Statistical Principles for Clinical Development

Mark Levenson, Ph.D.
Deputy Director
Office of Biostatistics
CDER | US FDA

Clinical Investigator Training Course (CITC) – 2024

Learning Objectives

- To understand basic statistical concepts such bias, variability, multiplicity

- To understand the role of statistical concepts in good study design, conduct, and analysis

Outline

- Quick review of good-study principles

- Statistical concepts and implications on study elements:

  - Bias v. variability

  - Variability, Z-statistic, sample size, and multiplicity

  - Bias types, sensitivity analysis

- Concluding statistical principles

Stages of a Study

- **Design**: The conception, planning, and specification of the study

- **Conduct**: The running of the study (recruitment, screening, intervention, concomitant care, outcome ascertainment, safety and study monitoring)

- **Analysis**: The analysis of the study

- **Reporting**

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

- Statistical concepts motivate good design and conduct

Study Design, Conduct, and Analysis Goals

- To address a specific clinical question

- To minimize bias and minimize variability

- To ethically, safely, and feasibly conduct the trial

- Note: There is some conflict among these goals

Enrichment v. Generalizability

FDA

- Interested in drug effect on cardiovascular events for all patients with Type II diabetes

- Event trials usually require a certain number of events to get necessary statistical precision

- You can get more events for the same number of patients if you study patients with additional CV risk factors

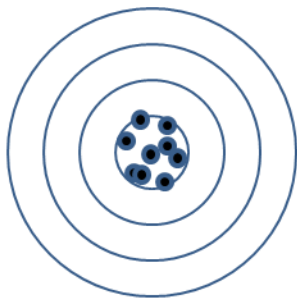- Balance of feasibility and answering clinical question

Outline

- Quick review of good-study principles

- **Statistical concepts and implications on study elements:**

  - **Bias v. variability**

  - Variability, Z-statistic, sample size, and multiplicity

  - Bias types, sensitivity analysis
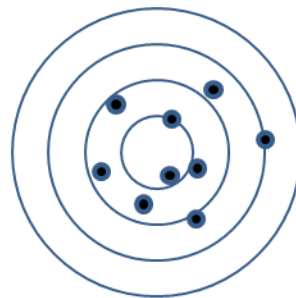
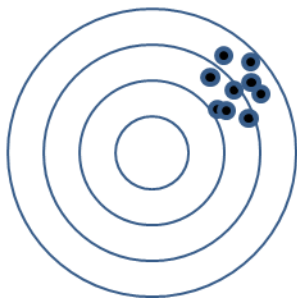- Concluding statistical principles

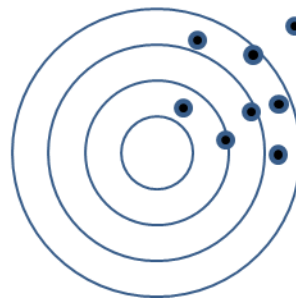# Variability v. Bias

Low Variability, Low Bias

High Variability, Low Bias

Low Variability, High Bias

High Variability, High Bias

Do These Data Have Low Bias?

Bias v. Variability

- We see and can measure variability

- We do not see bias

- Bias is mainly addressed in the design and conduct stages

Outline

- Quick review of good-study principles

- **Statistical concepts and implications on study elements:**

  - Bias v. variability

  - **Variability, Z-statistic, sample size, and multiplicity**

  - Bias types, sensitivity analysis

- Concluding statistical principles

Variability and Sample Size

FDA

- 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 50 attendees and calculate their average age

Variability and Sample Size

- $\hat{X}_5$ = mean of 5 people

- $\hat{X}_{50}$ = mean of 50 people

- $\hat{X}_{50}$ has less variability than $\hat{X}_5$

Variability and Sample Size

- $standard\ error(\hat{X}_{50})$ is a measure of the variation of $\hat{X}_{50}$

- $standard\ error(\hat{X}_{50}) = \dfrac{SD(X)}{\sqrt{N}}$

- As sample size $N \uparrow$, *standard error* $\downarrow$

A Little Statistics

- $Z = \dfrac{Effect\ Estimate}{Standard\ Error(Effect\ Estimate)}$

- Assume observed *Effect Estimate* is 0.5 m.

- As sample size ↑
  - Standard error ↓
  - Z ↑
  - **P-value goes down (more significant).**

Sample Size, Significance, Power

- Typically, if there is no drug effect, then probability of seeing $|Z| > 2 = 0.05$

- If we observe $|Z| > 2$, then we say the finding was significant at the 2-sided 0.05 level

- For a given real effect, say 0.5m, the probability of a significant effect increase with sample size

- Larger sample sizes have higher *power*

Multiplicity

- AKA: Multiple bites from the apple

- A study found a drug had a statistically significant effect for females aged 51- 60 years

Multiplicity

Normalized Effect Measures (Z-Statistic)

	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	90+
Male	0.67	0.40	-0.74	-0.29	0.15	1.15	-0.04	-1.70	-0.08	-0.94
Female	-0.47	-0.29	-1.01	0.57	-0.75	2.49	-0.54	0.45	1.27	1.11

- Each of these cells gives you a chance $|Z|>2$ even if there is no drug effect.
- It is not surprising that some were $|Z|>2$

Multiplicity

- Even if drug has no effect, if you conduct enough statistical tests (without proper adjustment), you will find some results "statistically significant"

- This is known as multiplicity.

Multiplicity

FDA

- Multiplicity can show up with multiple subgroups, endpoints, or analyses.

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

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

- Analyses: Multiple regression with different covariates

Multiplicity Solutions

- Prespecification: Tell the world ahead of time, what you will primarily look at

- Avoids fishing for p-values

- Protocols and Statistical Analysis Plans (SAP) are how that is done

- Statistical methods can handle multiple outcomes or subgroups

- Need to be prespecified too

Outline

- Quick review of good-study principles

- **Statistical concepts and implications on study elements:**

  – Bias v. variability

  – Variability, Z-statistic, sample size, and multiplicity

  – **Bias types, sensitivity analysis**

- Concluding statistical principles

A Little Statistics: Review

- $$Z = \frac{Effect\ Estimate}{Standard\ Error(Effect\ Estimate)}$$

- Assume observed *Effect Estimate* is 0.5 m.
- As sample size ↑
  - Standard error ↓
  - Z ↑
  - **P-value goes down (more significant).**

Bias Example

- Unblinded trial

- Effort dependent endpoint (6 min walk)

- Patients on experimental drug may try harder
  (let's say 0.5 meters on the average)

- Assume no drug effect.

- **This 0.5m bias will not likely be affected by sample size.**

- **P-values get smaller for larger sample size even though drug is not effective!**

Addressing Bias

- Sample size does not address bias

- Design and conduct should address bias

- Design methods: randomization, blinding, good outcome assessment

- Analysis cannot rescue poor design and conduct

- Analysis can address some bias and explore sensitivity to biases

Bias Types

- Confounding
  (Baseline differences between treatment groups)

  – Addressed by randomization

- Measurement
  (Endpoint ascertainment)

  – Addressed by blinding and reliable and valid measures

- Selection (baseline and post-baseline population issues)

Confounding

FDA

- 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

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.

515

Selection Bias Type 1

FDA

Baseline study population not representative of intended population

- Easy to understand

- Example: Few people with co-morbidities in study

- Affects generalizability (external validity)

Selection Bias Type 2

Analysis population based on post-baseline information

- Harder to understand

- Examples

  – Analyzing adherers only

  – Analyzing those with no missing data

  – Analyzing those with initial response to treatment

- Biases causal relationships (internal validity)

Adherence May Be Related to Outcome

Table 1. Five-Year Mortality in Patients Given Clofibrate or Placebo, According to Cumulative Adherence to Protocol Prescription.

ADHERENCE *	TREATMENT GROUP			
	CLOFIBRATE		PLACEBO	
	no. of patients	% mortality †	no. of patients	% mortality †
Total study group	1065	18.2±1.2 (18.0)	2695	19.4±0.8 (19.5)

*A patient's cumulative adherence was computed as the estimated number of capsules actually taken as a percentage of the number that should have been taken according to the protocol during the first five years of follow-up or until death (if death occurred during the first five years).

†The figures in parentheses are adjusted for 40 base-line characteristics. The figures given as percentages ±1 S.E. are unadjusted figures whose S.E.'s are correct to within 0.1 unit for the adjusted figures.

Adherence May Be Related to Outcome

Table 1. Five-Year Mortality in Patients Given Clofibrate or Placebo, According to Cumulative Adherence to Protocol Prescription.

ADHERENCE *	TREATMENT GROUP			
	CLOFIBRATE		PLACEBO	
	no. of patients	% mortality †	no. of patients	% mortality †
<80%	357	24.6±2.3 (22.5)		
≥80%	708	15.0±1.3 (15.7)		
Total study group	1065	18.2±1.2 (18.0)	2695	19.4±0.8 (19.5)

*A patient's cumulative adherence was computed as the estimated number of capsules actually taken as a percentage of the number that should have been taken according to the protocol during the first five years of follow-up or until death (if death occurred during the first five years).

†The figures in parentheses are adjusted for 40 base-line characteristics. The figures given as percentages ±1 S.E. are unadjusted figures whose S.E.'s are correct to within 0.1 unit for the adjusted figures.

Adherence May Be Related to Outcome

Table 1. Five-Year Mortality in Patients Given Clofibrate or Placebo, According to Cumulative Adherence to Protocol Prescription.

ADHERENCE *	TREATMENT GROUP			
	CLOFIBRATE		PLACEBO	
	no. of patients	% mortality †	no. of patients	% mortality †
<80%	357	24.6±2.3 (22.5)	882	28.2±1.5 (25.8)
≥80%	708	15.0±1.3 (15.7)	1813	15.1±0.8 (16.4)
Total study group	1065	18.2±1.2 (18.0)	2695	19.4±0.8 (19.5)

*A patient's cumulative adherence was computed as the estimated number of capsules actually taken as a percentage of the number that should have been taken according to the protocol during the first five years of follow-up or until death (if death occurred during the first five years).

†The figures in parentheses are adjusted for 40 base-line characteristics. The figures given as percentages ±1 S.E. are unadjusted figures whose S.E.'s are correct to within 0.1 unit for the adjusted figures.

Adherence and Missing Data: Biases

- Patients who complete therapy may be different from those who stop therapy

- Treated group patients who complete therapy may be different from control group patients who complete therapy

- Analyzing of only patients who complete therapy (or complete data) will likely produce biased results

Adherence and Missing Data: Types

- Patients may stop treatment

- Patients may leave study (no more follow-up)

- Patients may miss a study visit (treatment or assessment)

- These are all *intercurrent events* and should be handled appropriately and not often the same

Adherence and Missing Data: Causes

FDA

- Random
  (not related to patient characteristics or patient outcome)

- Frailty of patient

- Lack of efficacy

- Side effects

What is Adherence?

- Adherence is not forcing patients to do something

- That is not ethical

- Protocol should allow for the best of interest of the patient

- Adherence is following the protocol, even if that means stopping therapy

Treatment Policy/Intent-To-Treat (ITT)

FDA

- First, want to get follow-up data regardless of treatment adherence

- Analyze patients according to random assignment regardless of adherence and missing data

- Answers the question: What is the effect of being assigned a therapy?

- Maintains randomization

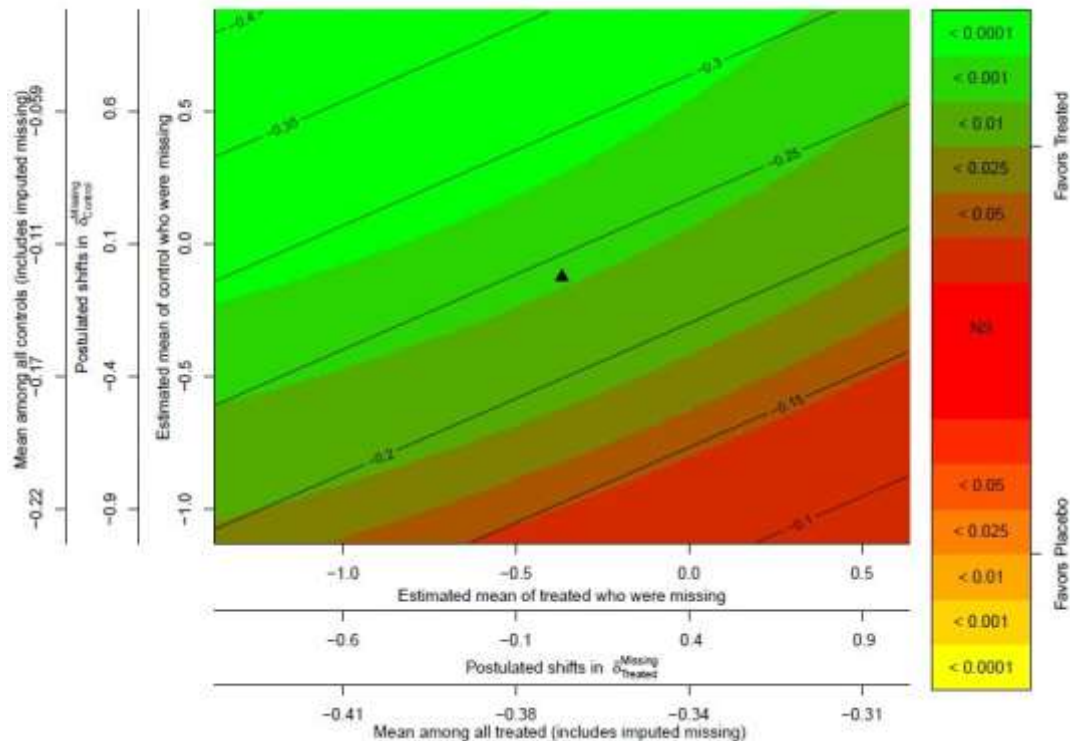# Treatment Policy/Intent-To-Treat (ITT)

- Problems

- Patients may stop therapy, cross over to other group's therapy, receive rescue therapy

- Still may have missing data

- Composite approach: stopping/changing/rescue therapy considered a failure in composite endpoint

- Approach should be tailored to specifics clinical question

Sensitivity Analysis

- Varying assumptions to explore potential biases

- Example: Drug effect for patients with missing data not the same as drug effect for patients with complete data

Sensitivity Analysis

Outline

- Quick review of good-study principles

- Statistical concepts and implications on study elements:

  – Bias v. variability

  – Variability, Z-statistic, sample size, and multiplicity

  – Bias types, sensitivity analysis

- **Concluding statistical principles**

Summary

- Careful design and conduct are needed for reliable results

- Randomization addresses baseline biases

- Blinding, good adherence, good follow-up, and good endpoint ascertainment address post-baseline biases

- Prespecification is important to address multiplicity

Take Home

- Address the clinical question

- Minimize bias and minimize variability

- Careful design, conduct, and analysis will promote valid and reliable results!

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

Thanks!