Biostatistics 226

- Design of RCTs: How to avoid errors and achieve valid results
- Identifying the research hypothesis and the corresponding study design
- Primary outcomes, Analysis plans, and Sample size calculations
- Monitoring trials for efficacy, futility, and patient safety
- Bayesian methods

Lecture 1

Learning Objectives:

Upon completion of this lesson, you should be able to do the following:

- State 6 statistical objectives that will be met with proper trial design.
- Distinguish between random variability and bias.
- Define 5 sources of potential bias in clinical studies and design strategies to reduce them.
- Identify the experimental unit in a proposed study.
- Recognize confounding in a clinical study proposal.
- Recognize features that should be described in a written protocol for an RCT.

Homework (due Monday following this class):

Read: Austin PC et al. J Clin Epidemiol. 2010 Feb;63(2):142-53. doi: 10.1016/j.jclinepi.2009.06.002.

Using the PBC dataset available at http://www.biostat.ucsf.edu.ucsf.idm.oclc.org/vgsm,

- Describe the population to whom trial results apply.
- Following Austin's guidance, prepare "Table 1" for this trial.
- Identify (i) the primary outcome, (ii) the intervention arms, and (iii) potential covariates.
- What strategies were employed to avoid bias?

Why study RCT methods?

Research is conducted in settings where need for improvement is recognized, the optimal intervention strategy is not certain, and ideal information needed to determine the optimal strategy is not available.

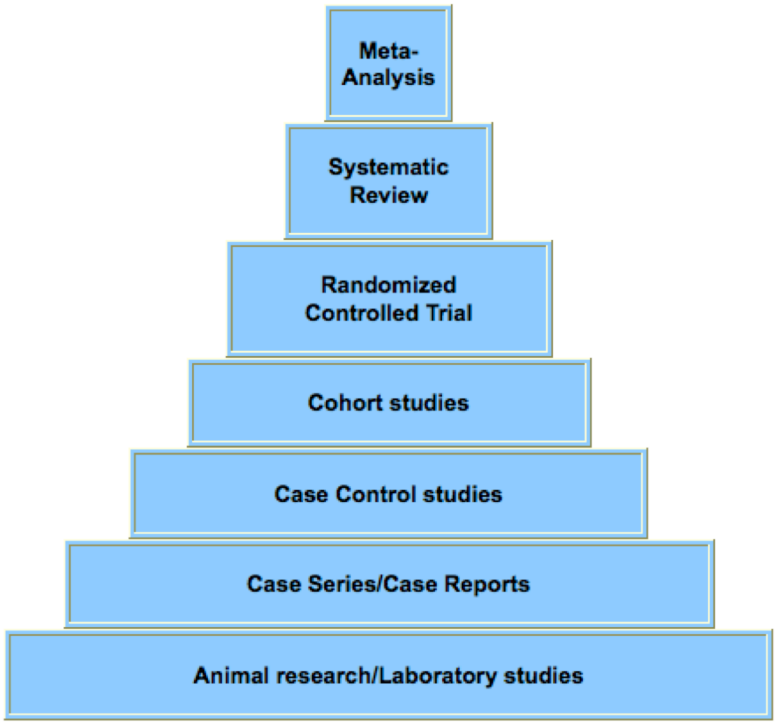
- We do not see all aspects of a disease course in a given individual – e.g., etiology, mechanisms, comorbidities, risk factors, disease progression
- We do not see breadth of disease manifestation in all individuals
- We do not see evolution of disease course over time – e.g., development of drug resistance

To address this lack, we generalize from samples of individuals.

Statistics *is the " theoretical science or formal study of the inferential process, especially the planning and analysis of experiments, surveys, and observational studies"*
(Piantadosi 2005).

Evidence Pyramid

Evolution of evidence generating and answering a health research question can be depicted by a pyramid, with earlier evidence gathered at the base and more definitive answers occurring at the top.



Pre-clinical (laboratory and animal) studies may seek to refine aspects of the research that will be studied at later stages (e.g., intervention dose).

Observational studies (Case reports, Case-control, Cohort studies, and database analyses) exposure / intervention options are not under the control of the investigator (e.g., smoking).

Experimental studies (*Randomized Controlled Trials*) exposure / intervention options are under the control of the investigator (e.g., diet, chemotherapy).

Systematic Reviews and Meta-analyses summarize evidence across studies. They appraise the quality of component trials and may make recommendations for practice.

Experimental Design Overview & Terminology

An ***experiment*** is a series of observations made under conditions controlled by the scientist.

- A randomized controlled (clinical) trial is an experiment testing (health) interventions on human subjects.

Design is the process or structure that isolates the factors of interest.

- **factor** – a variable that is controlled and varied during the course of the experiment (e.g., treatment). Usually few factors are controlled, with few levels per factor (e.g., Arms A and B; N.B. conventional labeling).

- **experimental unit** – the unit that is randomized to the study arm (factor level) and is directly exposed to the intervention

- **observational unit** – the unit on which outcomes are measured

    - In most RCTs, the experimental units and the observational units are one and the same, namely, the individual patient
    - In cluster-randomized trials, experimental and observational units differ (e.g., practice and patient, respectively)

- **primary outcome** – the intervention is hypothesized to modify this measure and the study is designed to detect the relative effects of interventions on this measure

    - summaries of the outcome data are called "statistics" (e.g., sample mean or proportion) and are represented by English alphabet (e.g., x-bar or p); data are expressed in terms of statistics

        - Analysis scale: natural; logarithm . . .

    - population analogs about which we want to learn are called "parameters" and are represented by Greek alphabet (e.g., μ or π); hypotheses are expressed in terms of parameters

	Random Assignment to Intervention Groups		
Group mean	Arm C (N_C =)	Arm E (N_E =)	Difference
Estimated (Sample)	$\bar{x}_C$	$\bar{x}_E$	$d = \bar{x}_E - \bar{x}_C$
True (Population)	μ_C	μ_E	$\delta = \mu_E - \mu_C$

Error is the difference between the true value of a measurement and the recorded value of a measurement.

- **Random error** is also known as random (or chance) variation); leads to an imprecise estimate. Differences between δ and *d* that reflect heterogeneity in the outcome among participants, in no particular direction (equal to 0 on average.

  → *Strategy to minimize: Increase N.*

- **Systematic error** is also known as **bias**; refers to deviations that are not due to chance alone. Differences between δ and *d* that reflect consistent over- or underestimation of the outcome among participants, in a particular direction (not equal to 0 on average).

  → *Strategy to minimize: Careful design; carefully executed.*

Validity

- **Internal validity** – the intervention effect (e.g., observed difference in outcome between the study groups) is real and not due to bias, chance, or confounding.

  → *Strategies to enhance: Randomization; placebo control; double blind.*

- **External validity** – generalizable to a broader population. (Irrelevant if internal validity is low.)

  → *Strategy to enhance: Use broad eligibility criteria when recruiting study participants*.

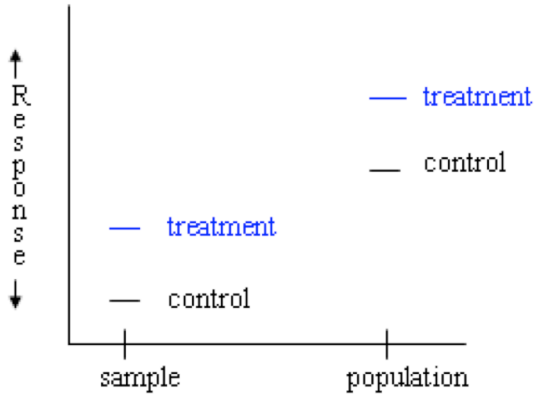
Statistical reasoning (Piantadosi, 2005)

1. Establishes an objective framework for conducting an investigation
2. Places data and theory on an equal scientific footing
3. Designs data production through experimentation
4. Quantifies the influence of chance
5. Estimates systematic and random effects
6. Combines theory and data using formal methods

Biases and Design Strategies to Control Them

1. **Selection bias** – Enrollment of a sample that is not representative of the target population.

   → *Note: Even if within-arm effects are biased, between-arm effect might remain unbiased.*



2. **Procedure-selection bias** – Occurs when an investigator systematically selects a certain type of patient for a particular treatment. *Solutions: Randomization; allocation concealment strategies (e.g., random block sizes)*

   - **Randomization** – A means of allocating experimental units to study arms that avoids systematic error (bias) and justifies type-1 error rate in hypothesis testing

   a. Example: Suppose randomize to *E*xperimental vs. *P*lacebo. If healthier patients are assigned to *E* and less healthy patients to *P*, the study could <u>invalidly</u> conclude that the experimental therapy is very effective.

b. **Blocking** – A strategy used to balance the number on each study arm in order to achieve planned allocation ratio (and level of power). *Addresses precision as well as bias. Used especially if planned N is small*.

   Blocking restricts randomization to ensure equal numbers of patients per treatment after a predetermined number (i.e., the block size) of randomizations.

   - Example: Blocks of 4 would mean that after 4, 8, 12, etc. patients enter the study, the numbers assigned to each treatment will be equal.

     i. NB: Revealing the block size to investigators involved with recruiting and interacting with participants could undermine allocation concealment!

3. **Confounding** – the effect of another factor on the outcome that may be incorrectly attributed to the intervention effect.

   a. Example 1: RCT will compare treatments A and B in patients between the ages of 18 and 65. Suppose that the younger patients tend to be healthier.

   → *Solutions:* ***Blocking*** *and* ***stratification*** *can be used to balance in the number on each treatment within each confounder level. (Goal: Control unwanted variation.)* At the design stage:

   i. Stratify by age. Construct age groups (e.g., 18-30, 31-50, and 51-65) and randomize patients to treatment within each age group.

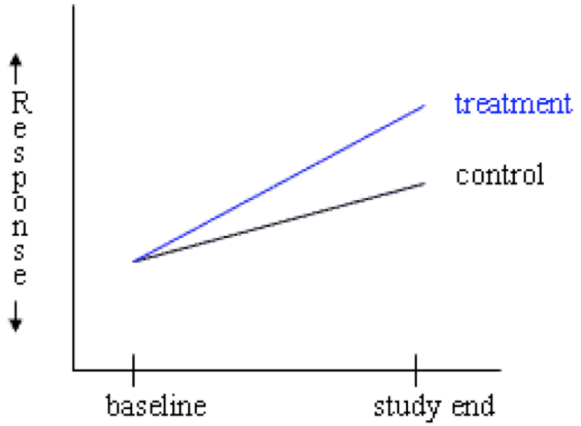
Age	Treatment A	Treatment B	All
18 - 30	12	13	25
31 - 50	23	23	46
51 - 65	6	7	13
	41	43	84

ii. If the expected stratum-specific sample-sizes are small: Block within strata. The goal is to ensure the numbers assigned to each treatment within each stratum are similar.

b. Example 2: If an RCT enrolls controls first then intervention participants, a significant difference between arms might be attributable to a change in the environment between the two treatment periods rather than to treatment – cannot know.

→ *Solutions: Use concurrent controls, with allocation achieved via unbiased randomization.*

This figure depicts groups studied concurrently; disease improved in both.
(How to interpret if control nonconcurrent?)



4. **Observer / Assessment Bias** – Failure to measure the study outcome in a consistent manner across all participants.

   → Solutions: Employ objective outcomes; Employ procedures to maximize objectivity of data – such as a detailed study protocol, study personnel with distinct roles, masking, and suitable control (e.g., placebo).

   a. **Treatment masking** or **blinding** – a strategy to ensure objectivity of the person measuring the outcome variables; not always possible (e.g., surgery).

      i. Especially important when the measurements are subjective or based on self-assessment.
      ii. **Single-masked trials** -- only patients are masked.
      iii. **Double-masked trials** -- both investigators and patients are masked to the treatment.
      iv. **Triple-masked trials** -- statisticians also are masked to treatment assignment (i.e., not knowing which group received the treatment and which is the control) until analyses have been interpreted.

   Investigators often underestimate the importance of masking as a design feature. This is because they believe that biases are small in relation to the magnitude of the treatment effects (when the converse usually is true), or that they can compensate for their prejudice and subjectivity.

- **Appropriate follow-up frequency and duration** -- Follow-up frequency and duration should be appropriate to the interventions, and completeness of follow-up should match between study arms.

    - Example: If half-life of a drug is 30 days, studying adverse events for 6 months will capture both treatment-related (desired) and disease-related (not desired) events.

b. **Placebo effect** – When an ineffective treatment appears beneficial in some patients due to their expectation of a positive response (rather than due to random fluctuations in their response or to variability in their disease).

   → *Solution: The placebo effect will affect outcomes both/all study arms. The between-arm effect will not be biased by the placebo effect.*

   i. True placebo: An inert or inactive treatment that mimics the route of administration of the real treatment (e.g., a sugar pill).
   ii. Disadvantage:
      1. Not acceptable ethically in many situations (e.g., when an accepted treatment already exists for a serious illness such as cancer, the control must be an active treatment).
      2. Not physically possible to attain (e.g., surgical control).

5. **Post-randomization Exclusion Bias**

a. Ineligible patients – As with Procedure-selection bias, this bias can occur differentially by arm, possibly leading to imbalances in unmeasured covariates (i.e., negates randomization) and to confounding.

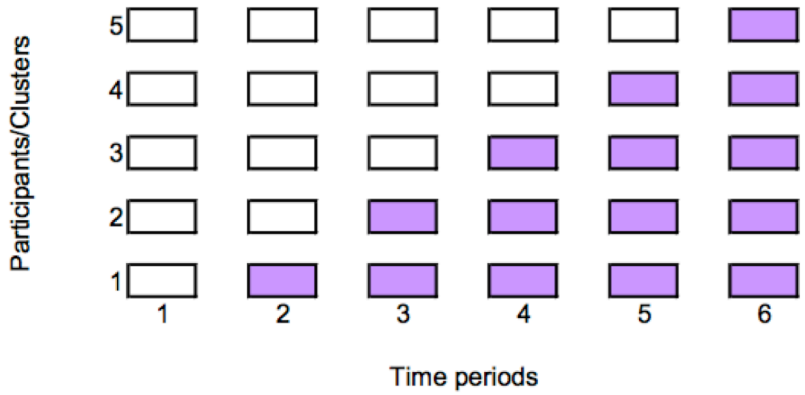
→ *Solutions: Avoid!*

b. Protocol violations (e.g., adding / deleting / changing / incompletely adhering to study medication that affects the outcome).

→ *Solutions: Avoid! Temptation is to conduct "per protocol" analysis an addition to or in place of intention-to-treat analysis but result may be biased.*

Intervention Schemes when 2+ factors are under study

- **Parallel design** – Each participant is exposed to one factor level during the trial
- **Cluster-randomized trial –** Groups of participants are exposed to one factor level
    - **Stepped-wedge design –** Rolls out an intervention in stages over time, with balance between Experimental and Control arms gradually increasing or decreasing



Shaded cells represent intervention periods
Blank cells represent control periods
Each cell represents a data collection point

- **Factorial design** – Patients are simultaneously exposed to at least two factors during the trial (e.g., chemo + radiation).

- **Crossover trial** – Patients are sequentially exposed to more than one treatment during the trial. Advantage: Each participant is exposed to each study arm. Disadvantage: Period effects can be hard to anticipate, prevent, and control.

    - Example: 12-month Dietary intervention vs. Gastric bypass surgery
    - Example: Colford J.

- **Adaptive Designs** – Depending on early responses, patients may be sequentially exposed to at least two treatments during the trial.

    - **SMART design: Sequential Multiple Assignment Randomized Trial**

    - **MAMS: Multi-Arm Multi-Stage trial**