# Statistics in Clinical Research

CH Yeh, Ph.D.
Cook Research Incorporated
Lafayette, IN

February 5, 2018

# Roles that Statistics Plays in Clinical Research

- Study design;
- Conduct of Study;
- Data analyses;

**Research Question → Form Hypothesis → Design Study → Data Analyses**

# Part I

- Overview of Clinical Research
- Statistical-Related Elements in Clinical Research
  - Study Design
    - Study Objectives
    - Statistical Hypothesis
    - Study Population
    - Selections of Study Groups
    - Efficacy and Safety Endpoints
    - Study Procedures and Schedules
    - Sample Size Determination
    - Applications of Meta Analyses in Clinical Studies
  - Conduct of Clinical Research
    - Randomization and Blinding
    - Interim Analyses and Unblinding
    - Confounding Effects on Study Outcome

# Part II

- Data Analyses
    - Subject Disposition
    - Analysis Datasets
    - Efficacy Endpoint(s)
    - Safety Endpoints
    - Subgroup Analyses
    - Missing Data and Outliers
    - Sensitivity Analyses
- Additional Topics on Data Analyses
    - Adaptive Study Designs
    - Interim Analyses
    - Exploratory Analyses

# From Seed to Plant

- Concept
- Plan
- Implement
- Conclude

# From Planning to Writing: Clinical Research Protocol

- Protocol provides the details of a proposed clinical study;
- NIH and FDA developed a clinical trial protocol template for NIH-funded studies or phase 2 and 3 clinical trials that require Investigational New Drug application (IND) or Investigational Device Exemption (IDE) applications:
  - https://osp.od.nih.gov/clinical-research/clinical-trials/
- Instructional and Examples texts are provided.
- Use it as a guideline/reference to cover all elements of Clinical Research are considered.

# Statistical-Related Elements in Clinical Research

## Study Design

- A newly proposed clinical research has a lot of unknowns → use information from published literature;
- Utilize relevant information to form close-to-correct assumptions and design a “best-available-plan” clinical research.

# Study Objectives

- Why and What is the purpose of this research?
  - “Does the new device have the same precision as the current one?” → Test for difference;
  - “Is TID more efficacious than BID?” → Test for superiority;
  - “Will the new treatment render less side effects?” → Test for inferiority;
- Types of Study Designs
  - Parallel
  - Cross-Over
  - Single Arm

# Statistical Hypothesis

- Null hypothesis: the one that the research attempts to dis-prove;
  Ho= Control group has lower BP than the new treatment
  → Ho: $BP_{control} \leq BP_{new\ treatment}$
- Alternative hypothesis: the one that the research attempts to prove;
  Ha= Control group has higher BP than the new treatment
  → Ha: $BP_{control} > BP_{new\ treatment}$
- Needs to set up correctly because it is related to type I and type II errors;
- Depending on the study objective(s), it can be ≠ (difference) , ≥ or ≤ (inferior or superior), ≤ and ≥ (equivalence), ≤ C or ≥ C (non-inferior);
- Specify type I error, # of sided test;
- Lists all key hypotheses will be tested in the study → penalty for multiple hypotheses;

# Study Population

- *Describe study participants : the population's characteristics under study should be clinically relevant to the research objectives ;*
- *Clearly define inclusion and exclusion criteria;*
- *Enrolling correct and willing population is essential to the outcome of the study;*
- *Stratification should be considered if the baseline characteristics of patients might have impact on outcome;*
  - *Severity of headache;*
  - *Male vs. female;*
  - *Elderly vs. young adults*

- Example, medication used to improve chronic limb ischemia(for lower extremity);
  - Use Rutherford score to define patient population

| Category | Clinical Description |
|---|---|
| 0 | Asymptomatic—no hemodynamically significant occlusive disease |
| 1 | Mild claudication |
| 2 | Moderate claudication |
| 3 | Severe claudication |
| 4 | Ischemic rest pain |
| 5 | Minor tissue loss—nonhealing ulcer, focal gangrene with diffuse pedal ischemia |
| 6 | Major tissue loss—extending above TM level, functional foot no longer salvageable |

# Selection of Study Groups

| | Pros | Cons |
|---|---|---|
| *Placebo Concurrent Control* | minimizes subject and investigator bias;<br>Establish placebo effect ;<br>Assess safety profiles; | Ethical concerns;<br>artificial environment from *real world;* |
| *Active (Positive) Concurrent Control* | Compares with current/standard treatment to assess additional clinical benefits;<br>Less ethical concerns; | Need to establish NI margins |
| External Control (historical data) | Common in medical device where no other device available;<br>Compare to a historical data; | study cannot be blinded; |

- Placebo modification
  - Add on Study, Placebo-Controlled; Replacement Study:
    - Standard trt (not fully efficacious) + (placebo, test) =improve clinical outcome = anticancer, antiepileptic, and heart failure drugs
  - Early *Escape*; Rescue
    - Treatment :prompt removal of subjects whose clinical status worsens or fails to improve;
  - Randomized Withdrawal
    - When long term placebo study is not feasible;
    - Wash out xx time period :  placebo →tested

# Study Endpoints

- A specific measurement or observation to address and correspond to the study's primary objective(s);
  - Treatment A works better than treatment B → define "*works better*" → patients will recover faster → "time (in days) to recover from surgery to able to walk on his/her own for 3 minutes";
- A study can have multiple endpoints, but should be prioritized and adjusted for multiplicity;
  - Effectiveness: reduce pain on knees, QOL,
  - Safety: headache, GI discomfort,
- Clearly specify the definitions of each endpoints (what measurement at what time and how to calculate, criteria qualify the endpoints)
- Types of Endpoints
  - Primary, secondary, *tertiary, surrogate (tumor size for cancer progression)*
  - Single measurement, composite variables (Death + MI + stroke)

# Treatment for Deep Vein Thrombosis (DVT)

Endpoints

- A significant reduction of *swelling (*tension-controlled tape)  of the affected DVT leg at 1MO;
- A significant reduction of *Pain at 1MO:* visual analogue scale;
- Functional status improvement as assessed by the walking impairment questionnaire at 1MO, 6MO;
- Improved signs and symptoms 12MO at : Villalta scale

# Study Procedures and Schedules

- Describe study intervention:
  - What product (medication, device) or procedure will be given by whom;
  - when/how/what data are collected;
  - Study schedules should be laid out clearly as they are relevant to study endpoints;
    - Data collected at fixed schedules: Blood pressure at 1 month after start of study; medical device migrates 2 years after implanted;
    - Time to event (survival study): time death after the CABG, time to open surgery → requires real-time follow-up

# Sample Size Determination

$$N = \frac{\sigma^2 \left( Z_{\alpha \, or \, \frac{\alpha}{2}} + Z_{\beta} \right)^2}{\Delta^2}$$

- Given α level (ex, 0.05 for 2-sided, $^{\alpha}/_{2}$ or α ) and power of study (80%, 90% $1 - \beta$);
- # of study participants is calculated base on the “primary hypothesis”:
  - Per “primary endpoint”: the rate/mean/time of control/test groups, variance for continuous endpoints;
  - The expected treatment benefits ($\Delta$ ):
    - minimal effect which has clinical relevance in the management of patients
    - a judgement concerning the anticipated effect of the new treatment
  - For NI or EQ studies, sample size will be larger;

- More often than not, this information is unknown;
- But most of time, similar info can be found in published literature or historical studies (relevant to the test group);
- Use Meta-analyses to provide estimates (assumptions) for sample size calculation;
  - investigate the sensitivity of the sample size estimate to a variety of deviation form the assumptions;

# Meta-Analyses

- A statistical method that ***systematically*** combines ***pertinent*** qualitative and quantitative study data from several ***selected*** studies to assess and develop a ***single conclusion*** that has greater statistical power.
- This conclusion is statistically stronger than the analysis of a single study, due to increased numbers of subjects, greater diversity among subjects, or accumulated effects and results.

# Meta-Analyses can be used

- To establish statistical significance with studies that have conflicting results
- *To develop a more correct estimate of effect magnitude*
- To provide a more complex analysis of harms, safety data, and benefits
- To examine subgroups with individual numbers that are not statistically significant

# Pitfalls of Meta-Analyses

- Difficult and time consuming to identify appropriate studies;
- Not all studies provide adequate data for inclusion and analysis;
- Requires advanced statistical techniques;
- Heterogeneity of study populations/designs/conducts;
  - Examination of heterogeneity is perhaps the most important task in meta-analysis.

# Systematic review for Meta Analyses

- Collect empirical evidence that fits prespecified eligibility criteria to answer a specific research question;
  - prespecify selection criteria will minimize selecting bias;
- Characteristics of a systematic review:
  - Define objectives of the review;
    - What is the benefit of aspirin in stroke/Afib? What is magnitude of the benefit? Better than new blood thinner (Eliquis, warfarin);
  - Eligibility criteria for studies;
    - Randomized active-controlled/placebo-control trails, double-blind, published (what journals, conferences);
  - an assessment of the validity of the findings of the included studies;
    - Is it peer-reviewed? Is the study well conducted? Sponsors of the study?

# Meta analyses and Sample Size Estimates

- If use meta-analyses to obtain the estimates for sample size calculation, be very cautious in selecting studies that are ***relevant*** to your study;
- ***Relevant*** means:
  - Study design:
    - parallel/cross-over, active/placebo control,
    - length of study period,
    - treatment groups: similar procedure/medication/doses/regimens,
    - efficacy endpoints: definition, calculation, timing, collecting tools,
    - safety endpoints: definition, timing,
    - study population: disease severity, baseline characteristics,  subsets,
    - Conduct of study: study procedure, where the study was conducted, medical practice,

# Be mindful of “Dis-similarities”

- If include “not so similar” studies, assess how that dis-similarity will impact the estimate;
  - Literature enrolled NYHA class II, your study enroll NYHA class III;
  - Drug XX at dose YY had treatment of 20%, your dose is twice higher;
  - Efficacy endpoint based on historical data at 18MO was 30%, your study will only go to 12MO;
  - Endpoint used in treatment peripheral vessel included major amputation and death, your endpoint include major amputation and death and wound healing;
  - Historical studies included patients did not respond to standard treatment (but did not know how long after treatment), your study include patients that are not responsive after 6MO;

# Meta-Analyses for sample size estimates

1. Clinical scientists review and provide the list of literature that are relevant to the planned clinical research;
2. Statisticians review the data/literature and perform statistical analyses for estimates (mean, variance, treatment effect);
3. Various statistical methods (for categorical, continuous, time to events endpoints) will be utilized to assess the estimates;
4. Based on the dis/similarities of the historical data and current study, clinical scientists and statisticians should assess how to adjust the estimates;
5. Assess the sensitivity/robustness of the estimates and sample sizes;

# Example

- Phase III new drug for peripheral artery disease/critical limb ischemia;
- Aim to widen blood vessels (vasodilator);
- Clinical benefits: reduce major amputation and death;
- Literature review;
- Estimate AFS+death rate;
- Assess sensitivity of the estimates: study length/Follow-up, population, procedure/SOC, treatment groups (dose, regimen, control group), endpoints ;

| | 6MO (95%CI) | 12MO | Overall |
|---|---|---|---|
| Test(%) | 69(45.53, 85) | 66.74 (56.94, 75.27) | 68(55.79, 78.16) |
| Control(%) | 49.54 (28.36, 70.89) | 31.37 (7, 73.77) | 42.73 (24.39, 63.3) |
| N (Δ) /group | 78 (20%) | 25 (35%) | 43 (26%) |
| | 99(17%) | 10(50%) | 28(31%) |

# Statistical-Related Elements in Clinical Research
## Conduct of Clinical Research

- Blinding : to reduce the occurrence of conscious and unconscious bias;
  - Interim Analyses and unblinding:  remaining unblinded after interim is critical to the integrity of the study;
- Randomization : provides a sound statistical basis for the quantitative evaluation;
  - Confounding Effects on Study Outcome : managed by randomization;

# Blinding

- Limit bias in *conduct* and *interpretation* of study;
- Double-blind is optimal;
  - but sometimes it is not possible to blind patients or investigators
    - stents from 2 device manufactures;
    - Treatment induced-effect is different (drug-induced rash);
- Single-blind or open-label study: effort should be made to limit/restrict investigators/staff/sponsors knowledge of treatment;
- Breaking the blind (for a single subject) should be considered only for the subject's care; any intentional or unintentional breaking of the blind should be reported;

# Interim Analyses and Unblinding
## (*conduct of study)*

- *Unplanned or impulsive analyses are not interim analyses;*

- *Should be pre-planned and described in protocol;*
  - *Types of interim; stop for futility/efficacy; sample size re-adjustment*
  - *Statistical algorithm/method; blinding issues; criteria for continuing/halting study; type I error adjustment; final p-value;*
  - *When to perform; frequency of analyses;*
  - *Who to perform; distributions of the interim results;*

- *Unblinding for interim analyses*
  - *it's workable, as long as it's documented;*
  - *Should minimize the number of people unblinded;*
  - *Reviewers/analysts who are unblinded should not be involved in study decision;*

# Randomization

- Tends to balance the baseline characteristics, medical history, disease severity of patient population between treatment groups → *isolate treatment effect*;

- Randomization schemes:
  - Fixed-randomization schemes = *Ratio (control : test) remains constant during study*
    - Blocked (by center, region, hospital) : block sizes
    - Stratified: variables that might potentially correlate with treatment effect

  - *Dynamic randomization = Ratio (control : test) modified during study*
    - For example, "*Play the Winner" scheme;*
    - Use in rare/critical illness or endpoint can be observed soon after treatment was given;
    - Logistics is complicated (blinding will be difficult);
    - But will give therapeutic superior test group a better chance to demonstrate efficacy;

# Confounding Effects on Study Outcome

- Mixing with test effect, causes “bias” that might preclude finding a true effect;
- Difficult to establish cause/effect link: may discredit the study outcome;
    - Treatment to avoid lower-limb amputation:

| Amputation at 1 year | | Control | Test |
|---|---|---|---|
| Yes | | 40/200 (20%) | 20/200 (10%) |
| Smoking | Yes | 35/150 (23%) | 6/30 (20%) |
| | No | 5/50 (10%) | 14/170 (8.2%) |
| No | | 160/200 | 180/200 |

- Important to identify and manage confounding factors when design the study;
    - Clinical: identify
    - Statistics: manage via design and analyses