



**PACU**

**PHARMACY & ACUTE CARE UNIVERSITY**



# **Biostatistics and Study Design for Pharmacists**

A. Shaun Rowe, PharmD, MS, BCCCP, FNCS, FCCP

# Objectives

- Describe differences in descriptive and inferential statistics
- Identify different level of measure
- Recall appropriate measure of central tendency and dispersion
- Identify appropriate inferential statistic test
- Recall an overview of common study designs
- Evaluate common calculations utilized in clinical trails
- Interpret clinical and statistical significance



PHARMACY & ACUTE CARE UNIVERSITY

# Levels of Measure and Distributions

and Distributions

# Level of Measurement

- Describes the nature of the data that you are viewing
- Provides a uniform description of how data is measured
- Four primary types of measurement
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Vary in level of precision
  - Can always decrease the level of precision if measured at a higher level of precision originally
- Must be able to identify the level of measurement before you can identify the right descriptive or inferential statistic

# Nominal

- Nominal variables answer a “yes/no” question
- The labeling and grouping do not have a hierarchy
- No measure of distance between values
- Examples
  - Mortality
  - Hair Color
  - Eye Color
  - Race

# Ordinal

- Items are labeled and grouped
- Grouping has an order / hierarchy
- Magnitude of difference between groups varies
- Examples
  - Pain Scale (mild, moderate, severe)
  - Hypertension Staging (normal, elevated, stage 1, stage 2)
  - Cancer Staging (stage 1, 2, 3, 4)
  - Education level (high school, BS, MS, PhD)

# Interval

- Numeric data
  - Measured on a continuum
- Scale has an order
- Scale has a uniform magnitude of difference between levels
- No set “zero” point
- Examples
  - Temperature in Fahrenheit and Celsius
  - pH



# Ratio

- Numeric data
- Scale has an order
- Scale has a uniform magnitude of difference between levels
- Scale has a true “zero” point
- Examples
  - Plasma drug levels
  - Temperature in kelvin

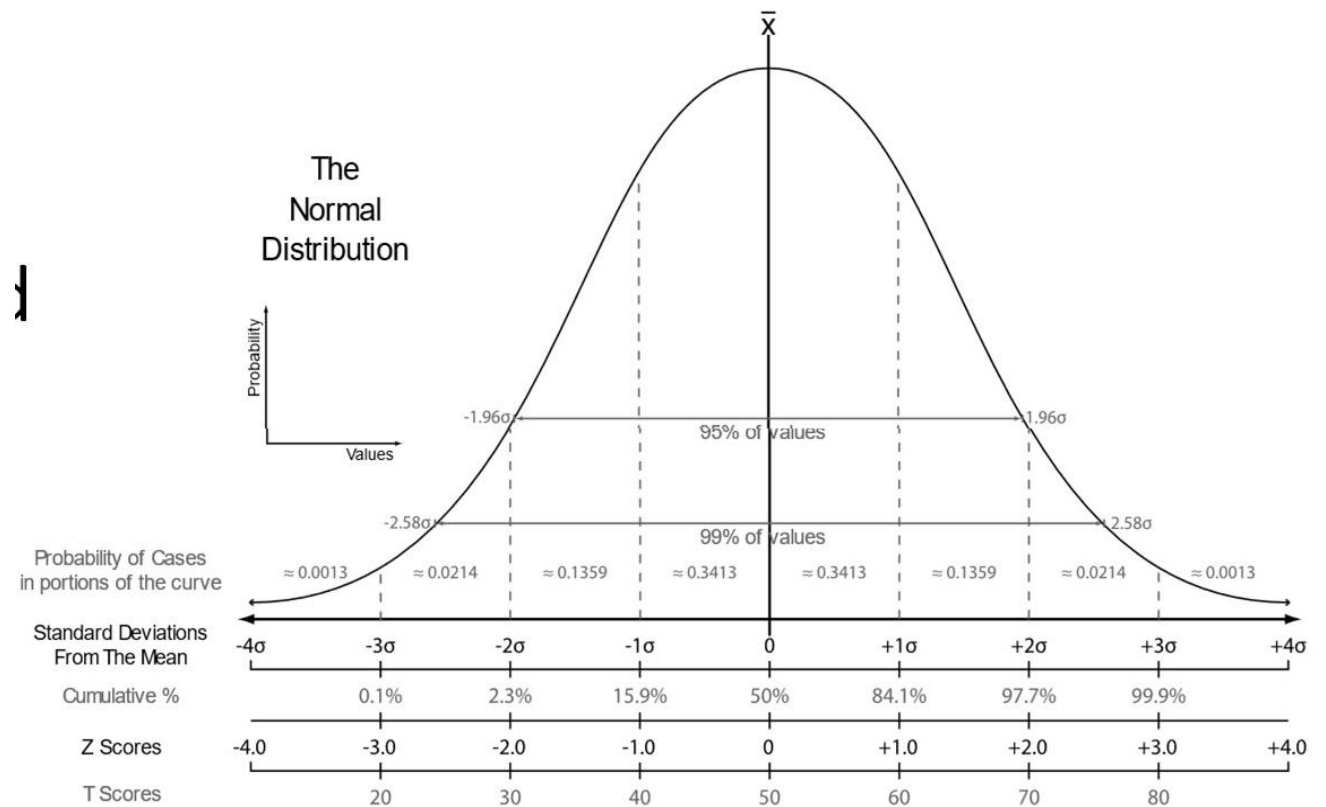
# Levels of Measurement Summary

	Nominal	Ordinal	Interval	Ratio
Labels and groups variables	X	X	X	X
Levels are ordered		X	X	X
Magnitude of difference is equal			X	X
True zero point				X

- Can move from a higher level of precision to a lower level of precision
  - Blood Pressure measured in mm Hg (Ratio) to Hypertension Staging (Ordinal) to Hypertension (Nominal)

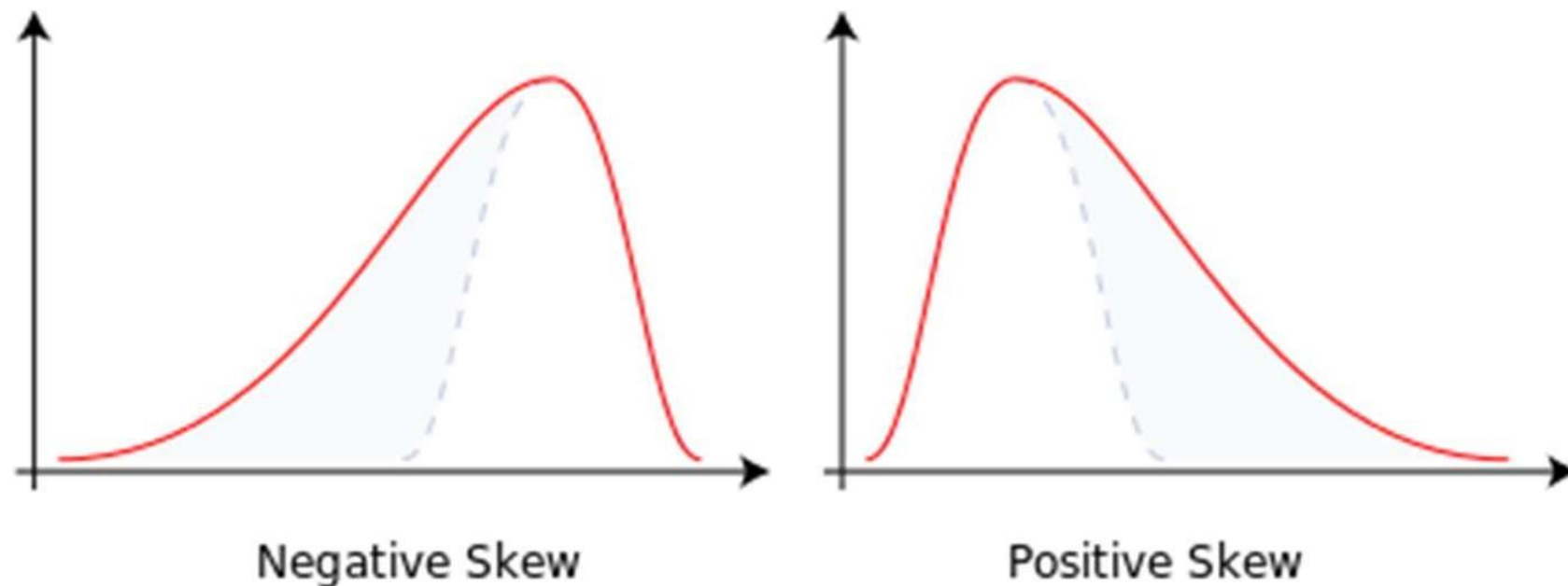
# Normal Distribution

- Commonly referred to as a Bell-shaped curve or Gaussian distribution
- Data is symmetrically distributed around the mean
- Mean, median and mode are equal
- Important for statistics but truly normal distributions are relatively rare in nature



# Skewness

- Follow the “tail” of the data to determine the direction of the skew
  - Left skew = lots of outliers toward the negative side of the chart
  - Right skew = lots of outliers toward the positive side of the chart
- Mean, Median and Mode are not equal



# Distribution summary

- Normally distributed numeric data is commonly referred to as parametric data
  - Does not contain outliers
  - Equal distribution of data points around the mean
- Non-normally distributed numeric data is commonly referred to as non-parametric data
  - Has many outliers
  - Commonly found in nature
  - Non-equal distribution of datapoints
  - Mean, median, and mode are not equal

# Descriptive Statistics

statistics

# Purpose of Descriptive Statistics

- Provides a uniform method of presenting large amounts of information
- Provides summary information
- Provides information about the center-point and dispersion of the data
  - Measures of central tendency
  - Measures of dispersion
  - Count and proportion
- May describe the relationship between two variables
  - Contingency Tables
  - Scatterplots
  - Correlation

# Measures of Central Tendency

- Mean
  - Most commonly used average
  - Add all observations and divide by the count of observations
  - Affected by outliers
- Median
  - Also known as the 50<sup>th</sup> percentile
  - Arrange observations from largest to smallest
  - Median is the middle-most value
  - Not affected by outliers
- Mode
  - Not commonly used as a descriptive statistic in biomedical literature
  - Most common value
  - Can have multiple modes



# Measures of Dispersion

- Also known as “Measure of Spread”
  - Help to describe the spread of the data
- Used in combination with measures of central tendency to describe the distribution of the data
- Most common measures of dispersion
  - Range
  - Interquartile range
  - Variance
  - Standard deviation
  - Confidence Interval

# Range

- Interval between the lowest and highest values
- Can also be reported as minimum and maximum values
- Affected by outliers
- Example:
  - 11 people reported their age on an interview questionnaire. The ages in years are 16, 16, 18, 21, 19, 22, 22, 21, 25, 45, 72
  - Range = 56
  - Minimum = 16, Maximum = 72

# Interquartile Range

- Interval between the 25<sup>th</sup> and 75<sup>th</sup> percentiles
- Can be reported as the 25<sup>th</sup> and 75<sup>th</sup> percentile
- Not affected by outliers
- Should accompany the median
- Example:
  - 11 people reported their age on an interview questionnaire. The ages in years are 16, 16, 18, 21, 19, 22, 22, 21, 25, 45, 72
  - Interquartile range = 7
  - 25th percentile = 18, 75th percentile = 25

# Variance

- Applicable to numeric data, but not commonly used for descriptive statistics.
- It's the sum of the squares of the individual deviances from the mean divided by the degrees of freedom
- Not very useful in day-to-day operations
- Usually, a large number
- Not in the same units as the mean

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

# Standard Deviation

- Describes the variability between individuals in the sample
  - Regardless of distribution, at least 75% of the observations will always be between the mean and 2 SD
  - If the distribution is normally distributed, see next slide
- Makes the variance useful
- Same units as the mean
- Useful only if the data is normally or near normally distributed
- Square root of the variance
- Should always accompany the mean

# Central Tendency and Dispersion

## Summary

- Mean is affected by outliers and should only be used when the data is near normal; otherwise, median should be used to describe the middle point of the data
- Mean should be accompanied by the Standard Deviation
- Median should be accompanied by the Interquartile Range
- Other measures of dispersion are used for specific statistical needs

# Confidence Intervals

- Estimates the range of values likely to include the true population mean
  - True population mean is an unknown and unknowable parameter
  - The CI gives a range of plausible values for the parameter
- A CI will be calculated at a *confidence level*
  - Confidence level is the probability that the interval contains the true population mean
  - Most commonly reported as the 95% CI but can be at any level desired
- Given a normal distribution, CI is calculated as:
  - 90% CI =  $\bar{X} \pm 1.645(\text{SEM})$
  - 95% CI =  $\bar{X} \pm 1.96(\text{SEM})$
  - 99% CI =  $\bar{X} \pm 2.58(\text{SEM})$

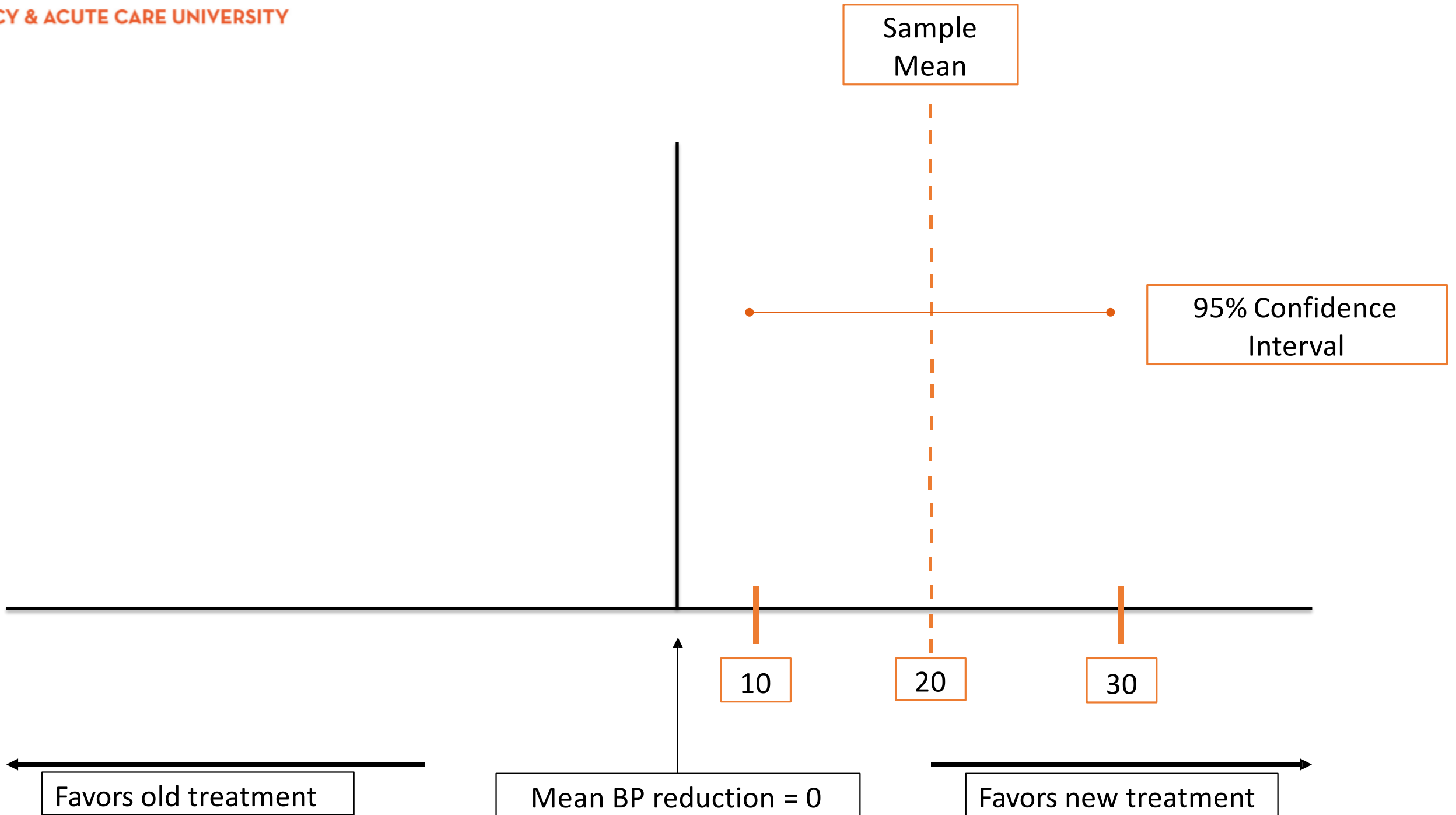
# Confidence Intervals

- Generally written as two numbers (i.e. 95% CI: 100 to 150)
  - Lower confidence limit and upper confidence limit
- Interval size is affected by sample size, size of SD, and confidence level
  - The larger the  $n$ , the smaller the CI, because the mean is more certain
  - The larger the SD, the larger the CI, because the mean is less certain
  - The larger the confidence level, the larger the interval, because you want to be more certain the true population mean is within the interval.
- Interpretation of statistical significance is dependent on if outcome is relative or absolute
  - Relative measure (e.g. Relative Risk, Odds Ratio) statistically significant if interval does not contain 1
  - Absolute measure (e.g. Absolute Risk Reduction, difference) statistically significant if interval does not contain 0



# Fictional Case

- Research question:
  - In adult patients with hypertension, does treatment with T-REXapril reduce blood pressure more than standard care?
- Hypothesis:
  - T-REXapril will reduce the blood pressure more than standard care in adult patients with hypertension.
- Investigators enrolled 1000 patients (500 in T-REXapril arm, 500 in standard care arm)
- Results:
  - The mean reduction in blood pressure between T-REXapril and standard care was 20 mm Hg (95% CI 10 – 30 mm Hg)



# Confidence Interval Interpretation

- Back to our case...
  - The mean reduction in blood pressure between T-REXapril and standard care was 20 mm Hg (95% CI 10 – 30 mm Hg)
- This means that we can say with 95% confidence that the TRUE POPULATION BLOOD PRESSURE REDUCTION is between 10 and 30 mm Hg.
- We can also say that as compared to a 0 mm Hg change in blood pressure, treatment with T-REXapril resulted in a statistically significant reduction in blood pressure



PHARMACY & ACUTE CARE UNIVERSITY

# Inferential Statistics

inferential

# Purpose of Inferential Statistics

- Move from the sample to the population
  - Descriptive statistics allow us to describe what we are seeing in the sample and inferential statistics allow us to make generalizations about the larger population
- Depend on the level of measurement, distribution of the data, and relationship of independent and dependent variables
  - An error in identifying any of these will lead to erroneous inferential results
- Can get complex, but it's very formulaic

# Types of Variables and Relationships

- Independent Variable
  - Generally, the grouping variable or exposure of interest
  - Variation does not depend on another variable
- Dependent Variable
  - Generally, the outcome of interest
  - Variation depends on another variable
- Unpaired samples (Independent Samples)
  - Groups are not related
- Paired samples
  - Groups are related

# Hypothesis Testing

- A hypothesis is an educated guess as to what will happen in an experiment
  - Can have an infinite number of outcomes making it difficult to statistically test all potential outcomes
- The null hypothesis is used in statistics because it allows for an evaluation of a single potential outcome
  - In a standard difference study, the null hypothesis is that there is no difference between the groups
  - This will change with non-inferiority studies

# Error

- Deducing the wrong conclusion
- Unable to appropriately infer from the sample to the population
- Type I error
  - False-positive
  - Reject the null hypothesis when it is true
  - You find a difference when one doesn't really exist
  - Alpha is the acceptable rate of a type one error – generally 0.05
- Type II error
  - False-negative
  - Accept the null hypothesis when it is false
  - Determine that there is no difference when one does exist
  - Beta is the acceptable rate of a type II error – generally 0.1 to 0.2



# P-values

- The calculated probability of a type I error
- If less than alpha we reject the null hypothesis
- Why 0.05?
  - Sir R.A. Fisher
  - A 1 in 20 chance of accepting a false positive is good enough and convenient
- Doesn't tell you if a result is **CLINICALLY** significant
- What if the p-value is greater than alpha?
  - Does not mean there is no association in the population – could be a type II error

# Parametric Tests

- Normally Distributed Interval and Ratio Data
- T-test
  - Paired
  - Unpaired
- Analysis of Variance (ANOVA)
  - One-way ANOVA
  - Two-way ANOVA
  - ANCOA
  - Repeated Measured ANOVA

# T-test

- Most commonly used inferential test
- Assumptions
  - Normally distributed data
  - Variance is near equal
- Method of choice when you are making a single comparison between two levels of an independent variable
- Compares two means

# T-Test examples

- Independent samples t-test
  - A researcher wishes evaluate the mean ICP reduction in patients who receive treatment. Patients are randomized to receive either mannitol or hypertonic saline at equimolar doses when their ICP gets above 20 mm HG.
  - In a cohort study, patients were divided into groups of based on if they received nicardipine or clevidipine. The primary outcome was mean blood pressure
- Paired t-test
  - A researcher wishes to evaluate the mean ICP reduction in patients who receive mannitol. The before treatment ICP is compared to the post treatment ICP.

# ANOVA

- Used to determine if there is a statistical difference when there are 2 or more levels of one or more independent variables
- Uses the F-test to determine if there is a statistical difference
- Same assumptions as t-test
- 1-way ANOVA = 1 independent variable with 2 or more levels
- 2-way ANOVA = 2 independent variables with 2 or more levels each
- ANCOVA = Adds confounders to the evaluation
- Repeated Measures ANOVA = Paired version of ANOVA

# ANOVA Examples

- One-way ANOVA
  - 3 different vancomycin dosing regimens were compared in a sample of patients. Vancomycin concentration was the primary outcome
- Two-way ANOVA (factorial design)
  - Researchers wish to evaluate the effect of epoprostenol and vitamin C had on mean post treatment P:F ratio in patients with ARDS.
- ANCOVA
  - Researchers wish to evaluate the effect of 3 different pain regimens on the mean post-treatment pain score. They wish to control for BMI as a potential confounder.

# Multiple Comparisons Methods

- Helps you determine where the difference is
- Pairwise comparisons that use corrections to keep the level of significance stable
  - The probability of obtaining a significance by chance is increased when making multiple comparisons of the same dataset
- To account for this increased probability of a type I error, there are multiple corrections that can be used
  - Bonferroni, Tukey, Scheffe's, and Dunnett's are the most common
  - The big take away is that these help you determine where the significance lies while taking into account the increased risk of a type I error

# Non-parametric tests

- Nominal
  - Independent Samples
    - Chi Square
    - Fisher's Exact
  - Paired Samples
    - McNemar
- Ordinal or Non-normal numeric
  - Independent Samples
    - Mann Whitney U
    - Kruskal Wallis
  - Paired Samples
    - Wilcoxon signed rank
    - Friedman



# Chi Square, Fisher's Exact, McNemar

- Evaluates if there is an association between two or more levels of a single independent variable with a nominal dependent variable
- Tests proportions rather than means
- Null hypothesis: proportions are equal
- Simplified Assumptions
  - Variables are nominal
  - Observations are independent
  - Expected cells should be 5 or more in 80% of cells
  - No cells contain a 0
- Violate any of those assumptions or the  $n < 20$ , use Fisher's exact

# Chi Square, Fisher's Exact, McNemar Examples

- Chi Square & Fisher's Exact
  - Patients are randomized to placebo or breathalol (new albuterol derivitaive). The primary outcome is the proportion of patients who become intubated.
  - Researchers wish to evaluate the occurrence of hypotension in patients who are treated with etomidate or ketamine following intubation.
- McNemar's Test
  - Researchers evaluate the occurrence of breathlessness before and after treatment with albuterol.

# T-test and ANOVA analogs

- Independent Samples T-test analog
  - Mann Whitney U
- Paired T-test analog
  - Wilcoxon signed rank
- ANOVA Analog
  - Kruskal Wallis
- Repeated Measures Analog
  - Friedman

# Mann Whitney U

- Also Known as Wilcoxon Rank Sum
- Compares the mean rank of the data
- Assumptions
  - Observations are independent
  - Responses are at least ordinal
- Generally, will see this reported as a comparison of medians
- Paired test = Wilcoxon signed rank

# Mann Whitney U Examples

- Researchers wish to evaluate RASS in patients treated with dexmedetomidine or bolus lorazepam
- Patients are randomized to morphine or hydromorphone at equianalgesic doses. Researchers compare post-treatment pain scores (1 to 10 visual analog scale) as the primary outcome.

# Inferential Test Summary

## Unpaired

- Parametric data
  - 2 levels of the independent variable
    - T-test
  - More than 2 levels of the independent variable
    - ANOVA
- Non-Parametric Data
  - Nominal
    - Chi Square or Fisher's Exact
  - Ordinal, interval, ratio
    - 2 levels of the independent variable
      - Mann Whitney U
    - More than 2 levels of the independent variable
      - Kruskal Wallis

## Paired

- Parametric data
  - 2 levels of the independent variable
    - Paired T-test
  - More than 2 levels of the independent variable
    - Repeated Measures ANOVA
- Non-Parametric Data
  - Nominal
    - McNemar's test
  - Ordinal, interval, ratio
    - 2 levels of the independent variable
      - Wilcoxon signed rank
    - More than 2 levels of the independent variable
      - Friedman

# Survival Analysis

- Survival doesn't have to mean mortality
  - Can be any nominal outcome
- If you see "TIME TO..." you should be thinking survival analysis
- Kaplan Meier curve is the most commonly used method for describing the time to an event in biomedical literature
  - Purely descriptive
- Log Rank is the inferential test used to compare the survival of two or more independent groups ... just gives you a p-value
- Cox Proportional Hazards is the multivariate way to compare survival in two independent groups ... provides a hazard ratio



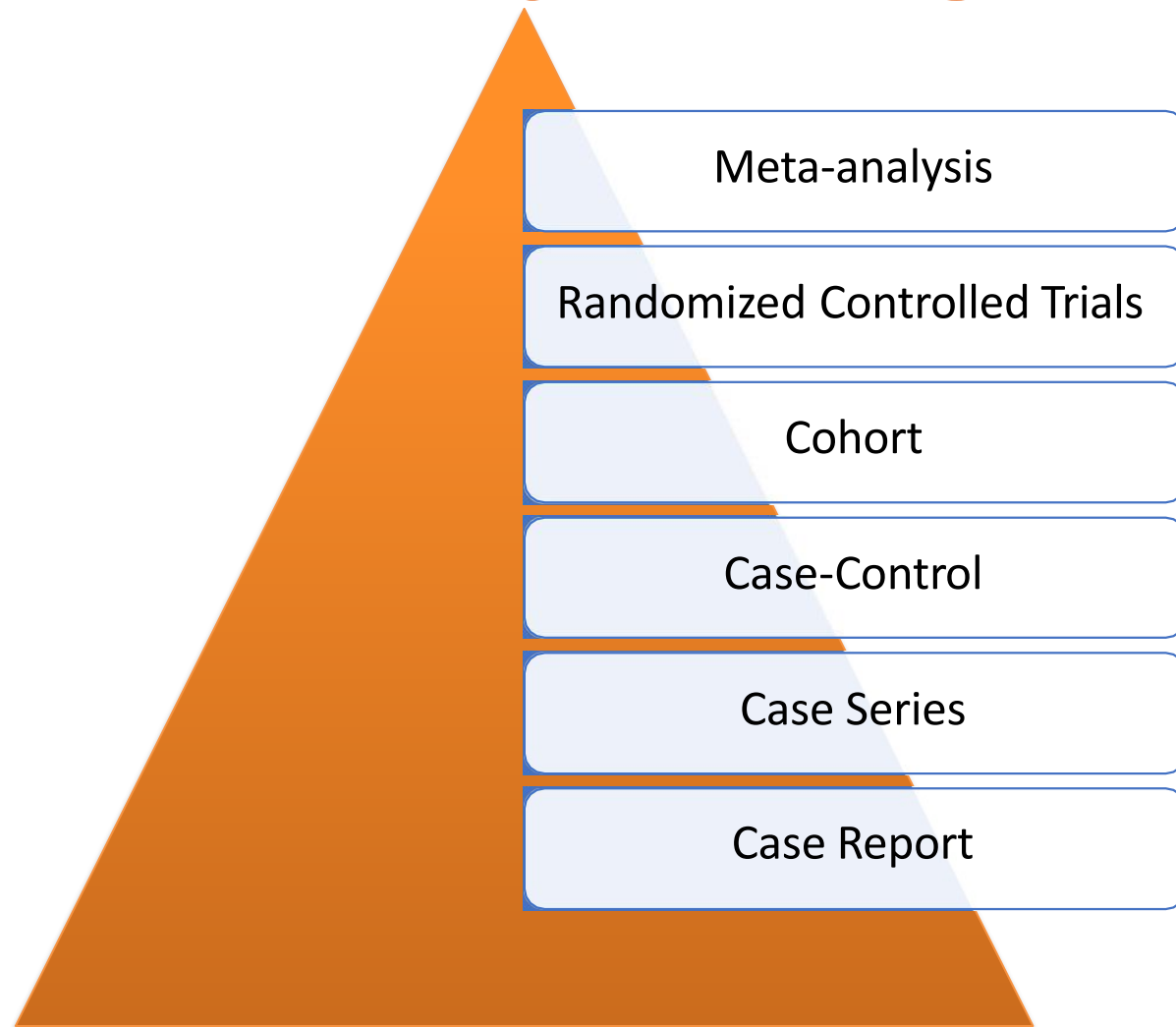
PHARMACY & ACUTE CARE UNIVERSITY

# Study Designs and Calculations

calculations



# All study designs are important



- Not everything fits into a randomized controlled trial format
  - Exposure to known harmful agents
  - Rare exposures
  - Rare outcomes

# Randomized Controlled Trials

- Best design for evaluating causality of an intervention
- Allows researcher to control for many potential biases and confounding through the design process
  - Blinding
  - Randomization
  - Subject Selection
- Commonly thought of as the strongest single study type

# Cohort Study

- Prospective or Retrospective in nature
- Patients are grouped based on EXPOSURE to something
  - Medication
  - Surgery
  - Environment
- Great for evaluating rare exposures
- Advantages
  - Cheaper than an RCT
  - Can study multiple outcomes
  - Strong observational study design

# Case Control Study

- Always retrospective
- Groups patients based on OUTCOME
- Great for evaluating rare outcomes
- Advantages
  - Quick and less expensive than a cohort and RCT
  - Good for rare outcomes
  - Great for outcomes that take a long time to develop

# Odds Ratio

- Defined as:
  - Odds of development of disease in the treatment group compared to the non-treatment group
- Used when incidence can not be calculated
  - Retrospective studies, when only prevalence is reported
- Relative Measure of Association
  - $OR > 1$ , the treatment has a positive association (increased odds) with disease
  - $OR < 1$  the exposure has a negative association (decreased odds) with disease
  - $OR = 1$ , no association between treatment and disease
- OR loses accuracy further away from 1.

# Odds Ratio Example Interpretation

**Table 2.** Unadjusted and adjusted odds ratios for variables in the final logistic regression model predicting for the development of contrast induced nephropathy

Variable	Unadjusted OR	Unadjusted 95% confidence limit	Adjusted OR	Adjusted 95% Wald confidence limits	Estimate	Standard error	Wald chi-square	P value
Intercept	-	-	-	-	-2.26	0.34	45.40	<.01
History of diabetes mellitus	3.62	1.64, 8.00	4.15	1.77, 9.75	1.42	0.44	10.65	<.01
No home medications	2.28	1.00, 5.19	3.56	1.42, 8.94	1.27	0.47	7.30	.01
History of smoking	0.22	0.06, 0.75	0.204	0.06, 0.72	-1.59	0.65	6.08	.01

Abbreviation: OR, odds ratio.

Deviance and Pearson  $P > .05$  ( $>.2151$  and  $.02384$ , respectively).

Hosmer and Lemeshow goodness-of-fit  $P = .3400$ .

- In a multivariate model predicting for development of contrast induced nephropathy in patients with stroke, a history of diabetes mellitus was associated with 4.15 times higher odds of developing CIN when controlling for the other variables in the model.

# Relative Risk and Absolute Risk

- Used for prospective studies
  - Can only be used when incidence is known or can be calculated
- Absolute Risk
  - Same as incidence
- Relative Risk
  - Compares the absolute risk of one group to the absolute risk of another group
  - Relative measure of association
    - Interpret the same as an Odds Ratio

# Relative Risk Example

Outcome*	Atorvastatin (N=2365)	Placebo (N=2366)	Unadjusted P Value†
	<i>no. (%)</i>		
<b>Primary outcome</b>			
Nonfatal or fatal stroke§	265 (11.2)	311 (13.1)	0.05
Nonfatal stroke	247 (10.4)	280 (11.8)	0.14
Fatal stroke	24 (1.0)	41 (1.7)	0.04

- The SPARCL trial was published in NEJM in 2006.
- 4731 patients with recent history of stroke were randomized to either 80 mg atorvastatin or placebo

$$AR_E = 265/2365 = 0.112 \times 100 = 11.2\%$$

$$AR_C = 311/2366 = 0.131 \times 100 = 13.1\%$$

$$RR = 0.11/0.131 = 0.84 \times 100 = 84\%$$

**Interpretation:** Those patients who received atorvastatin 80 mg once daily had 84% of (0.84 times) the risk of developing a nonfatal or fatal stroke patients as compared to those patients in the placebo group.



# Relative Risk Reduction

- When the Relative Risk is less than 1, it can be difficult to interpret what the number means in relation to treatment and non-treatment
- By subtracting the relative risk from 1, we now have a relative risk reduction
  - This tells you the amount by which the treatment reduced the risk as compared to the placebo
- Back to the SPARCL Example:
  - $RR = 0.84$
  - $RRR = 1 - 0.84 = 0.16 \times 100 = 16\%$
  - Interpretation: As compared to placebo, the use of atorvastatin 80 mg daily reduced the risk of developing a fatal or non-fatal stroke by 16%.

# Absolute Risk Reduction

- The difference between the  $AR_C$  and  $AR_E$ 
  - $ARR = AR_C - AR_E$
- An absolute measure of association
  - $ARR = 0$  – There is no difference in Risk Rate
  - $ARR > 0$  – The risk in the experimental group is reduced
  - $ARR < 0$  – The risk in the experimental group is increased
- Usually converted to a Number Needed to Treat or Number Needed to Harm

$$\text{Number Needed to Treat (NNT)} = \frac{1}{ARR} \quad \text{Number Needed to Harm (NNH)} = \frac{1}{ARR}$$

# Absolute Risk Reduction Example

Outcome*	Atorvastatin (N=2365)	Placebo (N=2366)	Unadjusted P Value†
	<i>no. (%)</i>		
<b>Primary outcome</b>			
Nonfatal or fatal stroke§	265 (11.2)	311 (13.1)	0.05
Nonfatal stroke	247 (10.4)	280 (11.8)	0.14
Fatal stroke	24 (1.0)	41 (1.7)	0.04

- The SPARCL trial was published in NEJM in 2006.
- 4731 patients with recent history of stroke were randomized to either 80 mg atorvastatin or placebo

## Fatal Stroke

$$AR_E = 265/2365 = 0.112 \times 100 = 11.2\%$$

$$AR_C = 311/2366 = 0.131 \times 100 = 13.1\%$$

$$ARR = 0.131 - 0.112 = 0.019 \times 100 = 1.9\%$$

$$NNT = 1/0.019 = 53 \text{ patients}$$

Interpretation: Those patients who received atorvastatin 80 mg daily had an absolute risk reduction of 1.9% of nonfatal or fatal stroke as compared to those who received placebo. This corresponds to a need to treat 53 patients with atorvastatin 80 mg daily to prevent 1 nonfatal or fatal stroke.

# Conclusions

- Biostatistics and interpretation of medical literature can be complex and intimidating, but it doesn't have to be
- Descriptive statistics are used to summarize the sample and inferential statistics are used to relate the sample back to the population
- Levels of measurement are commonly referred to as nominal, ordinal, interval, and ratio. Identification of the correct level of measurement is the first step in identifying the correct descriptive and inferential statistic
- Mean (SD) and Median (QR) = Peanut butter and jelly
- Identifying the correct inferential test relies on correct identification of level of measure, number of levels, and distribution of the data
- RCTS are the highest level of evidence but are not appropriate for all outcomes and designs
- Clinical significance is determined by you not the p-value; however, without a significant p-value you can't have clinical significance

# **Biostatistics and Study Design for Pharmacists**

A. Shaun Rowe, PharmD, MS, BCCCP, FNCS, FCCP