Introduction to Clinical Epidemiology

Kristen A. Stafford, MPH, PhD
Associate Professor
Department of Epidemiology and Public Health

Special thanks to Dr. Mona Baumgarten for the use of her slides

What is Epidemiology?

*Epidemiology is the **study** of the **distribution** and **determinants** of **health-related states or events** in **specified populations**, and the **application** of this study to the control of health problems*

Dictionary of epidemiology. 4th ed. 2001

What is Clinical Epidemiology?

Clinical Epidemiology is the application of principles of epidemiology to clinical medicine. While classical epidemiology is the study of the distribution and determinants of diseases in populations, clinical epidemiology is the application of the principles and methods of epidemiology to conduct, appraise or apply clinical research studies focusing on prevention, diagnosis, prognosis, and treatment of disease. Clinical Epidemiology is the basic science of Evidence-based Medicine.

Outline

- Measures of disease occurrence
- Measures of association

“We owe all the great advances in knowledge to those who endeavor to find out how much there is of anything.”

- James Maxwell, physicist (1831-1879)

“One’s knowledge of science begins when he can measure what he is speaking about and express it in numbers”

- Lord Kelvin, mathematician and physicist (1824 – 1907)

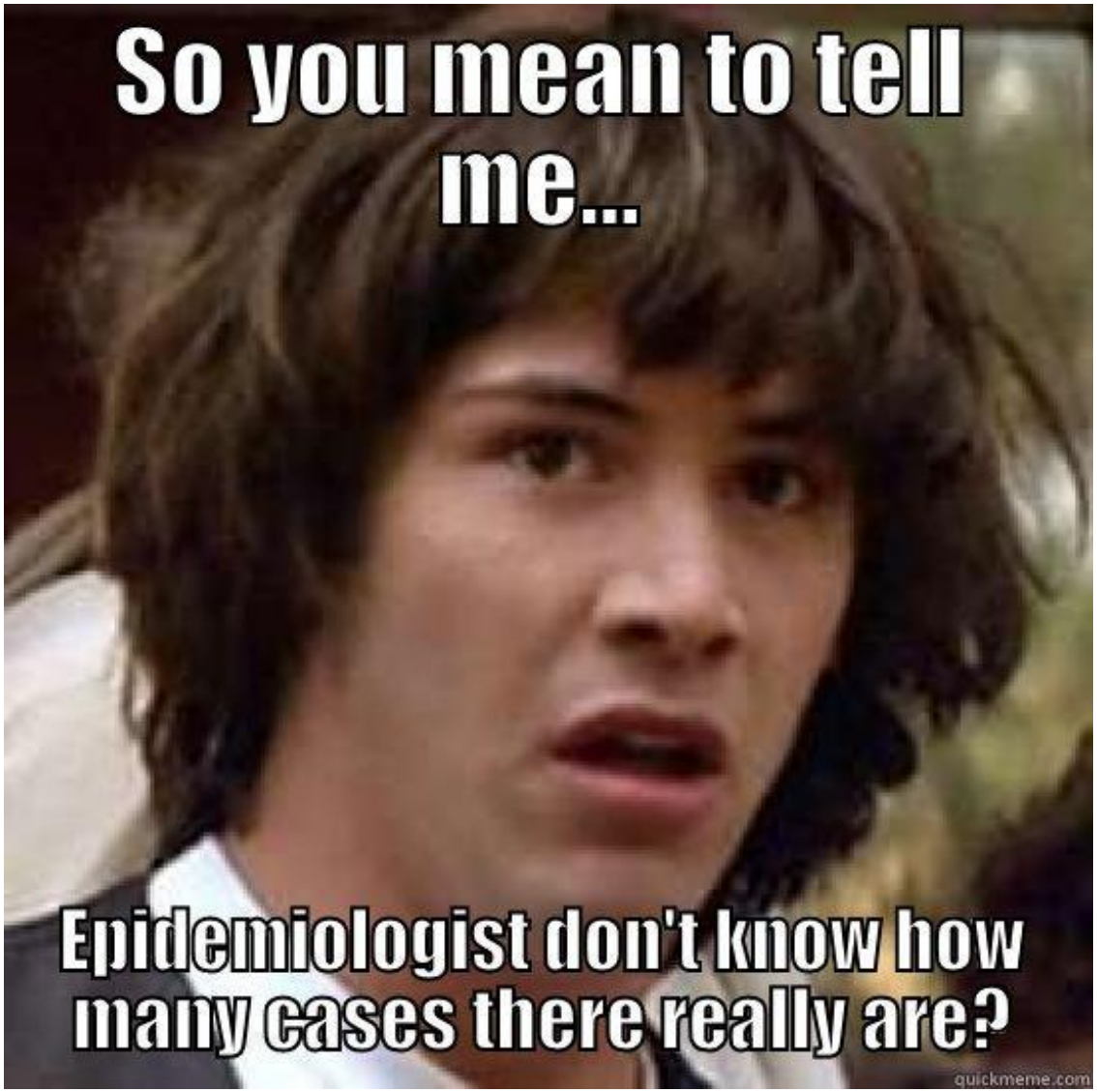
Measures of Disease Occurrence

- Prevalence
- Cumulative incidence
- Incidence rate

The occurrence of disease can be measured using rates or proportions

Rates tell us how “fast” a disease is occurring in a population

Proportions tell us what fraction of the population is affected


So you mean to tell me...
Epidemiologist don't know how many cases there really are?
quickmeme.com

Prevalence

$$\frac{\text{Number of diseased individuals in population at a specified time}}{\text{Total population at same specified time}}$$

Prevalence

- Is a proportion and therefore has no units
- Ranges from 0 to 1
- Numerator includes both new and ongoing cases of disease
- Represents a cross-sectional “snapshot” of the population

Prevalence

- Does not estimate risk of disease
- Is not useful for studies of risk factors
- Estimates burden of disease
- Is useful in planning of health services

Example

- 10,600 men age 50-59 years were examined in 2002 as part of a large heart health study
- 842 men were found to have coronary heart disease
- Prevalence of CHD =

$$0.079 \approx 8\%$$

Cumulative Incidence

$$\frac{\text{Number of new cases of disease during specified time period}}{\text{Number of individuals at risk of disease at start of time period}}$$

At Risk

Individuals are at risk of disease if they:

- Do not have the disease at the start of the follow-up period
- Are capable of developing the disease
  - Have the organ of interest

Cumulative Incidence

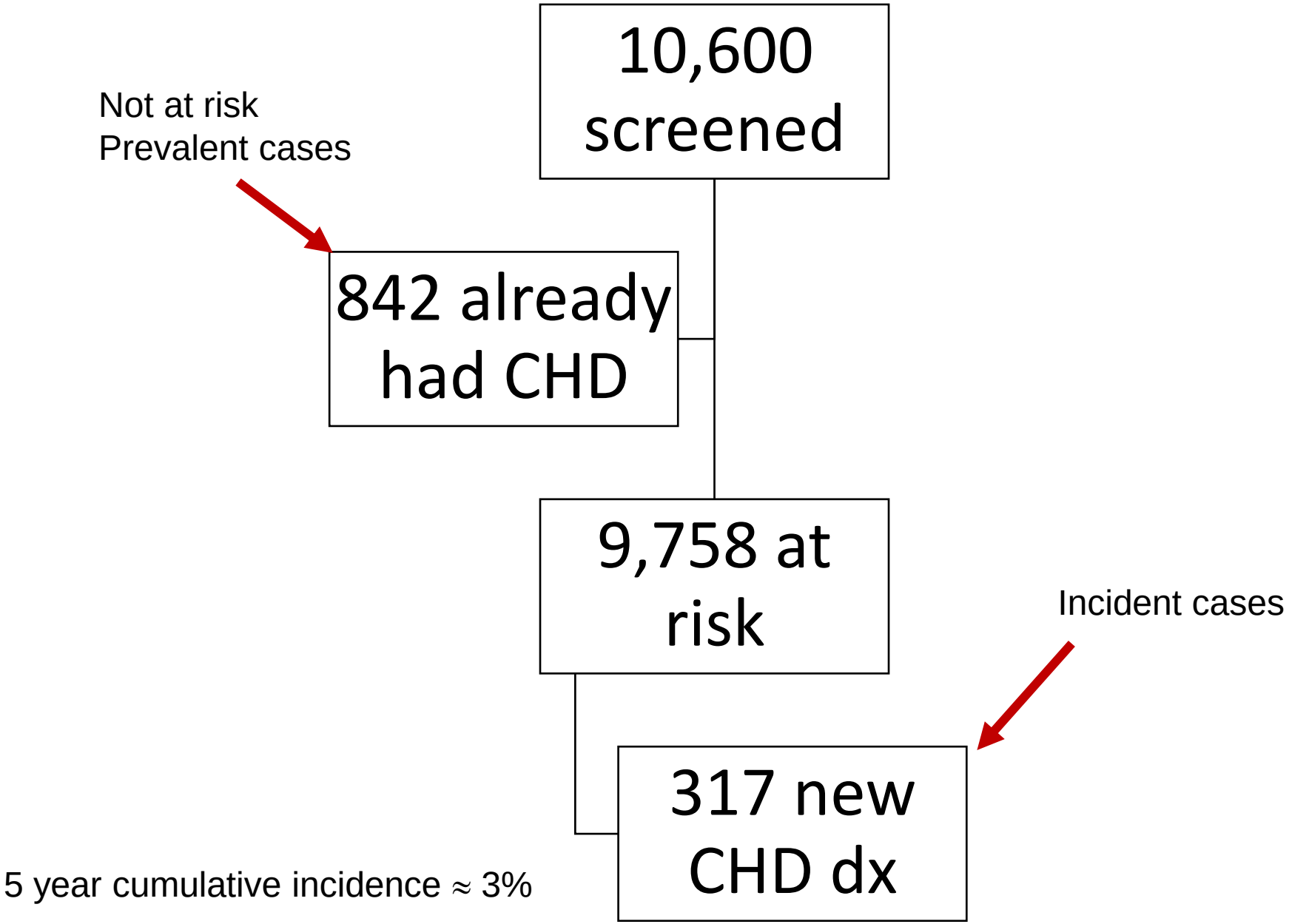
- Represents the probability that an individual will develop the disease over a specified time period
- Is a measure of disease risk
- Is a proportion and therefore has no units
- Ranges from 0 to 1

Cumulative Incidence

- Based on assumption that all at-risk individuals are followed until they develop the disease, or the observation period ends
- Does not reflect effect of differing lengths of follow-up
- *Syn.:* Incidence proportion

Example

- 10,600 men age 50-59 were examined in 2002 as part of a large heart health study
- 842 men were found to have coronary heart disease
- During the period 2002 to 2007, 317 men developed CHD
- Five-year cumulative incidence of CHD = ?


10,600 screened
Not at risk
Prevalent cases
842 already had CHD
9,758 at risk
Incident cases
317 new CHD dx
5 year cumulative incidence ≈ 3%

Incidence Rate

$$\frac{\text{Number of new cases of disease during specified time period}}{\text{Person-time of observation among people at risk during same time period}}$$

Incidence Rate

- Average rate at which a disease develops in a population over a specified time period
- Is a true rate and has the units of time
- Ranges from 0 to infinity
- Accounts for differing lengths of follow-up
- *Syn.:* Incidence density, hazard rate

Person-Time

Sum, over all individuals, of time at risk until the event of interest, loss to follow-up, or the end of the study

Example

Subject	Years of Follow-up	Got Disease
A	2	N
B	2	Y
C	1	N
D	1	N
E	1	Y
F	3	N
G	1	Y
H	1	Y
I	1	N
J	2	Y

Example

$$\text{Incidence rate} = \frac{5 \text{ cases of disease}}{15 \text{ person-years}}$$

$$= 0.33 \text{ cases per person-year}$$

$$= 33 \text{ cases per 100 person-years}$$

Challenge: person-years is epidemiologic jargon

33 new cases per 100 persons per year

Example

Interpretation:

New cases of the disease appear at the rate of

- 0.33 cases per person per year *or*
- 33 cases per 100 persons per year *or*
- 33 cases per 100 person-years

Incidence and Prevalence

- Change in incidence reflects change in etiologic factors (risk factors or protective factors)

- Change in prevalence reflects change in incidence or duration or both

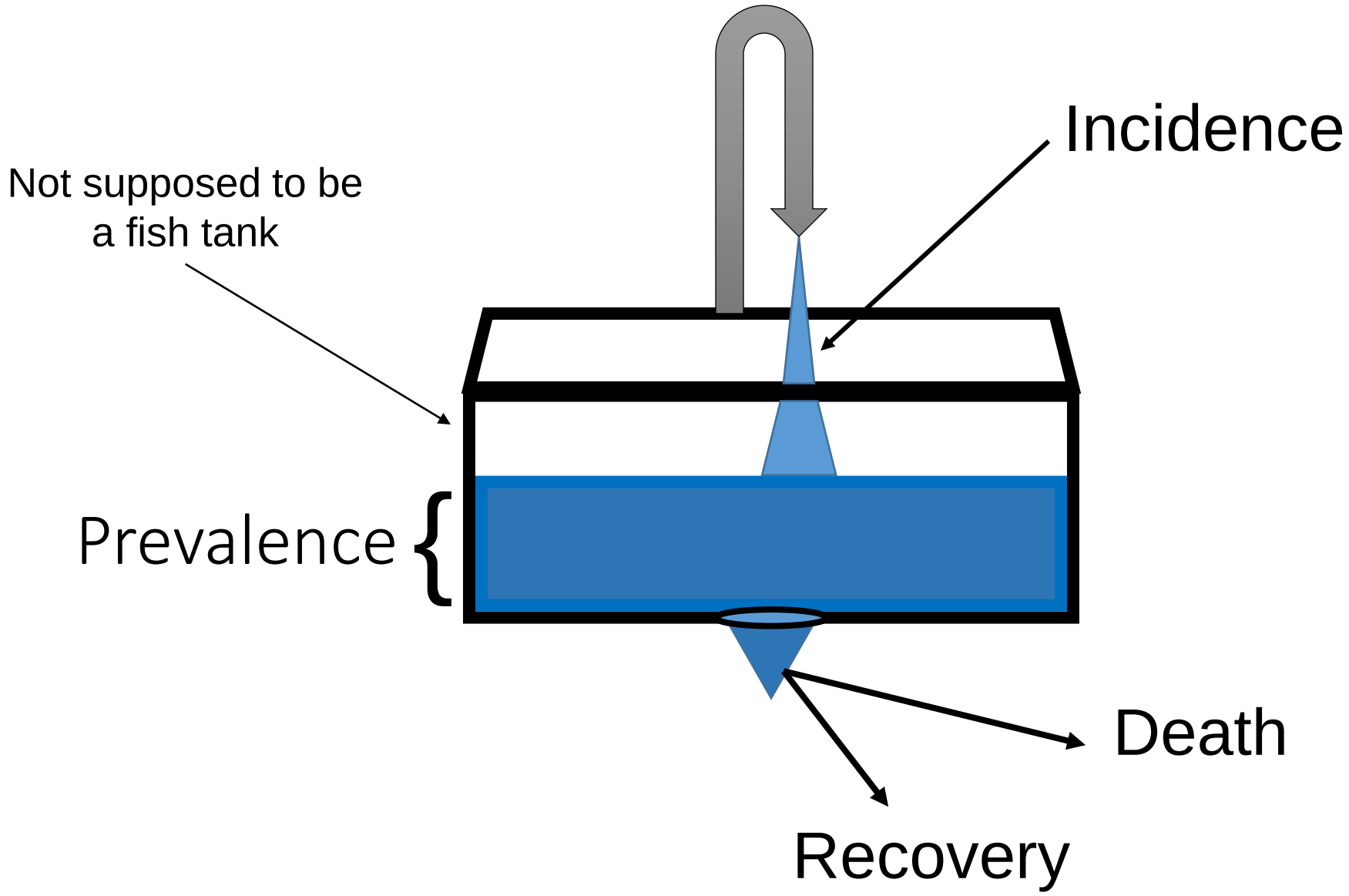
Incidence and Prevalence

$$\text{Prevalence} \approx \text{incidence rate} \times \text{average duration of disease}$$

- Assumes incidence, prevalence, duration are stable over time
- Assumes prevalence < 10%

In other words

- This definition is out the window for pandemics like SARS-CoV-2
  - You need SEIR models for that
    - Susceptible, Exposed, Infected, Recovered models


Not supposed to be
a fish tank
Incidence
Prevalence
Death
Recovery

Measures of Association

- Relative Risk (RR)
- Odds Ratio (OR)

Ways to express the results of an RCT

- Relative Risk Reduction (RRR)
- Absolute Risk Reduction (ARR)
- Number Needed to Treat (NNT)
- Number Needed to Harm (NNH)


Oh no-
not again!
When somebody
down there fools
people with biased
relative risks, an
angel loses its
wings
HEAVEN'S DEPARTMENT of EPIDEMIOLOGY
Statistically-funny.blogspot.com

Relative Risk

- Ratio of disease incidence among exposed individuals to disease incidence among unexposed individuals
- Useful in research on disease etiology
- Quantifies magnitude of the association between an exposure and a disease
- *Syn.:* Risk ratio

Relative Risk

		Outcome: Yes	Outcome: No	
Exposed	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

$$\text{Relative risk} = \frac{a/(a+b)}{c/(c+d)}$$

Relative Risk

- Varies from 0 to infinity
- When RR=1, there is no association between exposure and disease
- When RR > 1, the exposure is a risk factor for the disease, i.e., increases the risk of disease
- When RR < 1, the exposure is a protective factor for the disease, i.e., decreases the risk of disease

Example

		Outcome: Yes	Outcome: No	
Exposed	Yes	100	900	1,000
	No	20	980	1,000
		120	1,880	

Relative risk = ?

Relative Risk

- Can be calculated from cohort studies
  - Including RCTs

Odds

- Ratio of the probability that an event will occur to the probability that the event will not occur
- Risk = ratio of part to the whole
- Odds = ratio of part to the remainder
- Odds always higher than risk

Example

Rolling a die:

- Risk of rolling a 3 = 1/6 = 16.7%
- Odds of rolling a 3 = 1/5 = 20%

Example

- 20 smokers develop bronchitis while 30 do not
- Odds of bronchitis = ? 20/30 = 0.67
- Probability of bronchitis = ? 20/50 = 0.4

Odds Ratio

- Ratio of the odds of exposure among diseased to the odds of exposure among non-diseased

Odds Ratio

		Outcome: Yes	Outcome: No
Exposure	Yes	a	b
	No	c	d

$$\text{Odds ratio} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Odds Ratio

- Varies from 0 to infinity
- When OR=1, there is no association between exposure and disease
- When OR > 1, the exposure is a risk factor for the disease, i.e., increases the odds of disease
- When OR < 1, the exposure is a protective factor for the disease, i.e., decreases the odds of disease

Odds Ratio

- Only measure of association available from case-control studies
- Good estimate of the relative risk when the incidence is low (< 5% in the general population)

Example

		Outcome: Yes	Outcome: No
Exposed	Yes	25	10
	No	75	90
		100	100

Odds ratio = ?

Odds Ratio

- Only measure of association you can get from a case-control study
- Commonly presented in the literature from cohort studies with a dichotomous outcome
  - Because that's what logistic regression produces
    - There are ways around this with modern statistical packages

Expressing the results of an RCT

- Relative Risk Reduction
- Absolute Risk Reduction
- Number Needed to Treat
- Number Needed to Harm

Relative Risk Reduction

- By how much does the treatment reduce the risk of a bad outcome RELATIVE to the control group?

Outcome	**Control Event Rate (CER)**	**Experimental Event Rate (EER)**	**Relative Risk (RR)**	**Relative Risk Reduction (RRR)**
90-day mortality	138/239 = 0.58	122/238 = 0.54	EER/CER 0.54/0.58 = 0.93	1 – RR 1.0 – 0.93 = 0.07

Slide courtesy of Dr. Kerri Thom and the FRCT Epi Small Group

Absolute Risk Reduction

- Also called the risk difference. How many bad outcomes does the treatment prevent?

Outcome	**Control Event Rate (CER)**	**Experimental Event Rate (EER)**	**Relative Risk (RR)**	**Relative Risk Reduction (RRR)**	**Absolute Risk Reduction (ARR)**
90-day mortality	138/239 = 0.58	122/238 = 0.54	EER/CER 0.54/0.58 = 0.93	1 – RR 1.0 – 0.93 = 0.07	CER – EER 0.58 – 0.54 = 0.04

Slide courtesy of Dr. Kerri Thom and the FRCT Epi Small Group

Number Needed to Treat

- How many people do you need to treat for one person to benefit

Outcome	Control Event Rate (CER)	Experimental Event Rate (EER)	Relative Risk (RR)	Relative Risk Reduction (RRR)	Absolute Risk Reduction (ARR)	Number needed to Treat
90-day mortality	138/239 = 0.58	122/238 = 0.54	EER/CER 0.54/0.58 = 0.93	1 – RR 1.0 – 0.93 = 0.07	CER – EER 0.58 – 0.54 = 0.04	1/ARR 1/0.04 = 25

Slide courtesy of Dr. Kerri Thom and the FRCT Epi Small Group

Number Needed to Harm

- How many people need to be treated for one person to have an adverse event?
- You calculate one for each specific adverse event
- 1/absolute risk increase
- $1/(I_t - I_c)$ where t is treatment and c is control

Problems with NNH

- Subject to considerable error
- Should be presented with 95% confidence intervals for proper interpretation
- Does not take quality of life into account
- Limited value to patients
- Can help a practitioners estimate potential impact of a drug or intervention

Sensitivity and Specificity

		Gold Standard: Positive	Gold Standard: Negative
Test	Positive	75	10
	Negative	25	90
		100	100

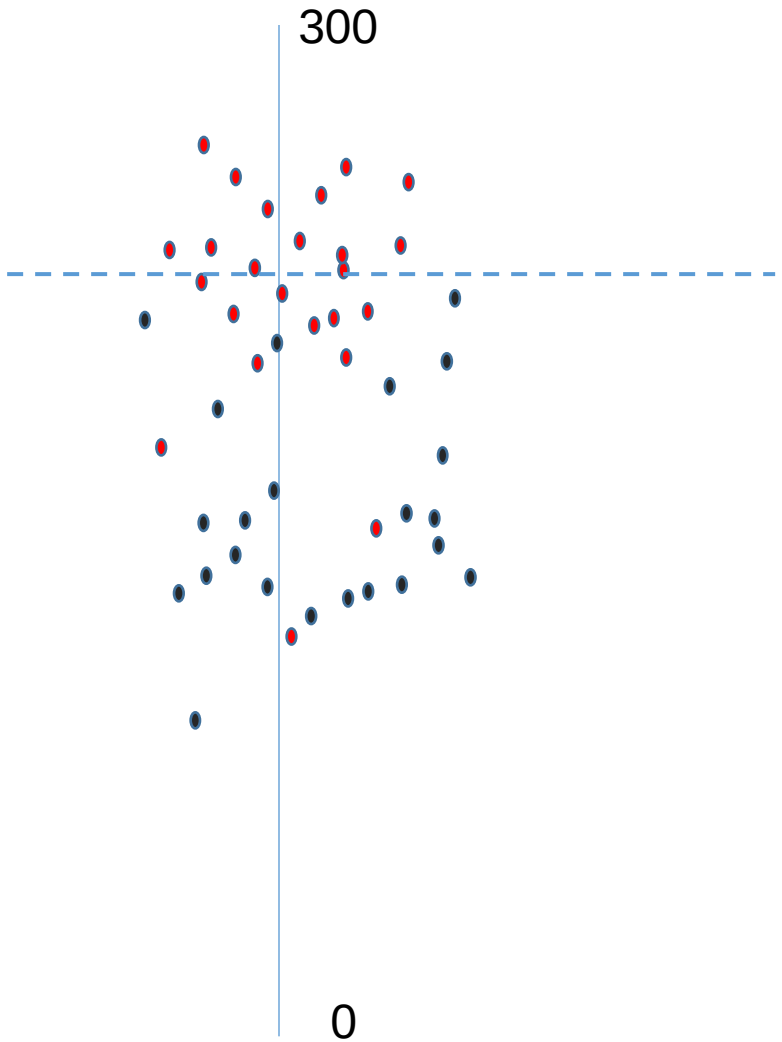
Sensitivity = 75/100 = 75%

Specificity = 90/100 = 90%

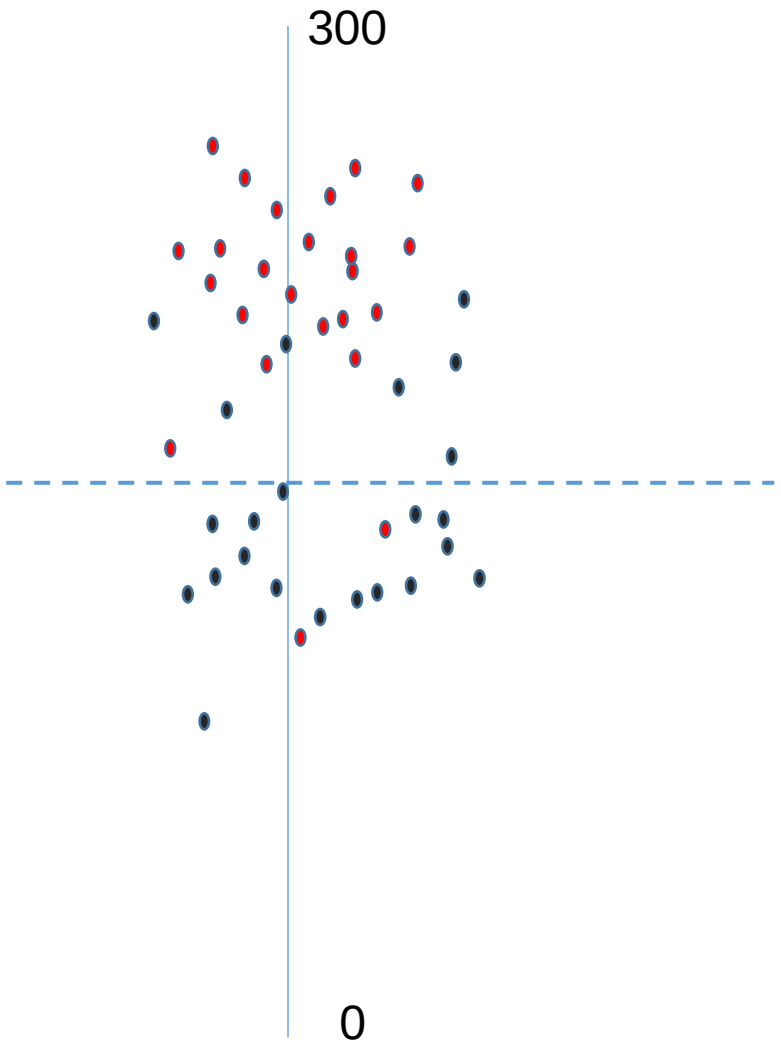
Sensitivity and Specificity

- Sensitivity
  - The probability of testing positive given the disease of interest is present
- Specificity
  - The probability of testing negative given the disease of interest is absent

Sensitivity and Specificity



Sensitivity and Specificity



Positive and Negative Predictive Value

		Gold Standard: Positive	Gold Standard: Negative
Test	Positive	75	10
	Negative	25	90
		100	100

PPV = 75/85 = 88%

NPV = 90/115 = 78%

Sens., Spec., PPV, NPV

- Sensitivity and Specificity are properties of the test itself
- PPV depends on the prevalence of disease in the population being tested

Prevalence of 10%

		Gold Standard: Positive	Gold Standard: Negative
Test	Positive	15	18
	Negative	5	162
		20	180

PPV = 15/33 = 45%

NPV = 162/167 = 97%

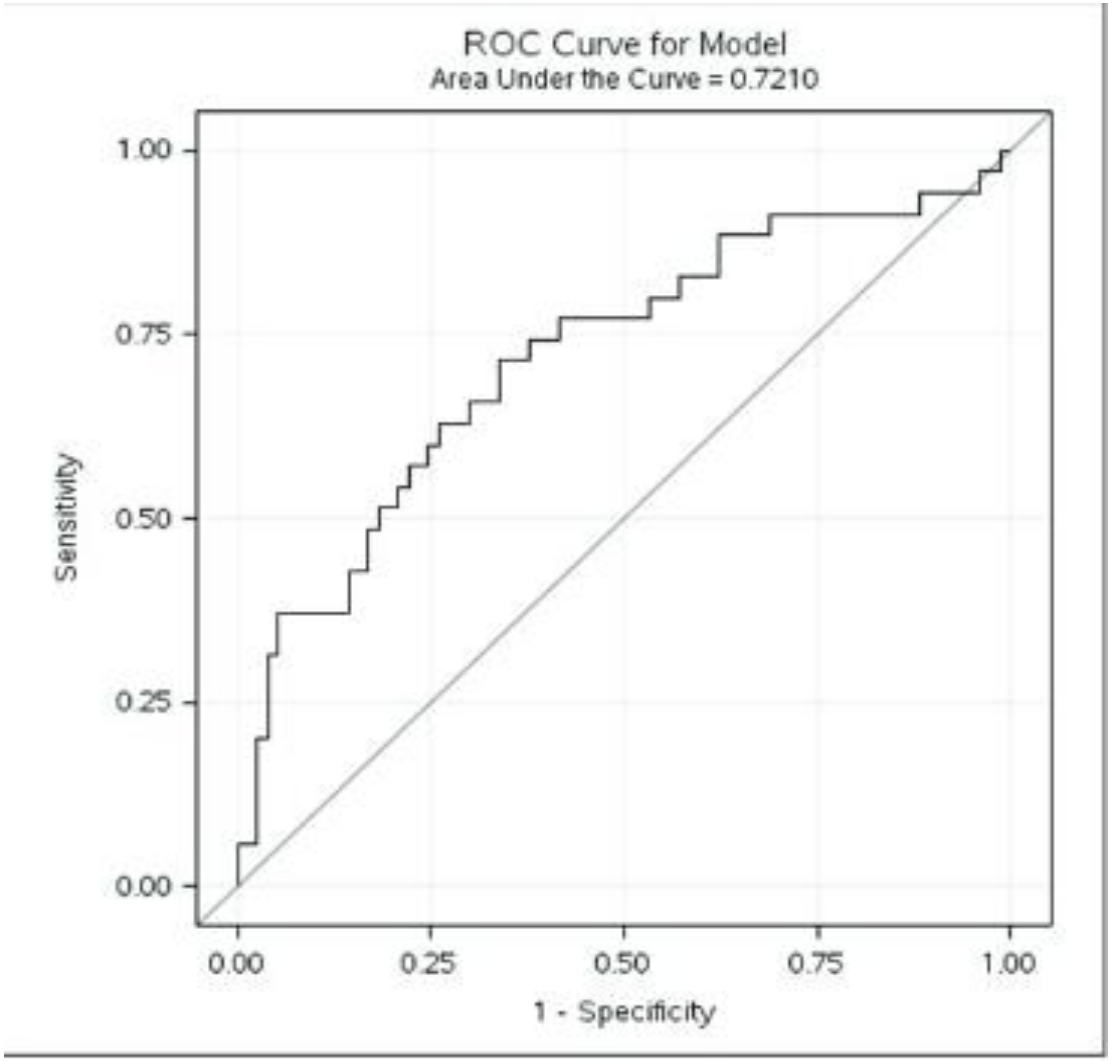
Receiver Operating Characteristic Curves

- Frequency used in clinical epidemiology to assess the ability of a risk factor to predict an outcome
  - Clinical characteristics
  - Lab values
  - Thresholds

Accuracy of prediction

- Calibration
  - How well do the predicted probabilities match up with (agree) observed probabilities
- Discrimination
  - How well does the model tell the difference (separate) between people with and without disease

ROC curves, AUC, and C-statistic



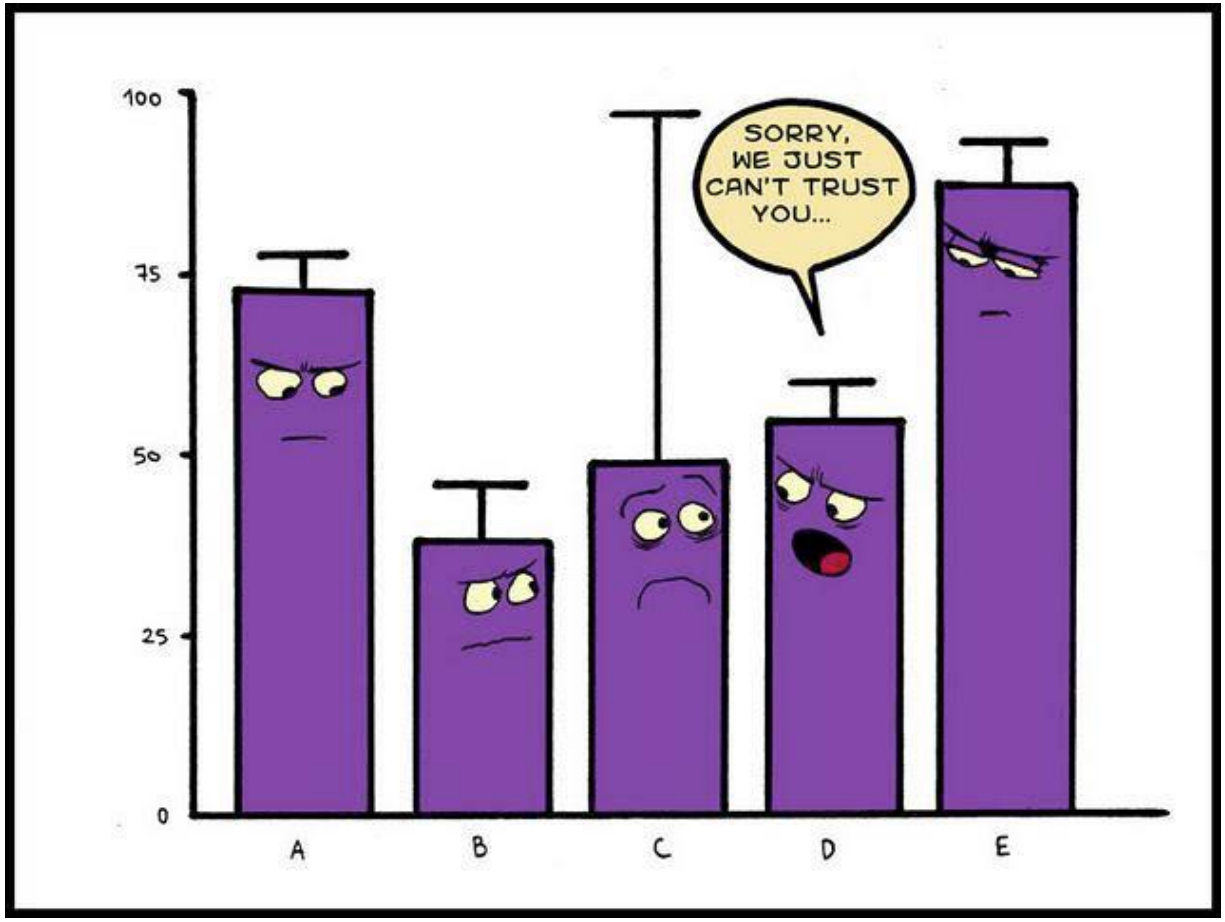
https://www1.mcw.edu/FileLibrary/Groups/Biostatistics/Datum-Newsletter/vo19no4.pdf

C-statistic

- Equal to the area under the curve
- Interpreted as the probability that a randomly selected individual with the outcome will have a higher predicted probability of having the outcome compared to a randomly selected individual who did not have the outcome

Other cool measures we don't have time to discuss

- Hazard Ratios
- Attributable risks and fractions


100
75
50
25
0
A
B
C
D
E
SORRY, WE JUST CAN'T TRUST YOU...