Epidemiology 101

Association and Causality

Types of Associations

- Variable – any quantity that varies
  - Independent – the risk factor
  - Dependent – the outcome
- Association relationship between variables
  - Positive – as one variable increases so does the other
  - Negative – as one variable increases the other decreases

Possible associations

- No association
- Associated
  - Non-causally
  - Causally

Determining associations

- Statistical tests which will identify associations
- Linear graphs
- Scatter plots
- Generally things are not perfectly correlated but evaluating the general pattern gives a sense of correlation and statistical tests determine if it is significant
- The closer the points lie with a straight line, the stronger the association

Non-linear associations

- U shaped curve
- Dose-response curve
- Multimodal curve

U shaped curves

- There is an increase in risk at both tails of the curve
- Commonly used in pregnancy studies
  - Higher risk for younger and older mothers

Dose-response curve

- Increases with increasing exposure and then reaches a threshold
- Studies of birth defects may find this as initially the rate of birth defects increases with different exposures and then at some point the exposure is so great women do not become pregnant

Multimodal curve

- Several peaks in the frequency of a condition
  - May result from changes in the immune system or a long latency period

Latency

- Time period between initial exposure and a measurable response
  - Can be very short or in the case of some chronic diseases (cancer, heart disease) may take years.

Epidemic Curves

- Graphic plotting of the distribution of cases by time of onset
- Unimodal
  - Goes up rapidly and then drops off
  - Evidence for a common exposure to a sigle etiologic agent
- Prolonged
  - Goes up rapidly but then continues
  - May include continuous exposure or person to person transmission

The Famous 2x2 table

- Also called a contingency table
- Tabulates data by two variables
  - Often exposure and outcome
  - Provides a description of the relationship between exposure and disease
  - Statistical tests can be performed to determine if the two variables are significantly associated

Exposure	Outcome: Yes	Outcome: No	TOTAL
Yes	A	B	A+B
No	C	D	C+D
TOTAL	A+C	B+D	

A= Exposure is present and outcome (disease) is present

B= Exposure is present and outcome (disease) is absent

C= Exposure is absent and outcome (disease) is present

D= Exposure is absent and outcome (disease) is absent

- Example

In a study among 200 students in an undergraduate class, 50 reported having had headaches in the last week. Of those with a headache, 25 reported staying up later than usual while 25 reported getting a normal amount of sleep. Of those 150 students without a headache, 30 reported getting less sleep and 120 reported getting a normal amount of sleep.

Create a 2x2 table showing this information.

Epidemiologic Research Strategies

- Epidemiologists seek to determine if a given exposure is causally associated with a given outcome.
- Epidemiologists develop a hypothesis.
  - Conjecture cast in a form that will allow it to be tested and refuted.

Ways to derive hypotheses

- Method of difference
  - Situation in which all the factors in two or more domains are the same except for a single factor. Differences in disease is hypothesized to result from differences in a risk factor.
- Method of concomitant variation
  - Situation in which the frequency of an outcome increases with the frequency of exposure to a factor. Dose-response relationship

Operationalization

- Refers to the process of defining measurement procedures to collect data on the variables used in a study
  - Developing and adminstering a survey
  - Conducting a medical record review

Causality in Epidemiologic Studies

- Central concern
- Determine if an association is causal
- If non-causal, it can be due to chance, errors in measurement, or due to the fact that a risk factor and outcome are both associated with another factor that causes the outcome.
- Developed criteria for determining causality

Determining Causality

- Challenging since many epidemiological studies are observational, and less precise than laboratory studies where the experimenter controls all aspects
- Developed guidelines for identifying causality
- Chapter 1, we discussed Koch's theory of causality

Criteria of causality defined by Bradford Hill

- Strength of the association
- Consistency
- Specificity
- Temporality
- Biologic gradient
- Plausibility
- Coherence
- Analogy

Strength of the association

- The stronger the association, the more likely something is causal
- Not an absolute as it could be that something that does not have a strong relationship still causes disease but just in a smaller number of exposed individuals

Consistency

- No single study proves causality
- It is a matter of looking at all the evidence and determining how consistent the relationship is

Specificity

- Is of somewhat limited value as it states that an exposure only causes one disease
- Somewhat more applicable in infectious disease but even then there is variation
- Some common exposures, smoking, high fat diet, sedentary lifestyle, result in several outcomes, cancer, heart disease, etc.

Temporality

- The cause needs to come before the effect
- Start with people who do not have the outcome before they have the exposure
- It can be difficult to sort out at times

Biologic gradient

- Dose-response curve
- Linear trend between the amount of exposure and the probability of an outcome

Plausibility

- Association must be biologically plausible, i.e., it must make sense
- There are cases in which our lack of information makes it difficult to determine this
  - Helicobacter pylori and stomach cancer

Coherence

- The proposed cause and effect should not seriously conflict with what is known about the natural history and biology of the disease

Analogy

- Relates to the correspondence between known associations and the one being evaluated for causality
- One type of exposure causes illness so it is possible others act similarly

Types of causality

- Single Cause of Disease
  - Influenza virus causes influenza
  - However, there are other important factors, such as the immune system function
- Multifactorial cause of disease
  - Many factors lead to a disease
  - More commonly seen in chronic disease
    - Heart disease
      - Genetics
      - Smoking
      - Exercise
      - Diet

Determining causality

- Epidemiologists use statistical techniques to determine how likely it is that associations are due to chance.
- Inference is using observations to generalize about an underlying population.
- First identify the best estimate of the association.
- Then you determine if that estimate is likely due to chance or if it is significant.

Example of point estimate

- First you obtain a sample. As you remember from your M&M exercise, each sample differs as a result of chance and possibly other factors.
- Point estimate refers to the single best estimate of an association between an exposure and a disease or the single best estimate of the prevalence of a condition.

Confidence Interval

- Range of possible values
- Each time you get a sample you get a slightly different answer
- We determine the range within which we are 95% certain the sample prevalence reflects the true prevalence.

Importance of Power

- Power is very important in determining if a result is significant
- The larger the sample size, the more power you have to find a significant difference
- But you also want the difference to be clinically significant.
  - Does it matter?