Epidemiology and Basic Statistics for the Boards

Sheela N. Magge, MD, MSCE

Professor of Pediatrics

Division Director, Pediatric Endocrinology and Diabetes

Lawson Wilkins Chair of Pediatric Endocrinology

Johns Hopkins University School of Medicine

Domain 19: Core Knowledge in Scholarly Activities – 4%

Topics Outline:

A. Principles of biostatistics in research
B. Principles of epidemiology and clinical research design
C. Ethics in research
D. Quality improvement and patient safety

Types of Variables

Nominal:

- Qualitative not quantitative
- Nominal/Categorical
  - Categories that have no particular order or rank
    - Eye color; ethnicity
  - Dichotomous
    - Yes/No; Dead/Alive

Ordinal:

- Inherent order among categories but differences may not be the same throughout the scale
  - Pain scale; Anxiety scale; Apgar score; GCS score
  - May approach an interval scale if there are enough categories – 100 point visual analog score (VAS) for pain

Types of Variables

Interval Variables:

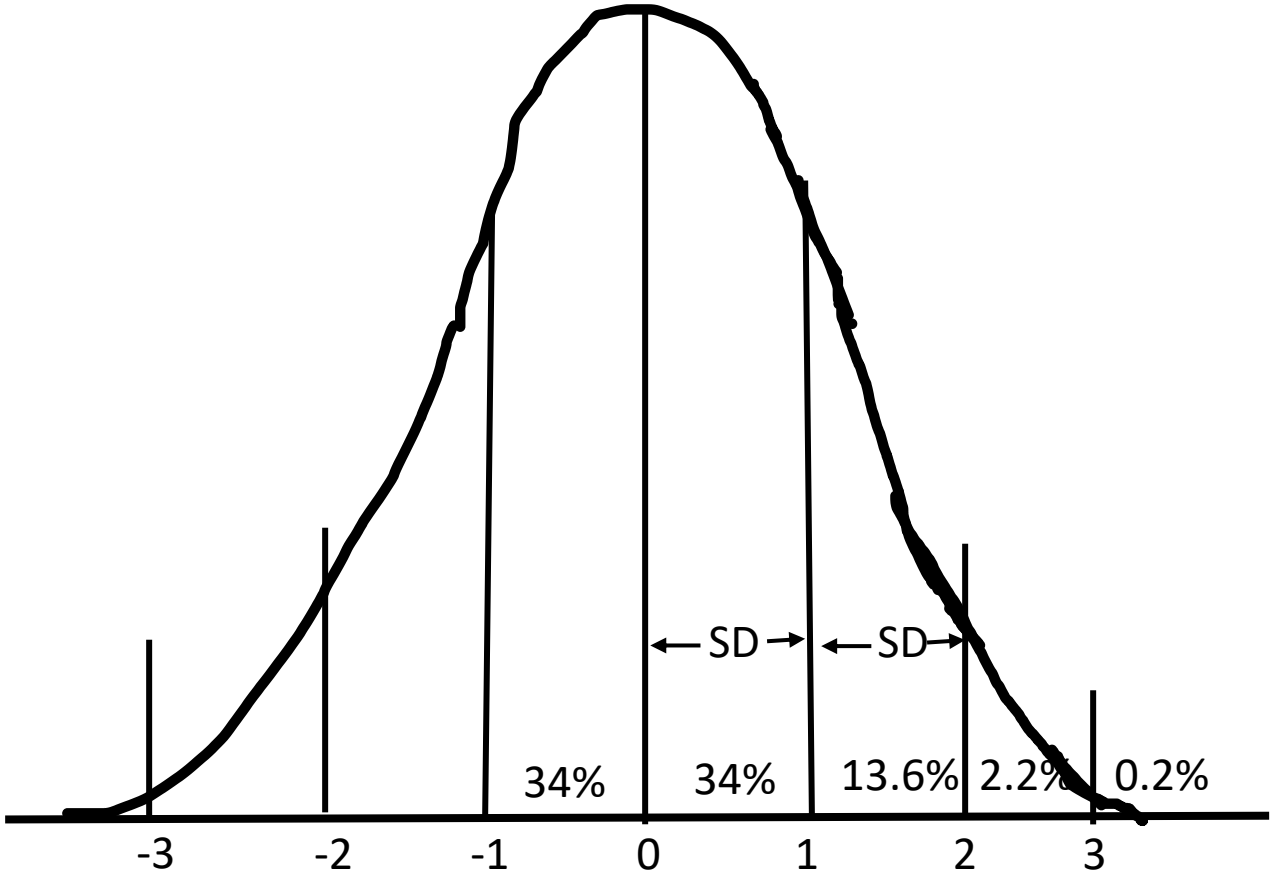
- Quantitative not qualitative
- Continuous: value can take any number in a range
  - Height, weight, length of stay, age, blood glucose
  - Some overlap with ordinal variables if that ordinal variable has enough categories (i.e. 100 pt pain scale)
  - Distribution can vary: symmetrical or skewed
  - More powerful

Types of Variables

Interval versus Ordinal:

- Interval Data
  - Number of Children in a Family
    - Values 0 1 2 3 4 5 . . .
    - 3 children are 3 times as many children as 1 child.
    - Does a difference of 1 child means the same throughout the range of possible values?
- Ordinal Data
  - Heart Murmur
    - Values I II III IV V VI
    - Is a IV/VI murmur twice as loud as a II/VI murmur?
    - Is the difference between I and II the same as the difference between III and IV?

Distribution of Data –
Normal Distribution

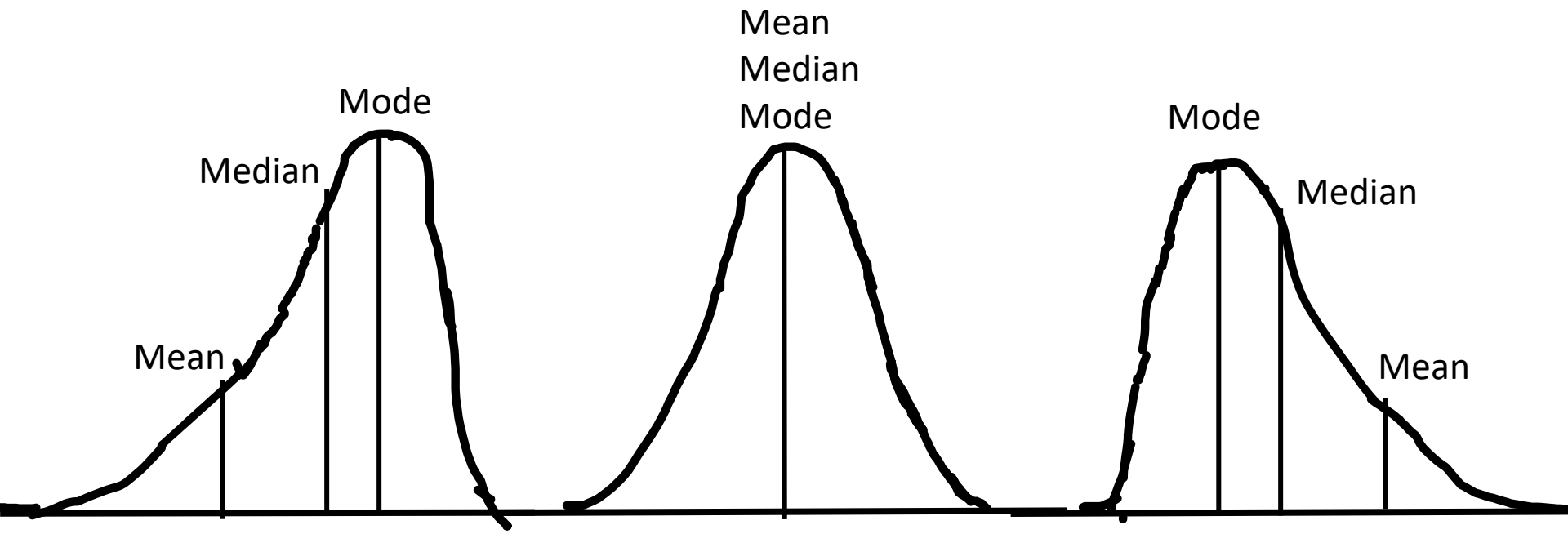


- What % fall within ± 1 standard deviation of mean? 68%
- What % fall within ± 2 standard deviations of mean? 95%
- What % fall within ± 3 standard deviations of mean? 99%

Distribution of Data
Measures of Central Tendency

- Mean = average
  - Reported and analyzed when distribution is normal/symmetrical/parametric
  - If there are outliers, the mean tends toward the outlier values
- Median = value of middle observation
  - Reported and analyzed when distribution is skewed or non-normal/nonparametric or ordinal data
- Mode = most common value

Distribution of Data

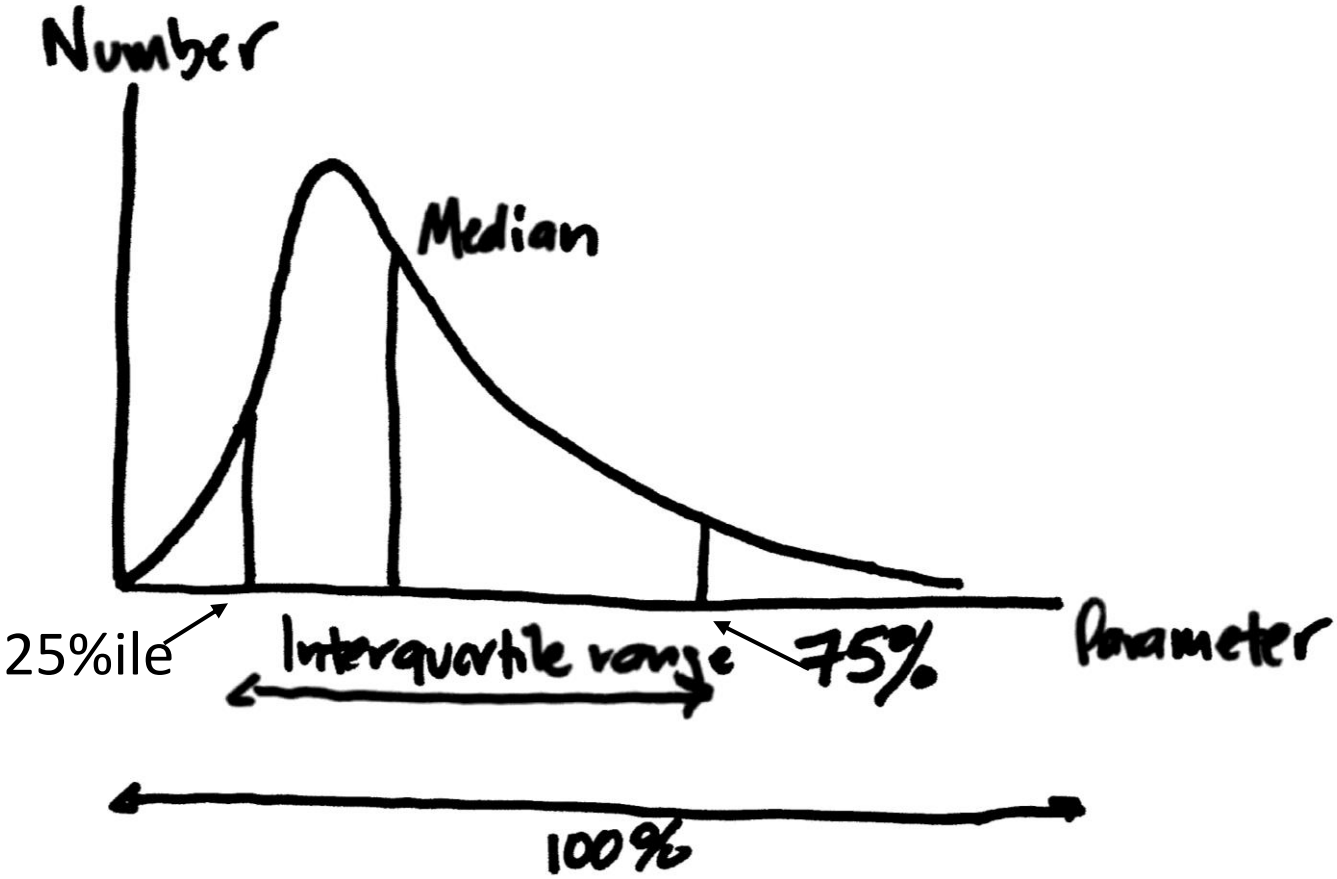


Negatively Skewed
= Left skewed

Normal Distribution

Positively Skewed
= Right skewed

Distribution of Data
Skewed (Non-Normal) Distribution



Descriptive Statistics

- Standard Deviation:
  - Quantifies the spread of individual observations of a value of a variable around the mean value of the sample
  - Add up the square of each difference from the mean, divide by number of data points, then take square root
  - Variance – measures the spread; SD= square root of variance

Descriptive Statistics

- Standard Error:
  - Measures the variability of the sample mean as an estimate of the true value of the population mean (from which the sample was drawn)
  - Indicates the degree to which the sample mean reflects the true population mean
  - SE= SD/square root of n

Distribution of Data

Synonyms:

- Normal distribution
  - Parametric
  - Bell-shaped
  - Symmetric
- Non-normal distribution
  - Non-parametric
  - Skewed
  - Not symmetric
  - How analyze? Can transform data

Hypothesis Testing

- Hypothesis Testing- test of statistical significance

- Null Hypothesis = H_0
  - The assertion that there is no relationship between the exposure and the disease

Hypothesis Testing

- One-tailed test: hypothesis is that one is greater than the other

- Two-tailed test: hypothesis is that the groups are different

Hypothesis Testing

- Type I error
  - Null hypothesis is rejected when it is actually true
  - False positive, α
- Type II error
  - Mistaken failure to reject the null hypothesis when the alternative hypothesis is true
  - False Negative, β

Hypothesis Testing

- Sample Size – What do you need to calculate?
  - Specify desired α and β
  - Proportion of baseline population that has the exposure or disease being studied
  - Magnitude of expected effect or detectable difference
  - If don't find a significant difference, need to see if study was adequately powered to find the difference
  - Larger sample size increases the power

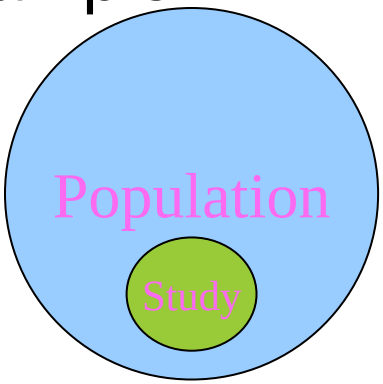
Hypothesis Testing
p-value

- Probability that the observed result (or results even more extreme) could have occurred simply by chance under the null hypothesis

- What if p>0.05? What does it mean - does it always mean that there is no difference?
  - No –study could be underpowered

Hypothesis Testing
Confidence Interval

- 95% confidence interval (CI): range of values which we can be 95% confident includes the population statistic from which the study sample was drawn
  - Mean
  - Proportion
  - RR or OR



- Relates to how precise the measurement is

- Things that affect the width of the confidence interval
  - Level of confidence
  - Sample size
  - Variability of the data (e.g. standard deviation)

Interpreting Confidence Intervals

- Means
  - Typically expressed as the 95% CI around the difference between means
  - Null value is 0
  - If 95% CI includes 0, not statistically significant

- OR/RR/Hazard Ratio
  - Null value is 1
  - If 95% CI includes 1, not statistically significant

P-value vs CI

- Confidence intervals versus p values
  - “Is the difference large enough to recommend a change in treatment”? (confidence intervals)
  - “Is there a difference?” (p values)

- Statistical significance can be associated with a clinically insignificant difference (large sample size).

- Clinical significance can be associated with a non-significant p-value (sample size too small).

Common Statistical Tests

- What test should you use to analyze your data?
  - Look at the type of variables you have – continuous vs categorical
  - Look at the distribution of the data you have – normally distributed (parametric) vs non-normally distributed (non-parametric)

Common Statistical Tests

- If you are comparing continuous outcome variables between 2 groups use….
  - Parametric distribution: t-test
    - **Unpaired or Student's T-test –** two independent samples
    - **Paired t-test –** not independent, ex: pre- and post- in same individual
  - Nonparametric distribution: Wilcoxon Rank Sum = Mann Whitney U test

Common Statistical Tests

- If you are comparing continuous outcome variables between >2 groups use….
  - Parametric distribution: **ANOVA**
    - Like a t-test for more than one group
    - If significant, only means that at least one of the groups is different than at least one other group
  - Nonparametric distribution: **ANCOVA**
    - Merger of ANOVA and regression for continuous variables

Common Statistical Tests

- If you are comparing a categorical outcome variable between 2 groups use….
  - Parametric: **Chi-Square Test**
  - Nonparametric: **Fisher exact test** (use if any cell has <5 participants)

Common Statistical Tests

- If you are comparing 2 continuous outcome variables, use….

- Correlation
  - Parametric: **Pearson** correlation
  - Nonparametric **Spearman** correlation
  - Obtain **correlation coefficient**, r (-1 to 1), indicates the strength of the relationship; p-value

- Linear Regression- can control for other variables
  - Fit to a line
  - R^2 = goodness of fit -> 1 is the best or maximum fit

Regression

- Statistical technique which focuses on the relationship between dependent variable (outcome) and ≥ 1 independent variables
- Allows to control for other variables
- Linear regression – dependent variable is continuous/interval variable
  - Fit to a line
  - R^2 = goodness of fit -> 1 is the best or maximum fit
- Logistic regression – dependent variable is dichotomous (yes/no, dead/alive)

Relative vs Absolute Risk

- **Relative Risk:** Assessing the risk for outcome in one group compared to another (ex. with or without a particular risk factor). Divide one risk by the other.

- **Absolute Risk:** Actual risk for an outcome in a group compared to actual risk for outcome in another. Difference in risk between groups.

Measures of Disease Risk

- Studies often focus on assessing the risk for disease with or without a particular trait
- Such measures of risk can be derived from the 2x2 contingency table with same Chi square statistic
- Two Measures
  - Relative Risk (RR)
  - Odds Ratio (OR)

Measures of Disease Risk - Difference Between Probability and Odds

- Probability can be expressed as a percentage
  - Relative Risk (RR) is a ratio of two proportions
- Odds ALWAYS implies a ratio of two probabilities
  - Probability of event happening over probability of event not happening
  - Odds Ratio (OR) is a ratio of two ratios.

Measures of Disease Risk
Probability and Odds

	Probability	Odds
You will fall asleep at your desk at least one night while studying for the boards	95%	95:5 =19:1 19
Passing your endocrinology board exam on the first try	80%	80:20 = 4:1 4
Learning statistics in this board review lecture	60%	60:40 = 3:2
Heads when tossing a coin	50%	50:50 = 1:1 1
Rolling 6 with a fair die	17%	1:5 (0.20)
Drawing an ace from a deck of cards	7.6%	1:12 (0.083)

As the probability decreases (< 10%), the probability and the odds converge.

Measures of Disease Risk: Relative Risk

		Disease		
		Yes (Disease +)	No (Disease -)	
Risk Factor	Present (RF +)	a	b	a + b
	Present (RF -)	c	d	c + d
		a + c	b + d	a + b + c + d = n

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

Risk of disease in exposed divided by risk of disease in unexposed

Measures of Disease Risk: Relative Risk

- Relative risk = risk ratio =

$$\frac{\text{\% of those with a risk factor who have the disease}}{\text{\% of those without the risk factor who have the disease}}$$

  Comparing probabilities of disease in each group:
  - RR = 1 (no difference or the null value)
  - RR > 1 (risk factor increases risk of disease)
  - RR < 1 (risk factor decreases risk of disease)

- Used in prospective cohort studies, randomized controlled clinical trials.
- Start with risk factor (exposure), not disease
- P-value obtained from the Chi-square test
- RR gives magnitude of the difference

Measures of Disease Risk: Odds Ratio

		Disease		
		Yes (Disease +)	No (Disease -)	
Risk Factor	Present (RF +)	a	b	a + b
	Present (RF -)	c	d	c + d
		a + c	b + d	a + b + c + d = n

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Odds of risk factor in those with disease over the odds of risk factor in those without disease

Measures of Disease Risk: Odds Ratio

- Odds ratio= OR = odds of having a risk factor in those with a disease compared to the odds of risk factor without disease. Comparing the odds of the risk factor in each group.
  - OR = 1 (no difference or the null value)
  - OR > 1 (odds of disease increased with risk factor)
  - OR < 1 (odds of disease decreased with risk factor)

- Used in case-control studies, outcome is chosen, so prevalence is artificial.
- Start with disease, look at risk factors.
- P-value is obtained from Chi-square test.
- OR gives magnitude of the difference

Relationship between RR and OR when Disease is Rare

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

		Disease		
		Yes (Disease +)	No (Disease -)	
Risk Factor	Present (RF +)	a	b	a + b
	Present (RF -)	c	d	c + d
		a + c	b + d	a + b + c + d = n

Titanic

	Dead	Alive	
Male	709	142	851
Female	154	308	462
	863	450	1313

- Males look like they more likely to die.
- RR for death if male (709/851) divided by (154/462) = 2.5
  - 2.5 greater probability of dying
- Odds ratio= (709/154) divided by (142/308) = (709)*(308)/(142)*(154)=10
  - 10 fold greater odds of dying if male

Which one is better?

- Relative Risk - is the most interpretable and consistent with the way people think
  - We like to think about the risk of disease associated with a particular exposure
  - Useful in large prospective cohort studies

- Odds Ratio - is calculated from case control studies, where the prevalence of disease is artificially created
  - Approaches the relative risk when outcome is uncommon (<10%)
  - Retrospective studies
  - Logistic regression software will give OR

Number Needed to Treat (NNT)

- NNT is the number of patients who need to be treated in order to prevent one additional "outcome"

- NNT = 1/ARR
  - Attributable (Absolute) Risk Reduction
  - ARR = risk of outcome in non-intervention group – risk of outcome in intervention group
  - Relative Risk Reduction = ARR/placebo or non-intervention group rate

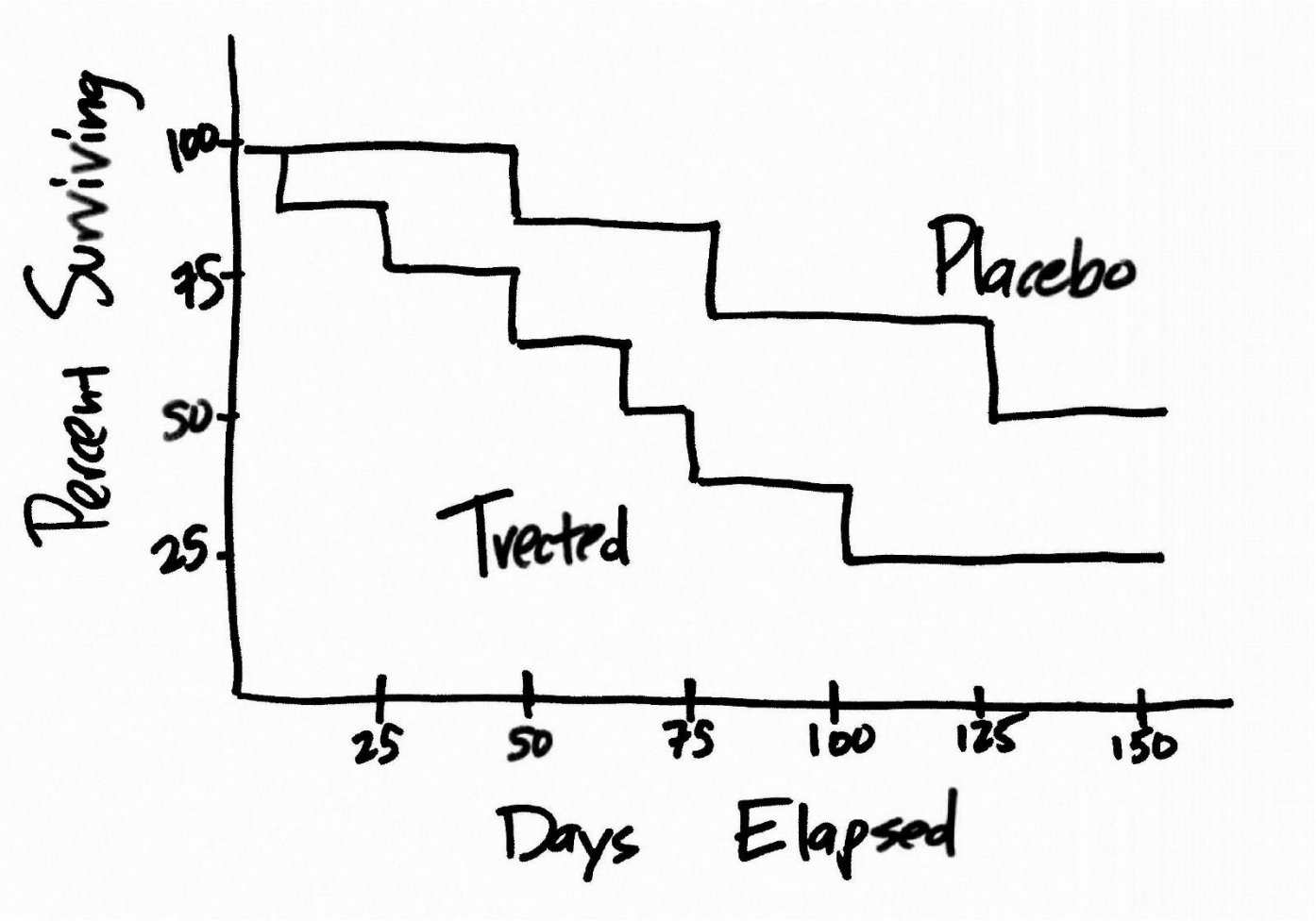
Survival Analysis

- Testing “whether” or “when” an event occurs
- Survival analysis – time to an event
  - Studying the occurrence and timing of events

- Related to logistic regression as the dependent variable is dichotomous (dead/alive)
  - Duration of follow-up not necessarily equal for all patients
  - Not all patients have to experience the event

- **Hazard Ratio-** probability of event in a treatment group relative to control group probability over unit of time – - measure of effect size for time to event data, estimates treatment effect

Survival Analysis
Kaplan-Meier Survival Curve



Statistics

- Problem of multiple comparisons- if you do many comparisons, the chance of finding a rare event increases, and the likelihood of incorrectly rejecting a null hypothesis (false positive, type 1 error) increases
  - Bonferroni adjustment: # tests/α (0.05) – use a more stringent p-value
  - ANOVA with posthoc pairwise testing

Diagnostic Tests

- When evaluating a diagnostic test, it is important to have an independent "gold standard"

- This is the current test felt to be the best currently available to diagnose a particular disease or condition under reasonable conditions

Diagnostic Tests

Test \ Disease	+	-
+	a	b
-	c	d

Sensitivity: Proportion of people with the disease, that have a positive test.

a/(a+c)

-useful for a rare disease that has high morbidity/mortality – want to detect all cases

– high sensitivity -> few false negatives

Specificity: Proportion of people without the disease who have a negative test.

d/(b+d)

-few false positives

Diagnostic Tests

Test \ Disease	+	-
+	a	b
-	c	d

Positive Predictive Value:

-Probability that if the test is positive, the person has the disease

a/(a+b)

Negative Predictive Value:

-Probability that if the test is negative, the person does not have the disease

d/(c+d)

Diagnostic Tests

HbA1c \ T2DM	+	-
+	100	20
-	25	125

Sensitivity:
80%

Specificity:
86%

Positive Predictive Value:
83%

Negative Predictive Value:
83%

Diagnostic Tests

Likelihood ratio =

$$\frac{\text{Probability of test result in the presence of disease}}{\text{Probability of test result in people without disease}}$$

Quantifies change from pre-test probability
to post-test probability

$$\text{Likelihood Ratio} = \frac{\text{Sensitivity}}{\text{1-Specificity}}$$

Diagnostic Tests

Test \ Disease	+	-
+	a	b
-	c	d

False Positive=
1 - specificity

False Negative=
1 – sensitivity

Likelihood Ratio=
sensitivity/ 1-specificity

Diagnostic Tests

HbA1c \ T2DM	+	-
+	100	20
-	25	125

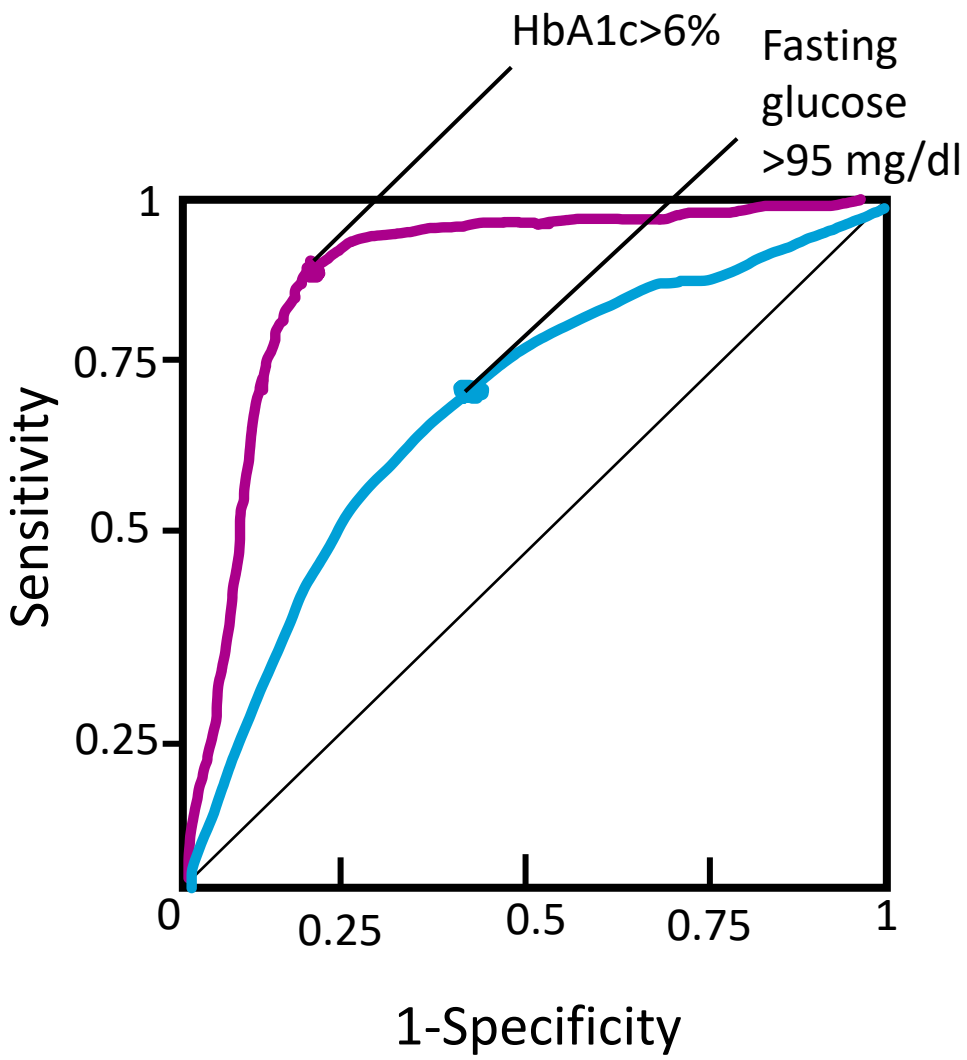
False Positive=
1 - specificity =14%

False Negative=
1 – sensitivity = 20%

Likelihood Ratio=
sensitivity/1-specificity = 5.7

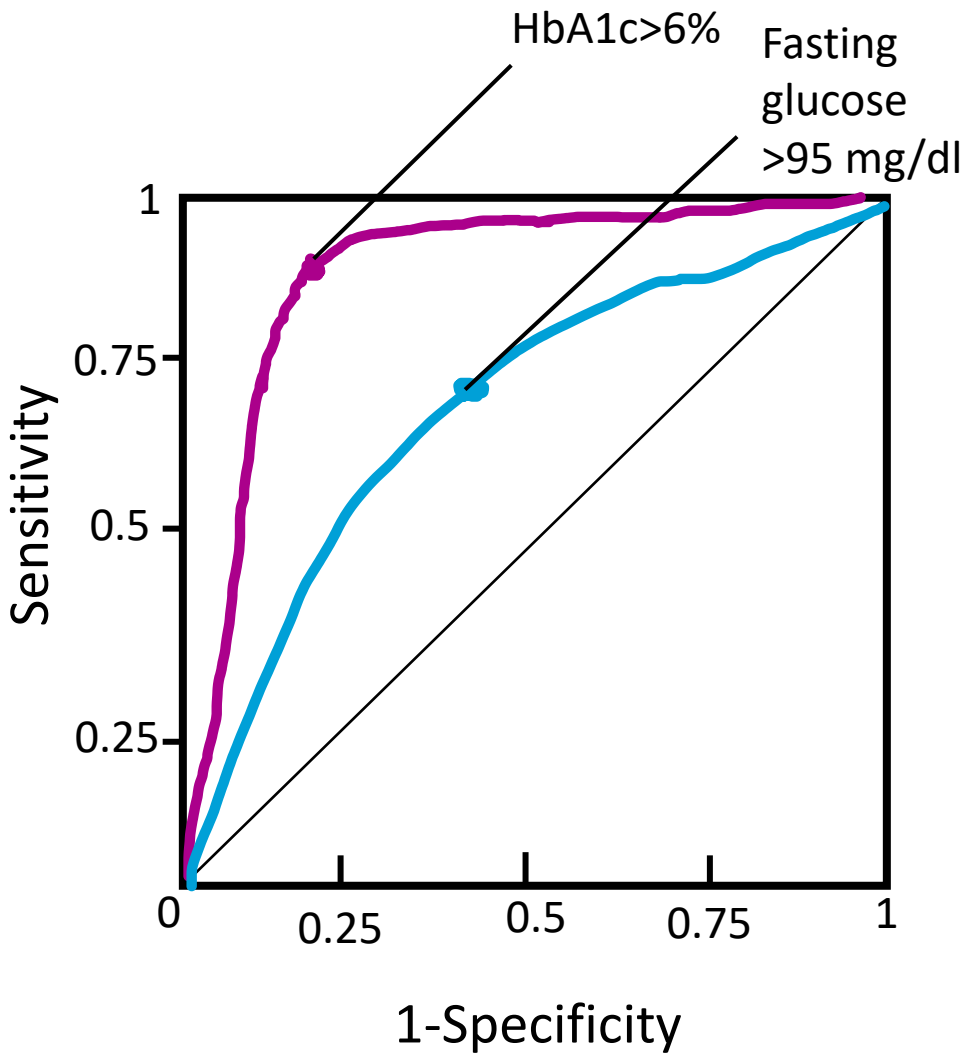
Diagnostic Tests – Receiver Operator Characteristic (ROC) Curves

- Ideal test would have a sensitivity and specificity of 1 (100%)
- Overall accuracy of test described by the area under the curve (the larger the area the better the test)
- Shows trade-off between sensitivity and specificity



Diagnostic Tests – Receiver Operator Characteristic (ROC) Curves

- Helpful for determining cut-offs for lab screening
- Test with curve that is straight line going through points (0,0) and (1,1) is no better than pure chance to detect the presence of the disease
- The top curve represents the best screening test



Systemic Review vs Meta-Analysis

- Systemic review answers a defined research question by collecting and summarizing all empirical evidence that fits pre-specified eligibility criteria.

- Meta-analysis is the use of statistical methods to summarize the results of these studies.

Domain 19: Core Knowledge in Scholarly Activities

Topics Outline:

A. Principles of biostatistics in research

B. Principles of epidemiology and clinical research design

C. Ethics in research

D. Quality improvement and patient safety

Study Design

- Types of Research Designs
  - Case Report
    - Use: hypothesis generating
  - Case Series
    - Use: Characterize disease and hypothesis generating
  - Analysis of Secular Trends / Cross-Sectional
    - Uses aggregate population data
    - Compare geographic or time trends of an illness to trend in risk factors

Study Designs

- Types of Research Designs
  - Case Control Study
    - Start with Disease status and then look back at whether had Exposure
    - Ex: study comparing two groups of infants – with and without HI; look to see who exposed to toxin in utero
  - Cohort Study
    - Start with Exposure status and then see who develops Disease
    - Ex: Compare two groups of children – one exposed to cigarette smoke, and one not; See who develops asthma.

Study Designs

- Case Control Study
  - Adv: Good for study of rare diseases, also can look at many different risk factors for a disease
  - Disadv: Can be prone to bias – ex. Recall bias
- Cohort Study
  - Adv: Can study many outcomes from a single exposure; Can use for more common diseases
  - Disadv: Long and expensive

Study Designs

- Clinical Trial
  - Risk factor or exposure is controlled by investigator
  - Experimental as opposed to observational
  - Intervention study
  - Adv: most convincing demonstration of causality
  - Disadv: Logistic and ethical problems dealing with human subjects

Bias

- Flaw in study design, or method of collecting or interpreting data, that can cause incorrect conclusions about what the study showed.

- Many types – some examples:
  - Selection bias
  - Misclassification bias
  - Recall bias
  - Reporting bias
  - Performance bias
  - Attrition bias

Study Designs

- Clinical Trial
  - **Randomization**:
    - Purpose: decrease bias by evenly distributing known and unknown confounders between the groups
    - Avoid selection bias
  - **Blinding:**
    - Single – study participant
    - Double – study participant and investigator
    - Triple – study participant, investigator, and DSMB/statistician/data collector, etc.
    - Purpose: Avoid ascertainment/information bias
    - Alternative: Open label study

Study Designs – Clinical Trial

- **Surrogate marker** is a biomarker that is intended to substitute for a clinical outcome.
  - Useful when the clinical outcome is undesired (ex. death) or in distant future (ex. diabetic complication in a child with diabetes)
  - Changes induced on the surrogate endpoint are meant to reflect changes on clinical outcome
  - May correlate with the treatment's effect on real clinical outcome but does not have guaranteed relationship, so may not predict an actual effect (=disadvantage)

Study Designs – Clinical Trial

- **Intention to Treat Analysis:** method to analyze prospective randomized study where all participants are included and analyzed in the groups they were originally assigned to, no matter what treatment they actually received – considered gold standard
  - Preserved benefits of randomization
  - Avoids bias
  - Opposite is **Per protocol** or **As treated analysis** – ex. May be randomized to tx but stop taking tx because it didn't work. If you do not include them in the tx group, the tx could look more effective than it was.

Study Design

- **Association vs Causality** - association between two variables implies that knowing the value of one variable provides information about the value of the other

- Does not imply causation (exposure produces the effect or outcome).

- How confirm causality?
  - Randomized controlled study

Study Designs

Prospective

Retrospective

Has the outcome of interest occurred at the time the investigator initiates the study???

Study Designs

Most case-control studies are retrospective and most cohort studies are prospective, BUT not always!

Study Designs

Longitudinal – over time

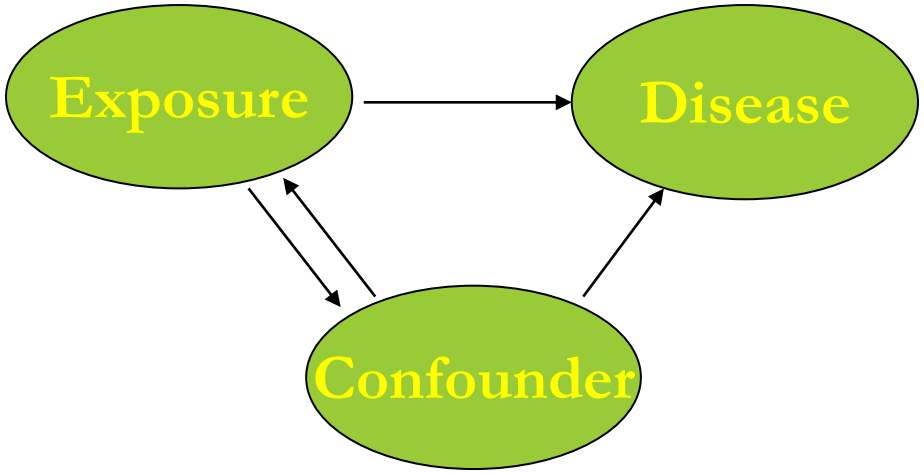
Cross-sectional – one slice of time; at that point

Study Design

- Control group
  - comparison group
  - both groups should be as similar as possible except for the variable being studied
  - “placebo-control” group
  - Randomized controls vs non-randomized controls
  - Historical vs concurrent controls

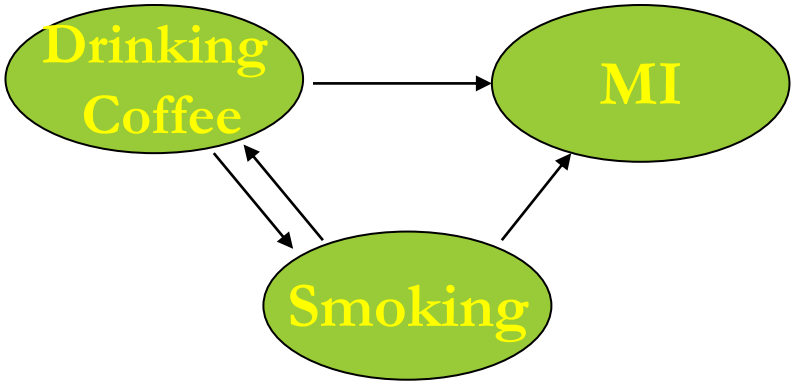
Study Design

- Confounding variables
  - Variable independently associated with both the exposure and the outcome of interest
  - Race/ethnicity, gender, SES can be confounders



Study Design

- Confounding variables
  - Ex. smoking in the relationship between drinking coffee and MI



  - As long as you measure it, you can adjust for a confounding variable, statistically

  - How to avoid confounding: Randomization, Restriction, Matching

Study Design

- Effect Modification (or Interaction)
  - Association between the exposure and disease under study varies by the level of a third factor
  - Different strata
  - Ex. Effect of physical activity on MI risk; EM by gender
  - Described and reported, but cannot control for EM

Study Design

- Generalizability = External validity
  - How applicable is it to the general population?
  - To other races, ethnicities
  - Non-representative samples can bias results
  - Source of data can affect results – ex. measuring the prevalence of dyslipidemia recruiting from Lipid Clinic vs recruiting from primary care. Need to know what question you are asking.

Diagnostic Tests – Prevalence vs Incidence

- Disease prevalence:
  - Number of existing cases at a specific time/number of people in the total population at that time
  - Cross-sectional measure

- Disease incidence:
  - Number of new cases over a period of time/ number of people at risk of developing the disease during that time
  - Time is in the denominator, ex. /Person-years

Qualitative Research Methods and Analysis - Measurement

- Validity
  - Face- does the content of the test appear appropriate for its aims
  - Construct- does it measure the concept it is supposed to
  - Criterion- do the results accurately measure the concrete outcome they are supposed to
  - Predictive- does the measure accurately predict outcomes
  - Content –whether the measure used covers all of the content in the underlying construct
- Reliability

Measurement

- Reliability
  - Consistency across
    - Raters: Intra-Rater and Inter-Rater
    - Time: Test-retest
    - Respondents
    - Items: internal consistency – looks at consistency within a set

- Absence of error, or extent to which random error is minimized

Measurement

- Inter-observer Reliability
  - Kappa
    - correlation-like measure that controls for the problem of inflated percent agreement due to chance
    - Ranges from +1 to -1
    - Needs to be >0.4, prefer between 0.4-0.7

Measurement

- Internal Consistency – how well is a sample of items representative of a domain
  - Cronbach's Alpha
    - Ranges 0-1
    - Tells proportion of a scale's true variance that is due to the true score on the measure (as opposed to error)
    - 0.7 is adequate, 0.8-0.85 is good, 0.9 indicates redundancy

Validity vs Reliability

- Reliability as consistency
- Can be consistent but not accurate
- Accuracy gets to validity

Cost-Effectiveness

- Cost versus charges
  - Cost – cost of services (equipment, personnel, etc.)
  - Charges – what the hospital bills the insurance and patient
- Cost-effectiveness ratio
  - Ratio of dollars expended to health care outcome obtained
- Quality-adjusted life years (QALY)
  - Fundamental component of cost-effectiveness research
  - Measures how well a tx lengthens and/or improves patients' lives – quality and quantity of life
  - To calculate: utility value or weight associated with a state of health x time lived in a state of health
  - One QALY represents 1 year in perfect health (0 death, 1 perfect health)
- Cost Benefit analysis – compare the costs and benefits of an intervention in monetary units

Domain 19: Core Knowledge in Scholarly Activities

Topics Outline:

A. Principles of biostatistics in research

B. Principles of epidemiology and clinical research design

C. Ethics in research

D. Quality improvement and patient safety

Professionalism and Misconduct in Research

- Identify and manage potential conflicts of interest in the funding, design, and/or execution of a research study

- Forms of Research Misconduct
  1. Plagiarism
  2. Falsification
  3. Fabrication

Professionalism and Misconduct in Research

- Misconduct is not making an honest mistake.

- Requires that there is a departure from accepted practice in the research community, that the misconduct was intentional, knowing, or reckless, and that it was proven by preponderance of evidence.

- Know how, and to whom, to report concerns of research misconduct - Office on Research Integrity and Committee on Publication Ethics

Ethics in Research

Research involving human subjects

- Nuremberg Code – 1947
  - 10 principles to satisfy ethical conduct for human experimentation
  - Includes voluntary consent
- Helsinki Declaration- 1964
  - Formal code of ethics for physicians involved in clinical research

Ethics - Research Involving Human Subjects

- Federal guidelines and Ethical safeguards
  - 1974 National Research Act – required Institutional Review Boards for all research funded in part by the government
  - 1978 Belmont Report- ethical principles and guidelines for protection of human subjects
    1. Respect for persons (individual autonomy)
    2. Beneficence
    3. Justice
  - Informed consent
  - Data safety monitoring boards – no vested interest in outcome

Ethics - Research Involving Human Subjects

- Unconventional patient care vs. research
- Ethical Considerations of Study Design
  - Placebo
  - Harm of intervention
  - Deception
  - Flawed design

IRB vs DSMB

- IRB – appropriately constituted group that has been formally designated to review and monitor biomedical research involving human subjects in accordance with FDA regulations. Has the authority to approve, require modification or disapprove research.

- DSMB – independent group of experts that prospectively identifies and documents activities to protect the safety of the subjects, the validity of the data and integrity of the research study.

Minimal Risk

- Risk where the probability and magnitude of harm or discomfort anticipate in the research are not greater in and of themselves than those normally encountered in daily life or during the performance of routine physical or psychological examinations or tests.

- Protections for vulnerable populations such as children

Informed Consent - Components

1. Description of the investigation – that it involves research, purpose of study, duration of participant's involvement, description of procedures and any that are experimental
2. Description of any foreseeable risks or discomforts
3. Description of potential benefits expected
4. Disclosure of alternative procedures or treatments that might be beneficial to person
5. Description of the extent that confidentiality will be maintained
6. If more than minimal risk, description of compensation and medical treatment in case of injury
7. Identification of contact people
8. That participation in voluntary, and that failure to participate will not involve any penalty or loss of benefits to which they are otherwise entitled

Consent vs Assent

- Consent may only be given by individuals who have reached the legal age of consent (typically 18 years of age).

- Assent is the agreement of someone not able to give legal consent to participate in the activity.

Domain 19: Core Knowledge in Scholarly Activities

Topics Outline:

A. Principles of biostatistics in research

B. Principles of epidemiology and clinical research design

C. Ethics in research

D. Quality improvement and patient safety

Quality Improvement Models and Tools

- Models provide systematic formal framework for creating QI processes in practice.
- Common QI models:
  - Model for improvement (Plan-Do- Study-Act (PDSA) cycles – combines Total Quality Management and Rapid Cycle Improvement models
  - Six Sigma – Method of improvement that tries to decrease variation and defects
  - Lean – Approach to drive out waste and improves efficiency in work so that all work adds value

Quality Improvement and Patient Safety
–
Project Design - PDSA

- Plan
  - Objectives, questions and predictions (why?)
  - Plan to carry out the cycle (who, what, where, when)
  - Plan for data collection
- Do
  - Carry out the plan, begin analysis of the data
  - Document problems & unexpected observations
- Study
  - Complete the analysis, compare data to predictions
  - Summarize what was learned
- Act
  - What changes need to be made?
  - Next Cycle?

Quality Improvement and Patient Safety

- Aim of quality improvement project should be specific, measurable, achievable, realistic, and time-limited

- Quality Improvement Tools

Quality Measures

Measure		Examples
Outcome Measure	What is your primary result?	Percentage of infants discharged from NICU who received HBV
Process Measure	Are the parts or steps in the system performing as planned? Early indicators of whether or not changes are improvement.	Pharmacy documentation in EHR regarding infant's HBV status? Percentage of nurses receiving education Percentage of parents receiving VIS to ensure consent
Balancing Measure	What happened to the system as we improved the outcome and process measure? Unintended consequences.	Anaphylaxis, fever/apnea after immunization, premies given too early

Common Cause vs Special Cause Variation

- Common cause variation – expected variation within a given system
  - Typical temperature range during a specific season 55-75F
- Special cause variation – changes that are unexpected or outside the norm
  - Temperatures outside of the normal range, ex. 90F, as a result of special cause variation such as heat wave

Run Chart

- Used to plot data over time, noting interval of time
- X-axis is time and Y-axis is the metric being studied (could be count or rate)

Statistical Process Control (SPC) Charts

- More useful way to show changes over time
- Central line is mean
- Study how a process changes over time – x axis always time
- Display the mean (center line) and upper and lower control limits, which are ~ 3 SD above and below mean

SPC Charts

- Spikes in control limits usually correspond to time periods with small sample sizes
- Values close to 0 or 100 will likely not have control limits calculated

- Compare current data to historical data

- Is process variation consistent or unpredictable?

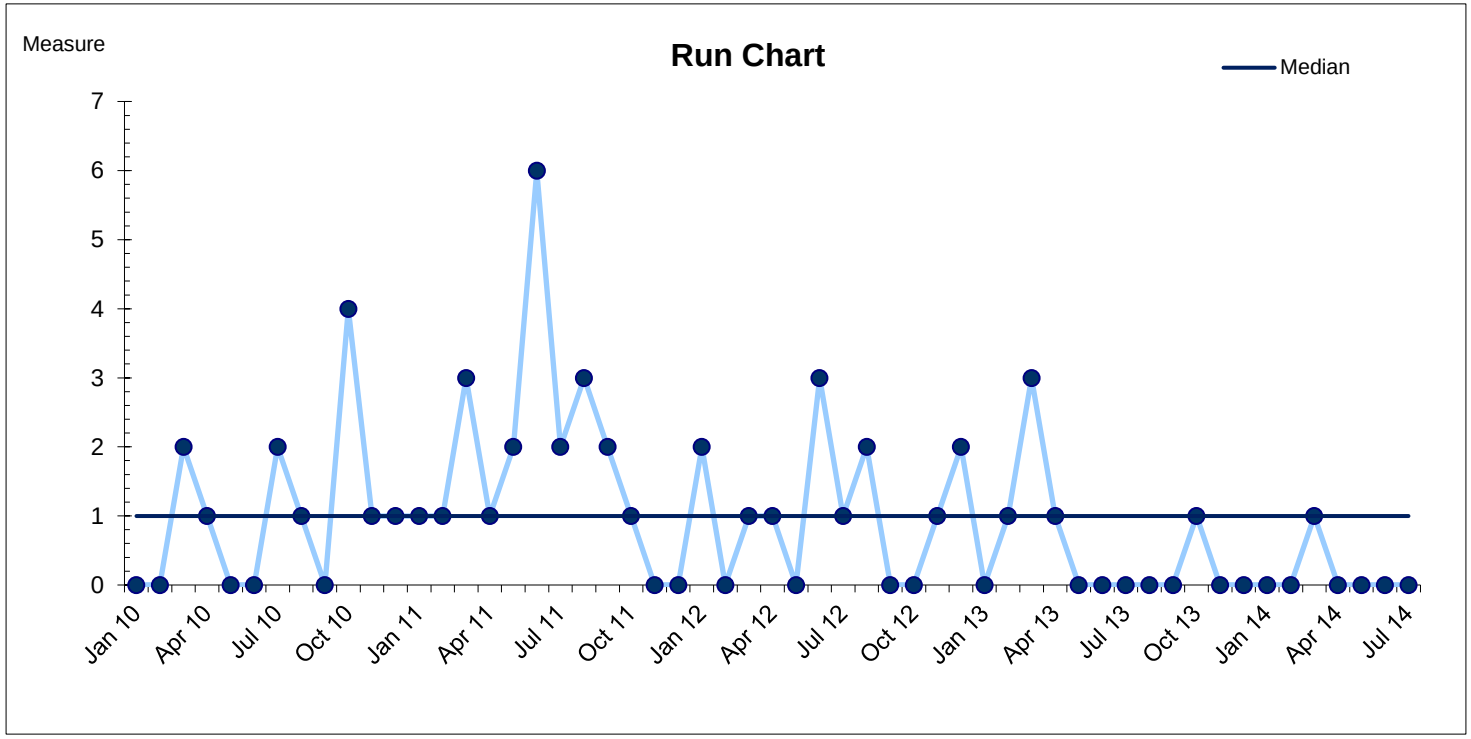
Difference Between a Run Chart and a SPC Chart

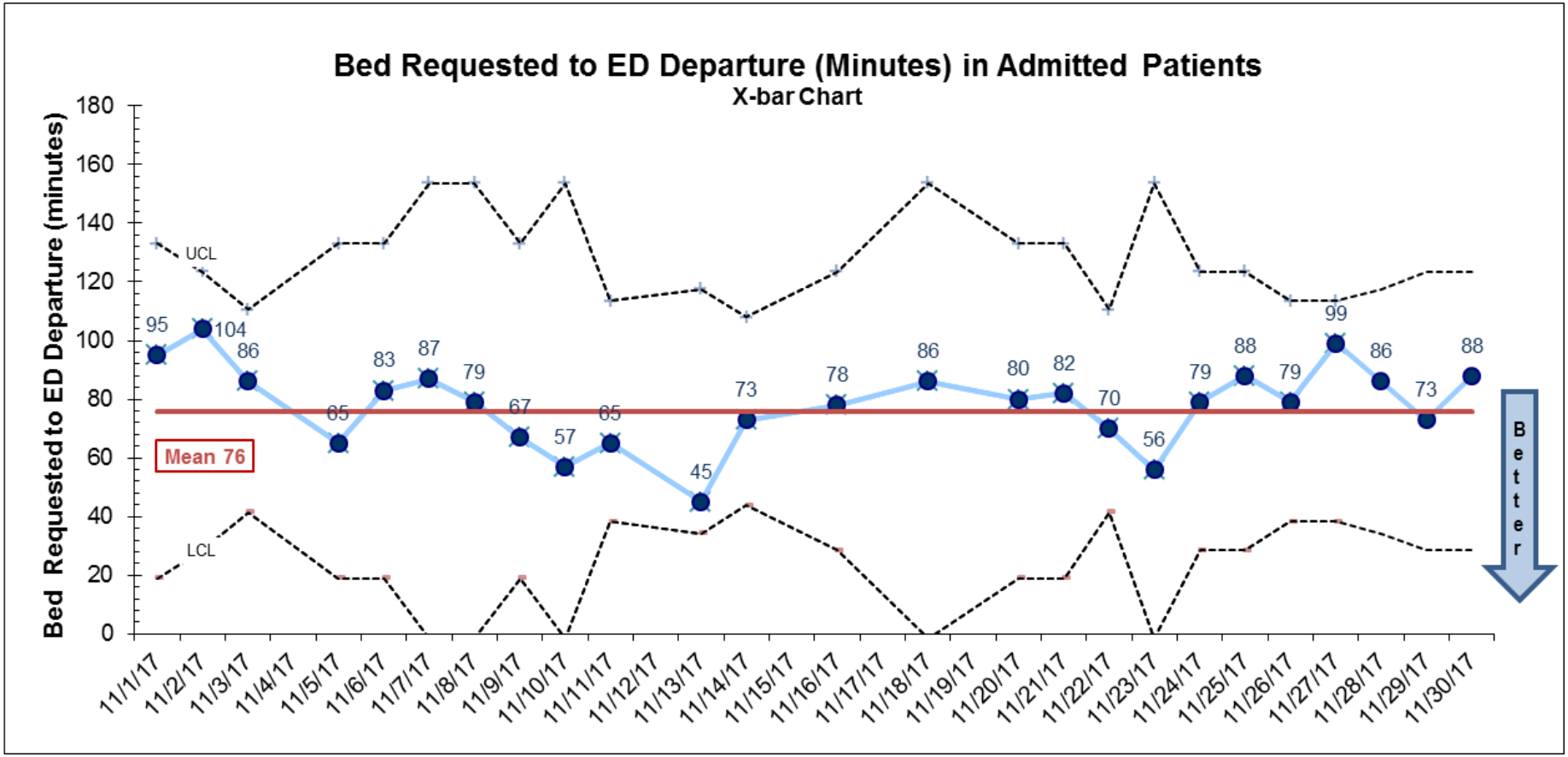
Run Chart

- Simple
- Measure of central tendency is median
- No measurement of variation
- First blush look at the data
- Investigative tool
- Different rules for special cause

SPC Chart

- More complex
- Measure of central tendency is mean
- Upper and lower control limits
- Provides more insight
- Use real time to compare to historical data
- Compare process before and after intervention


Measure
Run Chart
Median
7
6
5
4
3
2
1
0
Jan 10
Apr 10
Jul 10
Oct 10
Jan 11
Apr 11
Jul 11
Oct 11
Jan 12
Apr 12
Jul 12
Oct 12
Jan 13
Apr 13
Jul 13
Oct 13
Jan 14
Apr 14
Jul 14


Bed Requested to ED Departure (Minutes) in Admitted Patients
X-bar Chart
Bed Requested to ED Departure (minutes)
180
160
140
120
100
80
60
40
20
0
UCL
LCL
Mean 76
95
104
86
65
83
87
79
67
57
65
45
73
78
86
80
82
70
56
79
88
79
99
86
73
88
Better
11/1/17
11/2/17
11/3/17
11/4/17
11/5/17
11/6/17
11/7/17
11/8/17
11/9/17
11/10/17
11/11/17
11/12/17
11/13/17
11/14/17
11/15/17
11/16/17
11/17/17
11/18/17
11/19/17
11/20/17
11/21/17
11/22/17
11/23/17
11/24/17
11/25/17
11/26/17
11/27/17
11/28/17
11/29/17
11/30/17

How to Measure the Impact of an Intervention

- Should have 10-20 data points suggesting your process is in control
- Looking for special cause
  - Unusual, not previously observed, non-quantifiable variation
  - If special cause is present, you may reset the mean central line
- Upper and lower control limits play a role in determining special cause

Rules for Determining Special Cause

- Single point outside of the control limit
- Eight of more consecutive point above or below the centerline
- Six consecutive points increasing (trend up) or decreasing (trend down)
- Two out of three consecutive points near a control limit (outer on-third)
- Fifteen consecutive points close to the centerline (inner one-third)

GOOD LUCK!

Board Review Course Stats Questions

Question 1:

You are studying prolactin levels before and after weight loss in children with optic nerve dysplasia. Differences in prolactin levels in your study follow a normal distribution. Your primary outcome should be analyzed with which statistical test?

A. Independent or Student's t-test

B. Wilcoxon Signed Rank Test for Paired Samples

C. Mann Whitney U test

D. Chi Square test

E. Paired t test

ANSWER:

1. You are studying prolactin levels before and after weight loss in children with optic nerve dysplasia. Differences in prolactin levels follow a normal distribution. Your primary outcome should be analyzed with which statistical test?

A. Independent or Student's t-test (continuous or interval variable, independent, parametric, 2 different groups)

B. Wilcoxon Signed Rank Test for Paired Samples (paired data, continuous or interval variable, non-parametric, 2 groups)

C. Mann Whitney U test (continuous or interval variable, independent, non-parametric, 2 groups)

D. Chi Square test (nominal variable, cell size > 5)

E. Paired t test (paired data, continuous or interval variable, parametric, 2 groups).

Correct answer E

In this example you are comparing a normally-distributed, continuous variable (prolactin), so you need a parametric test that is used to compare a continuous variable between 2 groups. The variable is being compared in the same individuals, before and after an intervention, so you need a paired test. Thus, the paired t-test is the correct answer.

Question 2:

Which of the following is true of a power calculation?

A. As power decreases, the chances of a Type II error decreases.
B. The power of a study is usually set at 95%.
C. The power of a study is directly linked to the probability of making a Type I error.
D. Power analysis may be done a priori or post-hoc
E. The size of the difference to be detected can usually be determined by the statistician.

ANSWER:

2. Which of the following is true of a power calculation?

A. As power decreases, the chance of a Type II error decreases. (As power decreases, the chance of Type II error increases)

B. The power of a study is usually set at 95%. (Power is usually set at 80%.)

C. The power of a study is directly linked to the probability of making a Type I error. (Power of a study is directly linked to the probability of making a type II error)

D. Power analysis may be done a priori or post-hoc (this is true)

E. The size of the difference to be detected can usually be determined by the statistician. (The size of the difference to be detected by the investigator – it is a clinical decision based on prior literature or pilot studies)

Correct answer D. The power analysis can be done a priori (ahead of time) or post-hoc (after the study is done). However, if it is done post-hoc, and power is inadequate, you may have increased type II error.

Question 3:

You are reading the results of a recent clinical trial of growth hormone therapy in children between 1 and 3 years of age who are below the 5th percentile for height at enrollment. At the end of the 5 year treatment period, the treatment group mean change in height is 10.1 cm (sd 6 cm) and the mean change in the height in the placebo group is 9.7 cm (sd 5.9 cm), $p < 0.0001$. Which of the following most accurately describes the results?

A. Results are statistically significant and clinically relevant.
B. Results are statistically significant but not clinically relevant.
C. Results are not statistically significant but are clinically relevant.
D. Results are not statistically significant or clinically relevant.

ANSWER:

3. You are reading the results of a recent clinical trial of growth hormone therapy in children between 1 and 3 years of age who are below the 5th percentile for height at enrollment. At the end of the 5 year treatment period, the treatment group mean change in height is 10.1 cm (sd 6 cm) and the mean change in the height in the placebo group is 9.7 cm (sd 5.9 cm), $p < 0.0001$. Which of the following most accurately describes the results?

A. Results are statistically significant and clinically relevant.

B. Results are statistically significant but not clinically relevant. (this is true)

C. Results are not statistically significant but are clinically relevant.

D. Results are not statistically significant or clinically relevant.

Correct answer B

The p-value is <0.05, so it is statistically significant, but most clinicians would agree that 0.4 cm difference in the means is not clinically significant.

Question 4:

You conduct a 2-year, prospective, randomized, double-blinded, placebo-controlled clinical trial of a new medication X for youth-onset type 2 diabetes. By the end of the 2 year trial, out of the 300 randomized participants (150 per arm), 14 in the medication X arm say they didn't take the medication and 10 are lost to follow up, and 12 in the placebo arm say they didn't take their treatment and 15 were lost to follow-up. You now want to analyze the data in the most scientifically rigorous way. What should you do?

A. Compare the 126 remaining in the medication X arm to the 123 in the placebo arm that took their treatment and stayed in the trial, because then you will be comparing those that actually took the treatment they were supposed to take for the length of the trial.
B. Compare 140 in the medication X arm to 135 in the placebo arm, since these are the participants who remained in the study until the end.
C. Compare all of the 150 assigned to the medication X arm to all of the 150 assigned to the placebo arm.
D. Compare 136 in the medication X arm to the 138 in the placebo arm, who actually took the assigned treatments.

Question 5:

In Question 4, what is the most scientifically-sound analysis method called, and why is it considered the gold standard?

A. “As Treated” analysis because it represents what actually happened.
B. “Intention to Treat” analysis because it minimizes bias.
C. “As Treated” analysis because it minimizes bias.
D. “Intention to Treat” analysis because it preserves blinding.

ANSWERS to 4 and 5:

Question 4:

You conduct a 2-year, prospective, randomized, double-blinded, placebo-controlled clinical trial of a new medication X for youth-onset type 2 diabetes. By the end of the 2 year trial, out of the 300 randomized participants (150 per arm), 14 in the medication X arm say they didn't take the medication and 10 are lost to follow up, and 12 in the placebo arm say they didn't take their treatment and 15 were lost to follow-up. You now want to analyze the data in the most scientifically rigorous way. What should you do?

A. Compare the 126 remaining in the medication X arm to the 123 in the placebo arm that took their treatment and stayed in the trial, because then you will be comparing those that actually took the treatment they were supposed to take for the length of the trial.

B. Compare 140 in the medication X arm to 135 in the placebo arm, since these are the participants who remained in the study until the end.

C. Compare all of the 150 assigned to the medication X arm to all of the 150 assigned to the placebo arm. (This is correct)

D. Compare 136 in the medication X arm to the 138 in the placebo arm, who actually took the assigned treatments.

Question 5:

In Question 4, what is the most scientifically-sound analysis method called, and why is it considered the gold standard?

A. "As Treated" analysis because it represents what actually happened.

B. "Intention to Treat" analysis because it minimizes bias. (This is true.)

C. "As Treated" analysis because it minimizes bias.

D. "Intention to Treat" analysis because it preserves blinding.

Correct answers 4. C, 5. B.

Intention to Treat Analysis is considered the gold standard for analyzing prospective randomized treatment trials. In an intention to treat analysis, all participants are included and analyzed in the groups that they were originally assigned to, no matter what treatment they actually received or for how long. The idea is that using an intention to treat analysis preserves the benefits of randomization and avoids bias. For example, if someone is randomized to take the true medication but stops taking the medication because it didn't work, and you do not include them in the medication group in the analysis, the medication could look more effective than it actually was (because the denominator will be smaller). The opposite is a "per protocol" or "as treated" analysis. Sometimes a paper will report the results of an "intention to treat" analysis AND an "as treated" analysis, but the "intention to treat" is the gold standard.