Statistical Principles for Clinical Development

Mark Levenson, Ph.D.
Director Division of Biometrics VII
CDER | US FDA

Clinical Investigator Training Course (CITC) – December 7, 2022

Learning Objectives

- To understand basic statistical principles relevant to clinical studies
- To understand the concepts of bias and variability
- To correctly understand p-values and hypothesis testing
- To understand issues around multiplicity

Design and Conduct are more important than Analysis

- In other words: Analysis cannot make up for poor design and conduct

- Focus on good design and conduct, and analysis will be straightforward

Stages of a Study

- Design: The conception, planning, and specification of the study
- Conduct: The running of the study
- Analysis: The analysis of the study (Number crunching)
- Reporting

Adequate and Well-Controlled Study 21CFR314.126

- Clear objectives, summary of methods and results
- Design permits a valid comparison with a control
- Adequate selection of patients
- Assigning patients to treatment and control groups minimizes bias
- Adequate measures to minimize biases on subjects, observers, and analysts
- Well-defined and reliable assessment of subjects' responses
- Adequate analysis to assess drug results

Randomized v. Observational Studies

Randomized Study

Patient

Heads  Tails

Drug A

Drug B

Observational Study

Patient

Drug A

Drug B

- Lab tests
- Age
- Sex
- Race
- Medical history
- Family history
- Concomitant drug
- Insurance
- Convenience
- Geographic region
- Etc.

Randomized v. Observational Studies

- Randomized study:
  characteristics of patients receiving drug A are similar to characteristics of patients receiving drug B

- Observational study:
  characteristics of patients receiving drug A may not be similar to characteristics of patients receiving drug B

Confounding

- Without randomization there may be systematic differences (bias) when comparing people getting Drug A and people getting Drug B

  This is known as confounding

  Example: Drug A may be given to older sicker people.
  Even if there was no differences between the effects of Drug A and Drug B, the comparison may show Drug A has worse outcomes

  Note: There are other sources of biases to be concerned about that may exist even with randomization.
  Example: Bias on part of observers (lack of blinding)

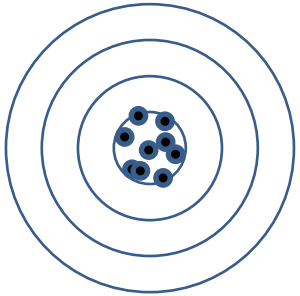
PERSONAL HEALTH

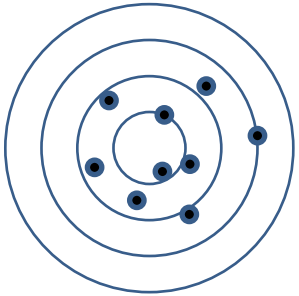
The Health Benefits of Coffee

Drinking coffee has been linked to a reduced risk of all kinds of ailments, including Parkinson's disease, melanoma, prostate cancer, even suicide.

Variability v. Bias

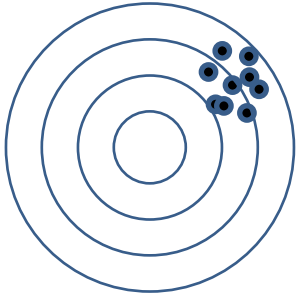
Low Variability,
Low Bias

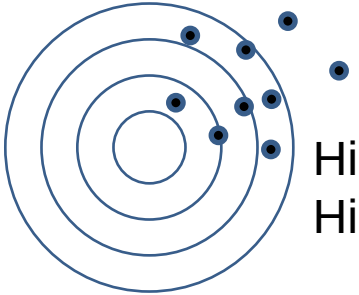




High Variability,
Low Bias

Low Variability,
High Bias





High Variability,
High Bias

Bias is worse than Variability

Variability v. Bias

- Statistics helps to quantify variability (in the analysis stage)
- Design and conduct generally reduce bias
  Examples: randomization, blinding
- Note: Statistics helps reduce bias and variability at each stage, design, conduct, and analysis

Reducing Variability with More Sample Size

- Sample size: number of people in study

- What is a better estimate of the average age of this session's attendees?

  A. Pick a random sample of 5 attendees and calculate their average age

  B. Pick a random sample of 20 attendees and calculate their average age

Hypothesis Testing

- Null Hypothesis: Typically, what you are trying to show is **not** true

  Ex: Drug A and placebo have the same effect (nothing)

- Alternative Hypothesis: Typically, what you are trying to show

  Ex: Drug A has a better effect than placebo

P-Values

- Probability of observing the effect or something more extreme, if the null hypothesis is true

  Example: Study estimated drug effect is a reduction of 5mm of diastolic blood pressure.
  P-value: probability of observing 5 or 6 or 7 or 8 …, if the drug had no effect

P-Values

- Small p-values are evidence **against** the null hypothesis (no drug effect)

  Example: p-value = 0.02. If there was no drug effect, the chance of seeing what we saw or more extreme is 0.02. This is small. Leading us to doubt that there is no drug effect

- P-value is **not** the probability that the null is true (no drug effect)

Decision Errors

- Type 1 error: Concluding the drug has an effect on when it does not
  (FDA and society's problem)
  (and drug company's)

- Type 2 error: Concluding drug does not have an effect when it does
  (Drug company's problem)
  (and FDA and society's)

Hypothesis Testing

- Reject null hypothesis if p-value < alpha.
  Equivalent to saying the drug has an effect.
  Example: alpha = 0.05

  (Recall small p-values are evidence against the null hypothesis.)

Sample Size and Power

- We can set the Type 1 error by choosing alpha
- We can limit the Type 2 error, by having larger sample size (more patients)
- More sample size = less variability = less likely to conclude drug has no effect when it really does (Type 2 error)

Multiplicity
(AKA: Multiple Bites from the Apple)

Example: Determining if a drug has effect

Drug has no effect

Study: Flip coin 4 times.

H = a good outcome, T = a bad outcome

If get 1, 2, or 3 H's, conclude drug has no effect.
If get 4 H's, conclude drug has effect.

Study 1: HHTH
Repeat study:
Study 2: HTTT
Study 3: THTH

Study 12: HHHH

Multiplicity

- If you do enough studies or if you look at data in many ways, you will see an effect (even when there is no effect)
- This is known as data dredging or p-value hacking. It is a known problem with science

Multiplicity

- Multiplicity can show up with multiple subgroups or endpoints

  Subgroups: effect on males, effect on females, effect on people over 65

  Endpoints: effect on blood pressure, effect on life expectancy, effect on happiness

- Prespecification: Tell the world ahead of time, what you will primarily look at
  Protocols and Statistical Analysis Plan (SAP) are how that is done.

Challenge Question #1

Which of the following does not reduce bias?

A. A larger sample size

B. Randomization

C. Blinding the knowledge of the drug from study participants and investigators

D. Prespecification in the protocol and statistical analysis plan

Challenge Question #2

Which of the following addresses multiplicity?

A. A larger sample size

B. Randomization

C. Blinding the knowledge of the drug from study participants and investigators

D. Prespecification in the protocol and statistical analysis plan

Resources

- Demonstrating Substantial Evidence of Effectiveness
- ICH E8(R1) General Considerations for Clinical Studies
- ICH E9 Statistical Principles for Clinical Trials
- ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials
- Multiple Endpoints in Clinical Trials
- Adaptive Designs for Clinical Trials
- Adjusting for Covariates in Randomized Clinical Trials

Thank You!


FDA
U.S. FOOD & DRUG
ADMINISTRATION