Analytic Epidemiology

Types of study design

Analytic studies

- Analytic studies serve to identify the causes of disease and thus are important as they can lead to the prevention of disease
- They are also useful for evaluating treatments and other health interventions
- They can be either observational or interventional
- They explore associations between exposures and outcomes

Factors to help identify the type of study design

- How individuals are selected for the study
- Number of observations
- Directionality of the exposure
- Data collection methods
- Timing of data collection
- Unit of observation
- Availability of subjects

The selection of study subjects

- Have a new disease or exposure
  - Case series, case report
- Representative of a group
  - Cross sectional study, ecological study
- Chosen based on having and not having a disease
  - Case-control study
- Chosen based on having and not having a certain exposure
  - Cohort study
- Chosen because they have (or are at risk of having) a condition and the researcher wants to evaluate a treatment
  - Randomized clinical trial, intervention study

Number of observations

- Only one data collection period
  - Generally a cross-sectional or ecologic study, and many case-control studies
- Multiple observations and followed over time
  - Cohort study, randomized clinical trial, and community intervention studies

Directionality of exposure

- Retrospective approach - a researcher obtains information on past exposures
  - Common in case-control studies
- Single point in time
  - Cross-sectional study, ecologic study
- Prospective approach - a researcher obtains information about the outcome that is collected in the future
  - Cohort studies, randomized clinical trials, intervention studies
Exposure  Disease  Population
Sun exposure  Skin cancer  US Southern vs. Northern states

Exposure  Disease  Population
Sun exposure  Skin cancer  US Southern vs. Northern states

Exposure  Disease  Population
Sun exposure  Skin cancer  US Southern vs. Northern states

Exposure  Disease  Population
Sun exposure  Skin cancer  US Southern vs. Northern states

Data collection methods
- Can use existing data
  - Advantage of existing data is that it is already available and this is usually less costly
  - Disadvantage of existing data is that the data collected might not include everything that a researcher needs
  - Need permission to use it
  - Existing data is often used in ecologic studies
- Can collect new data – interviews, examinations
  - Advantage is that you get what you need
  - Disadvantage is that it is expensive and time consuming

Timing of data collection
- Using data that occurs in past
- Concern about the quality of data
  - How often did you eat chicken when you were 10 years old?
  - What did you have for lunch 3 days ago?
- Currently collected data is less subject to problems with recall but if you follow people over time they may drop out of your study and if a large number drops out it may effect your results

Availability of subjects
- Some individuals are challenging to study for practical or ethical reasons
  - A study of people with dementia is difficult as they have trouble remembering things
  - Harder to enroll minority and lower SES populations in some studies
  - Protections against enrolling pregnant women, prisoners, young children

Ecologic study
- Group is the unit of analysis
- We do not have individual information
- Look for correlations between exposures and outcomes
  - Figure 6-2 and Table 6-2
- Ecologic fallacy
  - Just because there are correlations between exposures and outcomes in community-wide studies, the individuals with the outcome may not actually have the exposure

Ecologic assignment
- Review the material on ecologic studies and write down two examples of research questions that can be answered by an ecologic study and what population you would use. List the exposure, disease, population

Some important census bureau information
- The Census Bureau divides the country into different sized groups, starting with states, and then counties
- The next division is a census tract, with between 1,000 and 8,000 people
- The next division is a census block. There are many census blocks in a census tract
- Metropolitan areas – large population nucleus, together with adjacent communities that have a high degree of economic and social integration
Concerned that cigarette smoking causes an increase in headaches so I did a study with headaches and found that 20% of the smokers did. Do you think it is likely that cigarette smoking causes headaches?

**YES, NO, UNSURE**

What other piece of information might help you better answer that question?

**Smoking and headache solution**

- What you would want to know is how many people without headaches smoke.
- In the U.S. about 20% of the people smoke so it is not likely that the study I described showed that smoking causes headaches.

**Case-control studies**

- Use a retrospective approach
- Ask about past exposures
- Can look at many different exposures
- Researchers have a number of sources for cases and controls
- Cases can be hospital patients, listed in a disease registry
- Controls are more challenging to find as you want them similar to cases in many ways
  - Other hospital patient, community members

**Odds ratio**

- To determine if a disease is associated with an exposure, you do an odds ratio
  - Indirect measure of risk
  - Easy to calculate
  - See table 6-3

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

Odds ratio = \( \frac{A \times D}{B \times C} \)

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Lung Cancer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>80</td>
<td>30</td>
</tr>
<tr>
<td>No</td>
<td>20</td>
<td>70</td>
</tr>
</tbody>
</table>
Odds ratio answer

- 9.3
- This means that people with lung cancer are 9.3 times more likely to smoke than people without lung cancer

Advantages of case-control studies

- Can be used to study rare diseases
- Relatively quick to complete
- Usually less expensive
- May involve a smaller number of subjects

Disadvantages of case-control studies

- Measurement of exposure may be inaccurate
- Cases and controls may not fully represent the population
- Indirect estimate of risk
- It can be hard to identify the temporal relationship between the exposure and the outcome

Cohort study

- Individuals are selected by exposure
- They are free of disease at the start of the study
- Followed over time to see if they develop the illness
- Can calculate a direct measure of risk, called the relative risk
- You need non-exposed people so you can tell what the likelihood of the disease is in the general population

Calculating risk in a cohort study

- Calculate the incidence in the exposed group
- Divide that by the incidence in the non-exposed group
- See page 32 for a review of incidence

Relative Risk

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>C</td>
</tr>
</tbody>
</table>

Relative risk = \((a/b)/(c/d)\)
Relative Risk

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>80</td>
<td>300</td>
</tr>
<tr>
<td>No</td>
<td>20</td>
<td>700</td>
</tr>
</tbody>
</table>

Relative risk = \( \frac{80/80+300}{20/20+700} \)

Relative risk answer

- 10.5
- Used different numbers from the odds ratio example because lung cancer is fairly rare so we would not expect to have large numbers of people with lung cancer

Advantages of a cohort study

- Permit direct observation of risk
- Exposure factor is well defined
- Can study rare exposures
- The temporal relationship is known

Disadvantages of a cohort study

- Expensive
- Time consuming
- Complicated
- Subjects may be lost to follow up
- Difficult to follow people with diseases that take a long time to develop

Experimental studies

- Randomized clinical trials
- Used to test the effectiveness of treatments
- Measure the relative risk of disease between the treatment groups
- Often, though another analysis called survival analysis is used
  - Beyond the scope of this class

Quasi experimental studies

- Community interventions
- Program evaluations
Challenges to the validity of studies

- **External validity**
  - How generalizable are the results?
  - If you study heart disease in white men over age 50, what does it tell you about heart disease in black women under 50?
- **Sampling error**, by chance you get a sample that is very different than the population you want to study
  - Increased numbers help prevent this
- **Internal validity**
  - Are you measuring what you want to measure

Bias in studies

- Systematic deviation of results or inferences from the truth
- **Hawthorne effect**
- **Recall bias**
- **Selection Bias**
- **Healthy Worker effect**

Hawthorne effect

- Let's say you are in a study of your diet and you have to keep track of everything you eat for one day. You might be less likely to eat candy and more likely to eat fruit since you know someone is watching you.

Recall bias

- Two omen just had babies, one with a heart defect and one normal. They both took prenatal vitamins 5 days a week instead of 7.
  - The woman who has a baby with a heart defect is trying to think why the baby has this and she remembers all the time she skipped her vitamins so she tells a researcher she only took them occasionally
  - The woman who has the normal baby isn’t even thinking about that so she told the researcher she took them most of the time

Selection bias

- There are differences in how people enroll in your study.
  - You do a study of people with cancer treated at Moffitt and your controls are patients at Tampa General
  - Moffitt patients are richer than the Tampa General patients
  - You find the Moffitt patients were less likely to smoke but this could be because richer people smoke less than poorer people and was not related to the disease

Healthy Worker Effect

- People who are employed overall are healthier than those who don’t so if you compare people working at a construction site with the general population some of your comparison subjects may be unemployed because they were not healthy enough to work at the construction site
Confounding

- Distortion of a measurement due to the association of the exposure with some other factor that is associated with the disease

- People who carry matches have higher rates of lung cancer than those who do not carry matches
- TAB and bladder cancer

- What other possible examples can you think of?