Class 1: Introduction, Concepts of Study Design

Introduction
Types of Studies
Case-Control, Cohort, Cross-sectional study,
Twin concordance study, Adoption study

- **Epidemiology**
  - 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.
  - Science of making predictions about individual patients from clinical events in similar patients
  - Using sound scientific methods from studies of populations to ensure the predictions are accurate
  - Avoids being misled by systematic error and decreases random error
  - Study techniques and statistical methods produce findings that are valid (accurate) and reliable (precise)
  - Distribution (descriptive) and Determinants (analytic)
    - HOW MANY- Distribution
    - WHO- Distribution
    - WHERE- Distribution
    - WHEN- Distribution
    - WHY- Determinants
    - HOW- Determinants

DESCRIPTIVE
1. Case series
2. Cross-sectional/survey

ANALYTICAL OR DESCRIPTIVE
1. Prospective or retrospective cohort

ANALYTICAL
1. Case-control

EXPERIMENTAL
1. Randomized Controlled Trial
2. Systematic review (meta analysis) of all high quality RCT's

- **Evidence-Based Medicine**
  - Integration of best research evidence with clinical expertise and patient values

- **Clinical experience**
  - Clinical skills and past experiences to work with individual patients

- **Patient preferences**
  - Unique concerns and expectations from each patient that must enter into shared decision making

- **Epidemiological studies**
  - Descriptive
    - Look at the world without trying to change it
    - Rely on existing data
    - Use surveys of large groups of people
    - Example:
  - Who is most likely to develop AIDS?
  - What do outbreaks of Legionnaires' disease have in common?
  - Hypothesis testing
    - Gathering more focused information
    - Still not intervening
    - Ex. Legionnaires' disease is cause by stagnant water in cooling systems, we can immediately test for water purity as soon as we hear about an outbreak.
  - Intervention studies
    - Experimental or quasi-experimental
  - Methodological studies
    - Developing new measures, analyses, etc.
  - All focus on populations of people/patients
  - Clinical aspect is application of these (good) studies to individual patients
  - **PICO: patient, intervention, comparison, outcome (measurable)**

- **First Line Health Databases:**
  - Medline
  - EMBase
  - CINAHL
  - Cochrane
- Other databases:
  - Web of Science
  - First Search
  - CRISP
  - Digital Dissertations

MeSH terms – medical subject headings, NLM's controlled vocab for indexing articles
Include MeSH term and common name: "otitis media[MeSH]"

- **A Typical Epidemiologic Approach**
  - Determine existence and magnitude of a problem
  - Describe who has the problem:
    - Time, place, and person
  - Develop hypotheses about why the problem is
    - occurring (risk factors)
  - Test these hypotheses using appropriate analytic study designs and statistical techniques
  - Based on the findings, develop interventions
  - Evaluate the effectiveness of the interventions
  - Sequence:
  - 1. Hypotheses generated
  - 2. Followed by case control or retrospective cohort studies
  - 3. Prospective cohort studies may be done
  - 4. RCT occasionally undertaken

- **Types of Studies**
  - **Descriptive studies** used to study distribution of health outcomes
    - Describe the patterns of disease occurrence by time, place and person
      - Represent the initial step in many epidemiologic investigations
      - Uses
        – Health planning and administration
        – Generation of hypotheses
      - **Case reports**
      - **Case series**
        - Group of patients with common disease then describing common clinical symptoms, features and expression of disease
        - Data sources: chart reviews, interviews
        - Important link between clinical medicine and epidemiology
        - Usu consist of numerator data and calculation of percentages
        - Advantages
          - Allows delineation of clinical picture and associated factors
          - Good for formulating hypotheses and examining new diseases
          - Convenient, cheap, quick
        - Disadvantages
          - No comparison group
          - No internal validity, simply descriptive
      - **Descriptive incidence studies**
        • Examine patterns in occurrence of incident (new) cases of a disease or condition in a defined population
        • Data sources: surveillance data, medical record reviews
        • Denominators usually based on census data
        Rates often calculated and compared
        Sources of bias:
        Case misclassification/measurement bias
        Misclassification in terms of exposure, risk factors, and outcomes
      - **Descriptive prevalence studies (surveys) aka Cross- Sectional Study**

• Describe the prevalence of characteristics of a population or sample of a population
• "Snapshot"
• May study exposures, outcomes, or both
• Usually involve collection of new data
• May be used for purely descriptive purposes, but often used to examine associations between exposures and outcomes

- At one point in time, the subjects are interviewed and/or examined to determine whether or not they were exposed to the agent and whether they have the outcome of interest.
- Only uses one group
- Exposure and cases are determined simultaneously i.e. A group of women are interviewed to determine their use of VDTs and whether they had a miscarriage
- Advantages
  - Relatively inexpensive, simple, no follow-up
  - No one exposed to proposed harmful agent or denied treatment
- Disadvantages
  - Cannot establish causation – only association
  - Subject to recall bias

- **Analytic studies** used to examine determinants of these outcomes
  - Used to test a hypothesis concerning the relationship between a risk factor and an outcome and to measure the magnitude of the association and its statistical significance
  - May be experimental (RCT) or observational (epidemiology study)
  - Most analytic studies at CDC are observational in nature
  - **Experimental Studies**
    - **Randomized Clinical Trials**
      - Allocation of subjects to experimental and control conditions is under the control of the investigator
      - Subjects are randomly assigned to condition
      - Subjects and investigators (and others) often blinded/masked to group assignment
      - Investigator randomizes subjects into exposed groups and follows them over time to determine their outcomes
      - Randomization helps ensure comparability
      - Considered the "gold standard"
      - May be costly, time-consuming, difficult to do if outcome is rare; may not be feasible or ethical
      - Importantly, eliminates a number of potential sources of bias (selection bias)
      - Advantages
        - Groups are more comparable because of randomization (confounders evened out)
        - Blinding can often be done, which lessens bias
        - Statistical tests rest on assumption of randomization
      - Disadvantages
        - Expensive
        - Volunteer bias
        - Ethically problematic at times – potentially beneficial treatment withheld from some, potentially hazardous exposure given to some
        - Results may not be available for years

  - **Observational Studies**
    - Involve no human intervention in assigning study groups; investigator simply observes the relationship between exposure and outcome
    - Subject to many potential biases, but by careful design and analysis, many of these may be avoided
    - Used to study *distribution* of health outcomes; analytic studies used to examine determinants of these outcomes

    - **Cohort Studies**
      - Resemble experimental studies in design in that they proceed conceptually from exposure to outcome

- Start with exposed and unexposed groups and follow them to see if the rates of outcome differ between the exposed and the unexposed
- Used to investigate outbreaks in well defined populations
- In this type of study, two or more groups (or "cohorts") of subjects are identified: one(s) that by choice, luck, or chance has been exposed to the clinical intervention or putative causal
- agent, and one that has not been exposed. Our researcher, for example, can try to locate a group of clam juice drinkers and compare the proportion ofthose who have psoriasis in that group with the proportion of those with psoriasis in a group of abstainers.
- Exposure is not under the control of the researcher
- Usually prospective
- Can also choose groups that were formed in the past and follow them forward to the present (historical cohort
- Advantages
    - Treatment is not withheld nor is any participant subjected to potential hazards
    - Clear temporal sequence of exposure and outcome
    - Good for studying natural history of a disease or rare exposures
    - Allow study of multiple outcomes of same exposure
    - Can directly estimate and compare the rate of outcomes between the groups
    - Subjects can be matched for possible confounders
    - Easier and cheaper than RCT
    - Can establish directionality and timing of events, which is crucial to determining causation
- Disadvantages
    - If prospective, potentially long follow up: maintaining difficult
    - If retrospective: may be difficult to correctly determine exposure status
    - Might be difficult to obtain controls if therapy is popular or most people have been exposed
    - Potential for confounding
    - Blinding among subjects and investigators almost impossible to achieve
    - Expensive to do well
    - For rare disorders, it requires large sample sizes and long follow-up periods

Prospective: exposure has occurred, outcome has not
Retrospective: both exposure and outcomes have occurred before the study begins

- Sources of Bias:
    - Misclassification
    - Confounding – observe a true association and derive a causal inference when in fact *the relationship is not causal*
    - Selection bias
    - Differential loss to follow up before development of the outcome of interest
    - For retrospective studies: recall bias
- Strategies to reduce Bias:
    - Obtain exposure information from prospectively collected database where collection of data is by protocol
    - Define carefully the population and include in the cohort a representative sample of the population reflecting full spectrum of exposure and of sufficient size (power)
    - Ensure follow-up on all subjects and follow individuals for sufficient duration for disease to become manifest
    - Test *a priori* hypotheses and not post hoc ones

- **Case Control Study**
    - Groups identified on basis of *outcomes* and the search for the exposure is retrospective – In a case-control study, a group of people who already have the outcome(cases) is matched with a group who do not (controls), and the researcher determines the proportion in each who had previously been exposed to the suspected causal factor. For instance, in one study, the use of exogenous estrogens in one group of women with endometrial cancer was compared to that in a disease-free control group.
    - Collect exposure information without knowledge of case or control status
    - Advantages

- Quick, cheap
- Allow multiple exposures for a single outcome to be assessed
- Feasible for rare disorders or those with long latency periods
- Usually fewer subjects than cross-sectional studies

- Disadvantages
  - NOT suitable for studying rare exposures
  - Subject to bias if control group not carefully chosen
  - Relies on recall or records – inaccuracies abound
  - Subject to confounding – exposure might be caused by some other factor correlated with outcome
  - Difficult to select and find appropriate control group
  - If the group with the outcome is aware of the hypothesis, possibility of recall bias
  - Does not permit direct measurement of outcome incidence
  - Temporal relationship between exposure and outcome may be unclear
- Strategies to reduce bias:
  - Obtain exposure info from a prospectively collected database
  - Verify recalled information using administrative data or multiple surveys
  - Draw controls from the same population

Hierarchy of Epidemiologic Research Strategies

particular type of proportion, proportions are a particular type of ratio, rate – proportional ratio that measures an event in a population over time

- **Ratio**
  - Expression of the relationship between 2 quantities, not out of the total like in proportion
  - Quantities may or may not be related
  - Example
    - A computer camp has an enrollment of 20 students with 5 faculty members. There are 8 male students and 12 female student
    - Calculate the student to faculty ratio:
    - 20/5 = 4/1
- **Proportion (proportional ratio)**
  - A ratio in which the numerator is included in the denominator, out of the total
  - Example
    - A computer camp has an enrollment of 20 students and 5 faculty members. There are 8 male students and 12 female students
    - Calculate the proportion of female students:
    - 12/20 = 0.60 x 100 = 60%
- **Rate (proportional ratio of population over a specific time)**

- Relationship between the numerator and denominator
- Often a proportion with an added dimension - measures the occurrence of an event in a population over time
- Expressed as:
  - Number of cases or events occurring during a given time period / Population at risk during the same time period x 10 to the nth.
- Person in the denominator must reflect the population from which the cases in the numerator arose
  - Counts in the numerator and denominator should cover the same time period
  - In theory, the persons in the denominator must be "at risk" for the event, that is, it should have been possible for them to experience the event.
- Example
  - A cheerleading camp was held from June 9 -11. A group of teenagers became ill. For the investigation, illness was defined as either diarrhea (3 or more loose stools during any 24 hour period) accompanied by abdominal cramps or bloody diarrhea alone, occurring within 14 days after the start of the camp.
  - 67 of the 600 interviewed campers had illnesses that met the case definition.
  - Calculate the rate of interviewed teenagers who became ill according to the case definition within 14 days after the start of camp.
  - 67/600 = 0.1117 x 100 = 11.17%

- **Morbidity Frequency Measures**
  - To describe
    - presence of disease in a population
    - Probability (risk) of its occurrence
  - Disease: Illness, injury- including car crash, disability
  - Include:
    - Incidence (Rate)- number of NEW cases within a given time period divided by the average population (those at risk). How many new cases? Depends on how big population is- population wide is usually 10 to the 5^{th} = per 100,000. Or 10 to the 2^{nd} = per 100 people. Choose an n with the lowest whole number.
    - Prevalence- number of total cases, new and established at a specific time point divided by the population.
      - Period prevalence- new and old cases that occur within a time period over a time period- generally use the population at a midpoint. Example the prevalence of avian flu for a year. Add up new and old; horizontal line represents duration of illness
      - Point prevalence - # current cases, new and old, of a specified disease at a given point in time divided by estimated population at the same point in time
    - Attack Rate- number of new cases of a specific disease during an epidemic divided by the population at risk at the BEGINNING of the epidemic or study.

- **Prevalence and Incidence**
  - Prevalence is based on both incidence (risk) and duration of illness.
  - High prevalence of disease within a population may reflect:
    - high risk, or
    - prolonged survival without a cure
  - Low prevalence may indicate
    - low incidence,
    - a rapidly fatal process, or (example: gunshot wounds to the head)
    - rapid recovery
  - The prevalence (P) is proportional to the product of the incidence rate (I) and the mean duration of disease (D).
  - If the prevalence is small and the incidence rate and the distribution of duration are constant over time, then the prevalence can be approximated by the product of the incidence rate and the mean duration of the disease. P = I x D
  - P > I for chronic diseases (Diabetes mellitus)
  - P = I for acute diseases (Common cold)

Incidence: the probability (or "risk") that a given person will develop the outcome (i.e., the flu) over a defined period of time.
I=number of new cases/number of individuals <u>at risk</u> (remember age/sex specific conditions…testicular vs. breast cancers) for specified time period.
Prevalent cases at the start don't count (they aren't at risk, they already have it)

Prevalence: the proportion in the population at *any given time* that has the factor of interest
P= number of current cases/population

Prevalence = Incidence X duration
About 42 million flu cases occur in the US per year.
US population ~ 281 million

Incidence ~ 42 million/281 million ~ 150/1,000 per year
42 million cases, each lasting about 2 weeks
42 million x 2/52 ~ 1,615,400 ill at any given time
**remember to half your population if they throw a sex-specific condition at us!*

Prevalence = 1.6 million/281 million ~ 6 per thousand ill at any give time
Assumes even distribution throughout the year

Mortality rate = number of deaths ÷ number of people at risk [for dying of that particular cause] (eg, alive) over a specified period of time.

- **Screening Tests**
  - Identification of unrecognized disease via rapidly conducted procedures (test, exams)
  - Deals with healthy population looking for asymptomatic disease
  - Persons with suspicious findings are referred for follow-up diagnostic tests and treatment
  - Screening = public health intervention among populations
  - Diagnosis = clinical interventions applied to Individuals

- **Successful Screening Program**
  - Condition is an important health problem
  - Accepted treatment for disease should be available
  - Available facilities for diagnosis and treatment
  - Recognizable latent or early symptomatic stage
  - Screening test should be suitable (sensitive and specific)
  - The test should be acceptable to the population and entail minimal risk.
  - Diagnostic work-up for a positive test result must have acceptable morbidity given the number of false-positive results.
  - The cost should be economically balanced in relation to possible expenditure on medical care as a whole.
  - Case-finding should be a continuing process and not a once and for all project.

- **Potential Disadvantages of a Screening Program**
  - Longer morbidity for cases where prognosis is unaltered- screening someone at a young age and now they now at a young age they have a fatal disease and have to deal with it longer. Example is Huntington's Disease- with 1/3 suicide rate. Ethical issues.
  - Over treatment of questionable abnormalities
  - False reassurance to persons with false negative results. Example is latent period for HIV+ patient. Biological onset is not yet detectable.
  - Anxiety and morbidity for persons with false positive results

- **Sensitivity**
  - %people with disease who test positive
  - a/a + c; want a high number, low c = false negatives
  - SNOUT: a test with high sensitivity, if negative, rules disease OUT
  - Sensitivity of 97% means that 97% of pts with dz had positive test

- **Specificity**
  - Opposite
  - d/b + d; want a high number, low b = false positives
  - SPIN: a test with high specificity, if positive, rules disease IN
  - Specificity of 72% means that 28% of pts without dz had positive test

- **Right test for each purpose**
  - Screening – high sensitivity
  - Ruling out disease – high sensitivity
  - Making definitive diagnosis – high specificity
  - Managing a disease process – optimal balance of both

- **PPV**

- Percent of positive tests that are actually positive.
- a/ a + b

- **NPV**
  - Percent of negative tests that are actually negative.
  - d/ d + c

a	**b**
c	**d**

- **Measure of Association**
  - Assess the strength or magnitude of the association between the exposure and disease of interest
  - Cohort study
    - Relative risk- rate of illness after an exposure
    - Incidence of disease in the exposed group divided by the incidence of disease in the unexposed group
    - Calculated as percent with disease in exposed group divided by percent with disease in the unexposed group
    - Ratio of the probability of the event occurring in the exposed versus the non-exposed
    - Relative Risk (Risk Ratio)
    - = [a / (a + b)] / [c / (c + d)]
    - Risk of developing disease divided by all those exposed/ Risk of developing disease divided by all those not exposed.
    - The ___ people were ___x more likely to develop ___ then ___.
  - Case-control study
    - Begin with the selection of cases and a group of controls
    - Unable to determine disease rates because these groups were NOT determined by nature, but by the investigators' selection criteria.
    - Thus, an incidence rate of disease for exposed and non-exposed can not be computed.
    - Thus, **can not** determine relative risk
    - Odds Ratio
    - Can calculate relative frequency of exposure among cases and controls.
    - Odds that a case is exposed divided by the odds that a control is exposed.
    - Odds of Exposure
    - Odds of Exposure Ratio = Odds Ratio = Cross Tabulation
    - (a/c) / (b/d) = ad / bc
    - If the frequency of exposure is higher among cases, the OR will be > 1, indicating increased risk.
    - The stronger the association between the exposure anddisease, the higher the OR.
    - If the frequency of the exposure is lower among cases, the OR will < 1, indicating protection.
    - For rare diseases, the OR approximately equals the relative risk that would have been found if a cohort study had been conducted.
    - Some report odds ratio as "estimated relative risk"

Class 3: Measures of Central Location and Dispersion

Statistical Distribution
Standard deviation vs. standard error
Confidence interval
Bias, Correlation

- "Variable"
  - Anything that is measured in a study
  - Independent variable
    - varied by and under the control of the experimenter or measure that is used to "predict" an outcome
  - Dependent variable
    - outcomes from the experiment or the study

- Qualitative vs. Quantitative
  - **Qualitative**
    - non-numeric or categorical
    - These vary in kind rather than in magnitude

        - Ex: marital status, gender, race
    - **Quantitative**
        - Numeric
        - These vary in magnitude rather than in kind
        - Ex: age, income, temperature, serum creatinine

- Classification of Data
    - **Discrete**
        - Having a fixed number of values
        - Qualitative data is always discrete
    - Examples: marital status, blood type, number of children
    - **Continuous**
        - Having an infinite number of values
        - Quantitative data can be discrete or continuous
    - Examples: height, weight, temperature

- **Qualitative Data: Nominal**
    - Data which fall into mutually exclusive (discrete) categories for which there is no natural order; each is discretely-example red is not better then blue; No set order.
        - Examples:
            - Favorite Color
            - Race/ethnicity
            - Gender
            - Marital status
            - ICD-10 codes
            - Dichotomous data such as HIV+ or HIV-; yes or no
            - State of origin

- **Qualitative Data: Ordinal**
    - Data which fall into mutually exclusive (discrete) categories but have a rank or graded order. Set order.
        - Examples:
            - Grades
            - Socioeconomic status
            - Stage of disease
            - Low, medium, high
            - Pain scales

- **Quantitative Data: Interval**
    - Data which are measured by standard units
    - The scale measures not only that one data point is different than another, but by how much
        - Examples:
            - Number of days since onset of illness (discrete)
            - Temperature in Fahrenheit or Celsius (continuous)

- **Quantitative Data: Ratio**
    - Data which are measured in standard units where a true zero represents total absence of that unit
        - Examples
            - Number of children (discrete)
            - Temperature in Kelvin (continuous)

- Statistics
    - Inferential statistics – see lecture 4
    - Descriptive statistics:
        - Methods of producing quantitative summaries of information
        - Trying to describe a set of data
        - Measures of central tendency
        - Measures of dispersion
        - Inferential statistics
        - Methods of making generalizations about a larger group based on information about a subset (sample) of that group

- Measures of central tendency: (don't inform us of the shape of the data)
  A) ***Mode***: observed value that occurs with greatest frequency (most common value).
  -can state the mode of qualitative variables
  -value is NOT influenced by extreme scores
  B) ***Median***: value that divides the frequency in half (value in middle, equal number of values above and below).
  -also valid for qualitative variables
  -also NOT sensitive to extreme values, but less sensitive to actual values of other data points
  -put in rank sum and if odd it's the middle value and if even take average of middle
  C) ***Mean***: sum of all the elements divided by the elements in the distribution (sum of all values divided by number of values).
  -valid only for quantitative variables
  -is VERY sensitive to extreme values
  ~~$\bar{X}$~~ (sample); μ (population)

- Measures of dispersion: used to describe variability of spread of observations in a sample
  - Range- difference of the largest value from the smallest value; very sensitive to extreme observations; and generally as you have more people in your study the larger your range.
  - Percentiles- 50th % is the middle person. Less sensitive to outliers or sample size.
    - Median is always the 50th percentile; *p*th percentile is the Vp at which p percent of the points are less than or equal to p
  - Dispersion of values around the mean (*average deviation*)
  - Variance: s^2 (sample); σ^2 (population)
    - average deviation of individual values about the mean
    - if close to variance is small, and if large variance is large
    - $\frac{\sum(x_i - \bar{x})^2}{n}$
  - standard deviation
    - square root of variance puts it back in original units of measurement
    - $\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$

    $sd = \sqrt{\sum(x - \bar{X})^2 / n}$

–Example: If we are given the variance of something, we simply take the √ of the given variance to find sd.

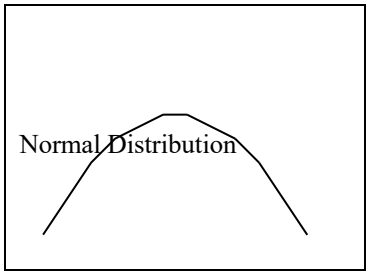
→The variance of # of cigars smoked per day is found to be 16.

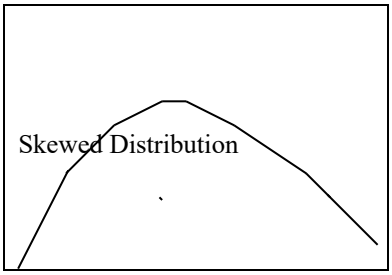
Therefore the sd would have to be √16, or a value of 4.

- Not every sample is the same
  - Before we select a sample, the sample mean X is a random variable.
  - We don't know what value it will take.
  - The value varies from sample to sample.
  - The population mean μ is a single fixed number.
  - There can be many different samples but there's a single population.
  - For random sampling, the sample mean X can fall either above or below the population mean μ.
  - The mean of the distribution of the sample means X equals the population mean μ.

If one tail is longer than the other, the mean is always toward the longer tail, the mode toward the shorter tail, and the median between the two
Negative skew – left tail is longer; mass of distribution on right side of figure (hump)

Positive skew is more common in biological systems





Normal Distribution:
-mean, median, and mode all in the middle of the curve.

Skewed Distribution:
-if one tail is longer than the other, the mean is always toward the longer tail, the mode is always toward the shorter tail, and the median between the two.
-*Positively skewed* have large number of low scores and small number of very high scores.
-*Negatively skewed* have relatively large number of high scores and small number of low scores.

Standardized Scores:
-help compare measures with different scales on a similar metric.

Deviation score:
-difference between the individual value and the mean value.
-Example: if a patient smokes 24 cigars per day, and the mean value of cigars smoked per day was 16, then the deviation score would be 24-16 = 8.

Z Scores:
-how many standard deviations a score lies from the mean

$$z = x\text{-}\overline{X} / sd$$

-gives standard deviation units
→Because z scores are in standard deviation units, they can be related to normal curve
→allow us to predict the proportion above or below a given score

-We can use the information learned about "normal" distributions and z scores to find the probability that a random ***sample*** will have a mean above or below a certain value (probability of random error)
-z scores allow us to predict the proportion above or below a given score

Commonly used measures of health and disease:

Incidence of a disease
Prevalence of a disease
Mortality rate
- Crude
- Proportional
- Age-specific
- standardized
- **Incidence:** the probability (or "risk") that a given person will develop the outcome (i.e., the flu) over a defined period of time.
  - *I=number of new cases/number of individuals at risk* (remember age/sex specific conditions…testicular vs. breast cancers) for specified time period.
  - Prevalent cases at the start don't count (they aren't at risk, they already have it)
- **Prevalence**: the proportion in the population at *any given time* that has the factor of interest; existing cases in a population
  - P= number of current cases/population
- ***Prevalence = Incidence X duration***

        - About 42 million flu cases occur in the US per year.
        - US population ~ 281 million
    - **Incidence** ~ 42 million/281 million ~ 150/1,000 per year
        - 42 million cases, each lasting about 2 weeks
        - 42 million x 2/52 ~ 1,615,400 ill at any given time
        - **remember to half your population if they throw a sex-specific condition at us!*
    - **Prevalence** = 1.6 million/281 million ~ 6 per thousand ill at any give time
        - Assumes even distribution throughout the year
    - **Mortality rate** = number of deaths ÷ number of people at risk [for dying of that particular cause] (eg, alive) over a specified period of time.
    - **Case fatality rate** = number of deaths among the diseased ÷ number of diseased people over a specified period of time
    - **Proportional mortality rate (PMR)** = number of deaths due to a specific disease ÷ total number of deaths over a specified period of time
        - From the CDC: "36,000 Americans die each year from influenza"
        - Mortality rate: 36,000 deaths ÷ 281 million ~ 13 deaths per 100,000 people in the US per year
        - Case fatality rate: 36,000 deaths ÷ 42 million cases =~ 86 deaths from flu per 100,000 cases per year
        - PMR: 36,000 deaths from flu ÷ 2.4 million deaths overall =~ 1.5% of deaths are due to influenza
        - Age-specific mortality rate = number of deaths in a particular age range ÷ number of people in that age range
        - Standardized mortality rate (SMR) = number of deaths that would occur in a standardized population ÷ number of people in the standardized population
    - **<u>Understand this concept</u> (we won't have to calculate it): <u>SMR</u> is the only way to compare rates across different populations…i.e. to compare death from flu in Florida to death from flu in Mississippi.**
        - Age specific mortality rate: "over 90% of deaths occur in people over 65yrs"
        - 90% of 36,000 = 32,400 deaths
        - There are 35 million people > 65yrs in the US
        - 92 deaths from influenza per 100,000 older people per year (compared to 13/100,000 overall)
        - Influenza vaccine: efficacy varies from year to year, but is usually 70-90%. This year, it may be as low as 15%.
    - **Measurements of Association: Between two groups, not just the outcome of a single group.**
    - **RR** = Incidence of outcome among all the exposed/ incidence of the outcome among all the non-exposed
    - By convention, we count the adverse outcomes, so that a beneficial exposure gives a RR < 1, a harmful exposure gives a RR > 1.

	Flu	No Flu	
Oseltamivir (Tamiflu)	4	458	462
Placebo	34	459	493

RR = risk in treated ÷ risk in non-treated
RR = 4/462 ÷ 34/493 = 0.009 ÷ 0.069 = 0.13
Thus, the risk of flu in the treated group was 13% of the risk in the control group
Relative risk reduction RRR:
(1 – RR) x 100 = %RRR…Same as: 100 – RR as a percent…Same as:
<u>Risk in control group – risk in treatment group</u>
Risk in control group
If RR is 0.13, then RRR is 87%
The risk of getting the flu from an infected family member was reduced by 87%

- **Odds Ratio OR:**
    - Compares (a *ratio*) the odds of the outcome in the exposed vs the non-exposed, or odds of the exposure in the diseased vs. the non-diseased (it comes out the same either way)

- ***A:B ÷ C:D = AD/BC***

- Estimates the RR, but *always* further from 1 than the RR (either <1 or >1), estimates best if the rate of the outcome is very low, <1%.
- The only measure that can be obtained from case-control studies, and using complex (multivariate) methods of analysis

- Commonly misinterpreted/misused as RR.

	Flu	No Flu	
Oseltamivir (Tamiflu)	4	458	462
Placebo	34	459	493

- ***OR = odds in treated ÷ odds in non-treated***
  - OR = 4 to 458 ÷ 34 to 459 = 0.009 ÷0.074= 0.11
  - Thus, the odds of flu in the treated group was 11% of the odds in the control group
- **Odds Ratio vs Relative Risk:**
  - "A 90% (71% to 96%) relative risk reduction was observed…"
  - But we found RR = 0.13 (13%); RRR = 87%
  - We found OR = 0.11 (11%)
  - OR *estimates* RR when outcome is rare
  - Odds ratio is always further from 1.0 (either above or below) than RR
  - OR gets further and further from RR as outcome gets more common
- **Risk Difference:**
  - (a *difference*, not a ratio)…aka **ARR, absolute risk reduction:**
  - Risk of the outcome in the non-exposed (ie, controls) minus the risk in the exposed (ie, the experimental group)
  - If positive, the exposure is beneficial (ie, fewer adverse outcomes)
  - An absolute difference, not a relative measure
- **Absolute Risk Reduction:**

	Flu	No Flu	
Oseltamivir	4	458	462
Placebo	34	459	493

ARR = risk in treated – risk in non-treated

ARR = 4/462 (0.9%) – 34/493 (6.9%) = - 6%

Thus, the risk of flu was decreased by 6% among those who got the drug

RRR vs ARR

Relative risk reduction shows decrease in one group RELATIVE to the other.

Absolute risk reduction shows the absolute difference in risk between groups.

RRR is nearly always more impressive than ARR, thus is usually what drug ads show.

Number needed to treat NNT:

The number of patients you would need to treat so that one patient would benefit:

1 ÷ absolute risk reduction (100%/ARR%)

If a disease is uniformly fatal and a new disease (I think she meant to type "drug") saves 50% of the patients, what is the NNT?

What is the NNT for an ARR of 10%? (answer: 10) 1%? (answer: 100)

NNT for Oseltamivir (Tamiflu) to prevent one case of flu in an exposed family member = 100%/6% = ~17

Thus, 17 flu-exposed people need to take Tamiflu (@~$100/person) to prevent one case of influenza. [In short, if Pt is +65 yrs old, its worth it, if Pt is 25 yrs old and healthy, not so much worth it based on what we've shown here.]

If baseline risk is high, then NNT is much lower for the same RRR (.20 reduces to .05 AR, .01 reduces to .0025 AR; the .20 group has a smaller NNT)

- **Correlation**

- Correlation DOES NOT equal causation
    - My office wall clock and watch are set 1 minute apart. The wall clock chimes every hour, exactly 1 minute before my watch beeps. These events are exactly correlated but the wall clock is not causing my watch to beep.
- Just because two things are correlated doesn't mean they are the same
    - USMLE scores of both a poor student and a good student may rise 5 points for every day spent studying, but their scores may still be very different. Their scores would be exactly correlated (r=1), but not equal.

- **Potential Pitfalls**
    - Before we can trust any perceived associations, we must look for hidden bias that may be responsible
        - Selection Bias (selected so that apparent association observed when in reality there was not one)
            - Often from nonresponse: must keep nonresponse to minimum and characterize to determine how different nonresponders were from other groups
        - Misclassification Bias
        - Measurement Bias

- **Selection (Sampling) Bias**
    - We assume that our sample is a random selection from all potential persons with a given condition (or exposure)
    - If persons selected for study are somehow *systematically* different from the overall population, we may find associations due to this unknown factor

- **Misclassification Bias**
    - One group systematically has their exposures and outcomes measured differently than another group.
    - Misclassification of exposure may occur such that cases are misclassified as exposed more often than controls
    - May lead to false associations: Smokers are more likely to be diagnosed with emphysema.
    - Are smokers are more likely to have emphysema or just more likely to have their symptoms evaluated?
    - Recall bias: women with baby with malformation tended to remember more mild infections during pregnancy than did mothers of normal infants
    - Nondifferential misclassification
        - Results from degree of inaccuracy that characterizes how information obtained from all study groups
        - Inherent in data collection methods
        - Tends to result in diluted OR or RR
        - Example would be if some cases classified as controls and some controls classified as cases resulting in smaller difference in exposure

- **Measurement Bias**
    - When one group has a better chance of having an outcome detected than the other groups.
    - Patients on IV Cytoxan therapy are seen twice as often as those receiving oral MMF therapy. One study group (Coumadin) has PTT drawn weekly, while the other (Lovenox) has no monitoring of PTT unless clinically bleeding.
- **Wish Bias**
    - Bias introduced by subjects who have developed a disease and seek to show, often unintentionally, that the disease is not their fault
    - May deny certain exposures related to lifestyle
    - Considered a type of reporting bias
- **Surveillance Bias**
    - Population monitored over period time, disease ascertainment may be better than in general population
    - Relationship of OCP and thrombophlebitis
    - Apparent association may be observed when none exists
- **Information Bias**
    - Means for obtaining info about the subjects in the study are inadequate so that some of the information about exposure and/or disease outcome is incorrect
    - Example would be misclassification bias
        - Cases classified as controls (vice versa)
        - May result from limited sensitivity of diagnostic test involved or from inadequacy of information derived from medical record
        - Exposure history (belief, memory, records)
    - Bias in abstracting records, interviewing, surrogate interviews, surveillance bias, recall bias (case control studies), reporting bias (wish bias)
- **Confounding**
    - Assuming a correlation shows causation
    - Generally considered a problem which should be addressed

  - May be helpful guide in screening populations even when we do not identify specific etiologic agent
  - Not an error in the study, rather a true phenomenon that describes relationship (in contrast to bias)
  - Approaches to handle confounding
    - In designing and carrying out study:
      - Individual matching
      - Group matching
    - In analysis of data
      - Stratification
      - Adjustment
- **Interaction**
  - When incidence rate of disease in presence of two or more risk factors differs from incidence rate expected to result from their individual effects; simultaneous influence of two variables on a third is not additive
  - e.g. *Interaction* between adding sugar to coffee and stirring the coffee. Neither of the two individual variables has much effect on sweetness but a combination of the two does.
  - can be greater (positive interaction) or less (negative interaction) than what is expected
  - Is there an association? If so, is it due to confounding?
  - Is there an association equally strong in strata formed on the basis of the third variable (across all strata)?
    - If no, then interaction present
    - If yes, then interaction not present

Class 4: Statistical Hypotheses

Statistical Hypothesis: Null (H0), Alternative (H1)
Type 1 error (alpha), Type II error (beta)
Power (1-beta)
t-test vs. ANOVA vs. chi-squared

Inferential Statistics – Makes generalizations on the entire population by making small samples of that population. Sampling error results because each sample population is not 100% representative of the entire population. Inferring from what you get from a small group of patients. Data is never taken from the entire population with disease, and hope they are representative of all patients.

Sample Error- random variation from your sample; control with stats.; use a tight protocol
Systematic Error- you cannot control with stats.

Definitions:

- **Standard Error of the Mean** – Standard Deviation of the distribution of mean values. Standard Deviation of the group is related to the standard deviation divided by the size of the sample. Larger sample size results in smaller error.

  Standard Error = $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

  Using Standard Error to find the z score of a population:
  -Population has mean resting HR=70, sd=10. What's the probability that a random sample of 25 people will have mean HR > 75?

  $SE = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$ Z Score = $\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{75 - 70}{2} = 2.5$

  Using a Z Score table we find that a Z Score equals a probability of 0.0062

- **Central Limit Theorem** – distribution of a sample mean approximates a normal curve regardless of shape of distribution from which it is sampled
- **Hypothesis Testing** – The null hypothesis (H_0) always assumes no difference between two samples. The alternative hypothesis (H_A) assumes a difference between the two samples.
  - **Risk Reduction = 0**
  - **RR = 1**
  - **OR = 1**

- **Null Hypothesis**- your hypothesis you start with; your expected outcome = there is no difference between two groups/they are essentially the same group. Assume to be true unless you have overwhelmed evidence that it's not true.

- **Alternative Hypothesis**- There is some difference between the two groups.

- ***p* value** – Probability that the data, of any chosen sample, to show a difference as large as our sample data. Happens because you can only sample a small portion of the population. Tells you about sampling error.
    - **Cliff Notes:** Probability you found a result that wasn't really there—that you reject your null hypothesis falsely. A *p* that = 0.1 means that 10% of the time you falsely find a difference and reject the Null Hypothesis.
    - *p* value is not an indication of sample quality, size or direction.
    - Normal value of 0.05.

- What is a **loose Protocol**?
    - Random error and random analysis performed
    - Not following a prospectively declared plan
    - Analyzing data based on the outcomes, not on what was intended
        - Data driven analysis- adding in random error/ systematic error. Weren't really looking for it, and found it so reported it.
        - "Data-mining"
        - Example: Mr. FIT study; error never reproduced. Reported as a finding, but erroneous. Post doc analysis requires a new study, it may be statistically different, but we know 5% of the time this occurs due to the normal p value of 0.05.
    - Correct reporting
        - Results: No significant difference was found between men and women's SBP in this trial.
        - Discussion: "Although the main outcome finding was negative, the significantly increased pulse pressure found among men is interesting, and may partially explain their odd behavior. This newly generated hypothesis should be studied further in a future trial…"

- **Hypothesis Testing: Review**
    - Start by assuming no difference between groups (Null hypothesis: H0)
    - Determine the likelihood that the observed differences could be from chance alone (random error)
    - If the probability that you falsely reject H0 is small enough (say 0.05), then "reject" H0
    - P values help determine effect of sampling error on the results we find
    - *p* value is the probability that the observed results could have occurred by chance alone if the H0 is true
    - Without sampling there is no need for *p* values.
    - *p* value does not mean that H0 is true.
    - *P* value means IF null hypothesis is true and I repeated this experiment hundreds of time, sampling error alone would produce a difference at least this large in only 5 of every 100 samples

1. Define hypotheses
2. Specify your null distribution
3. Do an experiment
4. Calculate the p-value of what you observed
5. Reject or fail to reject (accept) the null hypothesis

- **Type I Error** – Probability of incorrectly rejecting the null hypothesis (H_0). Normal value is usually 5% (alpha or p value = 0.05). The Null was true and there is no difference but we found the difference. Must have strong evidence and therefore it's 5%.

- **Alpha** –Probability of incorrectly rejecting the null hypothesis (H_0). **This is a type I error.** p value can be known as alpha or significance level. Probability of rejecting the Null when in fact Null is true and there is no difference.

- **Type II Error** – Probability of incorrectly accepting the null hypothesis (H_0). Normal value is usually 20%. There is a difference, but we didn't find it. Also known as Beta. Power of a study to find a difference. Power = 1-Beta.

- **Beta** – Probability of incorrectly accepting the null hypothesis (H_0). H1 is really correct.

- **Power** – Probability that a given difference could have been detected.
    - **Power = 1 - Beta**
    - Power increases as:
        - Sample size increases
        - Alpha increases
        - Random error decreased
    - More power is required to find smaller differences

        - Differences in blood pressure between infants and adults (a large value) require less power than detecting differences in blood pressure between men and women (a comparatively less value).

- To determine Sample Size
    - You need to provide a few guesses:
        - The event rate in the control population
        - Alpha
        - The minimum effect size you think is clinically important
        - The Power (technically the type II error you are willing to accept).
        - (You could also flip this and determine power by knowing 1, 2, and the available sample size)

- Clinical vs. Statistical Significance
        - With a large enough N, any difference will be labeled as statistically significant
        - The *p* value indicates the likelihood that you found a result that wasn't really there
            - Dichotomous answer
        - The *p* value tells you NOTHING about the actual size of the difference between group or strength of association!
            - It is ***NOT*** a measure of effect size
        - The *p* value doesn't even give you the direction of effect (if any)

	H_0 Actually True	H_0 Actually False
H_0 Accepted due to findings	Correct	Type II Error Beta
H_0 Rejected due to findings	Type I Error Alpha, p value	Correct

- **Confidence Interval** – A statistical range with a specified probability that a given parameter lies within the range. Gives a range of plausible values.
    - **95% CI = +/- z * SE**
        - Same as a p value of 0.05.
    - CI's are often used instead of p values.
    - Narrow CI indicates a more precise estimate.

    - A CI that includes the number zero in a difference score would mean equivalence in the population (no difference).
        - ***Insignificant Example:** Medicare beneficiaries with alcohol-related disease had 2.6 fold increase risk of hip fracture (95% CI = -1.6, 3.6) *compared* to those without alcohol-related disease.
        - The value 0 is included in the CI range. If it crosses zero that means there is a possibility that there is no difference between the groups. It includes the Null Point (H0) = 0.
        - If you were to decrease the CI to 90% it would not include the Null Value you can report that, but it's not significant based on 95%.

    - A CI that includes the number 1 in an odd ratio comparison would indicate equivalence in the population (no difference). Use 95% CI.
        - ***Significant Example:** The relative odds of Medicaid recipients with acute MI in teaching hospitals receiving cardiac catheterization is 4.5 (4.06,4.98) compared to those in non-teaching hospitals.
        - The value 1 is not included in the CI range.
        - Note: Odds Ratio Null Hypothesis = 1. No difference = 1.

- Hypothesis testing
    - The type of study required to answer a clinical question depends on the hypothesis being tested.
        - To compare 2 different treatments
            - Randomized Control Trial
        - To evaluate a new diagnostic test
            - Comparison to a Gold standard
        - To determine prognosis

        - Follow up of a cohort
    - To determine cause of illness
        - Cohort or case-control Study

- A Pitfall of Hypothesis testing
    - A pre-defined question (hypothesis) and a tight protocol insures that we properly accounted for sampling error
        - But, a large trial is expensive and it would be a shame to have just 1 analysis
        - Thus, often many secondary analyses and "post-hoc" evaluations are carried out
    - Especially true in the new field of genetics
        - Often over 100 thousand analyses performed
        - Gene chip with many sequences
    - Using a 95% cut-off means that only 5% of values are that value or higher.
        - Therefore, 5% of the values will be higher by chance alone.
        - Therefore, 5% of your tests will be significant by chance alone.
        - 1/20 of your lab results will be significant by chance.

- Young's False Positive Rules
    - 1. With enough testing, false positives *will* occur.
    - 2. Internal evidence will not contradict a false positive. Only someone else's study will reject. This is random due to your study.
    - 3. Good investigators will come up with a possible explanation.
    - 4. Type I errors only happen to the other person.

- Bonferroni Procedure- Very conservative, high standard; add up all *p* and
    - Composite *p* value
    - (FWER = the probability of getting one or more *false positive)*
    - Bonferroni procedure is a variation of family-wise error rate.
    - Bonferroni "experiment wise" error rate= α'
    - α' = α Pr(at least one test causes rejection | H0 is true)
    - α' = 1- Pr(all tests do *not* cause rejection | H0 is true)
    - α' = 1- [Pr(one test does *not* cause rejection | H0 is true)]m
    - α' = 1- (1-α)m
    - For 10 tests with an overall α'=FWER=0.05, the individual α for each test would need to be ≈0.05/10 = 0.005
    - 1000 tests, to get an α = 0.05, each test would need p< 0.00005
    - Concerns with Bonferroni
        - Bonferroni procedure assumes that the tests are independent
        - Correction is very conservative
        - Leads to potentially missing true positive results!

Class 5: Clinical Trials

Types of Clinical Trials

• Clinical Trials are performed mainly in
– Pharmaceutical Products (Synthetic "Drugs")
– Biotechnology Products (products made from human cells/tissues for example, such as vaccines, blood products)
– Devices (e.g. dialysis machines, IV access)

Phases of Clinical Trials
(Pharmaceutical & Biotechnology)

• Pre-Clinical (*in vivo animal, in vitro human*)
• Investigational New Drug (IND) Application
• ***Phase I - First time in humans***
• ***Phase II - Exploratory***
– ***Phase IIb***
– ***Phase II/III***
• ***Phase III – Confirmatory (RCT)***
• ***Phase IV - Post Marketing***

The Drug Development and

Approval Process

1. Early research and preclinical testing
2. IND application filed with FDA
3. Clinical trials (phases 1, 2, and 3)
4. NDA filed with FDA
5. FDA validates claim and approves drug

Phase 1

- Very Few (15-30) people
- Mainly toxicity (safety) and pharmacokinetic:
  - What dosage is safe? is toxic?
  - How should treatment be given?
  - How does treatment affect the body?
  - Raise dose every few subjects until they start to get sick from side effects.

Phase 2

- First dosage study
- Preliminary Efficacy study, does it seem to work?
- Usually small (< 100 people)
  - Does treatment do what it is supposed to?
- Gather even more safety data
  - How does treatment affect the body?

Phase 3

- **Large** Combined Safety and Efficacy study
  - Randomized Controlled trial
  - From 100 to many thousands of subjects
  - Compare new treatment with current standard or placebo

Phase 4

- Post Marketing studies, rare side effects
- From 100s to millions of people
- Required by FDA for surveillance, Example: Vioxx
  - Usually take place after drug is approved
  - Used to further evaluate long-term safety and effectiveness of new treatment
  - Search for rare side effects

Clinical Equipoise – genuine uncertainty if a treatment will be beneficial. Once it is certain that treatment affects participants, the trial must end.

Types of Clinical Trials

- **Treatment**
  - Test safety and effectiveness of new agents or interventions
  - Possible benefit:
    - Early access to new treatments
  - Possible risk:
    - Occurrence of unknown side effects
- **Prevention, example vaccine**
  - For people at risk of developing an illness
  - Action studies vs. agent studies
  - Possible benefit:
    - Early access to new interventions
  - Possible risk:
    - Unknown side effects and effectiveness
- **Screening and early detection**
  - Assess new means of detecting illness earlier in asymptomatic people
  - Possible benefit:
    - –Detecting disease at an earlier stage, resulting in improved outcomes
  - Possible risks:
    - Discomfort and inconvenience
    - If imaging technique is studied, exposure to x-rays or radioactive substances
    - –More confirmatory tests required

• **Diagnostic**, MRI better then CAT?
Develop better tools for classifying types of illness and managing patient care
• Possible benefits:
– New technology may be better and less invasive
– Earlier detection of recurrences
• Possible risk:
– May require people to take multiple tests

• **Genetics**
• These trials seek to:
–Determine how one's genetic makeup can influence detection, diagnosis, prognosis, and treatment
–Broaden understanding of causes of disease
–Develop targeted treatments based on the genetics

• **Quality-of-life / supportive care**
• Aim to improve quality of life for patients and their families
• Possible benefit:
–Early access to new treatment
• Possible risk:
–May not benefit from participation

Important Clinical Trial Concepts
• Having Written Procedures (Protocol)
• Why we are doing this… and how. Ensures trial is consistent. Includes eligibility, intervention, and endpoint.
• Randomization
• Prevent bias in research. Double-blind.
• Stratification
• Groups based on characteristic first before you randomize. Example: patients with diabetes make sure about even in each group. Randomize between control and investigation group.
• Categorizing subjects into subgroups by specific characteristics
• Enables researchers to look into separate subgroups to see whether differences exist: Gender, Site, specific co-morbidities
• Allocation Concealment
• **Before** patient is randomized know what they are randomized into. Don't know what she's randomized to, with no blinding.
• Blinding – researchers unaware of treatment allocation **after** randomization
– Single-blind – investigator knows treatment patient is receiving
– Open-label – patient and investigator know the treatment patient is receiving
• Such trials occur in cases when comparing, say, chemotherapy with surgery
• Have a blinded evaluation of patient outcome (someone not involved in treatment administration nor care of the patient).
• Intention to treat analysis
• Considering subgroup analyses

Protect Patient
- Nuremberg
- Tuskegee with Syphilis
- **National Commission of Protection of Human Subjects of Biomedical and Behavioral**
  - **IRB**
  - **DSMBS** with overwhelming benefit or harm, can stop early
  - **International Conference on Harmonization** Europe, Japan and US
  - **Office for Human Research Protections** – the Common Rule
- **Protect before trial:**
  - Scientific review by sponsoring organization
  - Institutional review board approval
  - Informed consent
- **During trial:**
  - IRBs
  - Data and Safety monitoring boards (DSMBs)
    - Minimize risks, ensure integrity of data

        - Can stop study if necessary
- **Adverse events**
    - Short term vs longer term
    - Longer term follow up in face of early benefit
    - Rare adverse effects may be seen only with very large numbers of exposed patients and long term follow up
- **Serious Adverse Events:** death, irreversible event, requires hospitalization
    - Must be reported to regulatory agencies and IRBs
-

Measures of Efficacy

- Clinical trials are designed to detect differences between treatment groups
  – relative risk (or relative risk reduction)
  – mean absolute risk reduction (relative to placebo)
- In clinical trials, the method of assessing the primary endpoints is usually pre-specified.
- Clinical trials are **not** designed to directly estimate the incidence in the population at risk.
- The population in a clinical trial may not completely represent the population to be treated

Characteristics of a Good Outcome Measure

- Easy to compute
- Easily understand by all (non-technical)
- Minimal variance across baseline characteristics
- Statistically sound

We need both measures RR and ARR

- Effectiveness
  – Related to RR
- Benefits
  – Related to absolute risk

Relative Risk Reduction

- Usually
  – Constant over baseline characteristic
  – Constant over study time
  – Easy test of interaction
  – When not constant it is usually piece-wise constant
  – Differences seen among different studies can be viewed as random
  – Good statistical models are available

Absolute Risk

- Unlikely
  – To be constant over time
  – To be constant over baseline characteristics
  – To be able to describe with simple models
- Consequences
  – Patients characteristics can change with study time
  – Differences among studies cannot be ignored

- If the RRR is constant and detailed information about the AR is provided both summary measures provide useful information about the effectiveness and benefits of treatment
- ARR is *not* a simple index of therapeutic effectiveness. It is a function of the incidence rate for the event of interest in the population studied and may not be reflective of the true ARR for the patient sitting before you.
- Before making a recommendation, one needs to know the risk profile of the patients to be treated

Primary vs. Secondary Question

- Primary
  – most important, central question
  – ideally, only one
  – stated in advance
  – basis for design and sample size
- Secondary
  – related to primary

– stated in advance
– limited in number

Serious Adverse Events (SAEs)- report to IRB
• Death
• Irreversible event
• Requires hospitalization

Subgroup Questions, due to stratification- effective overall, but does it show consistency?
• Questions about effect of therapy in a sub-population of subjects entered into the trial
• Assess internal consistency of results
• Confirm previous hypothesis
• Generate new hypotheses
NOTE: The greater the number of subgroups analyzed separately, the larger the probability of making false positive conclusions, increase alpha. SO DON'T MAKE TOO MANY. You must repeat study. It is just a hypothesis generator. Could be due to random sampling error- high placebo group in one set.

Subgroup Considerations
- Rules for Subgroups
  - Stated in advance (in protocol)
  - Limited in number
  - Interpreted cautiously, qualitatively
  - Look for consistency of results
- May be used to
  - Confirm or answer specific questions generated in a previous trial (e.g. Metroprolol <65 vs. >65 age total mortality
  - Generate new hypothesis to be tested in some future trial
  - Consistency of primary outcomes

Statistical Issues in Clinical Trials; center of where the treatment is carried out
• Single-center or multi-center
  – Choice often depends on logistical issues
• Can you obtain enough subjects in a single-center?
• If so, can you obtain enough in an allotted amount of time
  – Is a single center representative of the population as a whole?

Treatment-by-Center Interaction
– Why multiple centers?
  - Increase enrollment
  - Increase generalizability of results to population
– Potential issue with multiple centers
  - Center demographics/center study conduct may affect treatment performance
  - Treatment-by-center (or treatment-by-clinic) interaction: A treatment difference that varies significantly across study centers
  - FDA would like assurance such an interaction does not exist so patients across centers can be pooled into one group for analysis

Sources of Bias
1. Patient selection
2. Treatment assignment
3. Patient Evaluation
4. Data Analysis

Methods to Minimize Bias
1. Randomized Controls
2. Double blind (masked)
3. Analyze what is randomized

Cross Over
• Placebo patients taking active medicine
• Treatment patients not taking them
• "Waters down" the effect

• Always moves the groups closer together
• Need to expect some of this and increase sample size accordingly

Loss to Follow-Up
- Devastating to a clinical trial
  - can affect relationships between exposure/outcome
  - you don't know which way
- Two main reasons:
  - They feel wonderful and no longer see a need to visit a doctor
  - They feel terrible and can't get out of bed to go visit a doctor

Intention to Treat Analysis
- Regardless of cross over and loss to follow-up, subjects are analyzed in the group they were randomized into
  - In "real life" patients will also cross over and be non-compliant and will drop out
  - Allows researchers to answer a question slightly different than what is generally thought
- *Does the strategy of trying active treatment result in improvement over alternative strategy*

Exclusions – screened but not randomized; affects generalizability but validity OK

Withdrawals from Analysis – randomized, but not included in data analysis; possible to introduce bias

Surrogate outcomes
- Would like to see a trial report an outcome that has meaning to patients/society
  - Mortality, morbidity, cost savings, etc
- Sometimes this is difficult since "hard" outcomes are rare or distant events
  - ESRD & cardiac death are more rare than doubling of creatinine or change in BP
- Studies often report intermediate outcomes, or surrogates for the "hard" one

Surrogate Response Variables
- Used as alternative to desired or ideal clinical response
- Examples
  - Suppression of arrhythmia (sudden death)
  - T4 cell counts (AIDS or ARC)
- Used often - therapeutic exploratory (Phase I, Phase II)
- Use with caution - therapeutic confirmatory (Phase III)

Surrogate Response Variables
- Frequent Criticism of Clinical Trials
  - Too long
  - Too large
  - Too expensive
- Advantages
  - Perhaps smaller sample size
  - Detect earlier effect → shorter trial
  - Easier

What is a systematic review?
• A systematic review is *RESEARCH*
• The subject of the research is the medical literature
• As with all research, a SR should have clear, reproducible methods
• As with all research, a SR is subject to critical appraisal to determine if it is a valid study
- Summary of available scientific evidence
- Limited to examination of the evidence; do NOT consider expert opinion or make policy recommendations

Systematic Review (increase power, increase sample size)
• Should include
  – Clear research question
  – Criteria for inclusion/exclusion defined prior to study
  – Process used to identify studies
  – Methods of quality assessment
  – Methods to extract and summarize results

*UN*systematic Reviews
• Often found in throw-away journals
• Written by single "experts", not peer reviewed

• Often address a very broad subject area
- may be a good source of background information, especially if recent
- should not be used to guide therapeutic decision-making

Meta-Analysis Definition
• A scientific discipline which applies a protocol to critically evaluate and uses statistical methods to combine the results of (previous) research
• Provides a quantitative summary of the overall treatment effect (typically an overall estimate and confidence interval)

META-ANALYSIS
• The mathematical combination of the results of several studies to produce a single estimate of effect.
• Used to increase power, precision of point estimate.
• A subset of systematic reviews use meta-analysis
• May be used to point out areas of heterogeneity.

Clinical Applications of Meta-Analysis
• Interventions
• to estimate the effectiveness and risks of treatments
• Diagnostic Tests
• to provide more reliable estimates of the diagnostic accuracy of tests
• Epidemiology
• to provide more reliable estimates of risks

Challenges of Systematic Reviews
• Availability and quality of research to incorporate in syntheses
• Ability to find existing studies
• Difficulties with outcome measures
• Problems of production and use

Availability and quality of research
• Small sample size in many studies
• Few available studies
• Uneven study quality
• Mixed populations

Finding the evidence
• Inadequate indexing and filtering
• Lack of resources needed for comprehensive searching
• Delays in time to publication
• Publication bias - investigator and publisher factors; favors treatment for question under investigation because it shows that no significant results = less likely to be published UNLESS very large study
• Grey literature- but not peer reviewed and never published; thesis or monograph
If articles missed at random, total N will be smaller and the estimate of effect will be less exact
If articles are missed systematically (BIAS), the estimate of effect will be systematically wrong
Most common issue is publication bias

Sources of good systematic reviews:
• The Cochrane Library - Cochrane reviews have been shown to be of high quality
• Evidence Reports commissioned by AHCPR - use same methodology as CDSR
• Some practice guidelines – ex: AAP technical reports
• Many are in literature - *caveat emptor!*

Funnel Plot
- scatterplot of treatment effect against a measure of study size; used primarily as a visual aid to detect bias or systematic heterogeneity
- odds ratios comparing two groups plotted against standard error; open circle and horizontal line show point estimates and 95% CI

When to combine studies:
- When research question, populations and outcomes are comparable or sufficiently similar (not much heterogeneity)
- Valid if studies are

    - Clinically similar
    - Methodologically similar
    - Statistically similar
- Test heterogeneity using the Q statistic

Results of systematic reviews
- Estimate of effects
- Odds ratios – how these relate to relative risk
- Absolute risk different, NNT take care to use appropriate baseline risk
- "no good studies" – these reviews are important – point out areas of limited evidence, future research needs

Challenges of systematic reviews
- Availability and quality of research to incorporate in syntheses
    - Small sample size in many studies
    - Few available studies
    - Uneven study quality
    - Mixed populations
- Ability to find existing studies
    - Inadequate indexing and filtering
    - Lack of resources needed for comprehensive searching
    - Delays in time to publication
    - Publication bias, grey literature
- Difficulties with outcome measures
    - Rarity of dichotomous outcomes
    - Rarity of mortality as an outcome
    - Use of proxy measures
    - Variablility in how outcomes are measured
    - Many important outcomes are a long time in coming
    -
- Problems of production and use

Critical Appraisal Criteria
- Does the review address a focused clinical question?
- Were any relevant studies likely to be missed?
- Were inclusion criteria reasonable?
- Was the quality of the primary studies assessed objectively?

When is assignment not random?
    - "alternate" assignment
    - inclusion of all not on therapy in the control group
    - inclusion of only compliant patients to study group
    - selective exclusion of patients that are too sick

Considerations when critically appraising a study
- Patients who continue in a study are likely to do better than those who drop out
- If more than 10-20% of patients are lost to follow-up, the study may be flawed
- Assume all dropouts from control had a favorable outcome and assume all dropouts from experimental group had an adverse outcome; if benefit still demonstrated, drug is likely to be beneficial
    - Use ARR
- Analyze patients according to an "intention to treat"
- Validity asks: was the study conducted in a manner that is sufficiently rigorous to allow one to be reasonably certain that the results are likely to reflect the truth with a minimum of bias?
- Results: was the study of sufficient size to demonstrate the expected findings?

Types of conclusions from any study
- Correct positive or negative
- False negative (type II Error: risk = beta)
- False positive (type I error; risk = alpha = p)

See "commonly used measures of Health and Disease" for ARR/NNT formulas

Class 6: Disease Prevention/Reporting and Evidence Based Medicine

Disease Prevention

Reportable Diseases
EBM: Using Epidemiology in Clinical Medicine

Population and Clinical Prevention
Types of Clinical Prevention

- Immunization
  - Prevent complication of infections
  - Largely what drives schedule of pediatric visits during the first 18 months
  - Adolescent and adult vaccinations expanding
- Screening
  - Identification of asymptomatic disease or risk factors
  - Start in pre-natal period
    - Alpha-fetoprotein
  - Continue throughout life
    - Hearing in elderly
- Behavioral counseling- individual basis

- Chemoprevention
  - Use of drugs to prevent disease
  - Example: Ocular antibiotic prophylaxis of all newborns; prevent gonococcal ophthalmia neonatorum
  - Folate for child-bearing age women
  - Aspirin prophylaxis for MI
  - Statin treatment for
  - hypercholesterolemia

- Community
  - Immunization requirements for students
  - No-smoking regulations in public buildings
  - Legislation restricting firearm sales
  - Public education, regulations and taxes on tobacco
- Clinical
  - Screening for colorectal cancer
  - Help individual patients stop smoking
  - Incorporated into ongoing care (check blood pressure for patient complaining of sore throat)

Levels of Prevention
- Almost all activities in medicine could be labeled as prevention
- Clinician efforts are geared towards preventing the five (okay maybe 6) D's
  - Death
  - Disease
  - Disability
  - Discomfort
  - Dissatisfaction
  - Destitution

Levels of Disease Prevention
- Primary Prevention
  - Prevention of the disease
- Secondary Prevention
  - Detection; screen so that if effected we can be prepared for what is coming
- Tertiary Prevention
  - Reduce Disability; prevent from getting further

Reportable Diseases
- All states
  - AIDS, Varicella (chickenpox), gonorrhea, hepatitis A and B, measles, mumps, rubella, salmonella, shigella, syphilis, and TB
- Other infections vary by state
  - HIV; Texas does
- B.A. SSSMMART Chicken or you're Gone

– Hep B, Hep A, Salmonella, Shigella, Syphilis, Measles, Mumps, AIDS, Rubella, Tuberculosis, Chickenpox, Gonorrheah

PICO- EBM
- Population
- Intervention
- Comparison
- Outcome

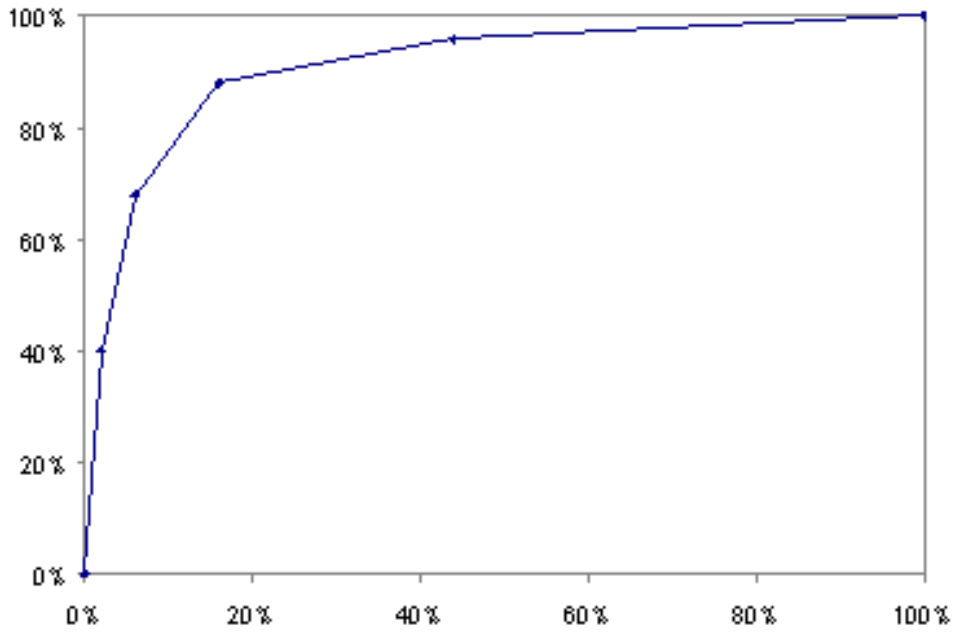
Class 8: Diagnostic Screening Tests

- Likelihood ratio – ratio of true positives:false positives or sensitivity/(1-specificity)
  - LRs can have multiple levels of results, not just "positive" or "negative"
- Use a nomogram
  - Start with baseline risk, then draw through LR to new post-test probability
- Sensitivity and Specificity – see lecture 2
  - Sensitivity: people with disease who have positive result; specificity: true negatives

	Patients with disease	Patients w/o disease	
Test "positive"	a	b	**Sensitivity** = proportion of patients with disease who have positive test **a/(a+c)**
Test "negative"	c	d	**Specificity** = proportion of patients without disease who have negative test **d/(b+d)**
Totals	100%	100%	

- ROC Curves
  - Receiver Operator (ing) Characteristic curve
  - Developed from radio operators to distinguish signal from noise
  - At each setting, sensitivity and specificity graphed

Sensitivity on Y axis



False positive rate (1-specificity)

- Assessing validity of studies about diagnostic tests
    - Independent, blind comparison to reference standard of diagnosis?
    - Was the test evaluated in an appropriate spectrum of patients?
    - Was the reference standard applied regardless of results of the test?
- Action threshold: at what level of probability are you willing to act?

Shoe Leather Epidemiology

- Epidemic Intelligence Service
    - Founded 1951 by Alexander Langmuir
    - Response to Korean War
    - First class: 22 members, Class of 2000: 65 members
- Scientific Method:
- Physician
    - H&P
    - Differential
    - Diagnostic Studies
    - Treatment
- Epidemiologist
    - Surveillance, descriptive epidemiology
    - Inference
    - Analytic epidemiology
    - Community intervention