

Statistics

Jason Ryan, MD, MPH



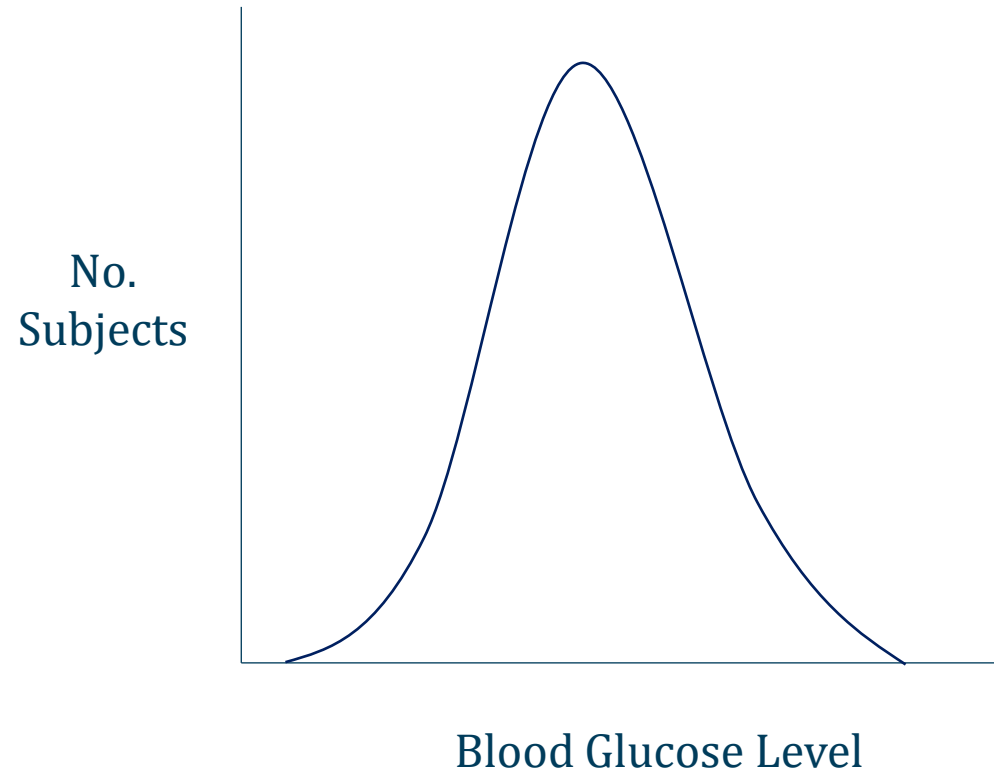
Statistical Distribution

Random Blood Glucose Healthy Subjects

90	115	90	115	90	115	90	115
87	112	87	112	87	112	87	112
101	101	101	101	101	101	101	101
110	92	110	92	110	92	110	92
105	85	105	85	105	85	105	85
93	79	93	79	93	79	93	79
92	100	92	100	92	100	92	100
95	99	95	99	95	99	95	99
88	86	88	86	88	86	88	86
112	102	112	102	112	102	112	102

Statistical Distribution

Normal or Gaussian Distribution



Central Tendency

- Center of normal distribution
- Three ways to characterize:
 - Mean: average of all numbers
 - Median: middle number of data set when all lined up in order
 - Mode: most commonly found number



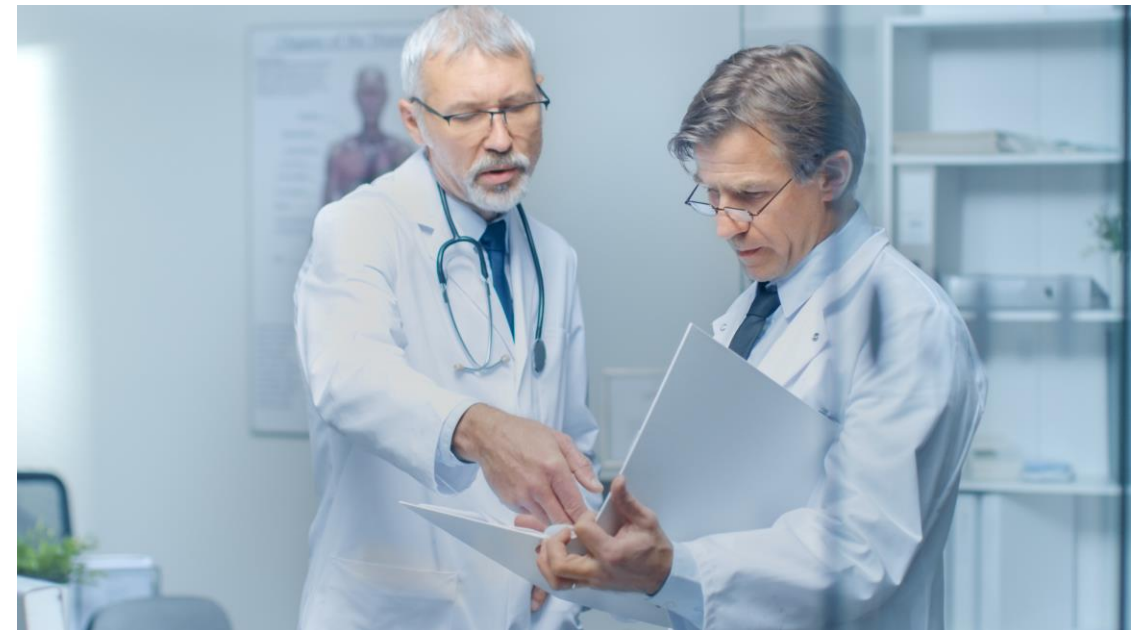
Mean and Mode

- Six blood pressure readings:
 - 90, 80, 80, 100, 110, 120
- Mean = $(90+80+80+100+110+120)/6 = 96.7$
- Mode is most frequent number = 80

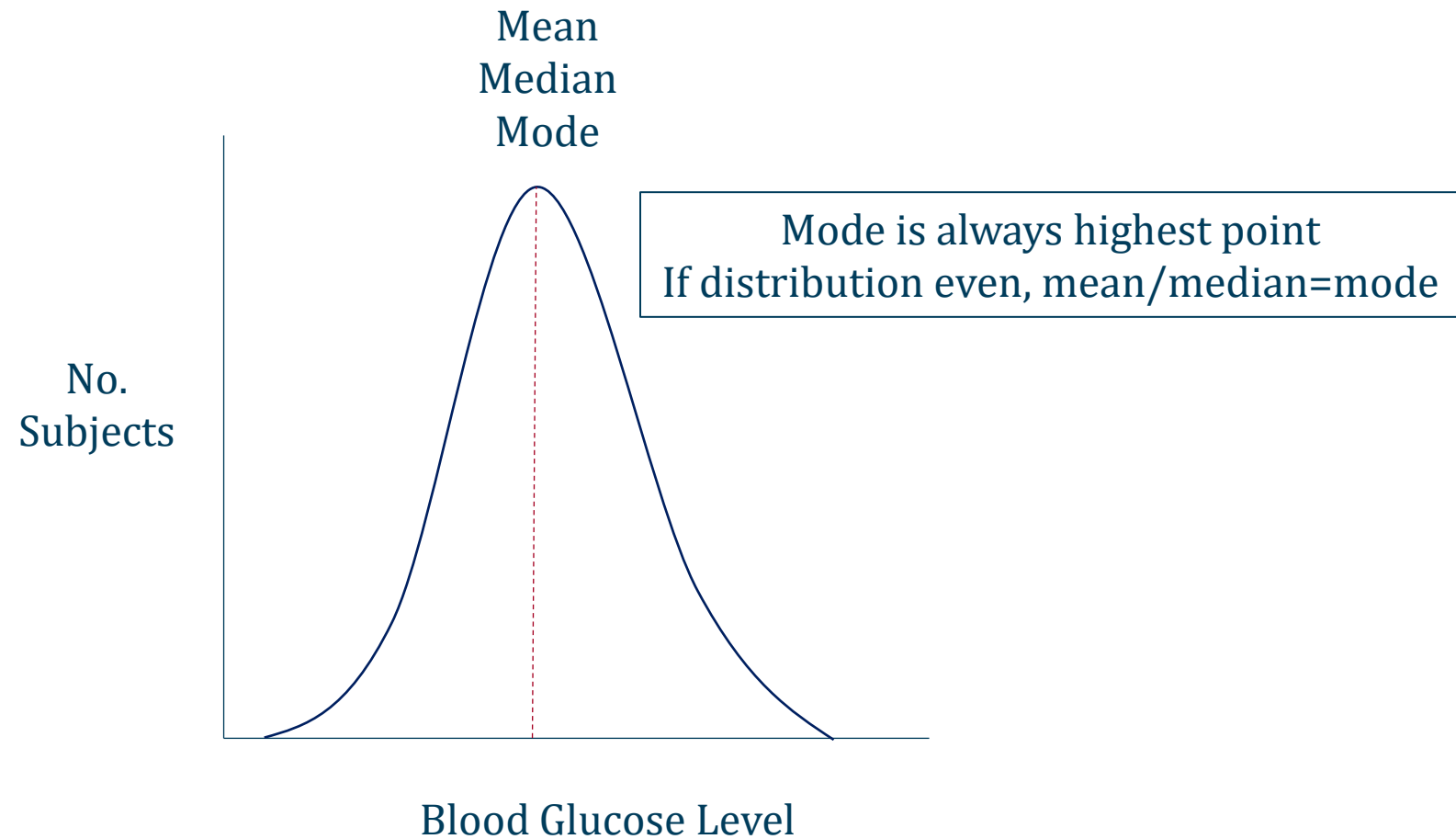


Median

- Odd number of data elements in set
 - 80-90-110
 - Middle number is median = 90
- Even number of data elements
 - 80-90-110-120
 - Halfway between middle pair is median = 100
- Must put data set in order to find median



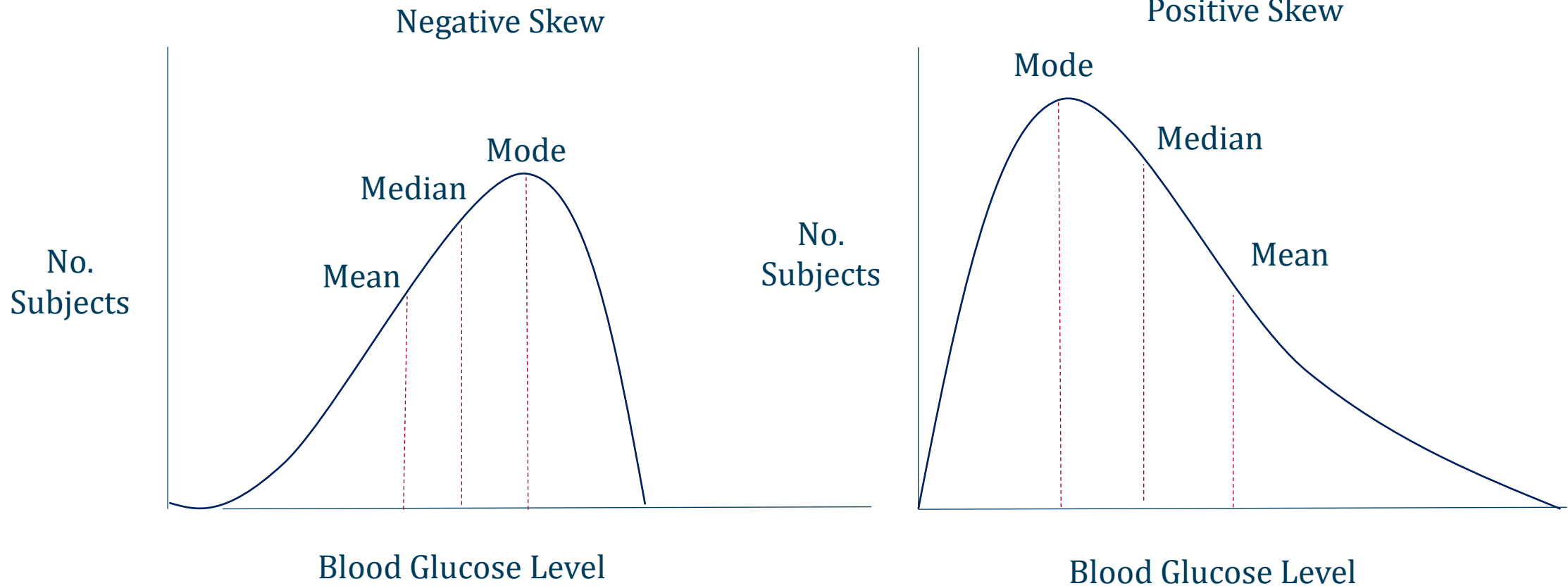
Central Tendency



Central Tendency

Skewness

- Positive or negative based on location of tail



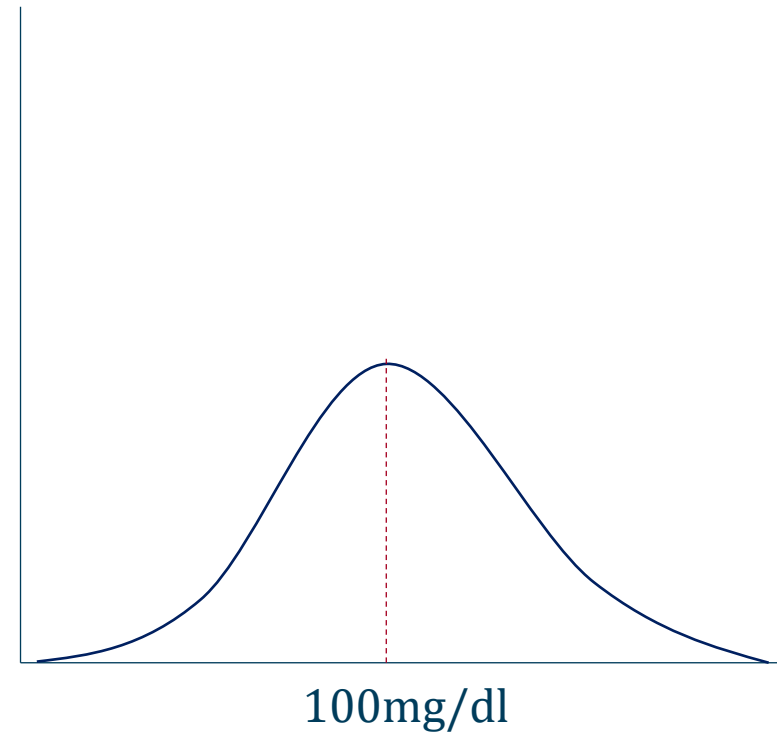
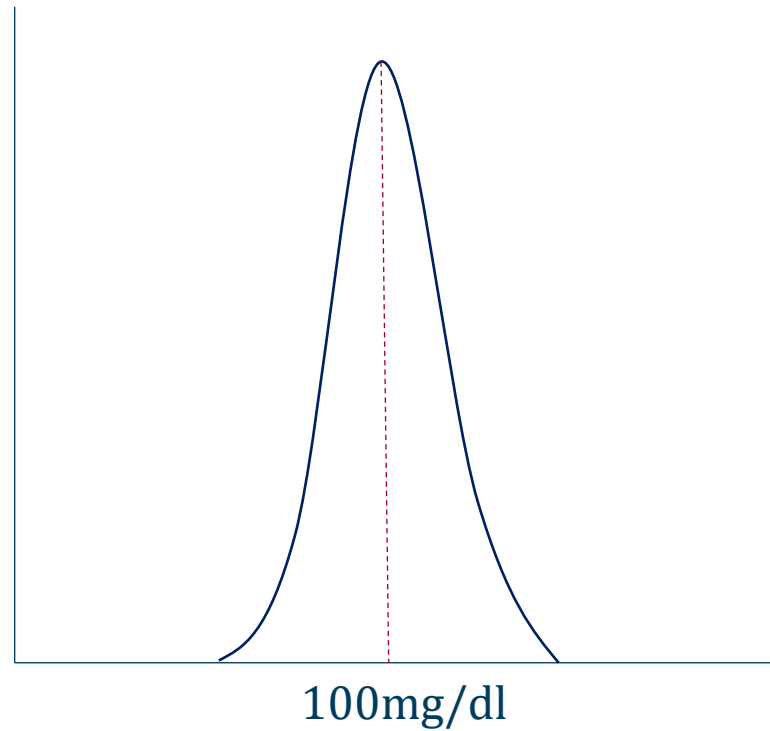
Central Tendency

Key Points

- If distribution is equal: $\text{mean} = \text{mode} = \text{median}$
- Mode is always at peak
- In skewed data:
 - Mean is always furthest away from mode toward tail
 - Median is between Mean/Mode
- Mode is least likely to be affected by outliers
 - Adding one outlier changes mean, median
 - Only affects mode if it changes most common number
 - One outlier unlikely to change most common number

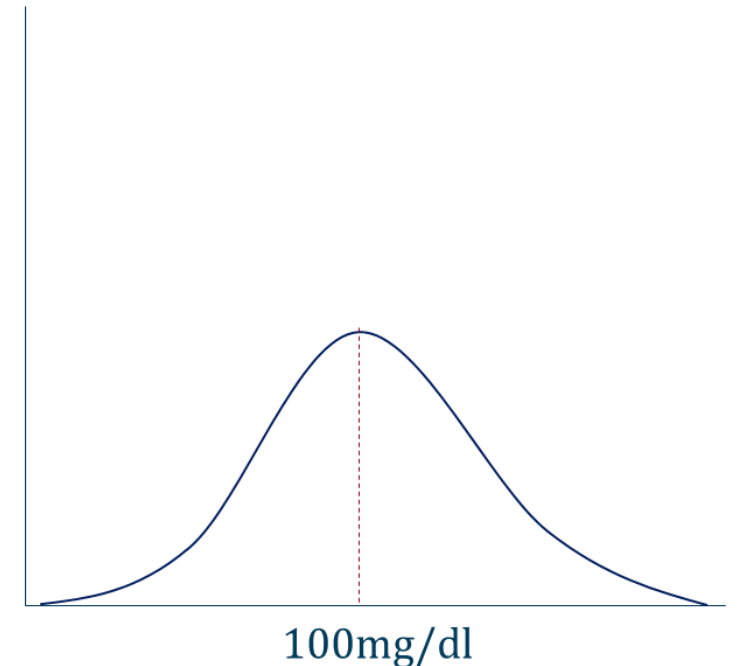
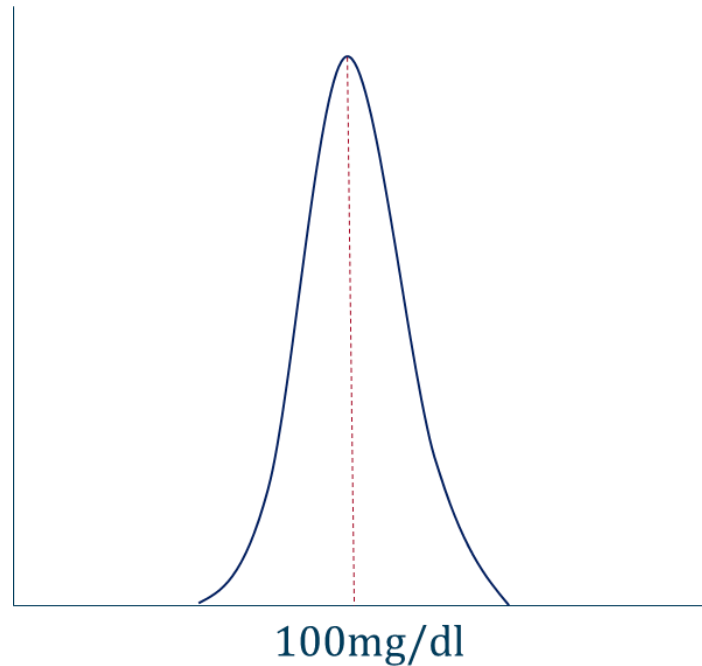


Dispersion



Dispersion Measures

- Used to describe **dispersion of data** in a **data set**
- Standard deviation
- Variance
- Z-score



Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$x-\bar{x}$ = difference between data point and mean

$\Sigma(x-\bar{x})$ = sum of differences

$\Sigma(x-\bar{x})^2$ = sum of differences squared

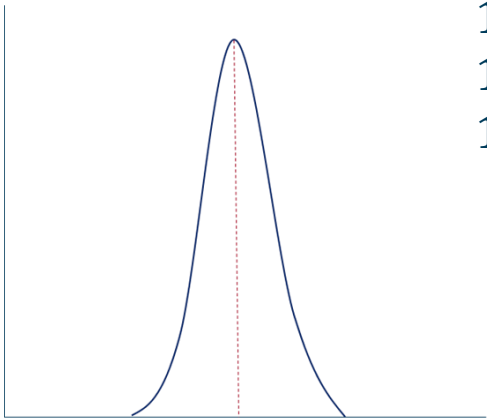
n = number of samples

Standard Deviation

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

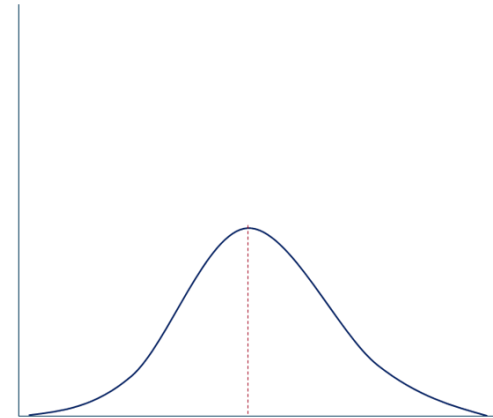
<u>Group 1</u> <u>(mean=10)</u>	<u>Difference</u> <u>from mean</u>	<u>Squared</u>
9	-1	1
8	-2	4
9	-1	1
10	0	0
11	1	1
12	2	4
10	0	0
10	0	0
		<hr/>
		11

$$\sigma = \sqrt{\frac{11}{7}} = 1.24$$



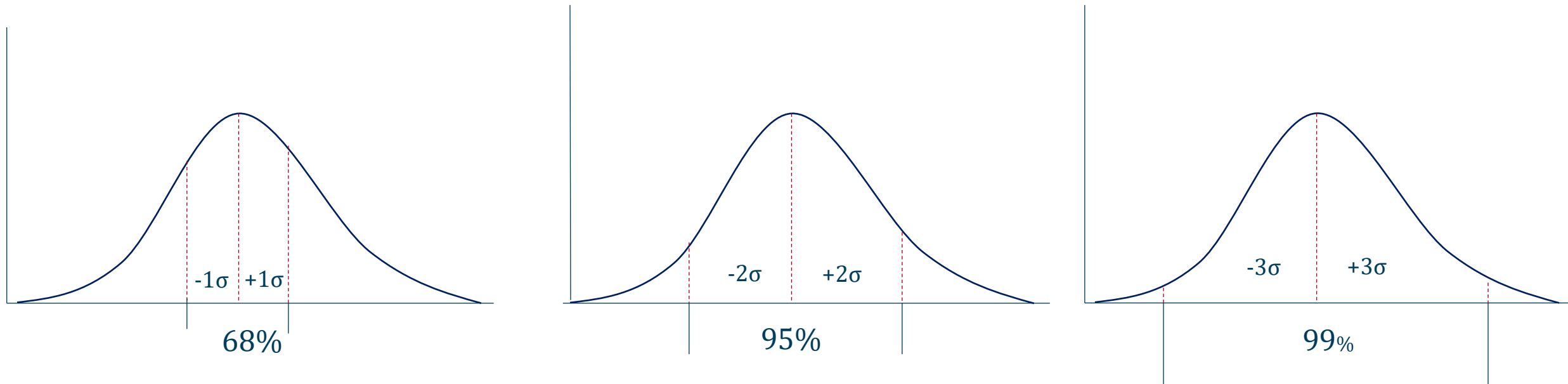
<u>Group 2</u> <u>(mean=10)</u>	<u>Difference</u> <u>from mean</u>	<u>Squared</u>
5	-5	25
6	-4	16
9	-1	1
10	0	0
12	2	4
13	3	9
15	5	25
14	4	16
		<hr/>
		96

$$\sigma = \sqrt{\frac{96}{7}} = 3.7$$

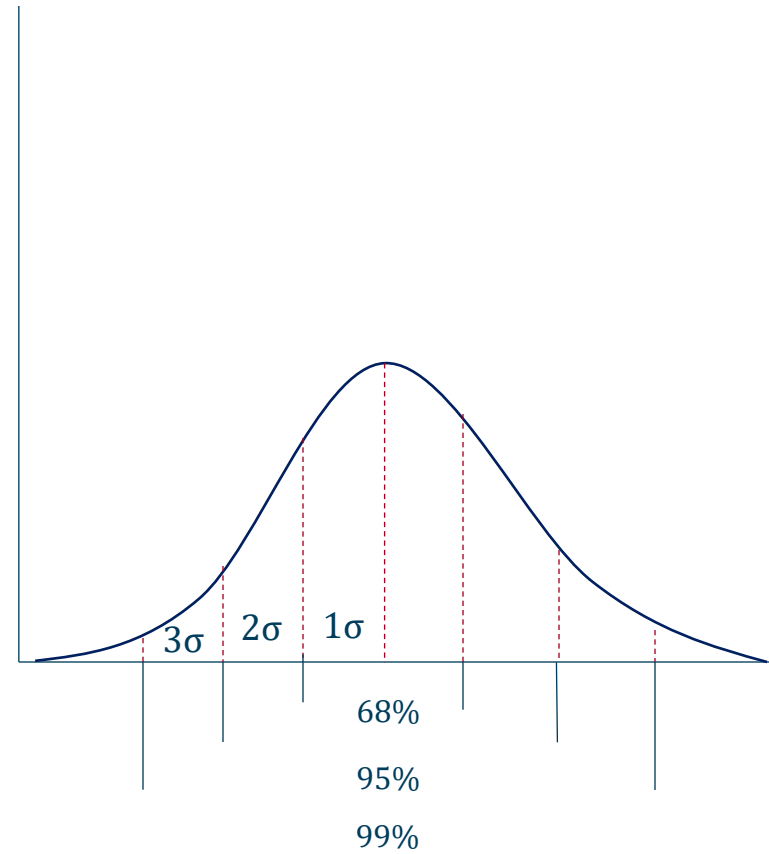
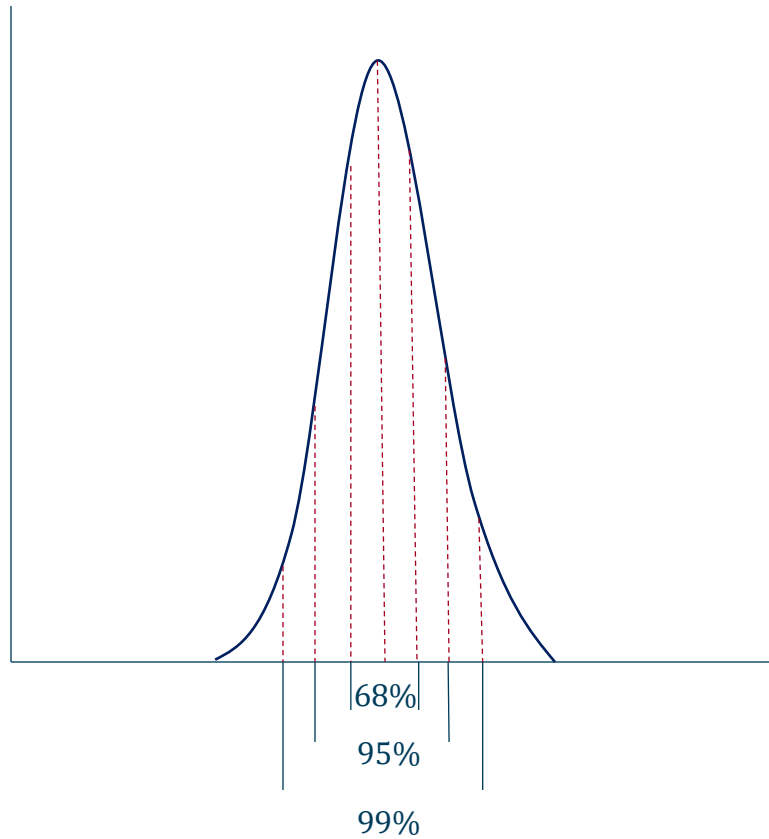


Standard Deviation

- Percentage within standard deviation

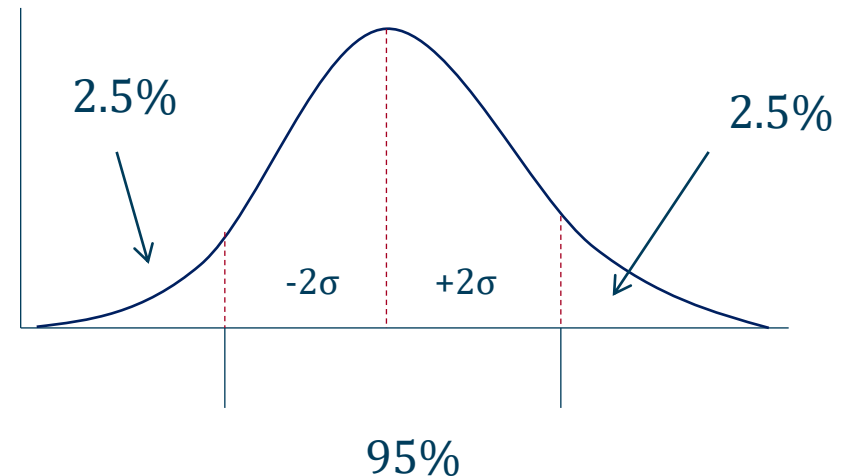


Standard Deviation



Sample Question

- Fasting glucose is measured in 200 medical students
- The mean value is 100 mg/dL with a standard deviation of 5 mg/dL
- The values are normally distributed
- How many students have a glucose > 110 mg/dL?
 - 110 is two standard deviations away from mean
 - 2.5% of students are in this range (1/2 of 5%)
 - 2.5% of 200 = 5 students



Variance

- Measure of dispersion in a data set
- Related to standard deviation
- Average degree to which each point differs from the mean

Standard Deviation

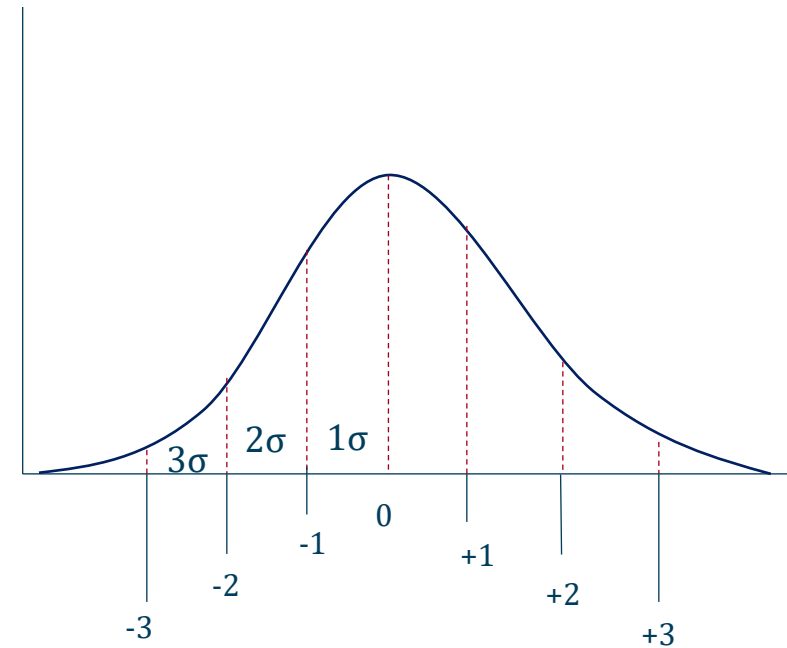
$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

Variance

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

Z score

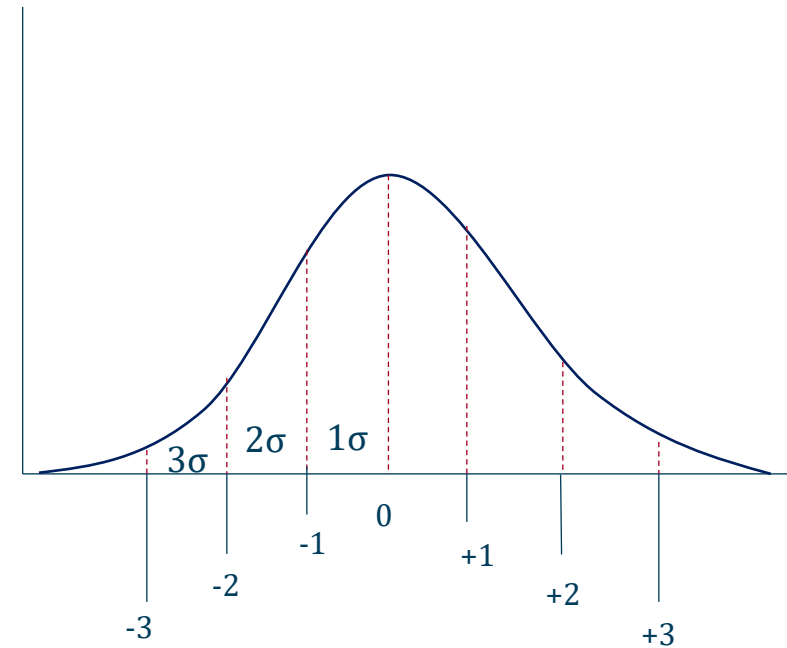
- Describes a **single data point**
- How far a data point is from the mean
- Z score of 0 is the mean
- Z score of +1 is 1SD above mean
- Z score of -1 is 1SD below mean



Z score

Example

- Suppose test grade average (mean) = 79
- Standard deviation = 5
- Your grade = 89
- Your Z score = $(89-79)/5 = +2$



True Population Mean

- Suppose 100 samples taken to estimate population mean glucose level
- Total population = 1M individuals
- How close is sample mean to true population mean?
- **Standard error of the mean**
- **Confidence intervals**

Standard Error of the Mean

- How far is **dataset mean** from **true population mean**
- Standard deviation divided by square root of n
- Small standard deviation → less SEM → closer to true mean
- More samples (n) → less SEM → closer to true mean

$$SEM = \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

- Range in which 95% of repeated measurements would be expected to fall
- 95% chance **true population mean** falls within this range
- Related to standard error of the mean (SEM)

$$CI_{95\%} = \text{Mean} \pm 1.96 * (\text{SEM})$$

Confidence Intervals

Example

- $n = 16$
- Mean = 10
- SD = 4
- $SEM = 4/\sqrt{16} = 4/4 = 1$
- $CI = 10 \pm 1.96*(1) = 10 \pm 2$
- 95% of repeated means fall between 8 and 12
- Upper confidence limit = 12
- Lower confidence limit = 8

$$CI_{95\%} = \text{Mean} \pm 1.96*(SEM)$$

Confidence Intervals

- Don't confuse standard deviation with confidence intervals
- Standard deviation is for a **dataset**
 - Suppose we have ten samples
 - These samples have a mean and standard deviation
 - 95% of *samples* fall between $\pm 2SD$
 - This is descriptive characteristic of the samples
- Confidence intervals
 - This does not describe the samples in data set
 - An inferred value of where the true mean lies for *population*

95%

- This value often confusing
- Read carefully: What are they asking for?
- Data set or true population mean?
- Range in which 95% of measurements **in a dataset** fall
 - Mean \pm 2SD
- Range in which **true population mean** likely falls?
 - Confidence interval of the mean
 - Mean \pm 1.96*SEM



Hypothesis Testing

Jason Ryan, MD, MPH



Hypothesis Testing

- Critical element of medical research
- Determination of whether results are meaningful
- Are observed study findings real or due to chance?
- Allows extrapolation of medical testing results to **general population**
 - All studies involve a subset of the total population
 - Hypothesis testing: do study findings represent reality in the total population?
- Usually involves comparison between different groups
- Are differences between groups real differences or due to chance?

Hypothesis Testing Example

- **MERIT-HF Trial**
- 3,991 patients with systolic heart failure
- Patients treated with metoprolol or placebo
- Primary endpoint: all-cause mortality
- Placebo group: all-cause mortality 11.0%
- Metoprolol group: all-cause mortality 7.2%
- Are these true differences or due to chance?
- Is metoprolol associated with lower mortality or is this due to chance?
- Hypothesis for testing: metoprolol is associated with lower mortality

Effects of Controlled-Release Metoprolol on Total Mortality, Hospitalizations, and Well-being in Patients With Heart Failure The Metoprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure (MERIT-HF)

Ake Hjalmarson, MD, PhD
 Sidney Goldstein, MD
 Björn Fagerberg, MD, PhD
 Hans Wedel, PhD
 Finn Wängstén, MD, PhD
 John Kjeldsen, MD, PhD
 John Wikstrand, MD, PhD
 Dia El Allaf, MD
 Jari Vitovec, MD, PhD
 Jan Aldershvile, MD, PhD
 Matti Hälönen, MD, PhD
 Rainer Dietz, MD
 Karl-Ludwig Neuhaus, MD
 Andrija Jansz, MD, DSc
 Gudmundur Thorgerisson, MD, PhD
 Peter H. J. M. Dusselmann, MD, PhD
 Lars Gullestad, MD
 Jerzy Kuch, MD
 Johan Herlitz, MD, PhD
 Peter Rickensbacher, MD
 Stephen Ball, MD, PhD
 Stephen Gottlieb, MD
 Prakash Deshmukh, MD
 for the MERIT-HF Study Group

CHRONIC HEART FAILURE IS A common disease that has a poor prognosis and periods of incapacitating symptoms necessitating recurrent hospital admis-

For editorial comment see p 1335.

Context Results from recent studies on the effects of β_1 -blockade in patients with heart failure demonstrated a 34% reduction in total mortality. However, the effect of β_1 -blockade on the frequency of hospitalizations, symptoms, and quality of life in patients with heart failure has not been fully explored.

Objective To examine the effects of the β_1 -blocker controlled-release/extended-release metoprolol succinate (metoprolol CR/XL) on mortality, hospitalization, symptoms, and quality of life in patients with heart failure.

Design Randomized, double-blind controlled trial, preceded by a 2-week single-blind placebo run-in period, conducted from February 14, 1997, to October 31, 1998, with a mean follow-up of 1 year.

Setting Three hundred thirteen sites in 14 countries.

Participants Patients (n = 3991) with chronic heart failure, New York Heart Association (NYHA) functional class II to IV, and ejection fraction of 0.40 or less who were stabilized with optimum standard therapy.

Interventions Patients were randomized to metoprolol CR/XL, 25 mg once per day (NYHA class II), or 12.5 mg once per day (NYHA class III or IV), titrated for 6 to 8 weeks up to a target dosage of 200 mg once per day (n = 1990), or matching placebo (n = 2001).

Main Outcome Measures Total mortality or any hospitalization (time to first event), number of hospitalizations for worsening heart failure, and change in NYHA class, by intervention group; quality of life was assessed in a substudy of 741 patients.

Results The incidence of all predefined end points was lower in the metoprolol CR/XL group than in the placebo group, including total mortality or all-cause hospitalizations (the prespecified second primary end point; 641 vs 767 events; risk reduction, 19%; 95% confidence interval [CI], 10%-27%; P < .001); total mortality or hospitalizations due to worsening heart failure (311 vs 439 events; risk reduction, 31%; 95% CI, 20%-40%; P < .001); number of hospitalizations due to worsening heart failure (317 vs 451; P < .001); and number of days in hospital due to worsening heart failure (3401 vs 5303 days; P < .001).

NYHA functional class, assessed by physicians, and McMaster Overall Treatment Evaluation score, assessed by patients, both improved in the metoprolol CR/XL group compared with the placebo group (P = .003 and P = .009, respectively).

Conclusions In this study of patients with symptomatic heart failure, metoprolol CR/XL improved survival, reduced the need for hospitalizations due to worsening heart failure, improved NYHA functional class, and had beneficial effects on patient well-being. JAMA. 2000;283:1295-1302.

www.jama.com

Author Affiliations and Financial Disclosures are listed at the end of this article.
Corresponding Author Reprints: Björn Fagerberg, MD, PhD, Wallenberg Laboratory for Cardiovascular Research, Sahlgrenska University Hospital, SE-413 45 Gothenburg, Sweden (e-mail: bjorn.fagerberg@medfak.gu.se).

©2000 American Medical Association. All rights reserved.

JAMA, March 8, 2000—Vol 283, No. 10 1295

Hypothesis Testing

Example

- **MIzyme protein level** may be elevated in myocardial infarction
- Can this level be used to detect myocardial infarction in ED?
- Samples of MIzyme level obtain in 100 normal subjects
- Samples of MIzyme level obtain in 100 subjects with myocardial infarction
- Mean level in normal subjects: 1 mg/dl
- Mean level in myocardial infarction patients: 10 mg/dl
- Can this test be used to detect myocardial infarction in the general population?

Hypothesis Testing

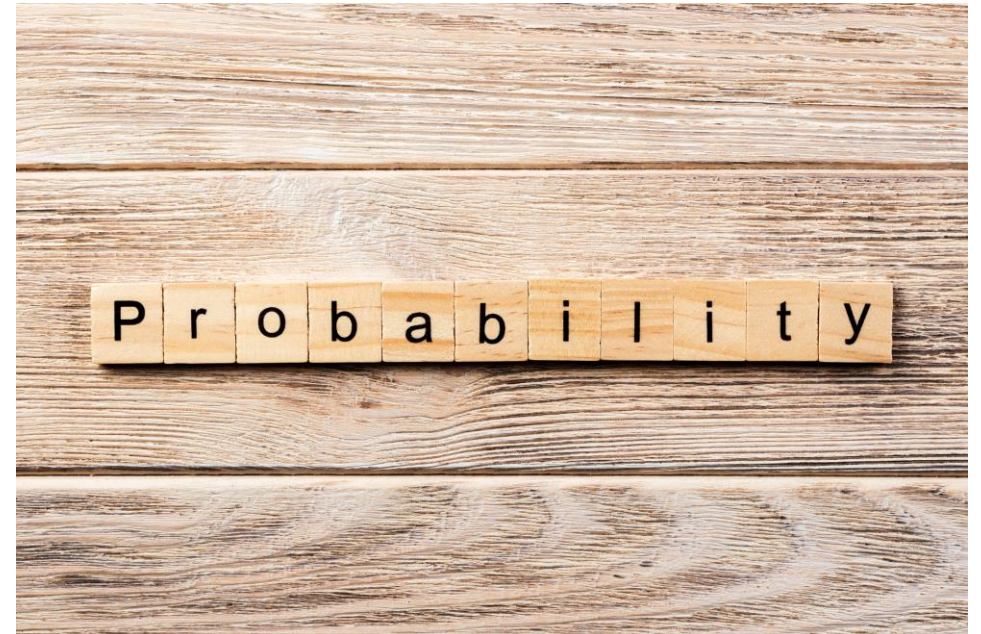
Example

- Is the mean value of MIzyme in myocardial infraction subjects truly different?
- Or was the difference in our experiment simply due to chance?
- Hypothesis for testing: MIzyme level is higher in myocardial infarction patients



Hypothesis Testing

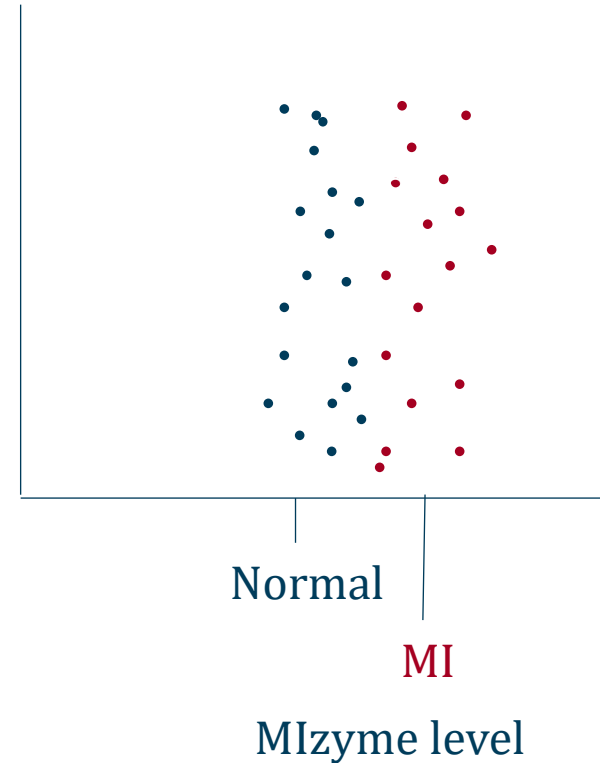
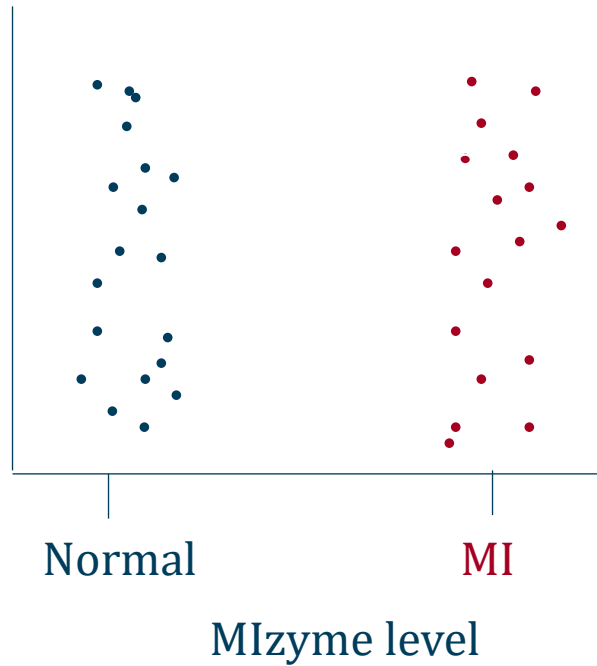
- Mathematically calculates **probabilities** (5%, 50%)
- Probability the two means truly different
- Probability the difference is due to chance in our experiment
- Math is complex (don't need to know)
- Probabilities by hypothesis testing depend on:
 - Difference between means normal/MI
 - Scatter of data
 - Number of subjects tested



Hypothesis Testing

Size of difference between groups

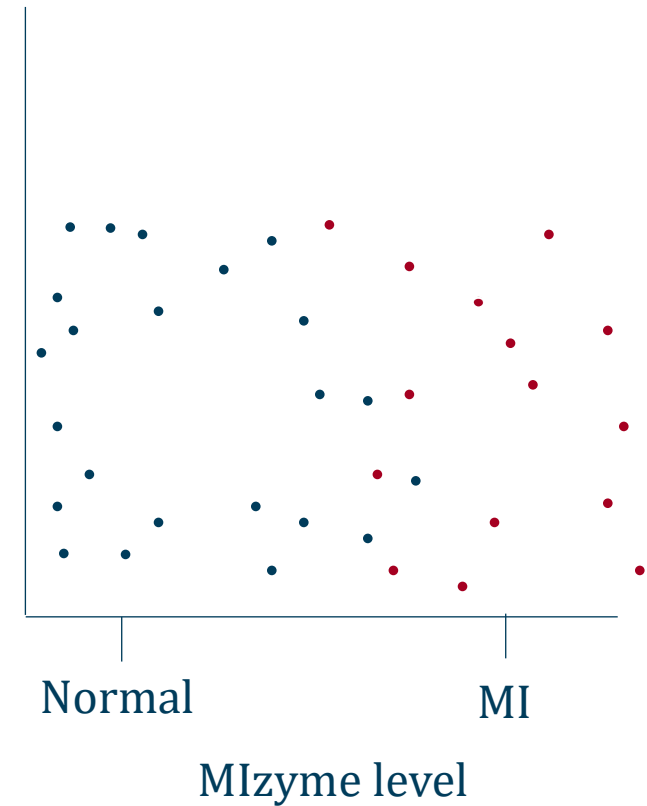
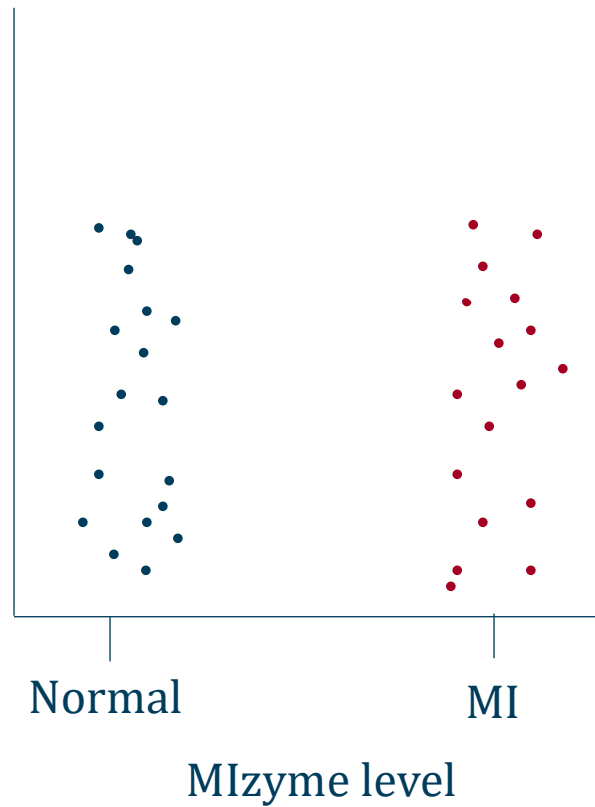
Magnitude of difference in means influences likelihood that difference between means is due to chance



Hypothesis Testing

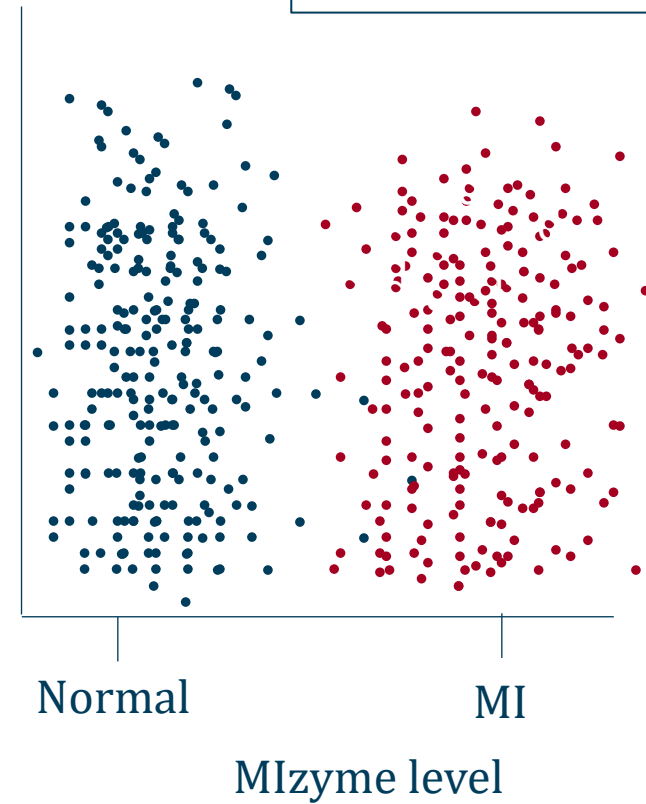
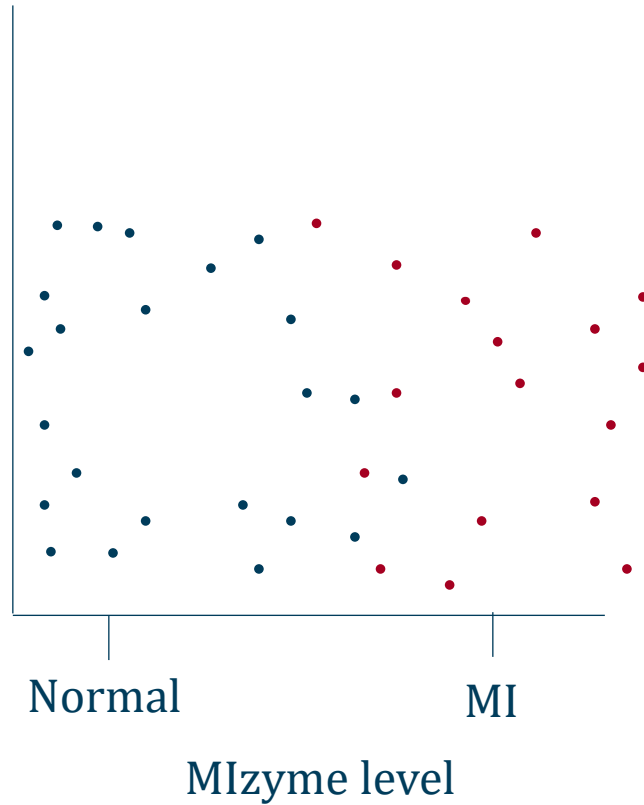
Scatter

Key Point: Scatter of data points influences the likelihood that there is a true difference between means



Hypothesis Testing

Number of samples



Key Point: Number of data points influences the likelihood that there is a true difference between means

Hypothesis Testing

- Two possibilities for MIzyme levels in normal and MI patients
- #1: MIzyme **DOES NOT** distinguish between normal/MI
 - Levels are the same in both groups
- #2: MIzyme **DOES** distinguish between normal/MI
 - Levels are not the same in both groups
- Null hypothesis (H_0) = #1
- Alternative hypothesis (H_1) = #2

H_0

Hypothesis Testing

- In reality, either H_0 or H_1 is correct
- In our experiment, either H_0 or H_1 will be deemed correct
- Hypothesis testing determines likelihood our experiment matches with reality

\mathcal{H}_0 \mathcal{H}_1

Hypothesis Testing

Four possible outcomes

- #1: There is a difference in reality and our experiment detects it
 - The alternative hypothesis (H_1) is found true by our study
- #2: There is no difference in reality and our experiment also finds no difference
 - The null hypothesis (H_0) is found true by our study
- #3: There is no difference in reality but our study finds a difference
 - This is an error! Type 1 (α) error
- #4: There is a difference in reality but our study misses it
 - This is also an error! Type 2 (β) error

Hypothesis Testing

Experiment	Reality	
	H_1	H_0
H_1		
H_0		

% Likelihood based on:
Difference between means normal/MI
Scatter of data
Number of subjects tested

Hypothesis Testing

% Likelihood based on:

Difference between means normal/MI

Scatter of data

Number of subjects tested

Experiment	Reality	
	H_1	H_0
	H_1	H_0
H_1	Power	α
H_0	β	H_0 Correct

Power = Chance of detecting difference

α = Chance of seeing difference that is not real

β = chance of missing a difference that is really there

Power = $1 - \beta$

P-value

- Used to accept or reject a hypothesis
- Calculated based on differences between groups
 - Magnitude of difference between groups
 - Scatter of data
 - Number of samples



Experiment

		Reality	
		H_1	H_0
H_1	Power	α	
H_0	β	H_0 Correct	

P-value

- Represents chance that null hypothesis is correct
 - No difference between means
- If **p < 0.05** we usually reject the null hypothesis
 - Difference in means is “statistically significant”

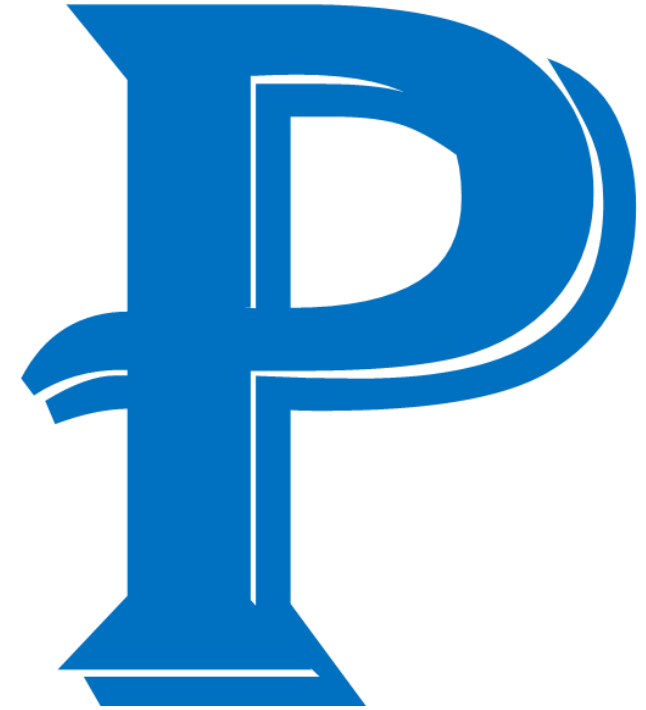
P

Experiment

		Reality	
		H ₁	H ₀
H ₁	Power		α
H ₀	β		H ₀ Correct

P-value

- Represents chance of “false positive” finding
- No difference in reality
- Study finds a difference that is not real
- Must be low (< 0.05) for findings to be positive
- Similar but different from α (significance level)
 - α set by study design
 - P value calculated by comparison of groups



Experiment

		Reality	
		H_1	H_0
H_1	Power	α	
H_0	β	H_0 Correct	

P-value

Example

- **MERIT-HF Trial**
- 3,991 patients with systolic heart failure
- Placebo group: all-cause mortality 11.0%
- Metoprolol group: all-cause mortality 7.2%
- P-value = 0.00009
- Reject null hypothesis that differences are due to chance
- Accept alternative hypothesis that metoprolol reduces mortality
- “Positive study”

Effects of Controlled-Release Metoprolol on Total Mortality, Hospitalizations, and Well-being in Patients With Heart Failure The Metoprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure (MERIT-HF)

Ake Hjalmarson, MD, PhD
Sidney Goldstein, MD
Björn Fagertberg, MD, PhD
Hans Wedel, PhD
Finn Waagstein, MD, PhD
John Kjekshus, MD, PhD
John Wikstrand, MD, PhD
Dia El Alhal, MD
Jiri Vitovec, MD, PhD
Jan Aldersville, MD, PhD
Matti Hjalinen, MD, PhD
Rainer Dietz, MD
Karl-Ludwig Neuhaus, MD
Andrie Jansz, MD, DSc
Günther Thorgeirsson, MD, PhD
Peter H. J. M. Dusselmann, MD, PhD
Lars Gullestad, MD
Jerzy Kuch, MD
Johan Herlitz, MD, PhD
Peter Rickenbacher, MD
Stephen Ball, MD, PhD
Stephen Gottlieb, MD
Prakash Deedwania, MD
for the MERIT-HF Study Group

CHRONIC HEART FAILURE IS A common disease that has a poor prognosis and periods of incapacitating symptoms necessitating recurrent hospital admissions.^{1,2}

For editorial comment see p 1335.

Context Results from recent studies on the effects of β_1 -blockade in patients with heart failure demonstrated a 34% reduction in total mortality. However, the effect of β_1 -blockade on the frequency of hospitalizations, symptoms, and quality of life in patients with heart failure has not been fully explored.

Objective To examine the effects of the β_1 -blocker controlled-release/extended-release metoprolol succinate (metoprolol CR/XL) on mortality, hospitalization, symptoms, and quality of life in patients with heart failure.

Design Randomized, double-blind controlled trial, preceded by a 2-week single-blind placebo run-in period, conducted from February 14, 1997, to October 31, 1998, with a mean follow-up of 1 year.

Setting Three hundred thirteen sites in 14 countries.

Participants Patients ($n = 3991$) with chronic heart failure, New York Heart Association (NYHA) functional class II to IV, and ejection fraction of 0.40 or less who were stabilized with optimum standard therapy.

Interventions Patients were randomized to metoprolol CR/XL, 25 mg once per day (NYHA class II), or 12.5 mg once per day (NYHA class III or IV), titrated for 6 to 8 weeks up to a target dosage of 200 mg once per day ($n = 1995$), or matching placebo ($n = 2001$).

Main Outcome Measures Total mortality or any hospitalization (time to first event), number of hospitalizations for worsening heart failure, and change in NYHA class, by intervention group; quality of life was assessed in a substudy of 741 patients.

Results The incidence of all predefined end points was lower in the metoprolol CR/XL group than in the placebo group, including total mortality or all-cause hospitalizations (the prespecified second primary end point; 641 vs 767 events; risk reduction, 19%; 95% confidence interval [CI], 10%-27%, $P < .001$); total mortality or hospitalizations due to worsening heart failure (311 vs 439 events; risk reduction, 31%; 95% CI, 20%-40%, $P < .001$); number of hospitalizations due to worsening heart failure (317 vs 451; $P < .001$); and number of days in hospital due to worsening heart failure (3401 vs 5303 days; $P < .001$). NYHA functional class, assessed by physician, and McMaster Overall Treatment Evaluation score, assessed by patients, both improved in the metoprolol CR/XL group compared with the placebo group ($P = .003$ and $P = .009$, respectively).

Conclusions In this study of patients with symptomatic heart failure, metoprolol CR/XL improved survival, reduced the need for hospitalizations due to worsening heart failure, improved NYHA functional class, and had beneficial effects on patient well-being. JAMA. 2000;283:1295-1302.

Power

- Chance of finding a difference when one exists
- Also called rejecting the null hypothesis (H_0)
- Power is increased when:
 - **Increased sample size**
 - Large difference of means
 - Less scatter of data

Experiment

		Reality	
		H ₁	H ₀
H ₁		Power	α
H ₀		β	H ₀ Correct

Power Calculation

- Studies try to **maximize power** to detect a true difference
- In study design, you have little/no control over:
 - Scatter of data
 - Difference between means
- You DO have control over **number of subjects**
- Number of subjects chosen to give a high power
- Commonly used power goal is 80%
- This is called a **power calculation**

Experiment

		Reality	
		H_1	H_0
H_1		Power	α
H_0		β	H_0 Correct

Detecting a Difference

- Study 1
 - Group A: 10%
 - Group B: 20%
 - P-value: 0.001
 - “Study detected a difference”
- Study 2
 - Group A: 10%
 - Group B: 20%
 - P-value: 0.25
 - “Study did not detect a difference”



Statistical Errors

Type 1 (α) error

- False positive
- Finding a difference or effect when there is none in reality
- Rejecting null hypothesis (H_0) when you should not have
- Example: researchers conclude a drug benefits patients, but it does not

Experiment	Reality	
	H_1	H_0
	H_1	H_0
H_1	Power	α
H_0	β	H_0 Correct

Type 1 Errors

Causes

- Example: researchers conclude a drug benefits patients, but it does not
- **Random chance**
 - Most studies: chance of alpha error 5%
 - One out of 20 times → error
- Improper research techniques



Statistical Errors

Type 2 (β) error

- False negative
- Finding no difference/effect when there is one in reality
- Accepting null hypothesis (H_0) when you should not have
 - Example: Researchers conclude a drug does not benefit patients ($p > 0.05$)
 - Subsequent study finds that it does

		Reality	
		H_1	H_0
Experiment	H_1	Power	α
	H_0	β	H_0 Correct

Statistical Errors

Type 2 (β) error

- Common cause: **too few patients**
- Need lots of patients to have sufficient power
- Especially when differences are small
- Significant p-value (< 0.05) more likely with:
 - Large difference between groups
 - Large number of patients

Experiment

		Reality	
		H_1	H_0
H_1	Power		α
H_0	β		H_0 Correct

Tests of Significance

Jason Ryan, MD, MPH



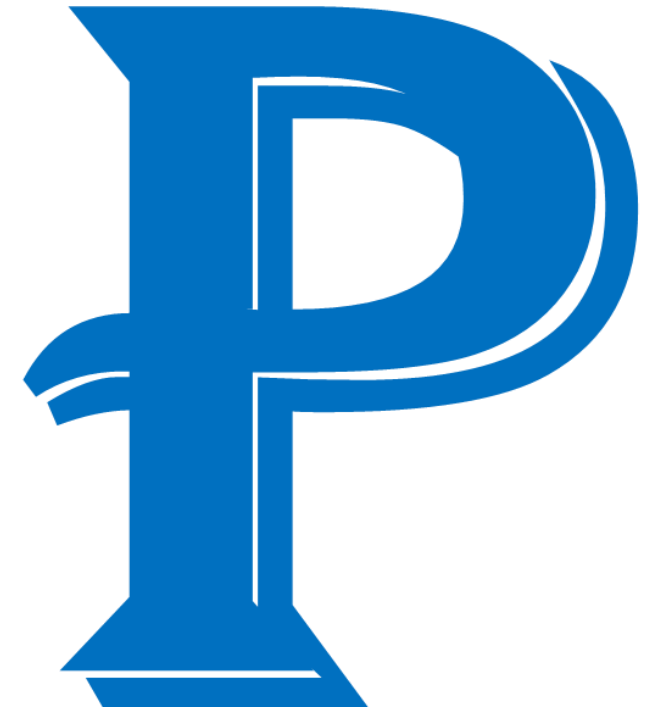
Comparing Groups

- Many clinical studies **compare groups**
- Often find differences between groups
 - Different mean ages
 - Different mean blood levels
- Are differences real or due to chance
- Key element of hypothesis testing
- Accept or reject null hypothesis



P-value

- Calculated based on differences between groups
- Represents chance that null hypothesis is correct
 - No difference between group values
- If **p < 0.05** we usually reject the null hypothesis
 - Reject hypothesis of no difference between groups
 - Accept alternative hypothesis of difference between groups
 - Difference between groups is “statistically significant”



Key Point

- P-value for group comparison calculated based on **group data**
- Magnitude of difference between groups
- Scatter of data points
- Number of data points
- Don't need to know the math
- Just understand principle



Comparing Groups

Determination of P values

- Three key tests
- T-test
- ANOVA
- Chi-square

Data Types

- **Quantitative** data/variables:
 - Have numerical values
 - 1, 2, 3, 4
- **Categorical** data/variables:
 - High, medium, low
 - Positive, negative
 - Yes, No



Data Types

- Quantitative variables often reported as **number**
 - Mean age was 62 years old
- Categorical variables often report as **percentage**
 - 40% of patients take drug A
 - 20% of patients are heavy exercisers



T-test

- Compares two **mean quantitative** values
- Yields a p-value



T-test

- A researcher studies plasma levels of sodium in patients with SIADH and normal patients. The mean value in SIADH patients is 128 mg/dl. The mean value in normal patients is 136 mg/dl.
- Common questions:
 - Which test to compare the means? (t-test)
 - What p-value indicates significance? (< 0.05)



T-test

- A researcher studies plasma levels of sodium in patients with SIADH and normal patients. The mean value in SIADH patients is 128 mg/dl. The mean value in normal patients is 136 mg/dl.
- A p-value of 0.01 is reported. What does this mean?
 - 1% probability that results are due to chance
 - 1% probability that there is no true difference in means



T-test

- A researcher studies plasma levels of sodium in patients with SIADH and normal patients. The mean value in SIADH patients is 128 mg/dl. The mean value in normal patients is 136 mg/dl.
- The p-value is 0.20 (non-significant) - why might that be the case?
 - Need more patients
 - Increase sample size → increase power to detect differences



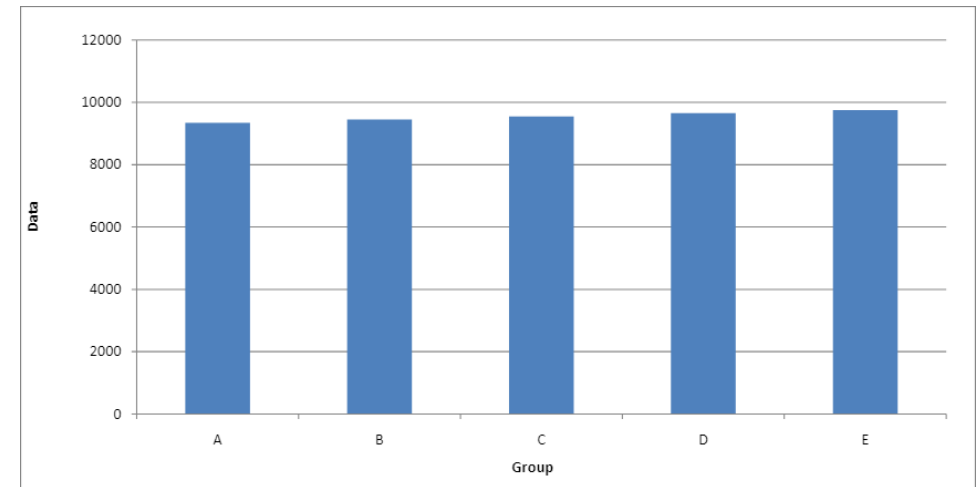
ANOVA

- Analysis of variance
- Used to compare **more than two** quantitative means
- Consider:
 - Plasma level of creatinine determined in non-pregnant, pregnant, and post-partum women
 - Three means determined
 - Cannot use t-test (two means only)
 - Use ANOVA
- Yields a p-value like t-tests



Chi-square

- Compares two or more **categorical variables**
- When asked to choose statistical test for a dataset: beware of percentages
 - Often categorical data
 - Always ask yourself whether data is quantitative or categorical



Confidence Intervals

- In scientific literature, means are reported with a **confidence interval**
- Study subjects: mean glucose was 90 ± 4
- If the study subjects were re-sampled
- Mean result would fall between 86 and 94 for 95% of re-samples
- For 5% of re-samples, the result would fall outside of the range of 86 to 94



Confidence Intervals

Group Comparisons

- If ranges overlap: no statistically significant difference
- Group 1 mean: 10 ± 5 ; Group 2 mean: 8 ± 4
 - Confidence intervals overlap
 - No significant difference between means
 - Similar to $p > 0.05$ for comparison of means
- Group 1 mean: 10 ± 5 ; Group 2 mean: 30 ± 4
 - Confidence intervals *do not* overlap
 - Significant difference between means
 - Similar to $p < 0.05$ for comparison of means

Confidence Intervals

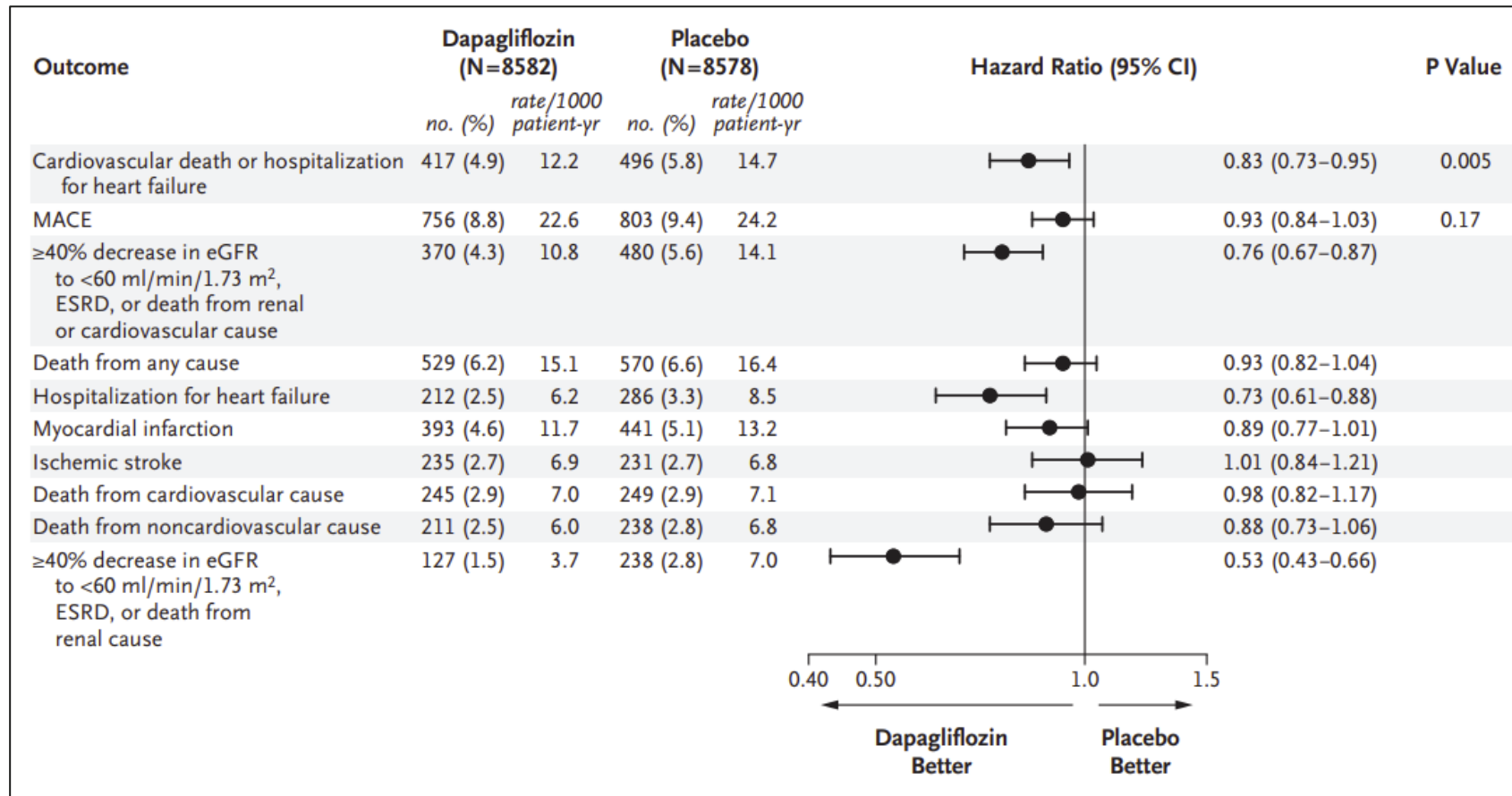
Group Comparisons

- Many studies report differences between groups
- Example: difference in systolic blood pressure between two groups = 10 mmHg
- Can calculate confidence intervals
- If range includes zero, no statistically significant difference
- Example:
 - Mean difference between two groups is 10.0 mmHg +/- 15.0 mmHg
 - Includes zero
 - No significant difference between groups
 - Similar to $p > 0.05$

Odds and Risk Ratios

- Some studies report odds or risk ratios with confidence intervals
- If range **includes 1.0** then exposure has no significant impact disease/outcome
- Example:
 - Risk of lung cancer among chemical workers studied
 - Risk ratio = 1.4 ± 0.5
 - Confidence interval includes 1.0
 - Chemical work not significantly associated with lung cancer
 - Similar to $p > 0.05$

Real World Example



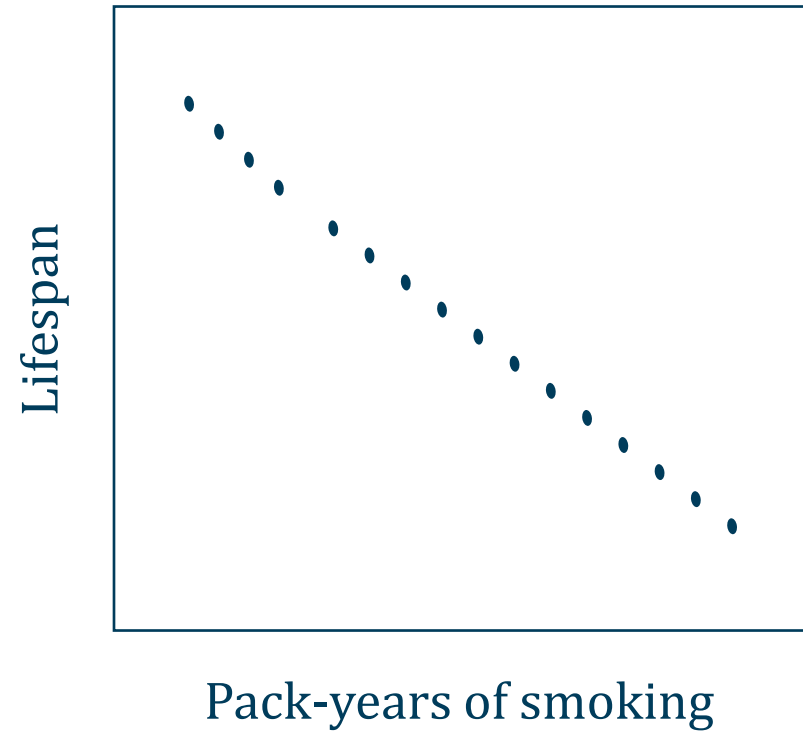
Correlations

Jason Ryan, MD, MPH



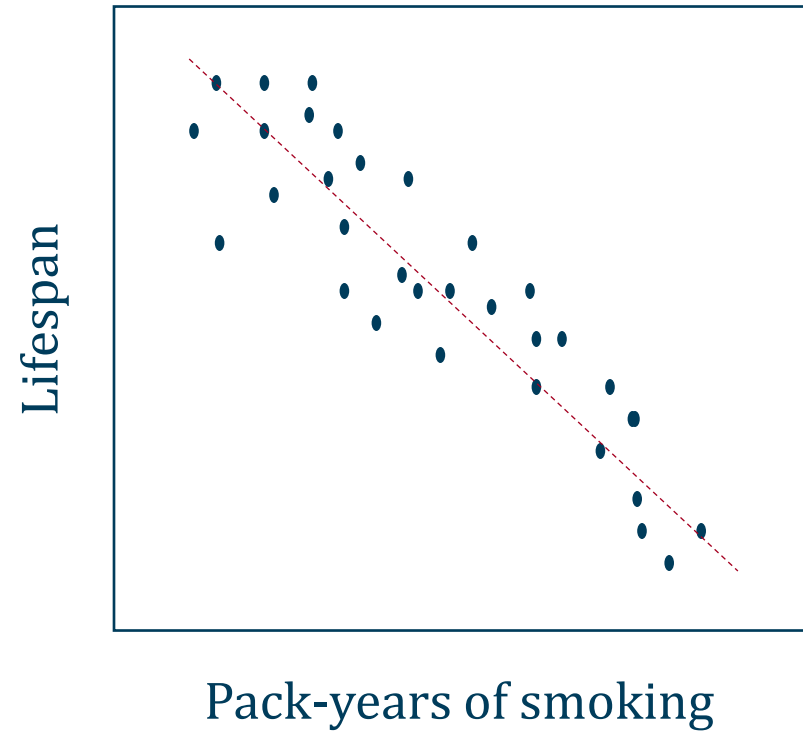
Correlation Coefficient

Pearson Coefficient



Correlation Coefficient

Pearson Coefficient



Correlation Coefficient

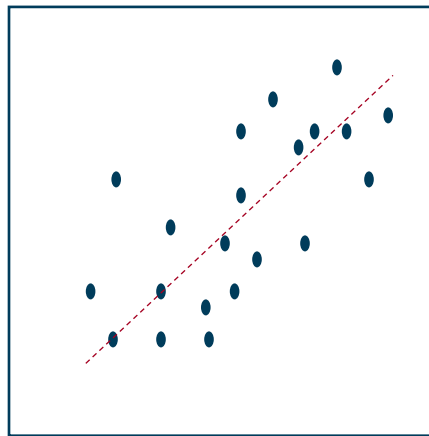
Pearson Coefficient

- Measure of linear correlation between two variables
- Represents strength of association of two variables
- Number from -1 to +1
- Closer to 1, stronger the relationship
- (-) number means inverse relationship
 - More smoking, less lifespan
- (+) number means positive relationship
 - More smoking, more lifespan
- 0 means no relationship

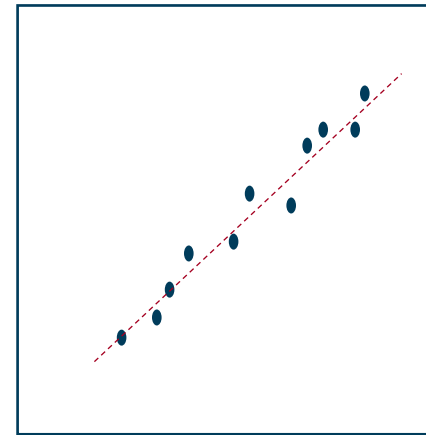
Correlation Coefficient

Pearson Coefficient

Strength of Relationship



$r = +0.5$

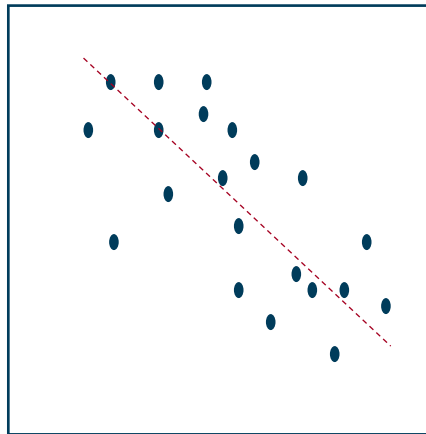


$r = +0.9$
(stronger relationship)

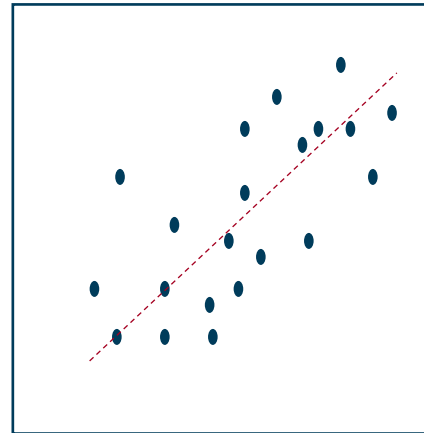
Correlation Coefficient

Pearson Coefficient

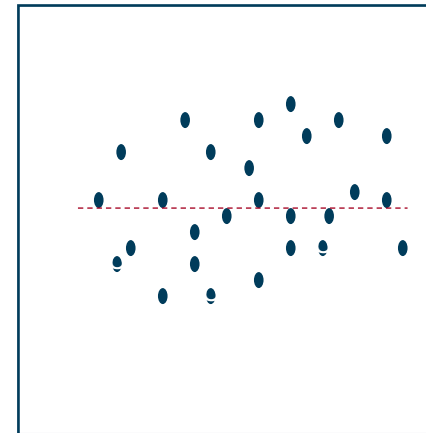
Direction of Relationship



$r = -0.5$
Negative



$r = +0.5$
Positive



$r = 0$
No relationship

Correlation Coefficient

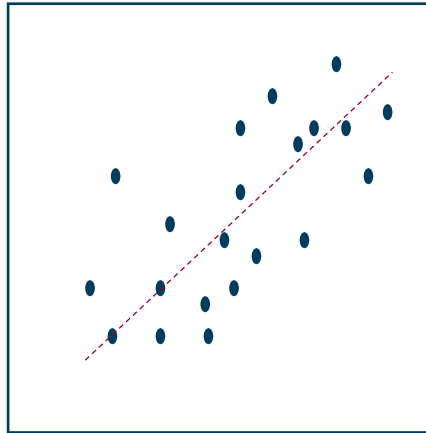
Pearson Coefficient

- Studies will report relationships with correlation coefficient
- Example:
 - Study of pneumonia patients
 - WBC on admission evaluated for relationship to LOS
 - $r = +0.5$
 - Higher WBC \rightarrow Higher LOS
- Sometimes p value is also reported
 - $P < 0.05$ indicates significant correlation
 - $P > 0.05$ indicates no significant correlation

Coefficient of Determination

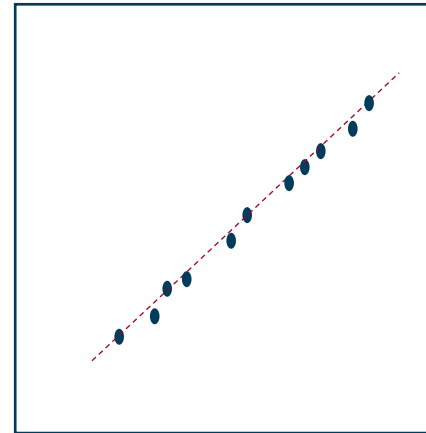
r^2

- Sometimes r^2 reported instead of r
- Always positive
- Indicates % of variation in y explained by x



$$r^2 = 0.6$$

(60% variation y explained by x)



$$r^2 = 1$$

(100% variation y explained by x)

Study Designs

Jason Ryan, MD, MPH



Epidemiology Studies

- Studies of populations
- Examine association of exposure with disease
- Many real world examples
 - Hypertension → stroke
 - Smoking → lung cancer
 - Toxic waste → leukemia
- Different from clinical/research studies
 - No control over exposure in epidemiology studies
 - Researchers control exposure in clinical studies



Types of Studies

Determine association of exposure with disease

- Cross-sectional study
- Cohort study (prospective/retrospective)
- Case-control study



Cross-Sectional Study

Prevalence Study

- Patients studied based on being **part of a group**
 - New Yorkers, women, tall people
- Frequency of disease and risk factors identified **at the same time**
 - How many have lung cancer?
 - How many smoke?
- Snapshot in time
 - Patients not followed for months/years
- Main outcome of study is **prevalence**
 - 50% of New Yorkers smoke
 - 25% of New Yorkers have lung cancer



Cross-Sectional Study

Prevalence Study

- Easy and quick to perform
- May have more than one group
 - 50% men have lung cancer, 25% of women have lung cancer
 - But groups not followed over time (e.g., years)
- Major disadvantage: can't determine causal relationships
 - How much smoking increases risk of lung cancer (RR)
 - Odds of getting lung cancer in smokers vs. non-smokers (OR)

Cross-Sectional Study

Prevalence Study

- New Yorkers were surveyed to determine whether they smoke and whether they have a morning cough. The study found a smoking prevalence of 50%. Among responders, 25% reported morning cough.
- Note the **absence of a time period**
 - Patients not followed for 1-year, etc.
- Likely questions:
 - Type of study? (cross-sectional)
 - What can be determined? (prevalence of disease)



Cross-Sectional Study

Prevalence Study

- Using a national US database, rates of lung cancer were determined among New Yorkers, Texans, and Californians. Lung cancer prevalence was 25% in New York, 30% in Texas, and 20% in California. The researchers concluded that living in Texas is associated with higher rates of lung cancer.
- Key points:
 - Presence of different groups could make you think of other study types
 - But note lack of time frame
 - Study is just a fancy description of disease prevalence

Cross-Sectional Study

Prevalence Study

- Researchers discover a gene that they believe leads to the development of diabetes. A sample of 1000 patients is randomly selected. All patients are screened for the gene. Presence or absence of diabetes is determined from a patient questionnaire. It is determined that the gene is strongly associated with diabetes.
- Key points:
 - Note lack of time frame
 - Patients not selected by disease or exposure (random)
 - Just a snapshot in time

Case Series

- Purely descriptive study (similar to cross-sectional)
- Often used in new diseases with unclear cause
- Multiple cases of a condition combined/analyzed
- Patient demographics (age, gender)
- Symptoms
- May identify clues about etiology
- No control group



Descriptive Studies

- Case reports
- Case series
- Cross-sectional studies

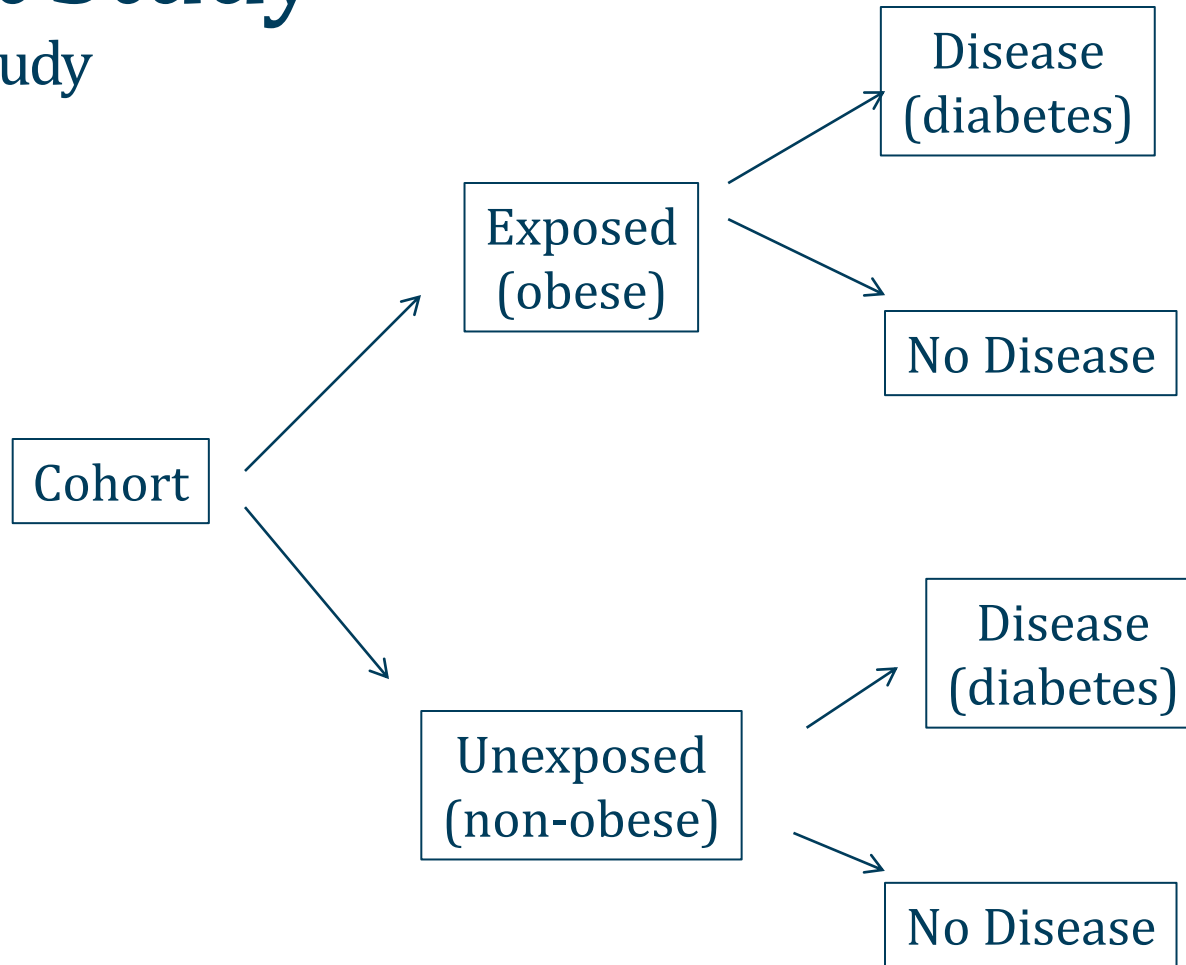
Cohort Study

Incidence Study

- Compares group with exposure to group without exposure
- Exposure determined **before outcome is known**
- Did exposure change likelihood of disease?
- Prospective: monitor groups over time going forward
- Retrospective: look back in time at groups over time
- Usually done for common diseases (e.g., diabetes)
- Easy to find cases in different groups
- Can establish **incidence** of disease in groups

Cohort Study

Incidence Study



Cohort Study

Incidence Study

- Main outcome measure is **relative risk (RR)**
 - How much does exposure increase risk of disease
- Example results
 - 50% smokers get lung cancer within 5 years
 - 10% non-smokers get lung cancer within 5 years
 - $RR = 50/10 = 5$
 - Smokers 5 times more likely to get lung cancer



Cohort Study

Incidence Study

- A group of 100 New Yorkers who smoke were identified based on a screening questionnaire at a local hospital. These patients were compared to another group that reported no smoking. Both groups received follow-up surveys asking about development of lung cancer annually for the next 3 years. The incidence of lung cancer was 25% among smokers and 5% among non-smokers.
- Likely questions:
 - Type of study? (*prospective* cohort)
 - What can be determined? (relative risk)

Cohort Study

Incidence Study

- A group of 100 New Yorkers who smoke were identified based on a screening questionnaire at a local hospital. These patients were compared to another group that reported no smoking. Hospital records were analyzed going back 5 years for all patients. The incidence of lung cancer was 25% among smokers and 5% among non-smokers.
- Likely questions:
 - Type of study? (*retrospective* cohort)
 - What can be determined? (relative risk)

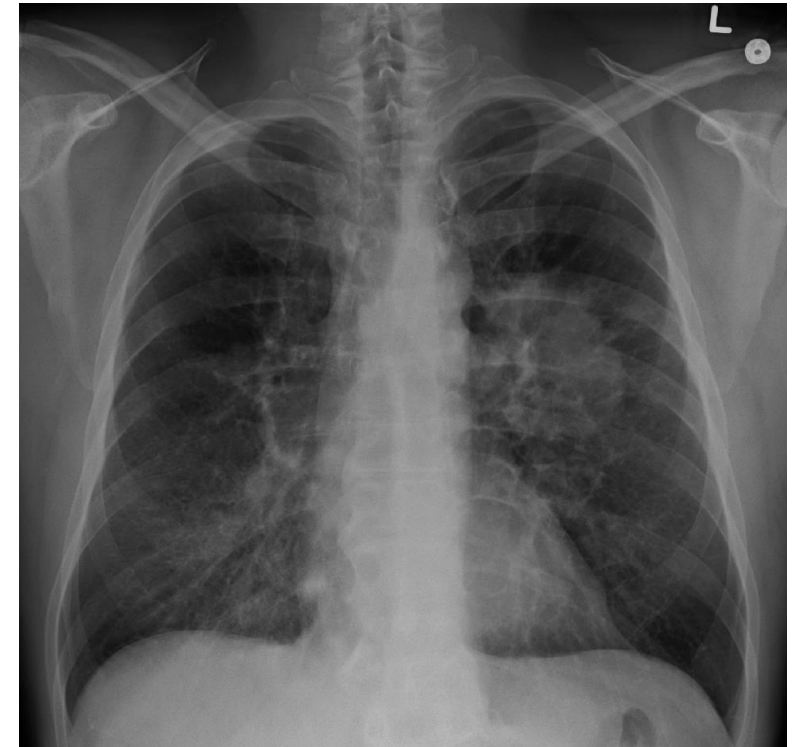
Retrospective Cohort Studies

- Challenging to identify
- Especially to distinguish from cross-sectional studies
- Patients identified “over a 5-year period”
- Cross-sectional study: outcome is **prevalence** of disease
 - How many patients at clinic over a 5-year period have hypertension
- Retrospective cohort: outcome is **incidence** of disease
 - Patients with exposure (e.g., smoking) identified
 - How many patients *developed* hypertension *over 5 years*

Cohort Study

Incidence Study

- Disadvantage: does not work with **rare diseases**
- Imagine:
 - 100 smokers, 100 non-smokers
 - Followed over 1 year
 - Zero cases of lung cancer both groups
- In rare diseases need LOTS of patients for LONG time
- Easier to find **cases** of lung cancer
- Then compare to controls without lung cancer

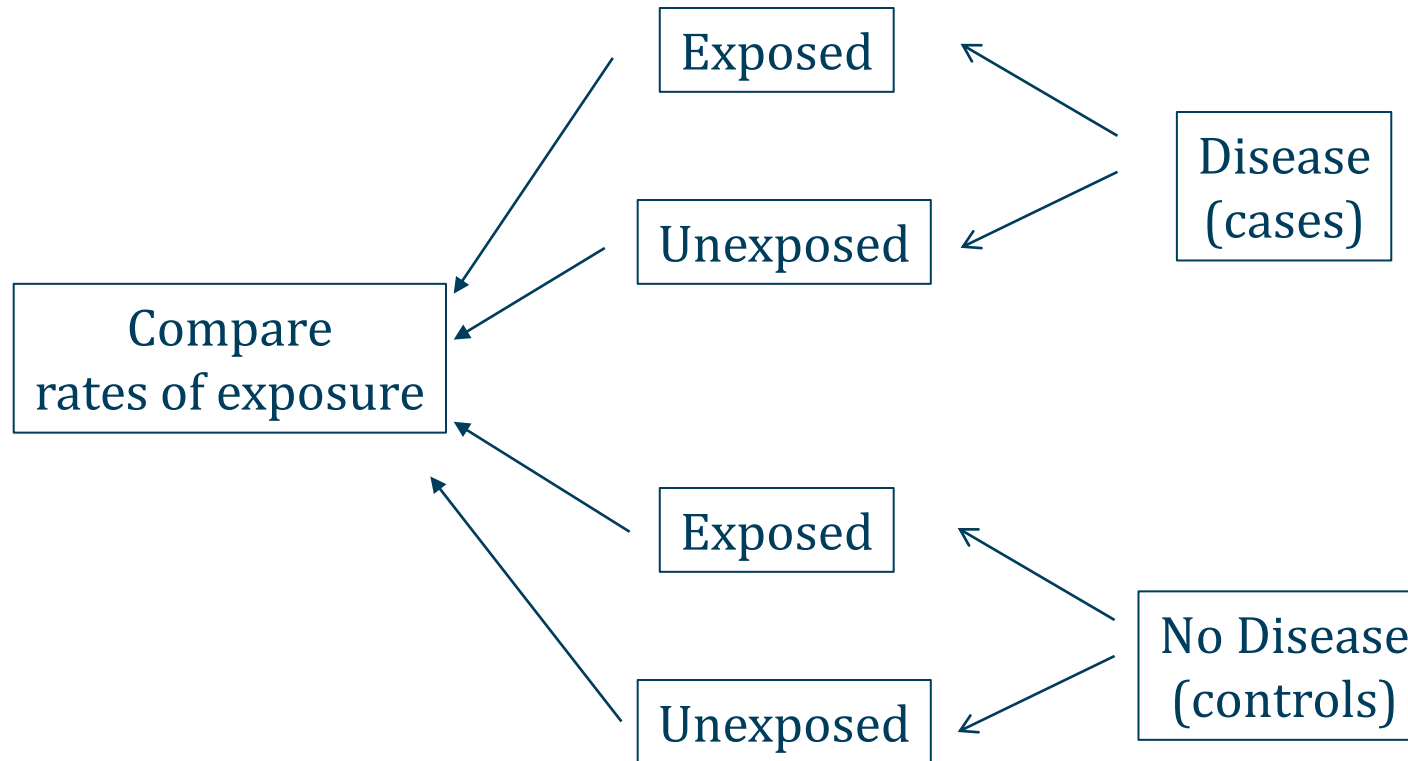


Case-Control Study

- Compares group with disease to group without
- Looks retrospectively over time for exposure or risk factors
- Opposite of cohort study
- Better for rare diseases



Case-Control Study



Case-Control Study

- Patients not followed over time for onset of disease
- Cannot determine incidence of disease
- Cannot determine relative risk
- Main outcome measure is **odds ratio**
- Odds of disease exposed/odds of disease unexposed



Case-Control Study

- A group of 100 New Yorkers with lung cancer were identified based on a screening questionnaire at a local hospital. These patients were compared to another group that reported no lung cancer. Both groups were questioned about smoking within the past 10 years. The prevalence of smoking was 25% among lung cancer patients and 5% among non-lung cancer patients.
- Likely questions:
 - Type of study? (case-control)
 - What can be determined? (odds ratio)

Matching

- Selection of control group (matching) key to getting good study results
- Controls should be as close to disease patients as possible
- Ideally, only difference between groups is presence or absence of disease
- This limits **confounding** in results



Randomized Trials

- Don't confuse with case-control
- Patients identified by disease like case-control
- Exposure controlled by researchers
- Exposure assigned **randomly**



How to Identify Study Types?

Cross-sectional	Cohort	Case-Control
Members of group Exposure/outcome same time Snapshot in time Prevalence	Selection by exposure status Risk ratio Incidence	Selection by disease status Odds ratio

Risk Quantification

Jason Ryan, MD, MPH



Risk of Disease

- Determined from **epidemiology studies**
- Cohort studies and case-control studies
- Smoking increases risk of lung cancer X percent
- Exercise decreases risk of heart disease Y percent



The 2 x 2 Table

- Derives from data from cohort or case control studies
- Patients with disease and exposure = A
- Patients with disease and no exposure = B
- Patients without disease and exposure = C
- Patients without disease and no exposure = D

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

The 2 x 2 Table

Uses

- Relative risk (risk ratio)
- Odds ratio
- Attributable risk
- Number needed to harm

Relative Risk

- Ratio of risk in exposed group to risk in unexposed group
- Established from a **cohort study**
- Risk in exposed group = $A/(A+B)$
- Risk in unexposed group = $C/(C+D)$
- Relative risk = risk exposed/risk unexposed

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

Relative Risk

- Risk of disease with exposure vs non-exposure
 - $RR = 5$
 - Smokers 5x more likely to get lung cancer than nonsmokers
- Ranges from zero to infinity
 - $RR = 1 \rightarrow$ No increased risk from exposure
 - $RR > 1 \rightarrow$ Exposure increases risk
 - $RR < 1 \rightarrow$ Exposure decreases risk

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

Relative Risk

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

$$RR = \frac{\text{Incidence Exposed}}{\text{Incidence Unexposed}} = \frac{A/(A+B)}{C/(C+D)}$$

Relative Risk

Example 1

- 10% smokers get lung cancer
- 10% nonsmokers get lung cancer
- $RR = 1$



Relative Risk

Example 2

- 50% smokers get lung cancer
- 10% nonsmokers get lung cancer
- $RR = 5$

EXAMPLE

Relative Risk

Example 3

- 10% smokers get lung cancer
- 50% nonsmokers get lung cancer
- $RR = 0.2$
- Smoking protective!



Relative Risk

- A group of 1000 college students is evaluated over ten years. Two hundred are smokers and 800 are non-smokers. Over the 10-year study period, 50 smokers get lung cancer compared with 10 non-smokers.

		Disease	
		+	-
Exposure	+		
	-		

$$RR = \frac{A/(A+B)}{C/(C+D)} = \underline{\hspace{2cm}}$$

Odds Ratio

- **Case control studies**
- Odds of exposure-disease/odds exposure-no-disease
- Ranges from zero to infinity
 - $OR = 1 \rightarrow$ Exposure equal among disease/no-disease
 - $OR > 1 \rightarrow$ Exposure increased among disease/no-disease
 - $OR < 1 \rightarrow$ Exposure decreased among disease/no-disease

Odds Ratio

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

$$OR = \frac{A/C}{B/D} = \frac{A*D}{B*C}$$

Risk Ratio versus Odds Ratio

- Risk ratio is the preferred metric
 - Easy to understand
 - Tells you how much exposure increase risk
- Not valid in case-control studies
 - RR is different depending on number cases you choose



Risk Ratio versus Odds Ratio

Suppose we find 100 cases and 200 controls

$$RR = \frac{50/100}{50/200} = 2.0$$

		Lung Cancer	
		+	-
Smoking	+	50	50
	-	50	150
		100	200

Risk Ratio versus Odds Ratio

Now suppose we find 200 cases and 200 controls

$$RR = \frac{100/150}{100/250} = 1.6$$

		Lung Cancer	
		+	-
Smoking	+	100	50
	-	100	150
		200	200

Risk Ratio versus Odds Ratio

OR does not change with case number

	+	-
+	50	50
-	50	150
	100	200

$$OR = \frac{50/50}{50/150} = 3.0$$

	+	-
+	100	50
-	100	150
	200	200

$$OR = \frac{100/100}{50/150} = 3.0$$

Risk Ratio versus Odds Ratio

- Risk ratio is dependent on **number of cases and controls**
- Invalid to use risk ratio in case-control studies
- Must use odds ratio instead



Rare Disease Assumption

- $OR = RR$
- Most exposed/unexposed have no disease (-)
- Few disease (+) among exposed/unexposed



Rare Disease Assumption

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

$$OR = \frac{A/C}{B/D} = \frac{A*D}{B*C}$$

$$RR = \frac{A/(A+B)}{C/(C+D)} = \frac{A/B}{C/D} = \frac{A*D}{B*C}$$

$$OR = RR$$

When $B \gg A$ and $D \gg C$

Rare Disease Assumption

- Allows use of a case-control study to determine RR
- Commonly accepted number is **prevalence < 10%**
- Case-control studies easier and less expensive
 - But odds ratio is a weak association
- Classic question:
 - Description of case-control study
 - RR reported
 - Is this valid?
 - Answer: only if disease is rare



Attributable Risk

- Suppose 1% incidence lung cancer in non-smokers
- Suppose 21% incidence in smokers
- Attributable risk = 20%
- Added risk due to exposure to smoking



Attributable Risk

		Disease	
		+	-
Exposure	+	A	B
	-	C	D

$$AR = A/(A+B) - C/(C+D)$$

$$RR = \frac{A/(A+B)}{C/(C+D)}$$

Attributable Risk

Study 1	Study 2
Risk exposed = 50% Risk unexposed = 25% Relative Risk = 2.0 Attributable Risk = 25%	Risk exposed = 10% Risk unexposed = 5% Relative Risk = 2.0 Attributable Risk = 5%

EXAMPLE

Attributable Risk Percentage

- Percent of disease explained by risk factor
- Ratio of attributable risk to risk in exposed
- Suppose attributable risk for smoking and lung cancer 25%
- Suppose risk in smokers is 50%
- Indicates 50% of lung cancers explained by smoking
- Can be calculated directly from attributable risk or relative risk

$$ARP = \frac{AR}{Re} = \frac{RR - 1}{RR}$$

Attributable Risk

Study 1	Study 2
Risk exposed = 50% Risk unexposed = 25% Relative Risk = 2.0 Attributable Risk = 25% Attributable Risk % = 50%	Risk exposed = 10% Risk unexposed = 5% Relative Risk = 2.0 Attributable Risk = 5% Attributable Risk % = 50%

$$ARP = \frac{AR}{Re} = \frac{RR - 1}{RR}$$

EXAMPLE

Number Need to Harm

- Number of patients exposed for one episode of disease to occur on average
- Number who need to smoke for one case of lung cancer to develop
- Equal to reciprocal of attributable risk
- If attributable risk to smoking is 20%, then NNH is $1/0.2 = 5$
- Similar to number needed to treat calculated from clinical trials

$$\text{NNH} = \frac{1}{\text{AR}}$$

Attributable Risk

Study 1	Study 2
Risk exposed = 50% Risk unexposed = 25% Relative Risk = 2.0 Attributable Risk = 25% Attributable Risk % = 50% NNH = 4	Risk exposed = 10% Risk unexposed = 5% Relative Risk = 2.0 Attributable Risk = 5% Attributable Risk % = 50% NNH = 20

$$NNH = \frac{1}{AR}$$

EXAMPLE

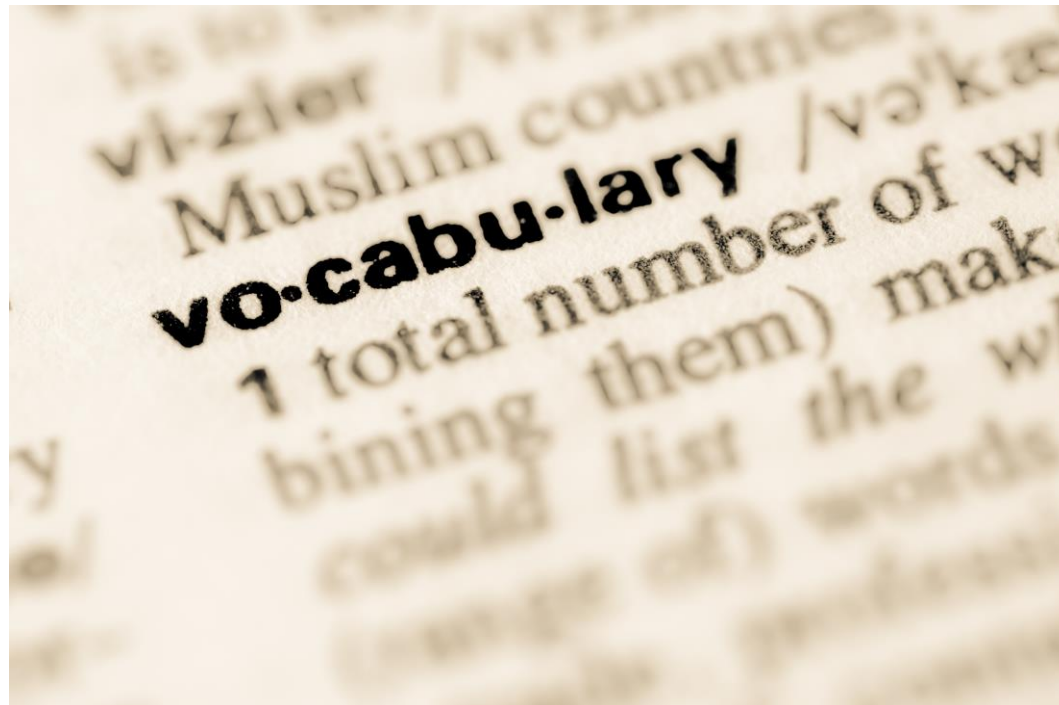
Sensitivity and Specificity

Jason Ryan, MD, MPH



Incidence and Prevalence

- **Incidence** of diabetes: 1,000 new cases diabetes *per year*
- **Prevalence** of diabetes: 100,000 cases at one point in time *in a population*



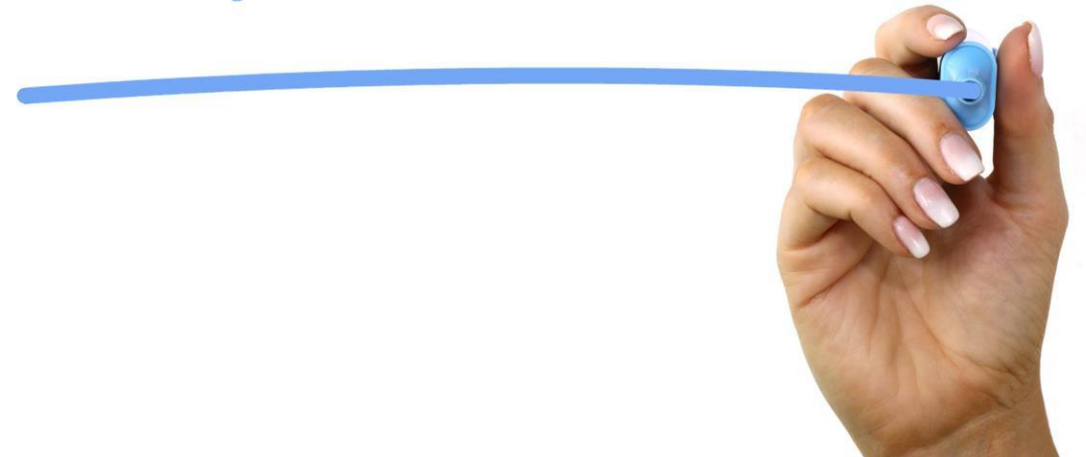
Incidence and Prevalence

- Incidence rate = **new cases / population at risk**
 - Determined for a **period of time** (e.g. one year)
 - Population at risk = total population – people with disease
 - 40,000 people
 - 10,000 with disease
 - 1,000 new cases per year
 - Incidence rate = $1,000 / (40k - 10k) = 1,000 \text{ cases} / 30,000$
- Prevalence rate = **number of cases / population at risk**
 - Entire population at risk

Incidence and Prevalence

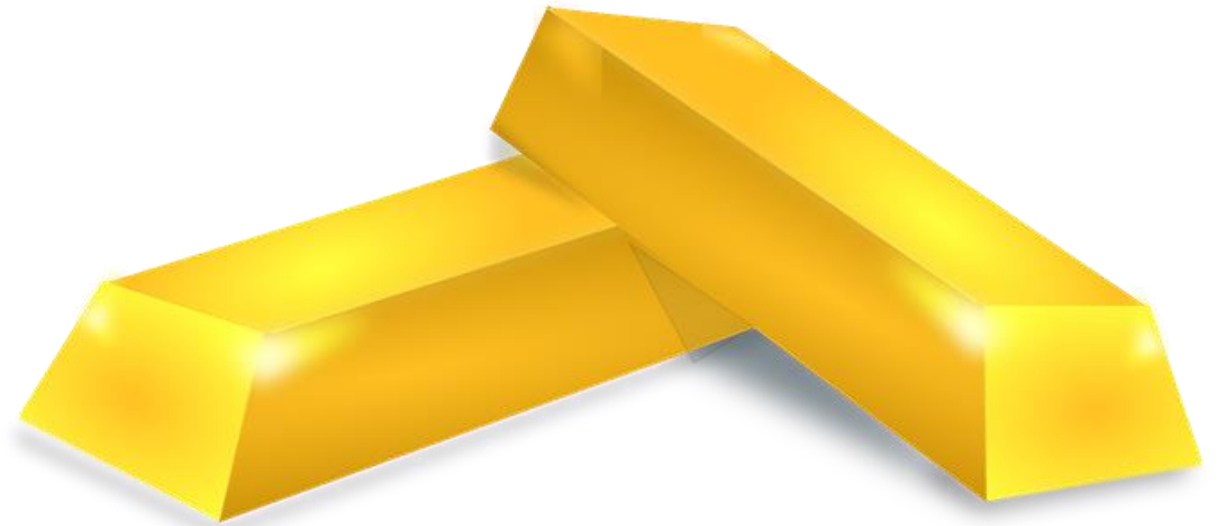
- For chronic diseases: prevalence \gg incidence
- For rapidly fatal diseases: incidence \sim prevalence
- New primary prevention programs:
 - Both incidence and prevalence fall
- New drugs that improve survival
 - Incidence unchanged
 - Prevalence increases

PREVENTION



Diagnostic Tests

- Used to identify individuals with and without disease
- **Gold standard** = best available test
- New tests compared to gold standard
- Described by test performance versus gold standard
- Key metrics: sensitivity and specificity



Diagnostic Tests

Blood Glucose Levels

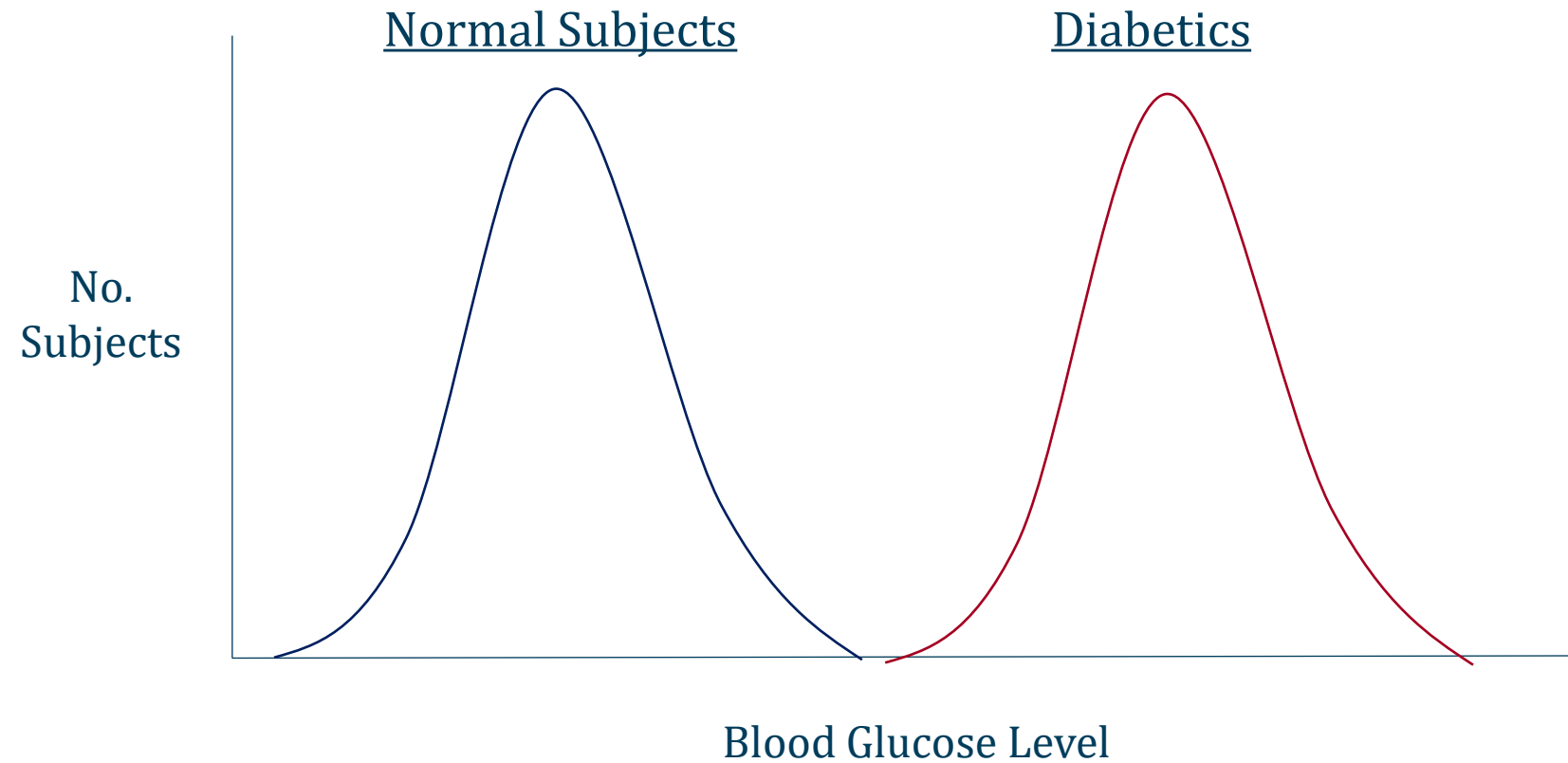
Normal Subjects

90	115	90
87	112	87
101	101	101
110	92	110
105	85	105
93	79	93
92	100	92
95	99	95
88	86	88
112	102	112

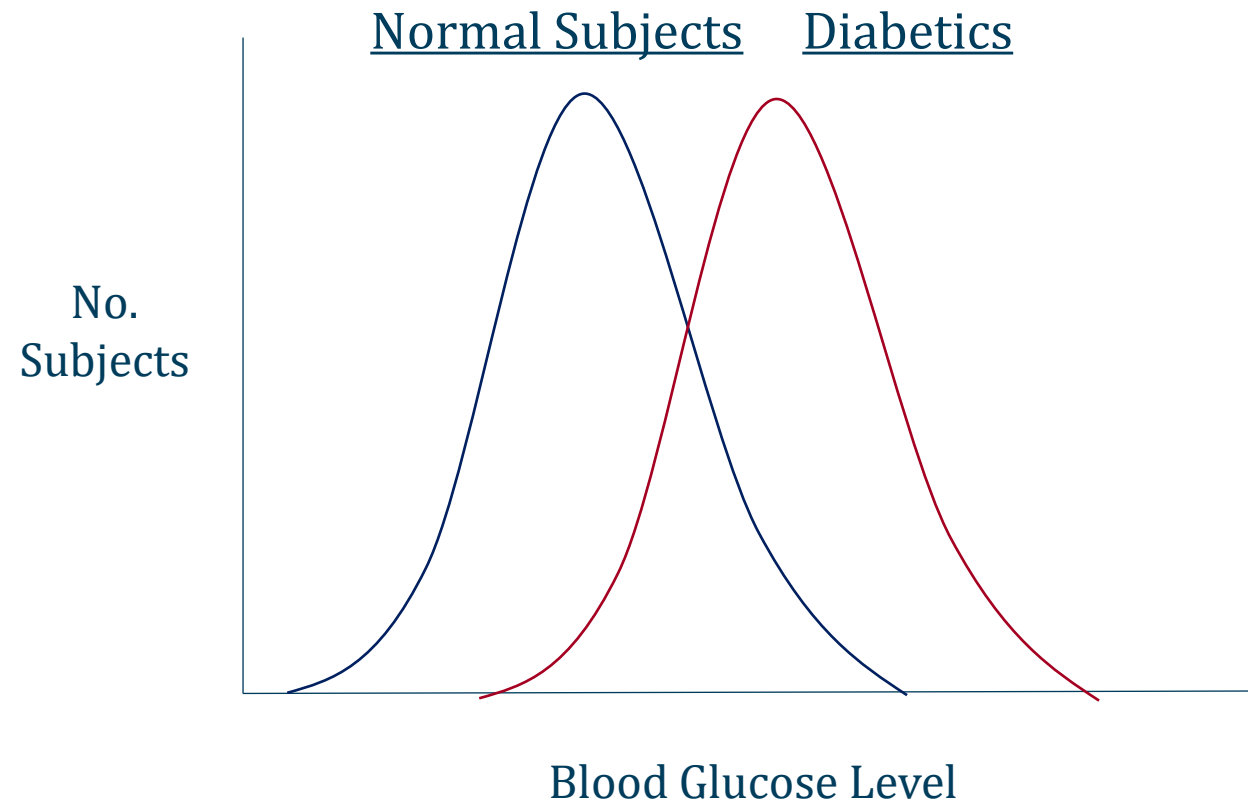
Diabetics

140	115	140
132	112	132
110	101	110
105	176	105
127	180	127
170	199	170
140	100	140
160	143	160
112	168	112
160	102	160

Diagnostic Tests



Diagnostic Tests



Diagnostic Tests

		Disease	
		+	-
Test	+	TP	FP
	-	FN	TN

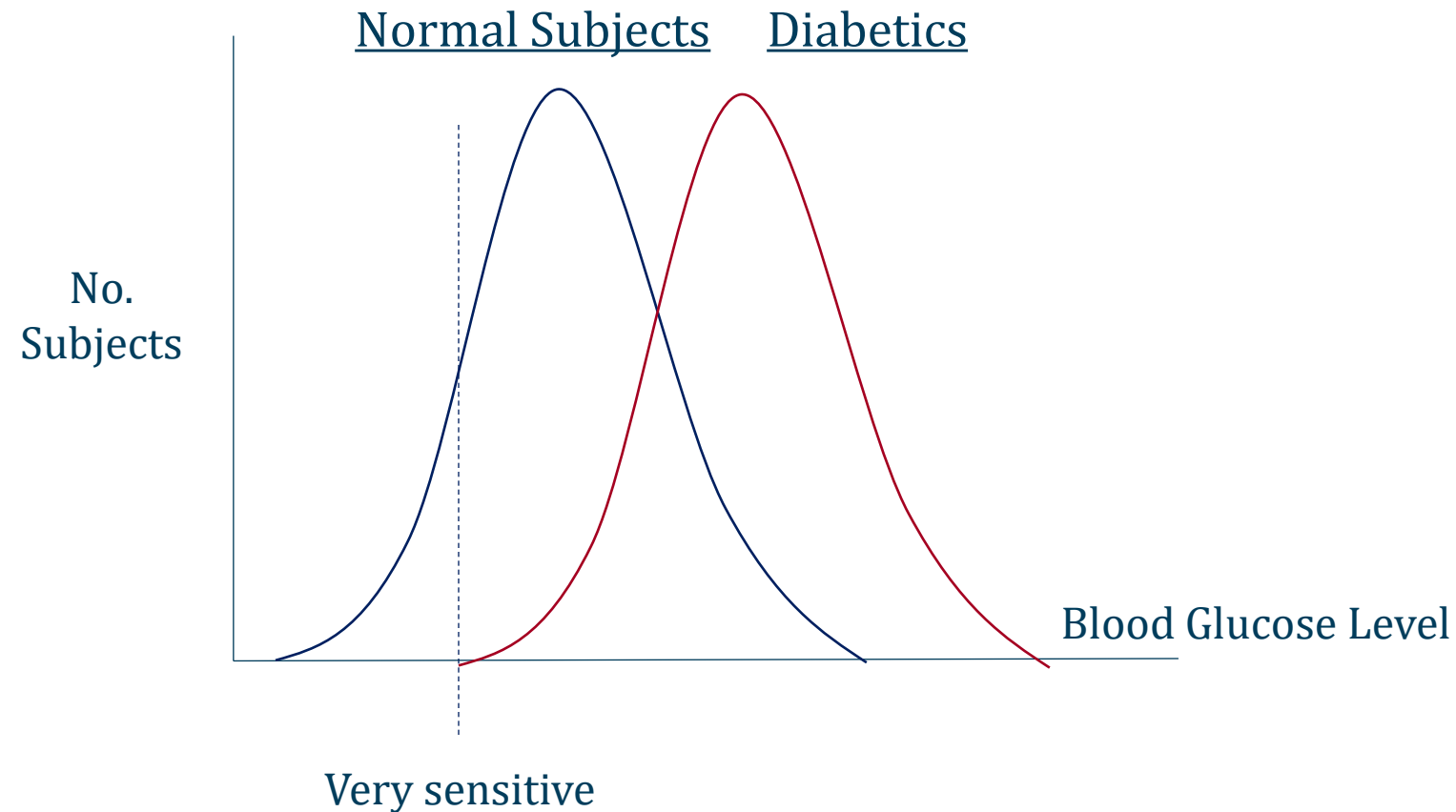
Sensitivity

		Disease	
		+	-
Test	+	TP	FP
	-	FN	TN

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

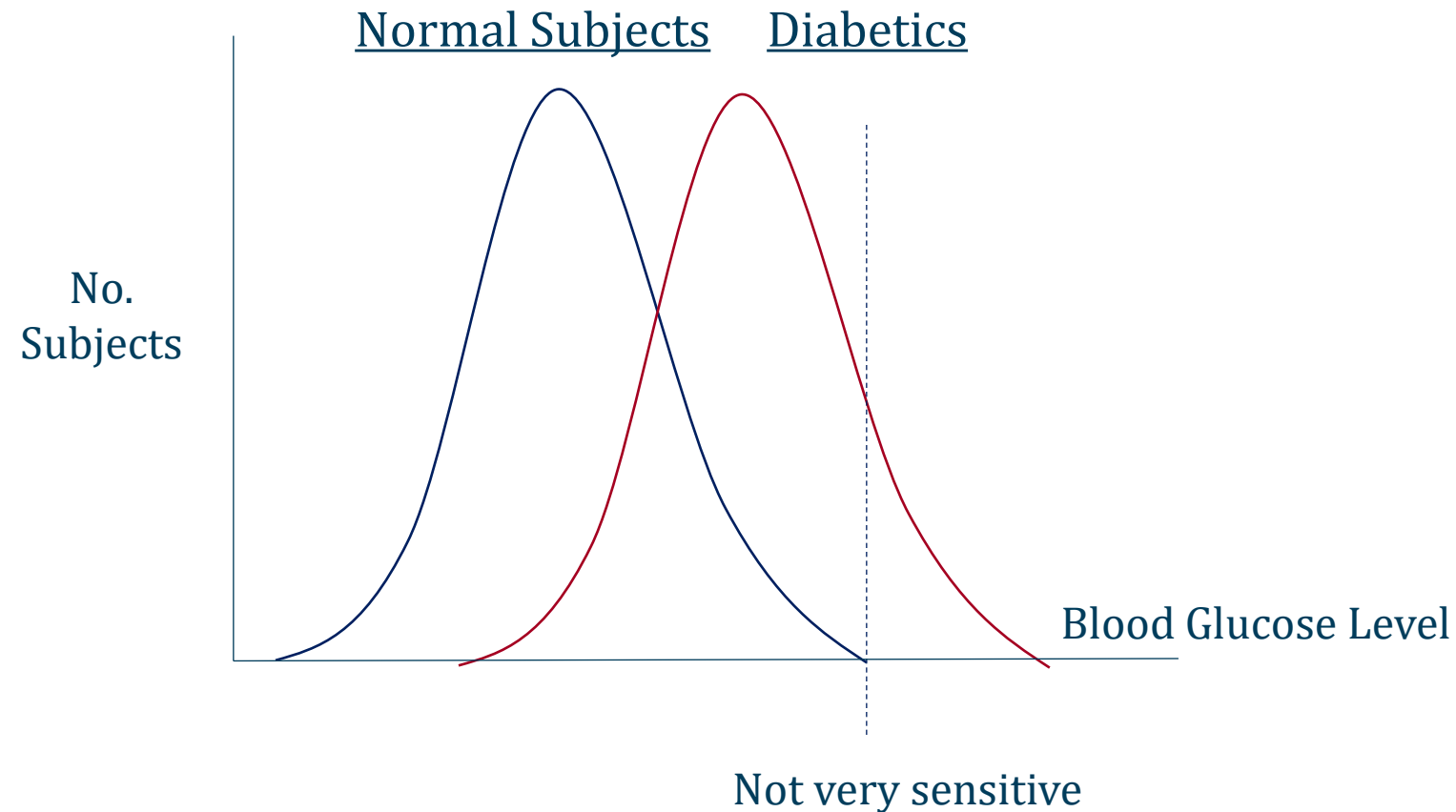
Sensitivity

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$



Sensitivity

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$



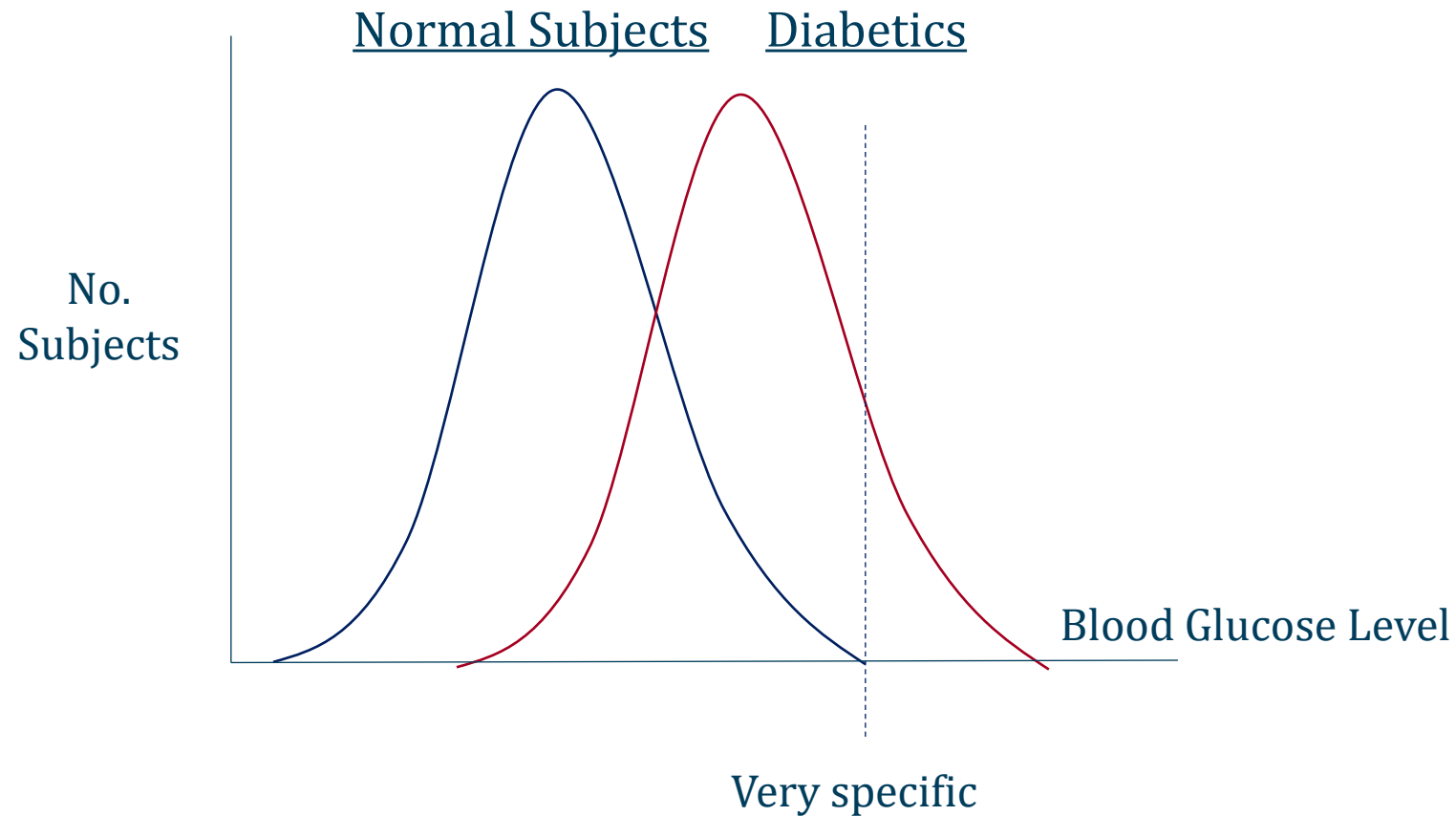
Specificity

		Disease	
		+	-
Test	+	TP	FP
	-	FN	TN

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

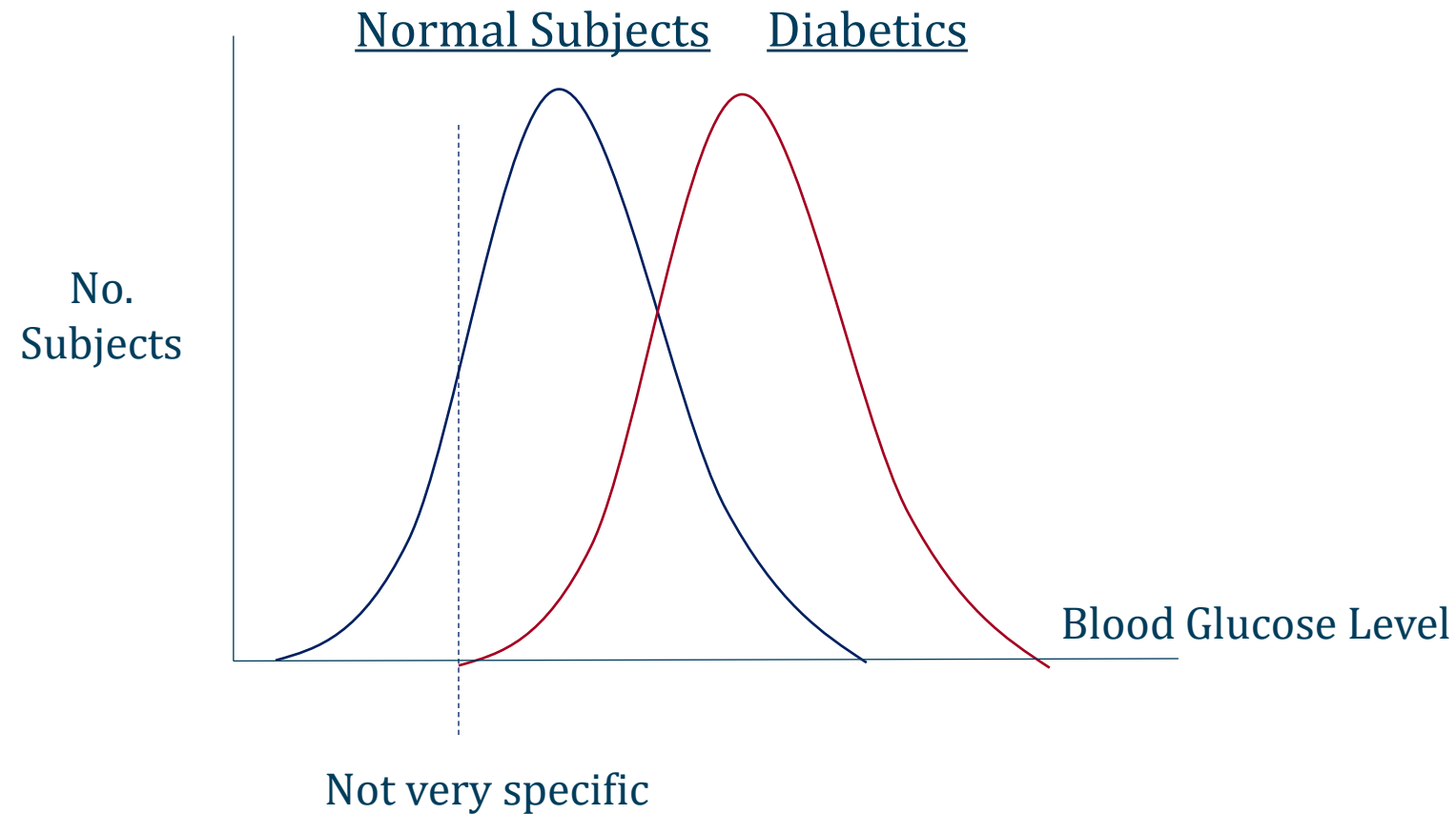
Specificity

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$



Specificity

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$



Sample Question

- The results below are obtained from a study of test X on patients with and without disease A. What is the sensitivity of test X?

		Disease A	
		+	-
Test X	+	25	10
	-	75	10

Sample Question

- The results below are obtained from a study of test X on patients with and without disease A. What is the specificity of test X?

		Disease A	
		+	-
Test X	+	25	10
	-	75	10

Key Point

- High sensitivity = good at ruling **OUT** disease
- High specificity = good at ruling **IN** disease
- **SnOUT and SpIN**



Finding Rare Disease

- Screen with a **SENSITIVE** test
 - Most people will be negative
 - Result is reliable because test is sensitive
- Follow up (+) screening tests with a **SPECIFIC** test
 - Sift through all the false/true positives

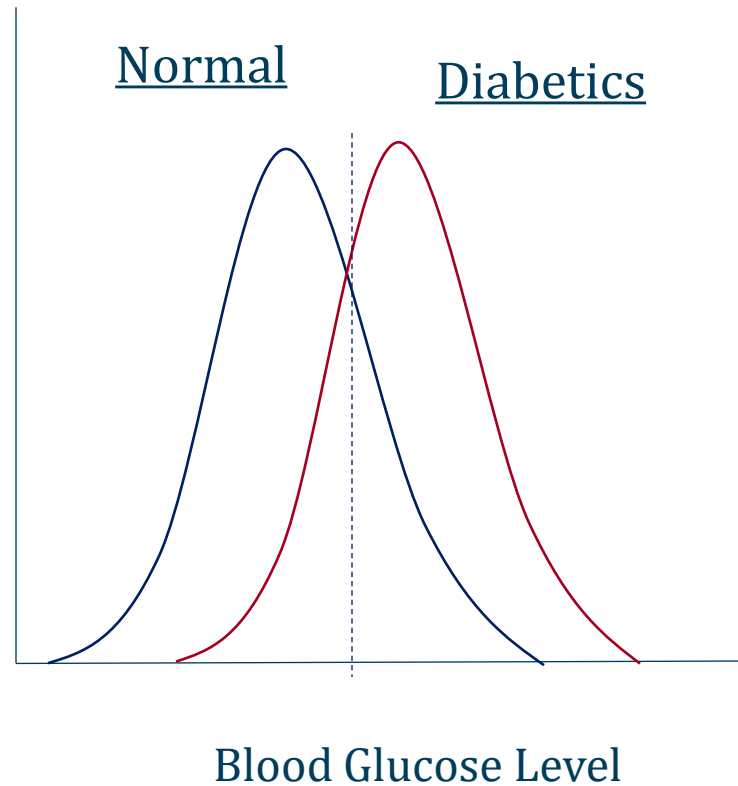
Sensitivity and Specificity

- Use sensitive tests when you don't want to miss cases
 - Captures many true positives (at the cost of false positives)
 - Screening of large populations
 - Severe diseases
- Use specific tests after sensitive tests
 - Confirmatory tests

RULES

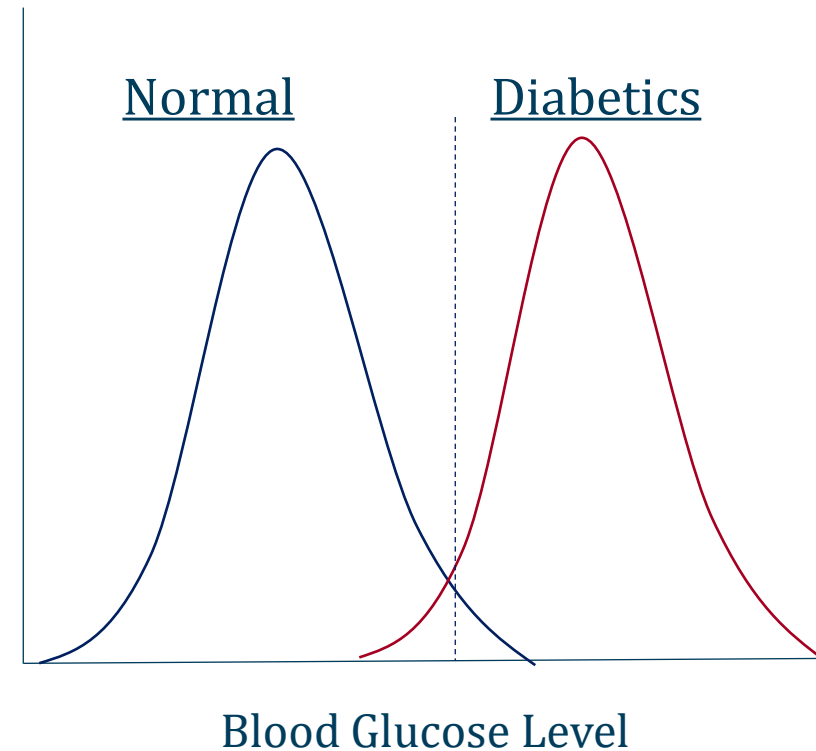
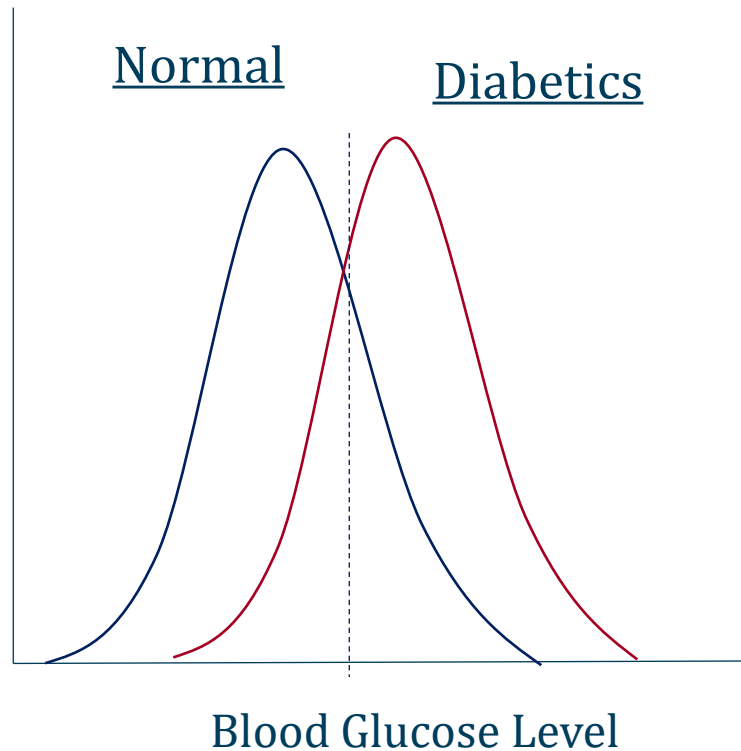
Sensitivity & Specificity

- Midpoint cutoff maximizes sensitivity/specificity



Sensitivity & Specificity

- Degree of overlap limits maximum combined sensitivity/specificity



Key Point

- Sensitivity and specificity are **characteristics of the test**
- Remain constant for **any prevalence** of disease



Sensitivity/Specificity

Test X
Sensitivity 80%
Specificity 50%

Group 1
Prevalence = 80%

		Disease	
		+	-
Test	+	64	10
	-	16	10
		80	20

Group 2
Prevalence = 20%

		Disease	
		+	-
Test	+	16	40
	-	4	40
		20	80

Sensitivity/Specificity

Group 1
Prevalence = 80%

		Disease	
		+	-
Test	+	64	10
	-	16	10

$$\text{Sens} = 64/80 = 80\%$$
$$\text{Spec} = 10/20 = 50\%$$

Group 2
Prevalence = 20%

		Disease	
		+	-
Test	+	16	40
	-	4	40

$$\text{Sens} = 16/20 = 80\%$$
$$\text{Spec} = 40/80 = 50\%$$

Sensitivity and Specificity

- “A test is negative in 80% of people who do not have the disease.”
 - True negatives; specificity
- “A test is positive in 50% of the people who do have the disease.”
 - True positives; sensitivity

		Disease	
		+	-
Test	+	TP	FP
	-	FN	TN

Predictive Values

Jason Ryan, MD, MPH



Predictive Values

- For diagnostic tests, what doctors and patients want to know is:
 - I have a positive result; what is the likelihood I have this disease?
 - I have a negative result; what is the likelihood I don't have this disease?
- Sensitivity and specificity do not answer these questions
- Need to use **positive and negative predictive values**

Positive Predictive Value

		Disease	
		+	-
Test	+	TP	FP
	-	FN	TN

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Negative Predictive Value

		Disease	
		+	-
Test	+	TP	FP
	-	FN	TN

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

Sample Question

- A test has a sensitivity of 80% and a specificity of 50%. The test is used in a population where disease prevalence is 40%. What is the positive predictive value?

		Disease A	
		+	-
Test X	+	32	30
	-	8	30
		40 patients	60 patients
		100 patients	

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{32}{62} = 52\%$$

Key Point

- Predictive values are dependent on the **prevalence of disease**
- Sensitivity and specificity are independent of prevalence of disease



Positive Predictive Value

Test X
Sensitivity 80%
Specificity 50%

Group 1
Prevalence = 80%

		Disease	
		+	-
Test	+	64	10
	-	16	10
		80	20

$$\text{PPV} = \frac{64}{74} = 86\%$$

Group 2
Prevalence = 20%

		Disease	
		+	-
Test	+	16	40
	-	4	40
		20	80

$$\text{PPV} = \frac{16}{56} = 29\%$$

Negative Predictive Value

Test X
Sensitivity 80%
Specificity 50%

Group 1
Prevalence = 80%

		Disease	
		+	-
Test	+	64	10
	-	16	10
		80	20

$$\text{NPV} = \frac{10}{26} = 38\%$$

Group 2
Prevalence = 20%

		Disease	
		+	-
Test	+	16	40
	-	4	40
		20	80

$$\text{NPV} = \frac{40}{44} = 91\%$$

Key Point

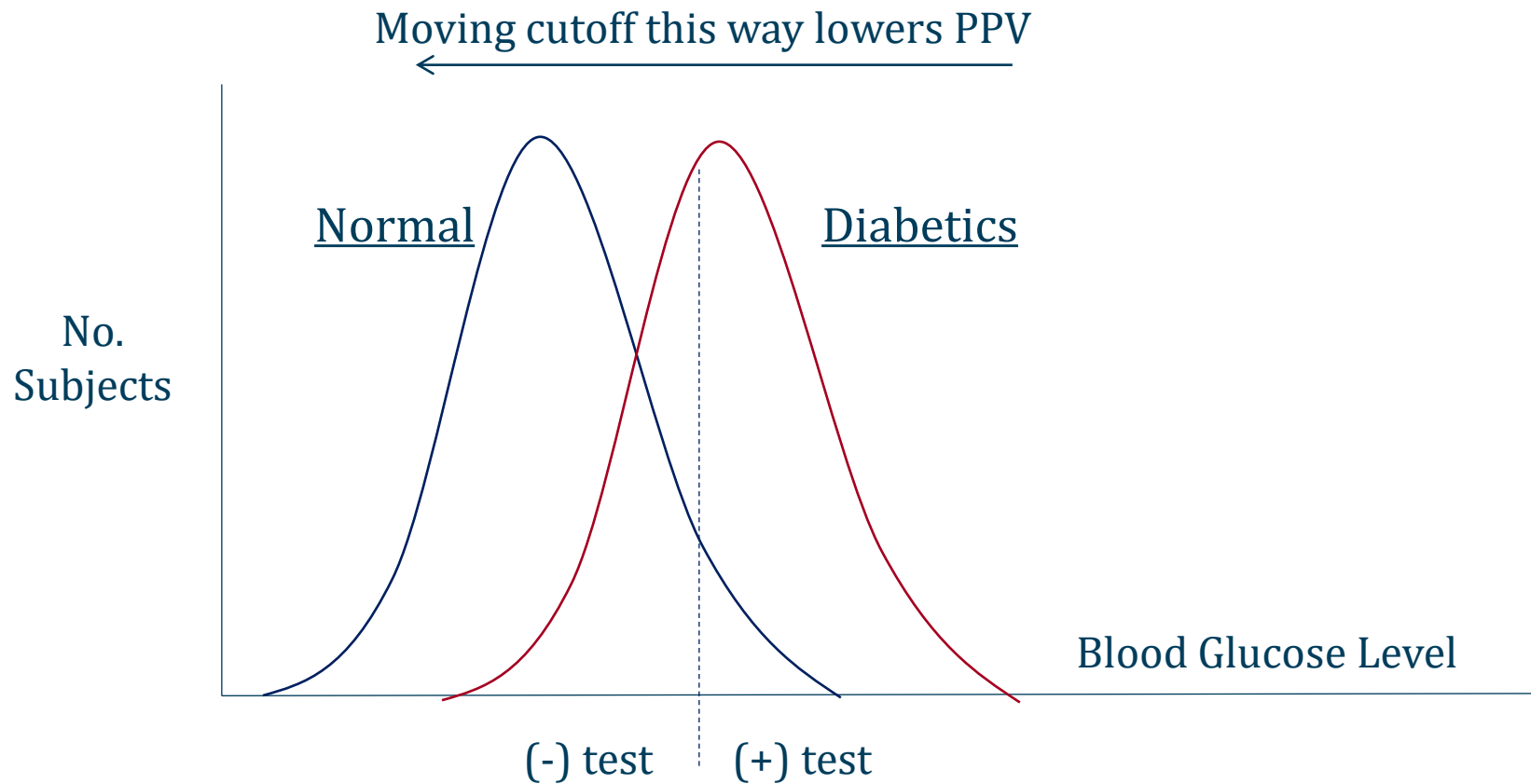
- PPV is higher when prevalence is higher
- NPV is high when prevalence is lower



Cutoff Point

Diagnostic Tests

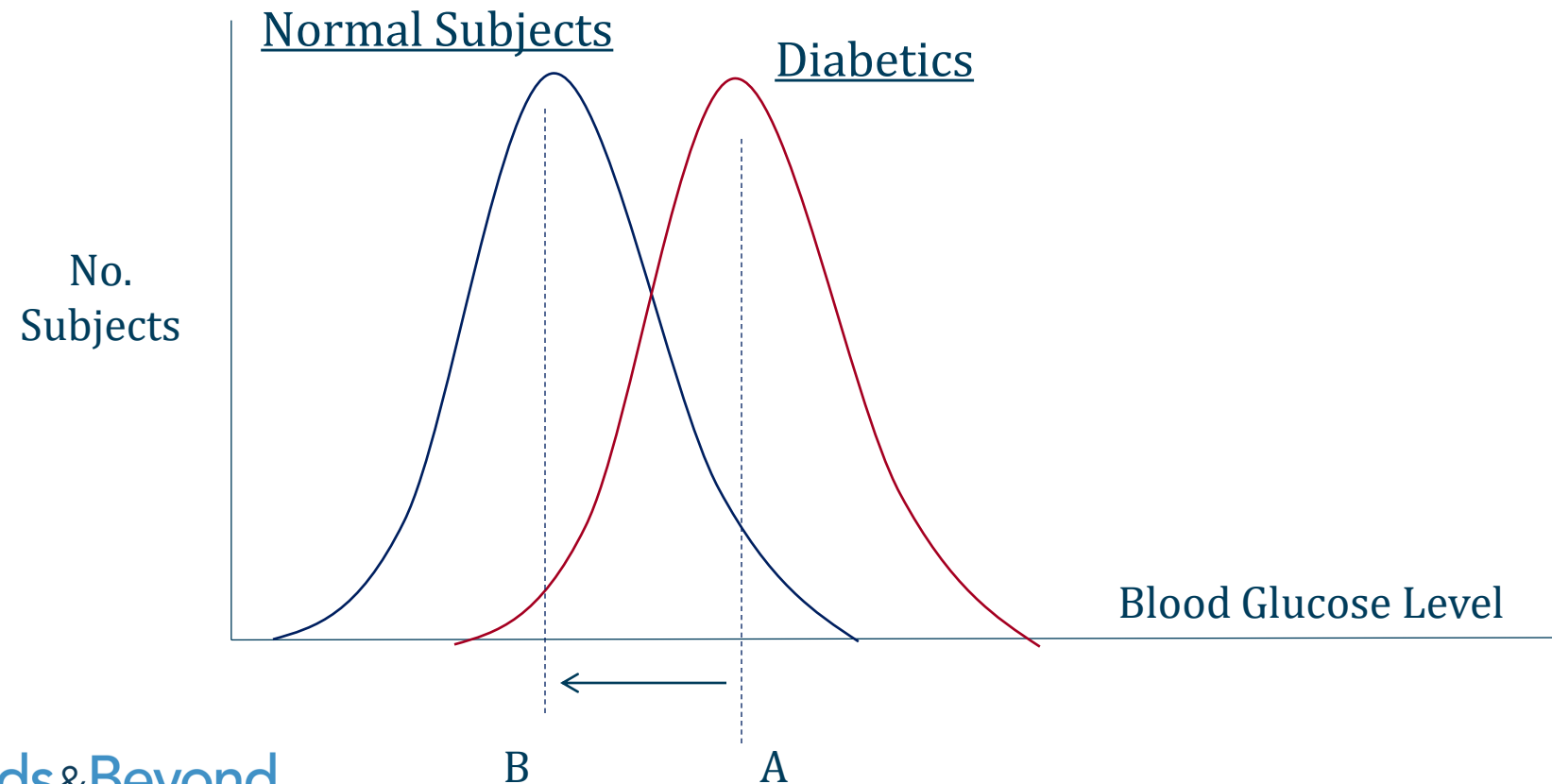
$$PPV = \frac{TP}{TP + FP}$$



$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Cutoff A
 TP = 10
 FP = 5
 PPV = 10/15
 = 66%

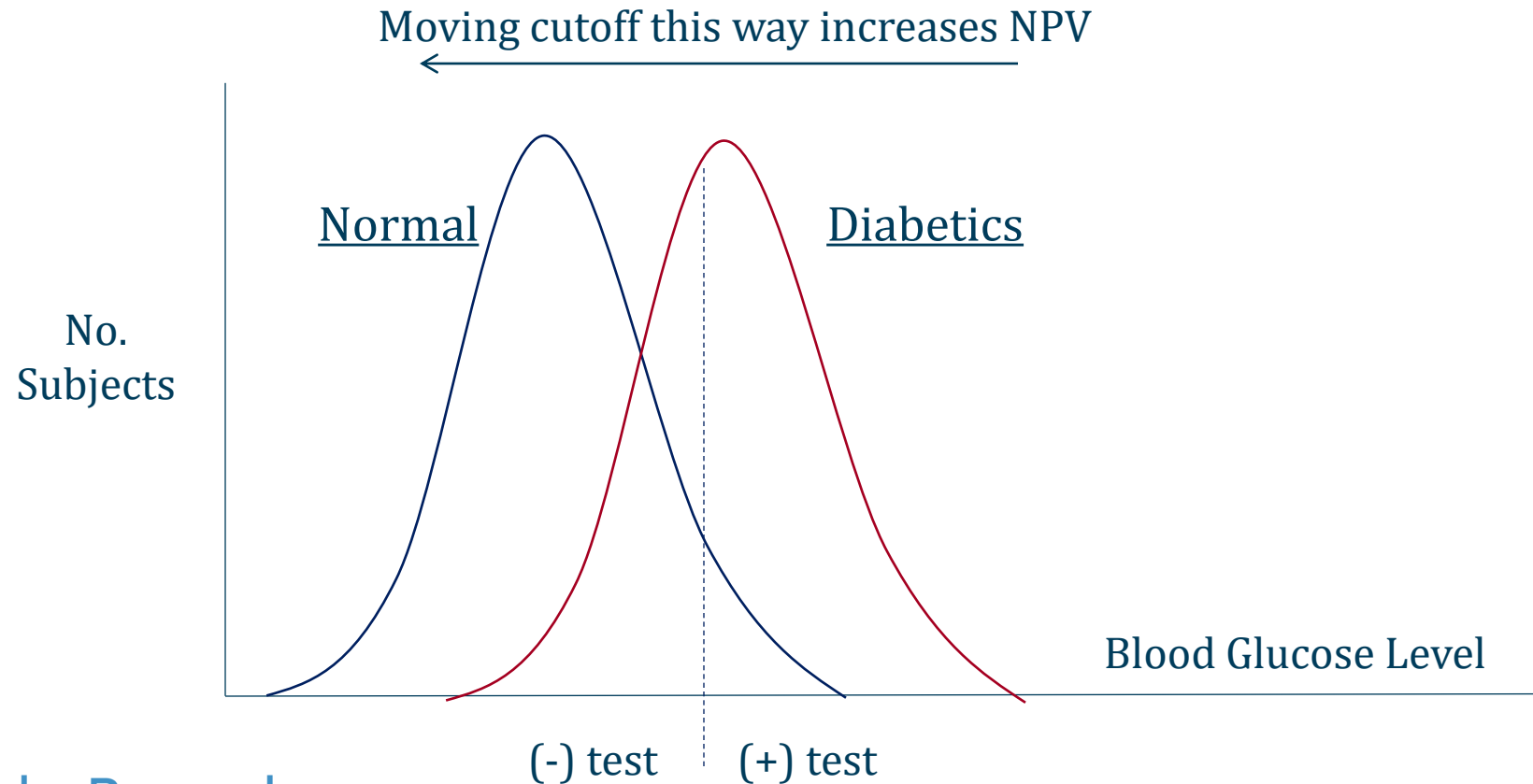
Cutoff B
 TP = 15
 FP = 10
 PPV = 15/25
 = 60%



Cutoff Point

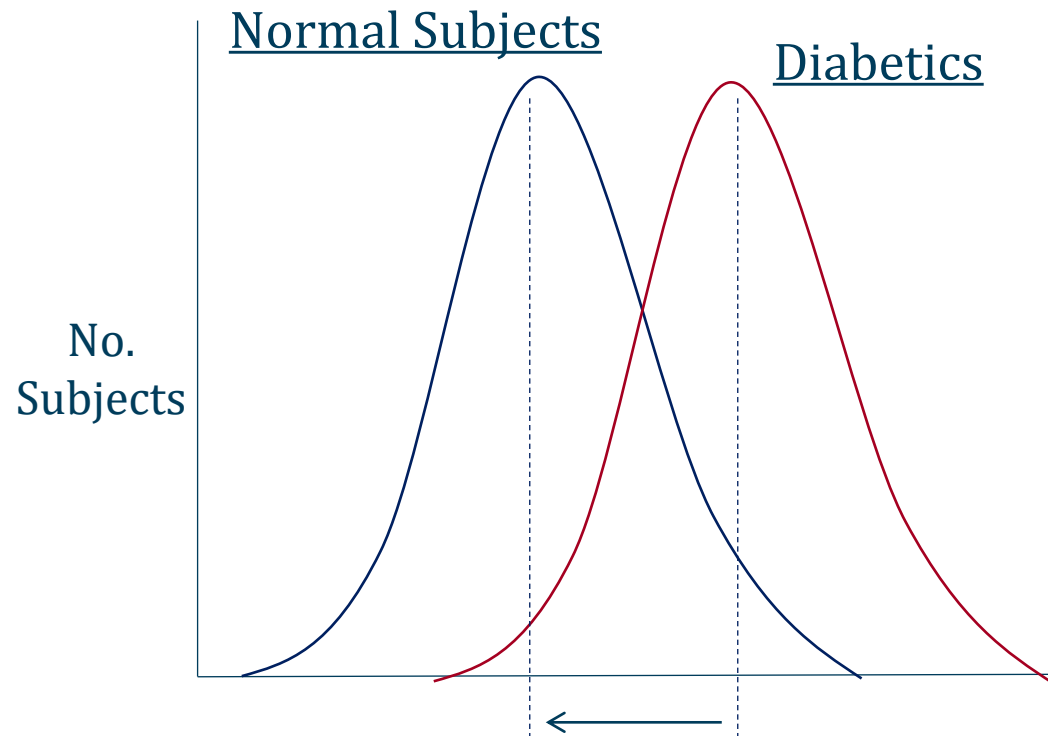
Diagnostic Tests

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$



Sample Question

- The American Diabetes Association proposes lowering the cutoff value for the fasting glucose level that indicates diabetes. How will this change affect sensitivity, specificity, PPV, and NPV?
 - Sensitivity: Increase
 - Specificity: Decrease
 - PPV: Decrease
 - NPV: Increase



Diagnostic Tests

Jason Ryan, MD, MPH



Diagnostic Tests

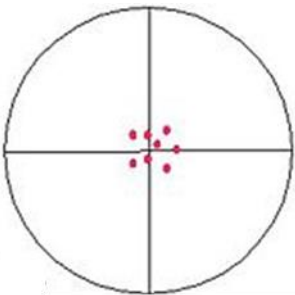
Special Topics

- Accuracy/Precision
- ROC Curves
- Likelihood ratios

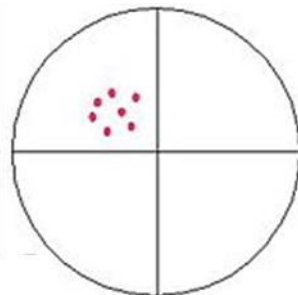


Accuracy and Precision

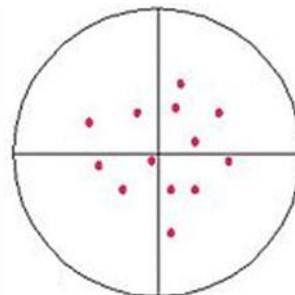
- Describe quality of **measurements** used as part of diagnostic test
- Accuracy (validity): how closely data matches reality
- Precision (reliability): how closely repeated measurements match each other
- Can have accuracy without precision (or vice versa)



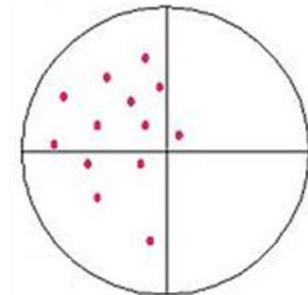
Precise and accurate



Precise not accurate



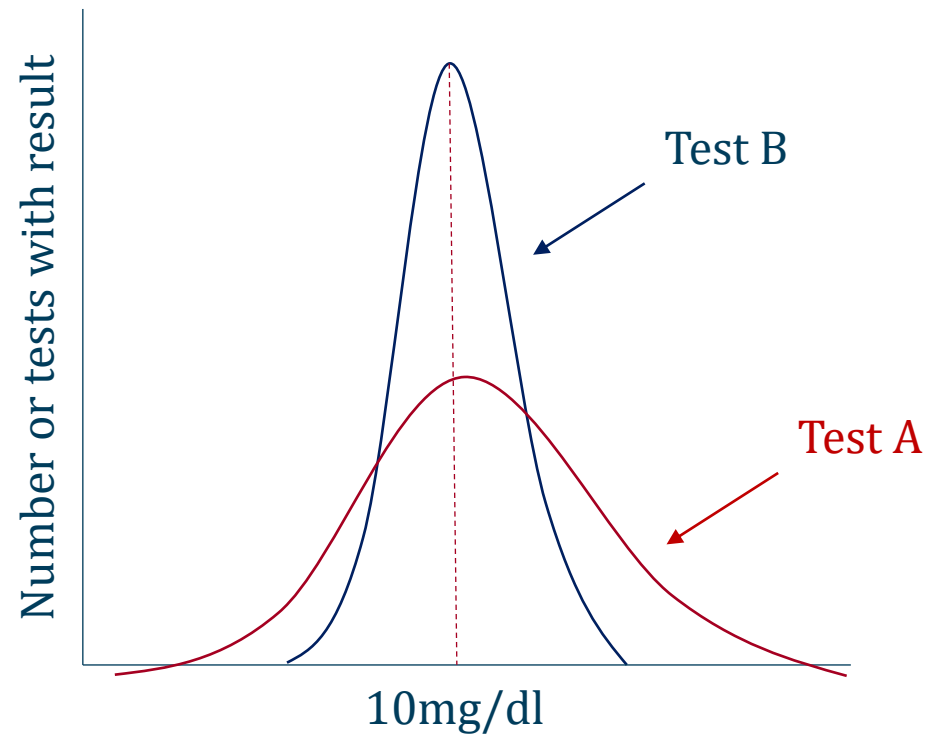
Accurate not precise



Not accurate or precise

Accuracy and Precision

- More precise tests have smaller standard deviations
- Less precise tests have larger standard deviations



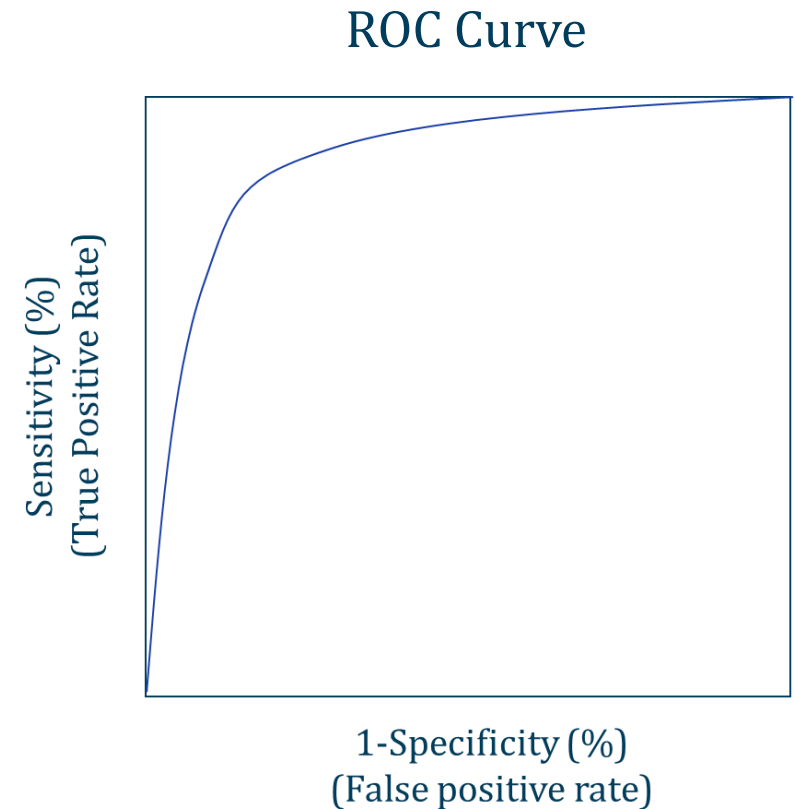
Accuracy and Precision

- **Random** measurement errors: reduce **precision**
 - Random error: some measurements okay, others bad
 - Accuracy may be maintained but lots of data scatter
- **Systemic** errors reduce **accuracy**
 - Imagine every BP measurement off by 10 mmHg due to wrong cuff size
 - Systemic error in data set (non-random error)
 - Precision maintained but accuracy poor

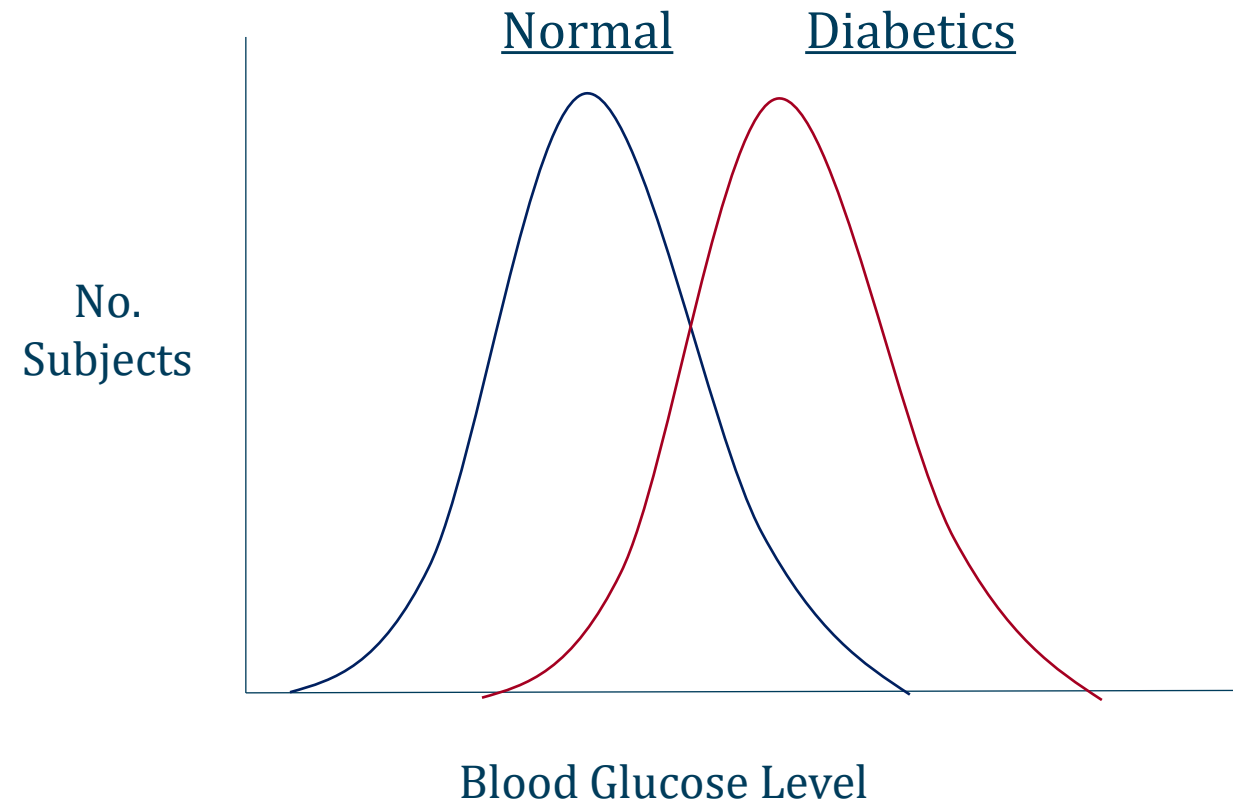
ROC Curves

Receiver Operating Characteristic

- Cutoff value for positive tests determines sensitivity/specificity
- Which cutoff value maximizes sensitivity/specificity?
- ROC curves answer this question



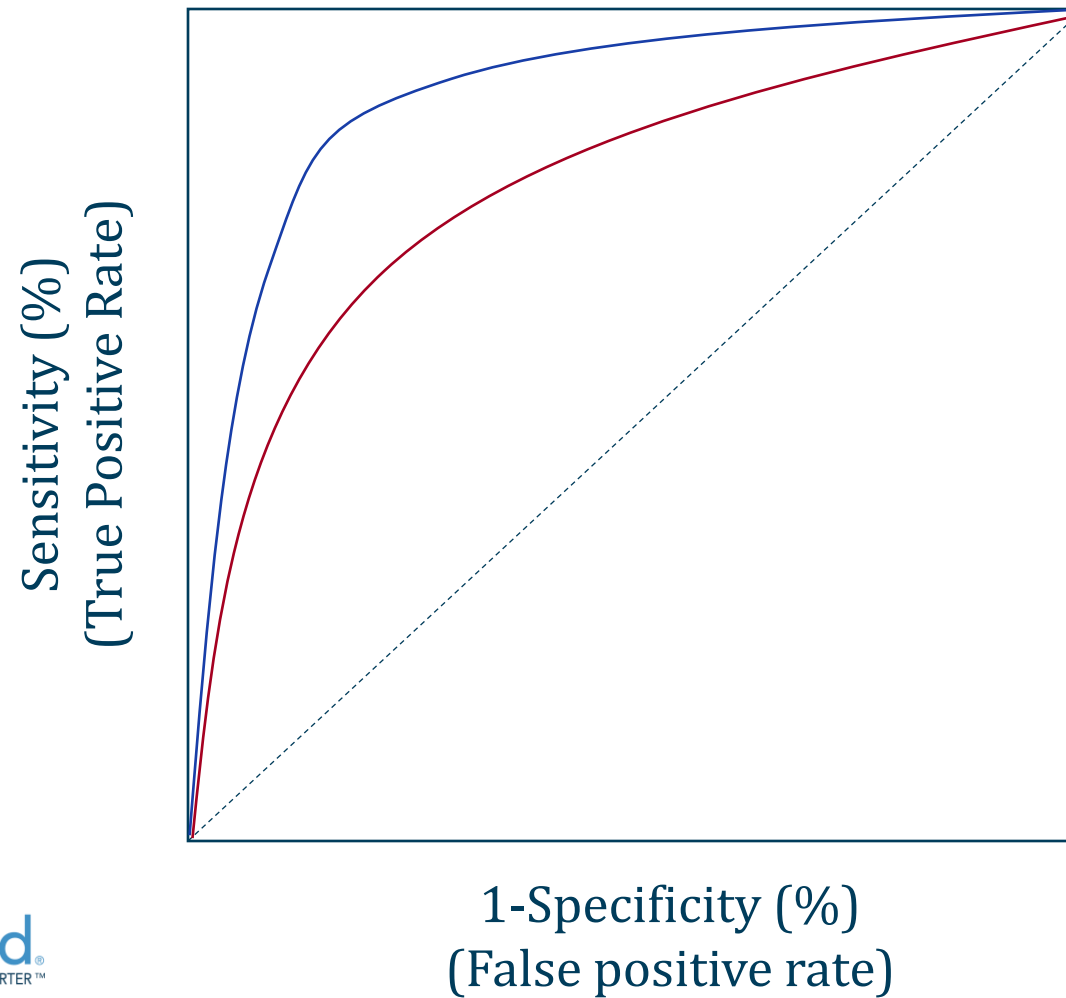
ROC Curve



ROC Curve

Cutoff (mg/dL)	Sensitivity (%)	Specificity (%)
100	80	20
110	65	42
120	54	63
130	42	75

ROC Curves

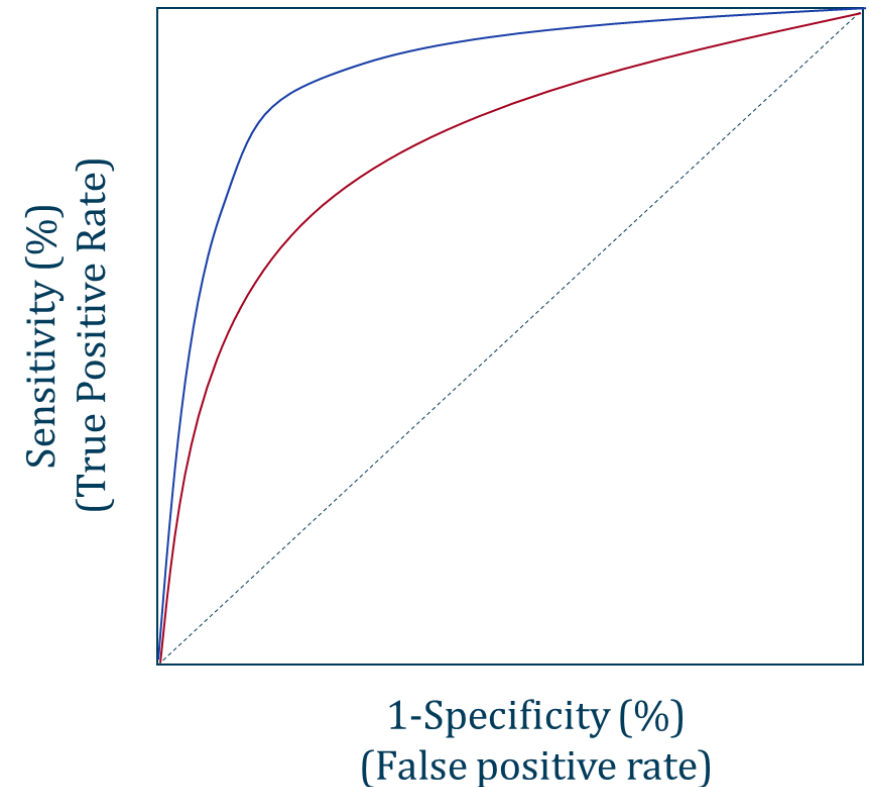


$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

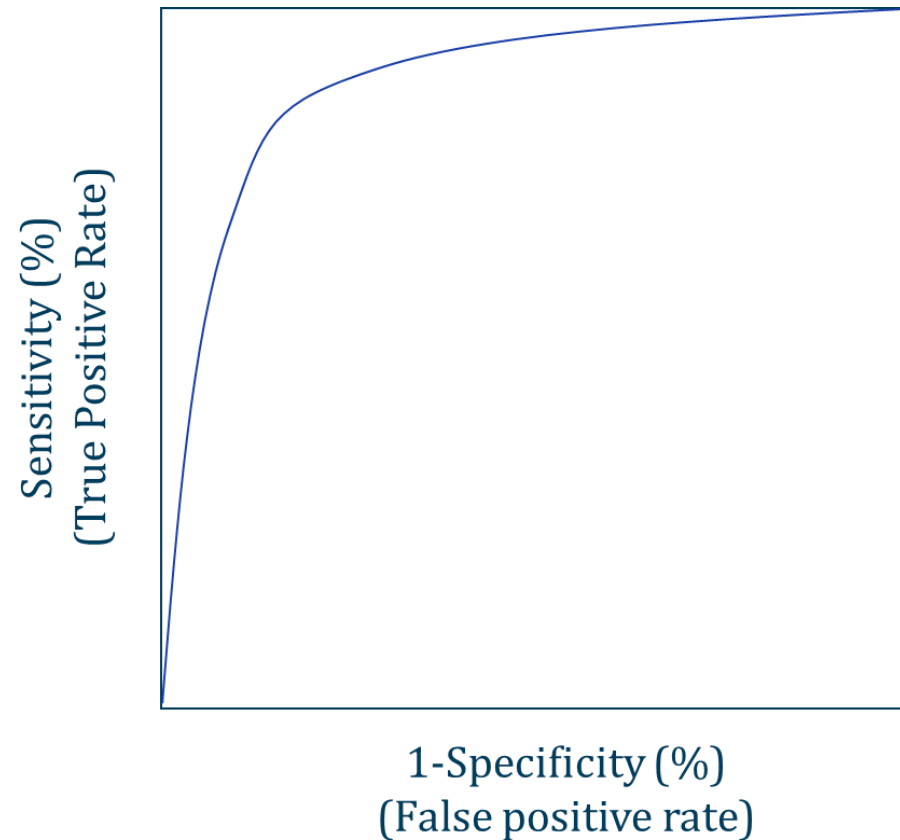
ROC Curves

- Straight line from bottom left to top right is a bad test
- Closer the curve is to a right angle, the better the test



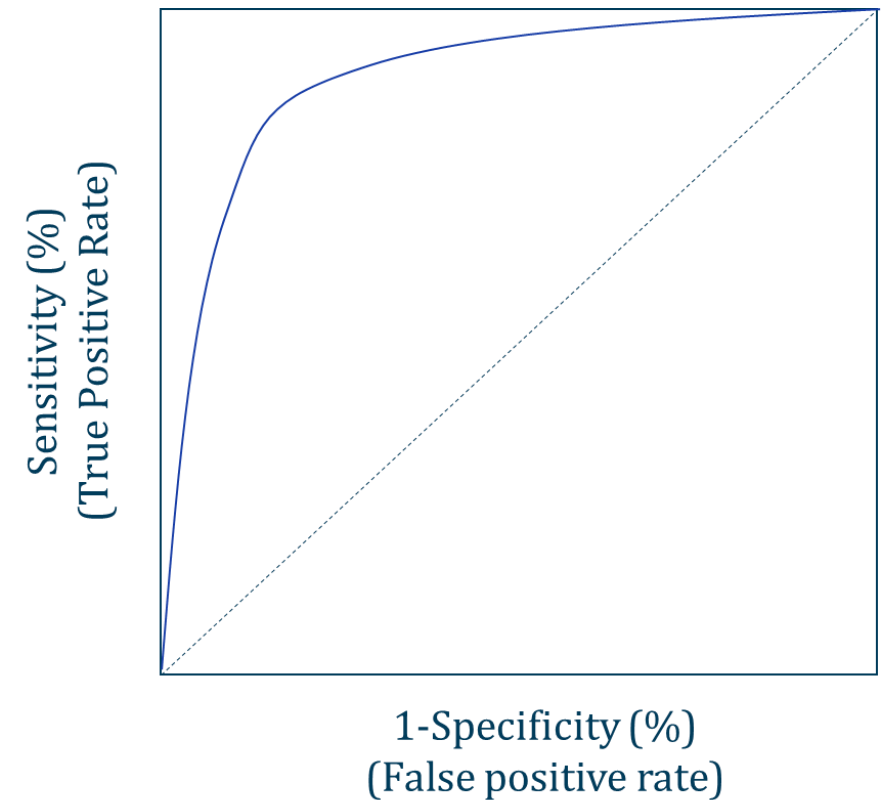
ROC Curves

- Point closest to top left corner is the best cutoff to maximize sensitivity/specificity

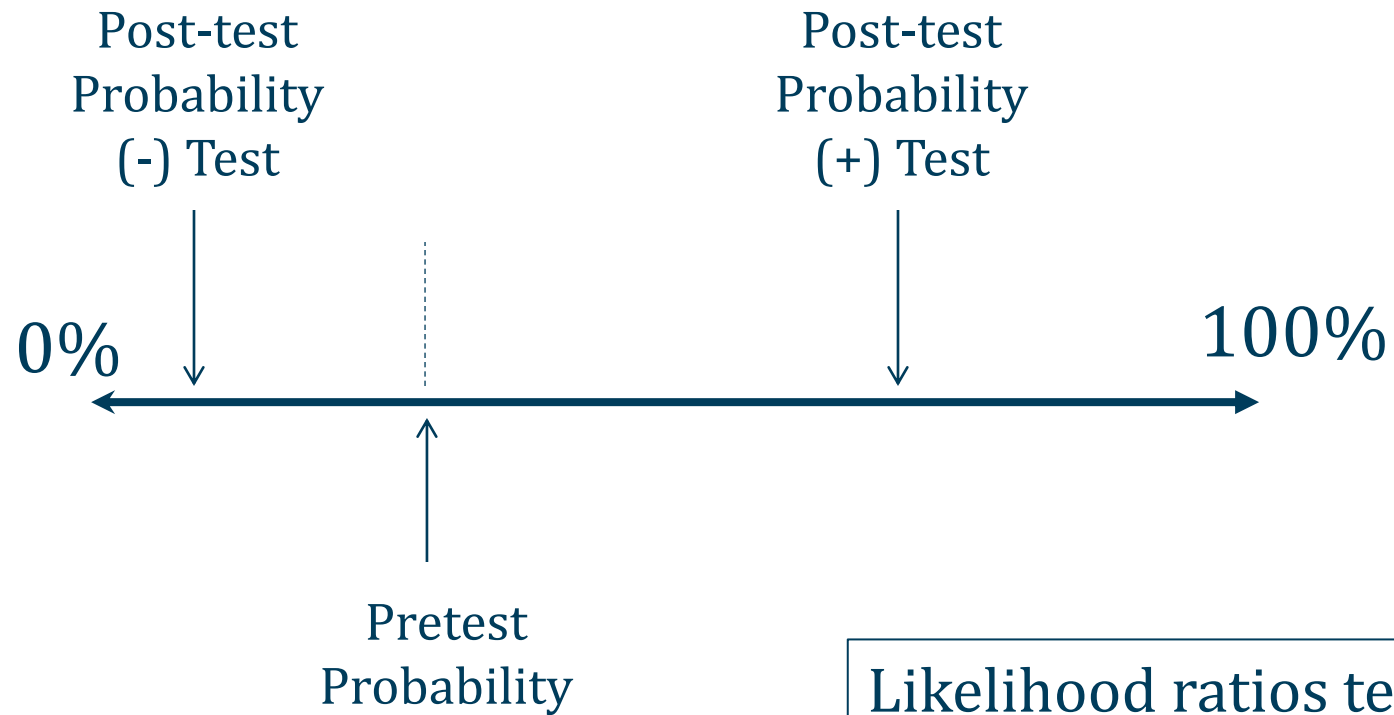


ROC: Area Under Curve

- Useless test has 0.5 (50%) area under curve
- Perfect test has 1.0 (100%) area under curve
- More area under curve = better test
- More ability to discriminate healthy from disease



Likelihood Ratios



Likelihood ratios tell us how much probability shifts with (+) or (-) test

Likelihood Ratios

$$LR^+ = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

$$LR^- = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

Characteristics of test like sensitivity/specificity
Do not vary with prevalence of disease

Likelihood Ratios

LR	Interpretation
> 10	Large increase in probability
1	No change in probability
< 0.1	Large decrease in probability

EXAMPLE

Likelihood Ratios

- Serum ferritin as a screening test for iron deficiency in children

Serum Ferritin (ng/dL)	Likelihood Ratio
< 15	51.8
15 to 24	8.8
25-34	2.5
45-100	0.5
>100	0.08

Term: “Likelihood”

- What is likelihood of disease in a person with (+) test?
 - Positive predictive value
- What is likelihood of disease in a person with (-) test?
 - Negative predictive value
- What is the positive likelihood ratio?
 - Calculated from sensitivity/specificity
- What is the negative likelihood ratio?
 - Calculated from sensitivity/specificity



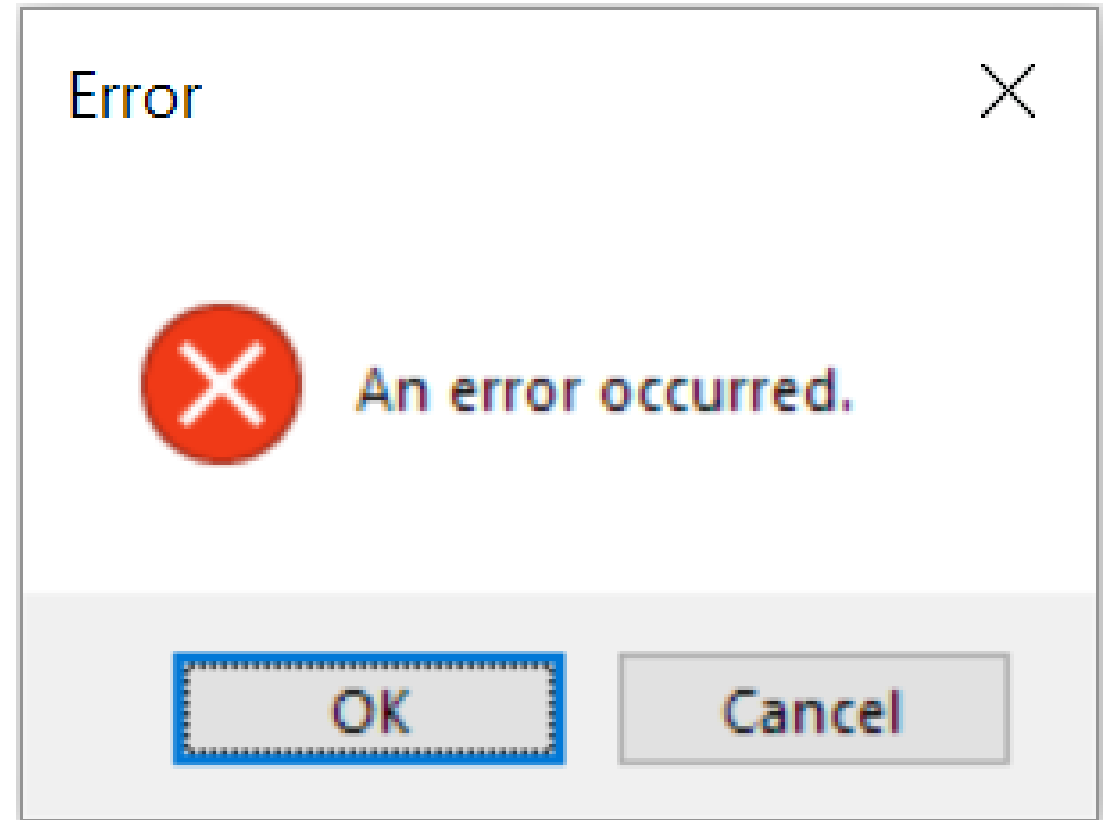
Bias

Jason Ryan, MD, MPH



Bias

- Systematic error in a study
- Group selection
- Measurement
 - Outcome assessment
 - Exposure assessment
- Confounding



Selection Bias

- Errors in **selection** or retention of study groups
- Usually used as a general term
- Many subtypes

Sampling Bias

- Subtype of selection bias
- Patients **not representative of actual practice**
- Results not generalizable to clinical practice
- Example: average age many heart failure trials = 65
- Average age actual heart failure patients = 80+
- Study results may not apply



Attrition Bias

- Subtype of selection bias
- Occurs in prospective studies
- Patients **lost to follow-up** unequally between groups
- Suppose 100 smokers lost to follow-up due to death
- Study may show smoking not harmful



Berkson's Bias

- Subtype of selection bias
- **Hospitalized patients** chosen as treatment or control arm
- May have more severe symptoms
- May have better access to care
- Alters results of study
- Example:
 - Study of hospitalized pneumonia patients
 - Shows many patients have history of stroke
 - Pneumonia associated with stroke
 - Similar study in community: no association



Nonresponse Bias

Participation Bias

- Subtype of selection bias
- Occurs with survey and questionnaire studies
- Non-responders not included
- Patients who respond may represent a selected group



Prevalence Bias

Neyman Bias

- Subtype of selection bias
- Exposure occurs long before disease assessment
- Patients exposed who die quickly not included
- Prevalence of disease based on select group of survivors

Exposure

Assessment



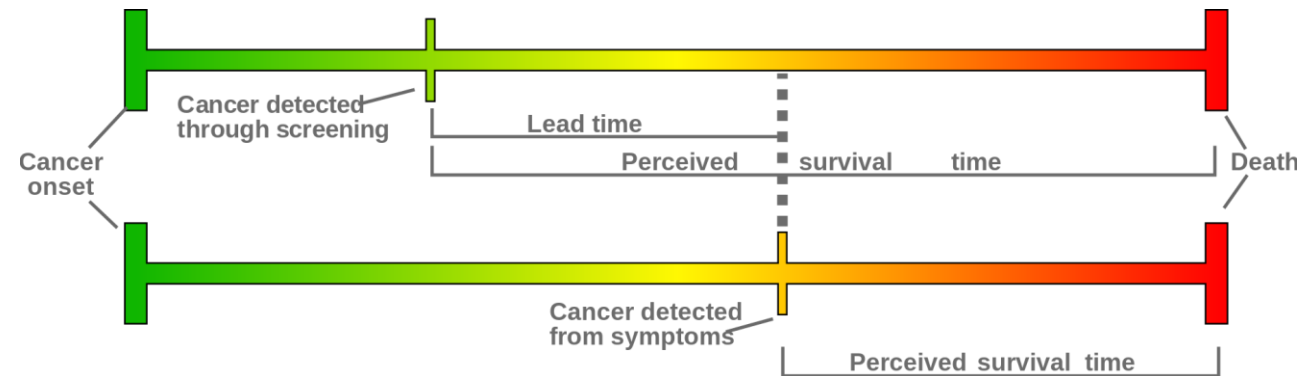
Length-time Bias

- Patients with severe disease do not get studied because they die
- Example: analysis of HIV+ patients shows the disease is asymptomatic
 - May overestimate survival because severe cases missed
- Example: screening program identifies only slow-growing tumors
 - Screening programs may appear more effective



Lead-Time Bias

- Screening test identifies disease earlier
- Survival appears longer when it is not
- Example:
 - Average time from detection of breast lump to death = 5 years
 - Screening test identifies cancer earlier
 - Time from detection to death = 7 years
- Avoided by controlling for disease severity
 - Expected survival based on stage at detection



Measurement Bias

- Sloppy research technique
- Blood pressure measured incorrectly in one arm
- Protocol not followed
- Avoided by **standardized data collection**
 - Objective, previously-tested methods
 - Carefully planned ahead of time



Recall Bias

- Form of **measurement bias**
- Inaccurate recall of past events by study subjects
- Common in **survey studies**
- Example:
 - Parents of disabled children asked about lifestyle during pregnancy
 - Pregnancy occurred many years ago
 - Poor recall leads to inaccurate findings
- Avoided by minimizing recall timeframe



Observer Bias

- Form of **measurement bias**
- Investigators know exposure status of patient
- Examples:
 - Cardiologist interprets EKGs knowing patients have CAD
 - Pathologist reviews specimens knowing patients have cancer
- Avoided by **blinding**



Procedure Bias

- One group receives procedure (e.g., surgery) and another does not
- More care and attention given to procedure patients
- Avoided by blinding (masking)
 - Care team unaware which patients had procedure
- Also avoided by using placebo
 - Sometimes sham surgery performed



Confounding Bias

- **Unmeasured factor** confounds study results
- Example:
 - Alcohol users more likely to develop lung cancer than non-users
 - Smoking much more prevalent among alcohol users
 - Smoking is true cause of more cancer
 - Smoking is a confounder of results

Stratified Analysis

Eliminates Confounding Bias

		Lung Cancer	
		+	-
Alcohol Use	+	50	50
	-	10	90

RR = 5

		Lung Cancer	
		+	-
Alcohol Use	+	15	35
	-	15	35

RR = 1

		Lung Cancer	
		+	-
Alcohol Use	+	15	35
	-	15	35

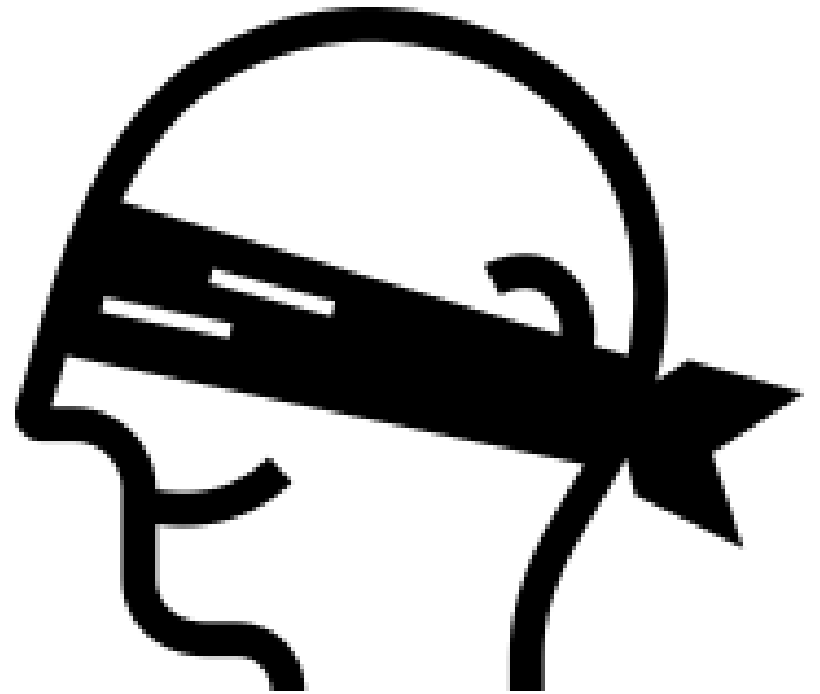
RR = 1

Controlling for Confounders

- **Randomization**
 - Ensures equal variables in both arms
- **Matching**
 - Case-control studies
 - Careful selection of control subjects
 - Goal is to match case subjects as closely as possible
 - Choose patients with same age, gender, etc.

Ways to Reduce Bias

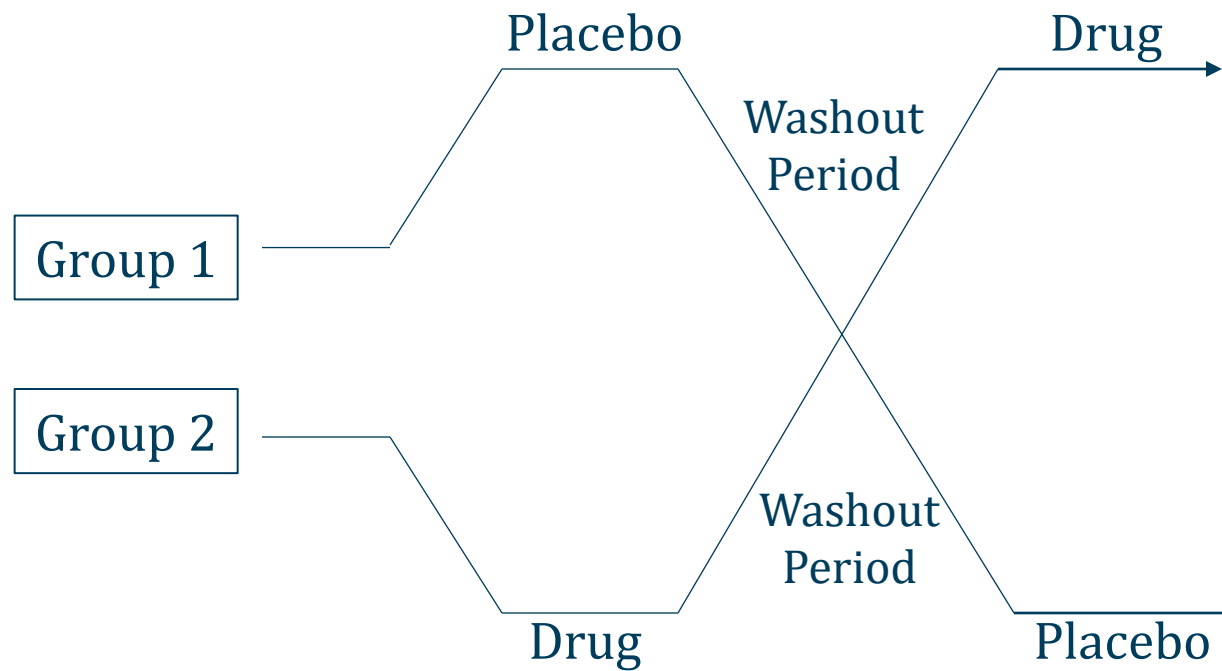
- Randomization - limits confounding and selection bias
- Matching of groups
- Blinding
- Crossover studies



Crossover Study

- Subjects randomly assigned to a **sequence** of treatments
- Group A: placebo 8 weeks → drug 8 weeks
- Group B: drug 8 weeks → placebo 8 weeks
- Subjects serve as their own control
- Avoids confounding (same subject!)
- Drawback is that effect can “carry over”
- Avoid by having a “wash out” period

Crossover Study



Effect Modification

- Not a type of bias (point of confusion)
- Occurs when third factor alters effect
- Consider:
 - Drug A is shown to increase risk of DVT
 - To cause DVT, Drug A *requires* Gene X
 - Gene X is an effect modifier

Effect Modification

Stratified Analysis

		DVT	
		+	-
Drug A	+	50	50
	-	10	90

RR = 5

Gene X (+)

		DVT	
		+	-
Drug A	+	25	25
	-	5	45

RR = 5

Gene X (-)

		DVT	
		+	-
Drug A	+	15	35
	-	15	35

RR = 1

Confounding vs. Effect Modification

- **Confounding:**
 - A 3rd variable *distorts* the effect on outcome
 - Smoking and alcohol
 - Alcohol appears associated with cancer (positive)
 - Real effect of exposure on outcome distorted by confounder
- **Effect modification:**
 - A 3rd variable *maintains* effect but only in one group
 - There is a real effect of exposure on outcome
 - Effect requires presence of 3rd variable

Confounding vs. Effect Modification

Example

- Patients taking drug A have increased rates of lung cancer
- Drug A is taken mostly by smokers
- Breakdown data into smokers and non-smokers:
 - NO relationship between Drug A and cancer
 - Smoking is the real cause
 - Drug A has no effect
 - This is confounding

Confounding vs. Effect Modification

Example

- Patients taking drug A have increased rates of lung cancer
- Drug A activates gene X to cause cancer
- Breakdown data into gene X (+) and (-)
 - Relationship exists between Drug A and cancer but only in gene X (+)
 - Drug A *does* have effect (different from confounding)
 - But drug A requires another factor (gene X)
 - This is effect modification (not a form of bias)

Hawthorne Effect

- Type of measurement bias
- Study patients improve because they are being studied
- Patients or providers change behavior based on being studied
- Common in studies of **behavioral patterns**
- Example:
 - Physicians know patients surveyed about vaccination status
 - Physicians vaccinate more often
- Example:
 - Patients being studied for exercise capacity
 - Patients exercise more often

Pygmalion Effect

Observer-expectancy effect

- Researcher believes in efficacy of treatment
- Influences outcome of study
- Example:
 - Creator of a new surgical device overseeing study
 - Creator assesses outcomes positively



Pygmalion vs. Hawthorne

- Pygmalion effect
 - Provider believes in treatment
 - Influences results to be positive
 - Pygmalion unique to *investigator driving positive benefit*
- Hawthorne Effect
 - Subjects/investigators behave differently because of study

Clinical Trials

Jason Ryan, MD, MPH



Clinical Trials

- Observational studies: no control over exposure
 - Cohort, case-control
- Experimental studies: researchers control exposure
- Goal is to determine benefit of therapy
- Drugs
- Surgery



Clinical Trial Features

- Control
- Randomization
- Blinding



Control

- One group receives therapy
- Other group no therapy (control group)
- Compared changes in therapy group to control group



Randomization

- Subjects **randomly assigned** to treatment or control
- All variables other than treatment should be equal
- Should eliminate confounding
 - All potential confounders (age, weight, blood levels) should be equal in both arms
- Limits selection bias
 - Patients cannot choose to be in drug arm of study
- Table 1 in most studies demonstrates randomization

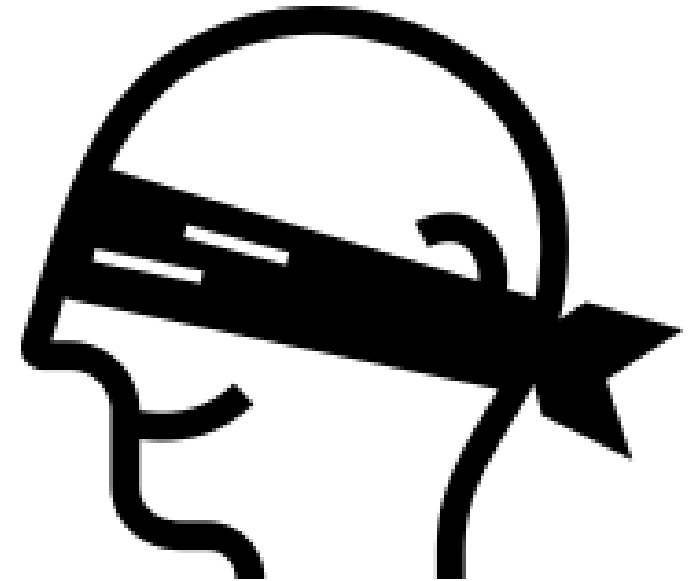
Table 1

- Non-significant p values indicated **randomization was successful**

	Intervention	Control	p value
Male (%)	49%	51%	NS
Age (mean)	64	65	NS
Diabetes (%)	10	11	NS
Systolic BP (mean)	121	119	NS

Blinding

- Treatment subjects given therapy/drug
- Control subjects given placebo
- Subjects **unaware** if they are getting treatment or not
- Single blind: subjects unaware
- Double blind: subjects and providers unaware
- Triple blind: subjects, providers, data analysts unaware



Clinical Trials

- Best evidence of efficacy comes from **randomized, controlled, blinded studies**
- Why not do these for everything?
- Takes a **long time**
- By end of study, new treatments sometimes have emerged
- Costs a **lot of money**



Parachute Example

- No clinical data exists showing parachutes are effective compared to placebo



Data from Clinical Trials

- Drug X → 30% mortality over 3 years
- Placebo → 50% mortality over 3 years
- Several ways to report this:

$$\text{Absolute Risk Reduction} = 50\% - 30\% = 20\%$$

$$\text{Relative Risk Reduction} = \frac{50\% - 30\%}{50\%} = 40\%$$

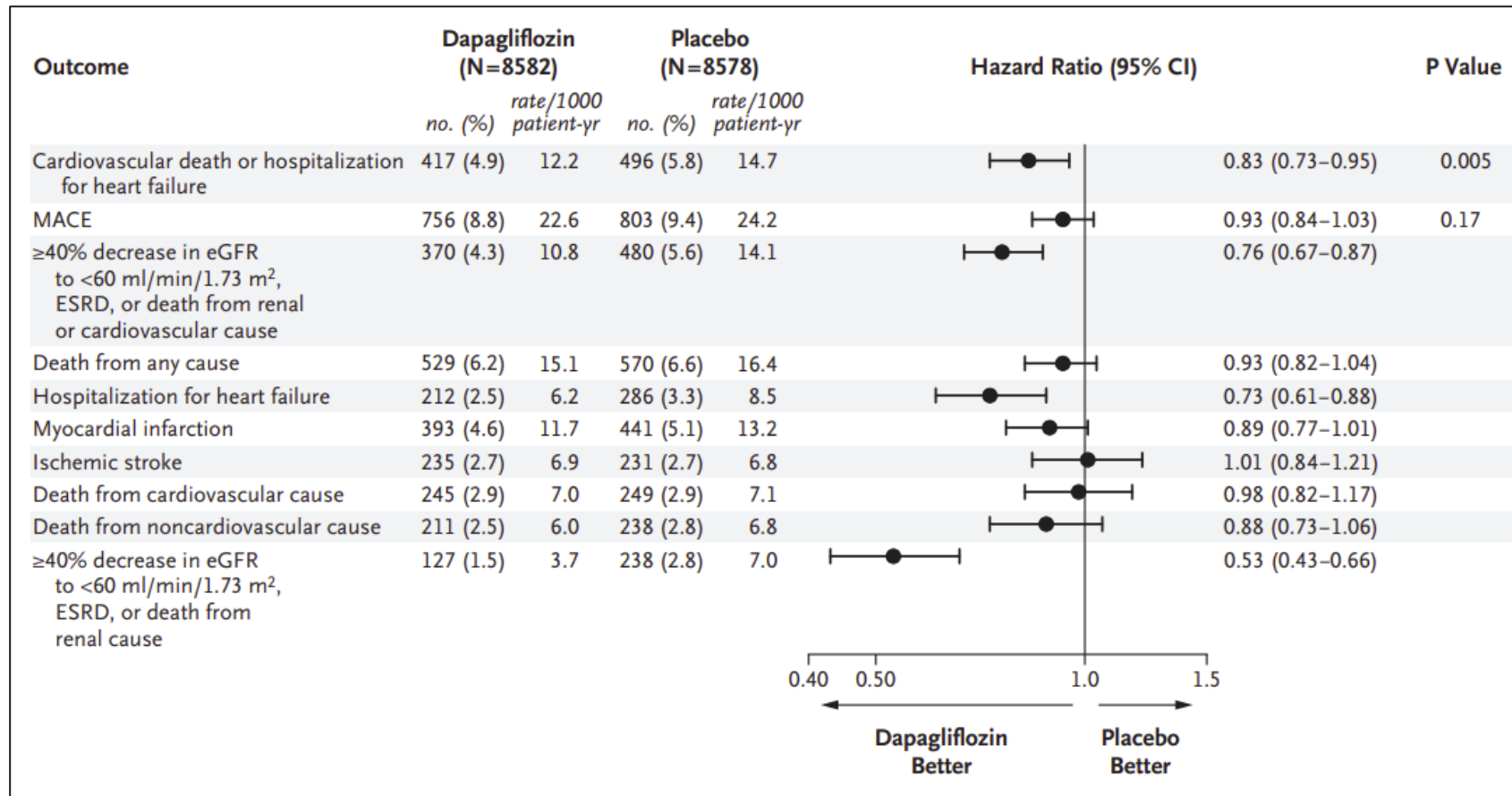
Number Needed to Treat

NNT

- Drug X → 30% mortality over 3 years
- Placebo → 50% mortality over 3 years
- Number need to treat to prevent **one outcome**

$$\text{Number Needed to Treat} = \frac{1}{\text{ARR}} = \frac{1}{0.2} = 5$$

Number Needed to Treat



Clinical Trial Results

Interpretation

- Is the result **clinically meaningful**?
 - Drug reduces blood pressure by 2 mmHg ($p < 0.05$)
 - Average HTN patient is 15 mmHg above goal
- Was the **population representative** of actual practice?
 - Patients described in Table 1 – are they similar to actual practice?
- Does the drug improve a **meaningful outcome**?
 - Drug reduces biologic activity of cancer cells
 - No change in mortality between groups

Clinical Trial Results

Outcomes

- Intervention group compared to control group
- What outcome (endpoint) will be assessed?
- Major element of interpreting study results

Clinical Trial

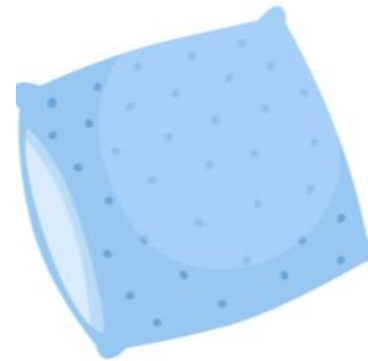
Outcomes

- **Continuous** – outcome exists on a continuum
 - Blood pressure, total cholesterol, weight
 - Mean values compared between groups
- **Categorical** – outcome exists in a category (yes/no)
 - Death
 - Hospitalization
 - Stroke
 - Myocardial infarction
 - Blood pressure < 140 mmHg
 - Compare percentage of patients with outcome in each group

Clinical Trial

Outcomes

- **Soft:** quality of life, reduction in pain
 - Not directly harmful
- **Hard:** hospitalization, stroke, myocardial infarction, death
 - Directly harmful



Soft



Hard

Surrogate Outcomes

- Soft outcomes that are ***predictive*** of hard outcomes
 - Systolic blood pressure → stroke
 - Hemoglobin a1c level → diabetes complications
- **Advantages**
 - Easier to obtain
- **Disadvantages**
 - May lead to erroneous findings



Clinical Trial

Outcomes

- **Combined endpoint**
 - Often used in trials with categorical outcomes (death, MI, stroke)
 - Imagine death as endpoint in trial of 2000 patients
 - After two years: 5 deaths total – no significant difference between groups
 - Imagine death, hospitalization and stroke as endpoint in trial of 2000 patients
 - After two years: 375 endpoints – significant difference between groups
- Many trials used combined endpoints to allow faster studies

Clinical Trial

Outcomes

- **Primary endpoint**
 - Outcome study is designed to evaluate
 - Endpoint used to determine power of trial
 - Trial power determines sample size needed
- **Secondary endpoint**
 - Outcomes of interest
 - Not used for power calculation or sample size determination
 - Interpret with caution



Negative Clinical Trials

- Difference between groups not statistically significant
- Must consider **power of study**
 - Power = chance of finding a difference when one exists
 - Or chance of rejecting no difference because there really is one
- Power increases with **sample size**
 - Small studies are “underpowered”
 - Cannot “detect small differences”
 - P value will be nonsignificant for small differences



Non-Inferiority Trials

- **New treatment** compared to **standard of care**
- Goal is to show new treatment has similar outcomes to standard of care
- Used when a placebo group may be unethical
- Standard treatment well-established and effective
- Or when improvement with new treatment may be small

Non-Inferiority Trials

- Study designed by choosing difference in outcomes that is acceptable (Δ)
- Null hypothesis: between group differences $> \Delta$ (one group superior)
- Alternative hypothesis: between group differences $\leq \Delta$ (non-inferiority)
- P-value < 0.05 = reject null hypothesis = new treatment non-inferior

Non-Inferiority Trials

- Apixaban versus dalteparin
- Recurrent thromboembolism:
 - 5.6% in apixaban group
 - 7.9% in dalteparin group
- $P < 0.001$ for noninferiority

ORIGINAL ARTICLE

Apixaban for the Treatment of Venous Thromboembolism Associated with Cancer

Giancarlo Agnelli, M.D., Cecilia Becattini, M.D., Guy Meyer, M.D., Andres Muñoz, M.D., Menno V. Huisman, M.D., Jean M. Connors, M.D., Alexander Cohen, M.D., Rupert Bauersachs, M.D., Benjamin Brenner, M.D., Adam Torbicki, M.D., Maria R. Sueiro, M.D., Catherine Lambert, M.D., Gualberto Gussoni, M.D., Mauro Campanini, M.D., Andrea Fontanella, M.D., Giorgio Vescovo, M.D., and Melina Verso, M.D.,
for the Caravaggio Investigators*

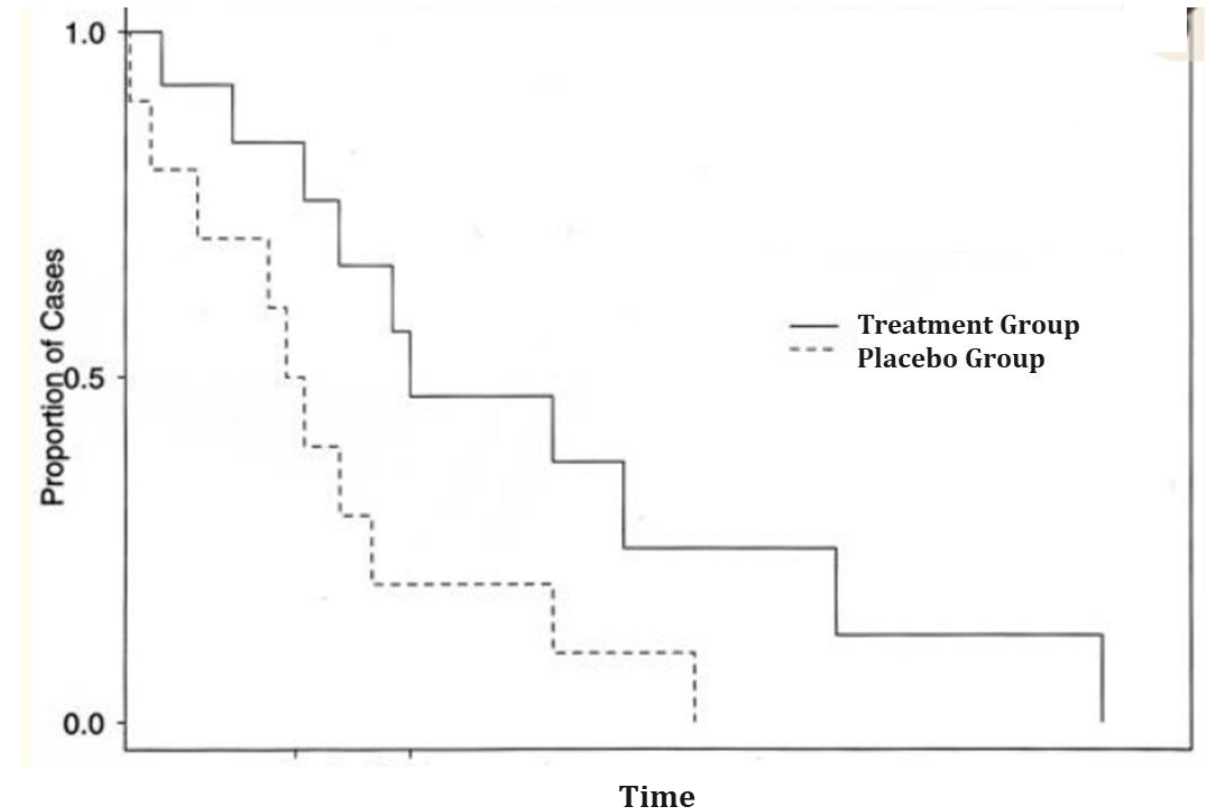
ABSTRACT

BACKGROUND

Recent guidelines recommend consideration of the use of oral edoxaban or rivaroxaban for the treatment of venous thromboembolism in patients with cancer. However, the benefit of these oral agents is limited by the increased risk of bleeding associated with their use.

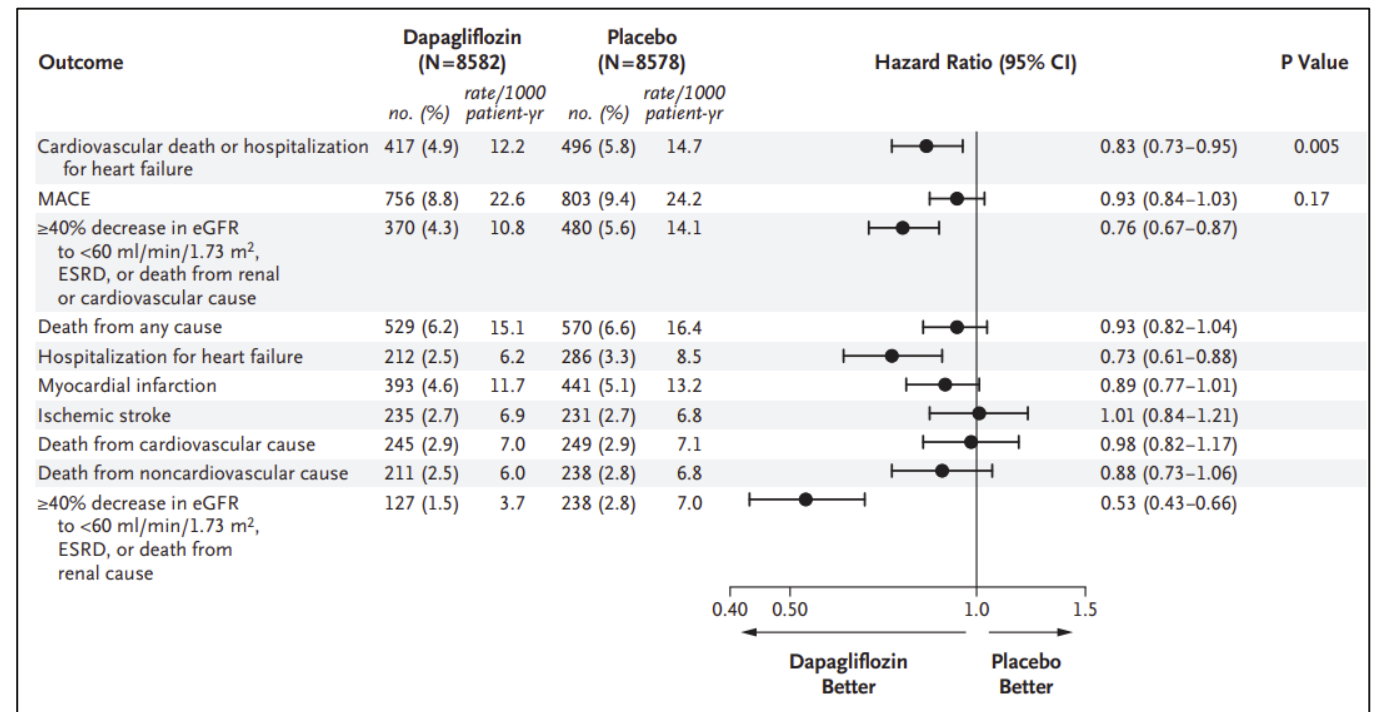
Kaplan-Meier Curves

- Time-to-event curves
- Proportion of patients without event over time



Hazard Ratios

- Probability of events in treatment group compared to control group
 - < 1 = event less likely in treatment group
 - > 1 = event more likely in treatment group

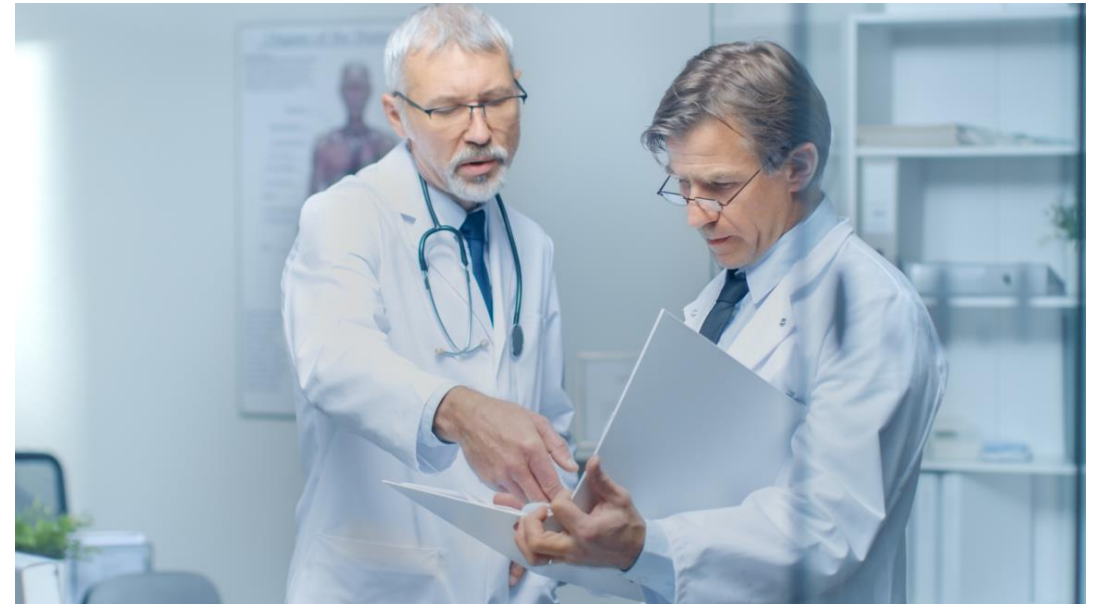


Intention to Treat Analysis

- Subjects analyzed according to the group they were **originally assigned**
- Regardless of whether they received treatment or not
- Not affected by crossover or dropout
- Patients in control group may require treatment (crossover)
 - Still analyzed as members of the control arm
- Patient in treatment group may be unable to comply with treatment (crossover)
 - Still analyzed as members of treatment arm
- Patients in treatment arm may dropout of study
 - Still analyzed as members of treatment arm

Meta Analyses

- Pools data from **several studies together**
- Increases number of subjects and controls
- Increases statistical power
- Limited because pooled studies often differ
 - Selection criteria
 - Exact treatment used



New Drug Approval

- Clinical trials conducted in phases
- **Phase 1**
 - Small number of healthy volunteers
 - Safety, toxicity, pharmacokinetics
- **Phase 2**
 - Small number of sick patients
 - Efficacy, dosing, side effects
 - Often placebo controlled, often blinded



New Drug Approval

- **Phase 3**
 - Large number of sick patients
 - Many patients, many centers
 - Randomized trials
 - Drug efficacy determined vs. placebo or standard care
- After phase 3, drug may be approved by FDA



Phase 4

- Post-marketing study
- After drug is on the market and being used
- Monitor for long-term effects



Evidence-Based Medicine

Jason Ryan, MD, MPH



Evidence-Based Medicine

- Caring for patients using best-available research
- Four basic elements:
 - Formulating a **clinical question**
 - Identifying best available evidence
 - Assessing **validity of evidence**
 - Applying the evidence in practice



Clinical Questions

- Should be focused
- Should be answerable from research literature
- **PICO** model
 - What is the **patient** population?
 - What **intervention** is being considered?
 - What is the **comparison** intervention or population?
 - What **outcomes** are important?

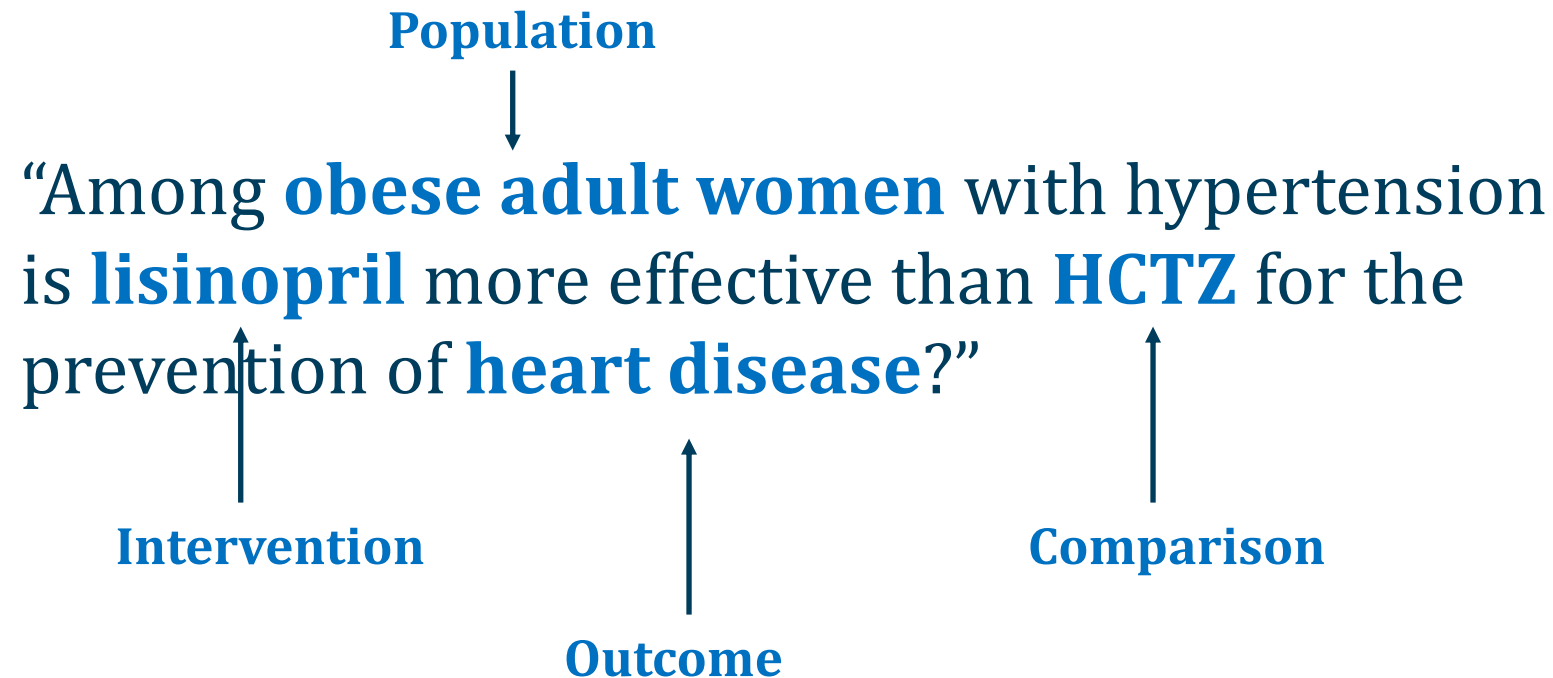
PICO

Bad Clinical Question

- “Do ACE inhibitors work for hypertension?”
- Vague
- No population
- No specific outcome

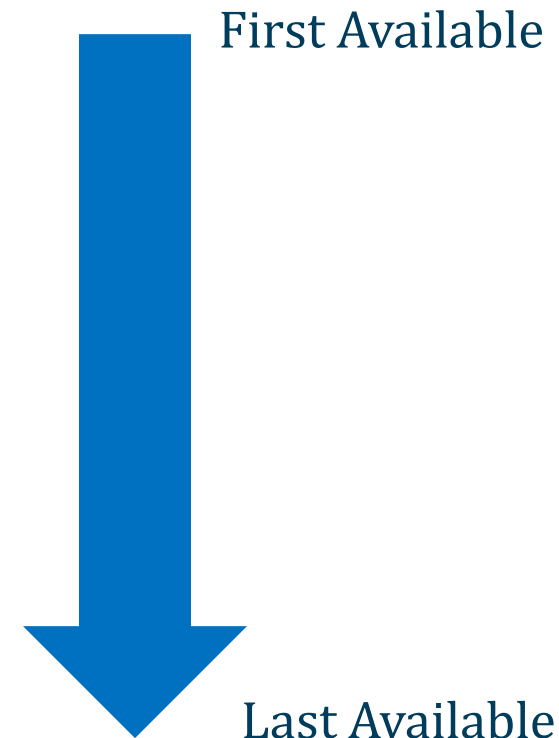


Good Clinical Question

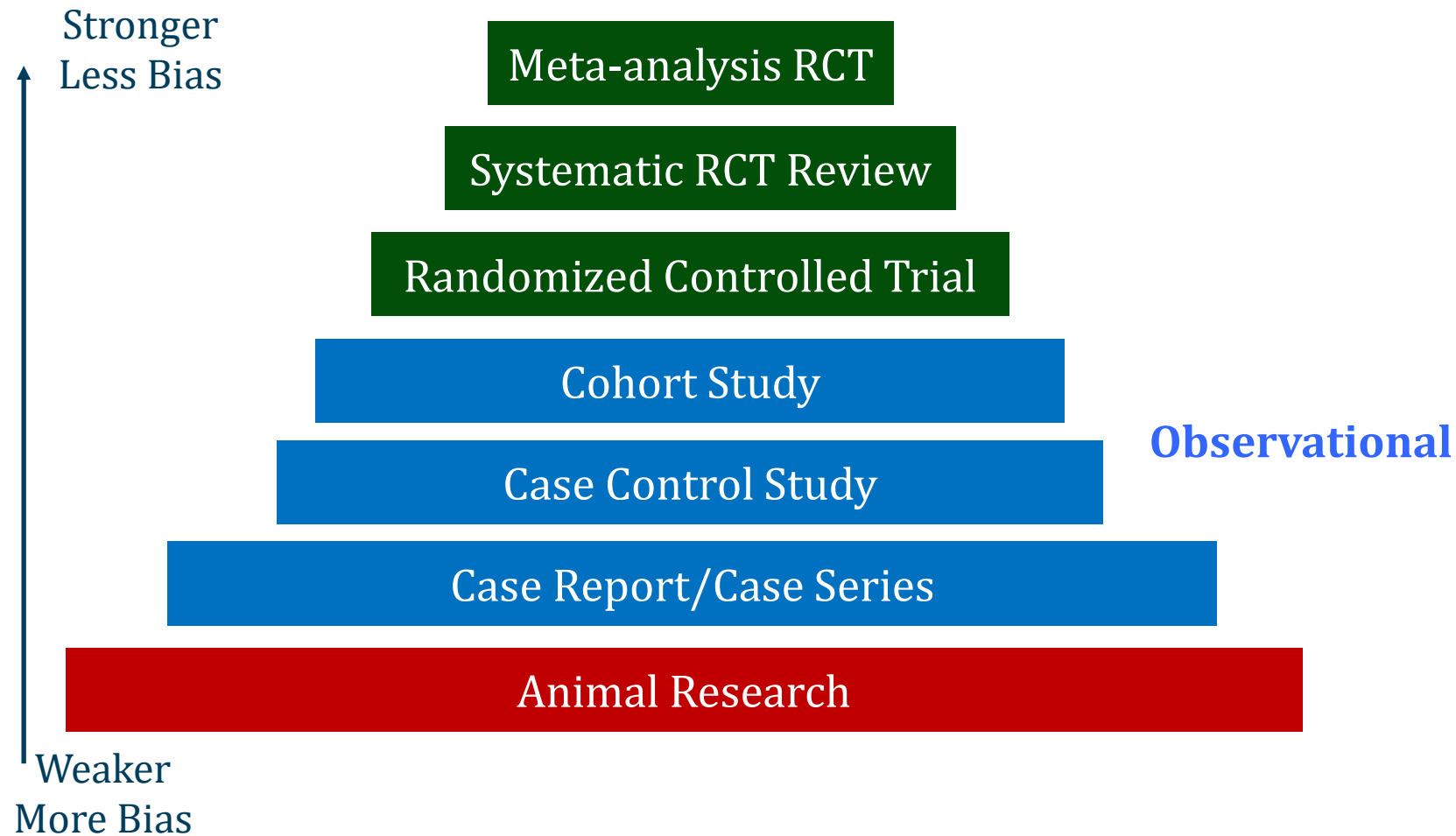


Types of Evidence

- **Primary resources**
 - Case reports/series
 - Observational studies
 - Randomized clinical trials (best)
- Systematic reviews/meta-analysis
 - Compilation of primary studies
- Society guidelines
 - Written based on multiple sources
 - Primary data, systematic reviews
 - Clinical expertise, patient preferences



Types of Evidence



Evaluating Evidence

- **Internal validity**
 - Was the research conducted properly?
 - Are the conclusions correct?
 - Is there bias?
 - Are results due to chance?



Evaluating Evidence

- **External validity**
 - Does the research apply to patients not in study?
 - Are study patients similar to real world patients?
 - Is the intervention similar to real world interventions?
 - Does this apply to the patient in my clinical question?



Evidence-Based Medicine

- Must also apply clinical expertise and patient's wishes

