ways to decrease sample size (for continuous outcome)

- Reduce variance of outcome measure
  - Change to more precise measurement method
  - Measure more than once and use average
  - Limit enrollment to patients with less variability
- Increase magnitude of treatment effect to be detected
- Minor: increase alpha or beta; change to one-sided hypothesis test

II-53

Number of treatment groups

- Increasing the number of treatment groups will increase the sample size
- Increase depends on # of specific group comparisons planned
- Increase depends on sizes of the detectable treatment effects

Example: For $\alpha=0.05$ and $\beta=0.20$,

No. of groups	2	3	
No. of comparisons	1	2	3
List of comparisons	Pc=0.5 v. Pt=0.3	Pc=0.5 v. Pt_1=0.3 Pc=0.5 v. Pt_2=0.3	Pc=0.5 v. Pt_1=0.3 Pc=0.5 v. Pt_2=0.2 Pt_2=0.2 v. Pt_1=0.3
Sample size	103 per group	123 per group	Nc=135 Nt_1=Nt_2=411

II-54

Sample Size - Summary

- Sample size and other scientific demands usually must be balanced with practical limitations of available funds and number of eligible patients
- Finding a satisfactory balance frequently involves modifying aspects of the study design
- Given the close link between study design and sample size, it is advisable to evaluate sample size requirements as early as possible in the planning process

II-55

Intention to Treat (ITT)

- Patients should be included in the group to which they were randomized for analysis, regardless of the treatment actually received.
  - Failure to adhere to or complete the assigned treatment is often due to side effects, perceived lack of efficacy, disease progression, i.e., it is at least partly an outcome of the assigned treatment
  - Failure to attribute these outcomes to the assigned treatment can introduce bias into the treatment comparison
- ITT can be viewed as a test of treatment policy, i.e. a test of outcomes given an intention to treat each patient using a particular therapy

II-56

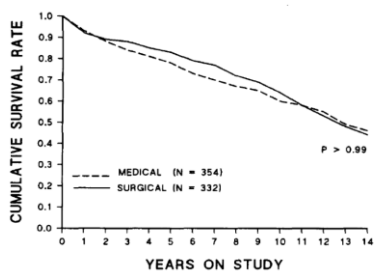
Example – Veterans Administration Cooperative Study of Coronary Artery Bypass Surgery*

- Medical therapy versus bypass surgery for CAD
  - 55% of medical therapy group received bypass surgery at some time during 14 years of follow up
  - Small % of surgery group refused surgery
- Compare 5 analysis methods:
  1. ITT ('as-randomized')
  2. Exclude treatment crossovers from analysis ('adherers-only')
  3. Include crossovers in alternate group ('treatment-received')
  4. Censor crossovers at time of treatment change ('censored')
  5. Transfer crossovers to alternate group at time of treatment change ('transition')

*Peduzzi et al. *Stat Med* 1993;12:1185-95.

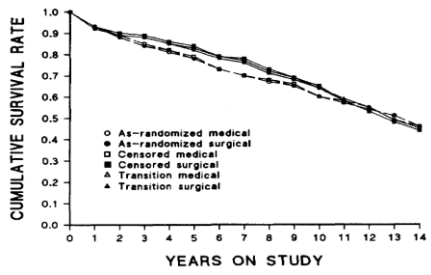
II-57

Results of ITT Analysis



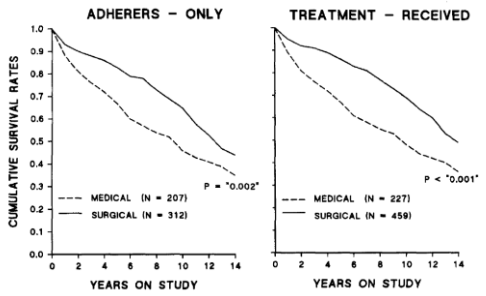
II-58

Results of ITT, Censored, and Transition Analyses



II-59

Results of Adherers-Only and Treatment-Received Analyses



II-60

Summary – Intention to Treat (ITT)

- ITT is the preferred analysis method in clinical trials as it avoids potential biases related to failure to adhere to assigned treatment
- ITT tests treatment strategy, rather than treatment received
  - Effect of following a treatment strategy is what is relevant when faced with a new patient
  - At time of initial treatment decision, it is unknown whether patient will adhere to treatment or whether other factors will intervene
  - Analyses based on knowledge of future events are not very relevant to current decision

II-61

Selected References

- Chow S-C, Liu J-P: Design and Analysis of Clinical Trials, 2nd ed. John Wiley and Sons, 2004.
- Meinert CLM: Clinical Trials: Design, Conduct, and Analysis. Oxford University Press, 1986.
- Piantadosi S: Clinical Trials: A Methodologic Perspective. John Wiley and Sons, 2005.
- Spilker, B: Guide to Clinical Trials. Raven Press, 1991.
- Controlled Clin Trials 1988; Volume 9, issue 4 has a series of articles on randomization in clinical trials by John Lachin.

Acknowledgments:

Chris Coffey, Sylvan Green, Elizabeth Lazar

II-62