Epidemiological Study Design: Basic Structures, Measures, and Limitations

Neal Simonsen

LSU Health Sciences Center
School of Public Health
&
Stanley S. Scott Cancer Center

Defining Epidemiology

“…the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to control of health problems.”

--Last, 1988, The Dictionary of Epidemiology

Another Definition of Epidemiology

➢ “The science of making the obvious obscure.”

--Anonymous Epidemiologist

Epidemiology Defining Itself

Etymological derivation: From the Greek

➢ “Epi” on/upon +

➢ “demos” the people +

➢ “logos” theory or study of

Characteristics of Epidemiology

➢ Concerned with the frequencies and types of illnesses, injuries or other outcomes in groups of people and the factors that influence their distribution.

(or whatever outcome is of interest…)

Characteristics (Continued)

➢ This implies that disease is NOT randomly distributed throughout a population, but rather that subgroups differ in the frequency of different diseases.

➢ Knowledge of this uneven distribution can be used to investigate causal factors and thus to lay the groundwork for programs of prevention and control.

➢ Can similarly be used to study consequences of different treatments
Study Designs In Epidemiology...

Basic Considerations; Fundamental Designs

How do we know if an observed association reflects a causal relationship?

Exploring Disease Etiology

Dose/Treatment

Environmental Exposure -> Cancer

Experimental Study Design

Exposed

Disease nonoccurrence

Not exposed

OK

Unethical to perform experiments on people if exposure is harmful

Exposed

Not OK

The next step in determining causation:

Conducting Studies in Human Populations

- Observational Epidemiology often key here....
- Allows capitalization on “natural” or “unplanned” experiments.
- Take advantage of groups who have been exposed for non-study purposes.

Conventional Hierarchy of Study Designs

Clinical observations (case series)

Ecological or Cross-Sectional studies

Case - control studies

Cohort studies

Randomized trials*

*(if potential beneficial intervention identified)
**Ecologic Study**
- Units of analysis are populations or groups of people, rather than individuals.
  - Often exploit pre-existing data collected for other purposes
  - Efficient and economical design

**Cross-sectional Study**
- Draw sample from population of interest at particular time
- Identify cases and non-cases of disease
- Measure characteristics (exposures)
- Examine associations between characteristics and disease

**Key potential limitation: The ecologic fallacy**
- Attributing to members of a group characteristics that they do not possess as individuals
  - E.g., only know average values of fat consumption by country
  - Don’t know if individuals with breast cancer actually had higher fat intake

**Example: Is stress associated with symptoms of TMD?**
- Random sample of population (N=680)
- Interviewed re: symptoms of TMD (pain, joint sounds, limited opening)
- Measure of life stress (e.g., Life Events and Difficulties Schedule (LEDS), Daily Life Experiences Checklist)

**Stress and TMD**

Percent reporting frequent stress:

<table>
<thead>
<tr>
<th>With TMD symptoms</th>
<th>Without symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>56%</td>
<td>21%</td>
</tr>
</tbody>
</table>

P<0.05
Cross-sectional studies:

- Can assess associations
- Cannot establish correct temporal relationship for inferring causation
  - Why?
  - Factor and disease measured at same point in time

Cohort Study

Exposed
  - Develop Disease
  - Do Not Develop Disease

Not Exposed
  - Develop Disease
  - Do Not Develop Disease

The Cohort Concept

Following the cohort through time...

...to the end of the study period.

Analytical Design of a Cohort Study

<table>
<thead>
<tr>
<th></th>
<th>Disease develops</th>
<th>Disease does not develop</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>First select</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposed</td>
<td>a     b</td>
<td>a + b</td>
<td></td>
</tr>
<tr>
<td>Not exposed</td>
<td>c     d</td>
<td>c + d</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a + b</td>
<td>a + b</td>
<td></td>
</tr>
</tbody>
</table>

End of Follow-Up
Relative Risk

Risk in exposed
Risk in non-exposed

Relative Risk Calculation for Cohort Study

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Do Not</th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke Cigarettes</td>
<td>84</td>
<td>2,916</td>
</tr>
<tr>
<td>Do not Smoke Cigarettes</td>
<td>87</td>
<td>4,913</td>
</tr>
</tbody>
</table>

RR = 28.0 / 17.4 = 1.61

Advantages of cohort studies
- Temporal relationship more certain
- Less opportunity for distortion of exposure data
- Can examine multiple disease outcomes

Disadvantages of cohort studies
- Can be time consuming and expensive – follow large group over long period
- Potential bias due to drop-outs from study

The Two Major Flavors of Cohort Studies: It’s All in the Timing

Concurrent Cohort Study Begun in 1995

- Defined Population
  - Non-Randomized (Occurs naturally)
  - Exposed
    - Develop Disease
    - Do Not Develop Disease
  - Not Exposed
    - Develop Disease
    - Do Not Develop Disease
Retrospective Cohort Study Begun in 1995
[AKA Historical Cohort Study]

Defined Population
Non-Randomized

1975

1985

1995

Exposed
Do Not Develop Disease

Develop Disease

Do Not Develop Disease

Not Exposed

Exposed
Do Not Develop Disease

Develop Disease

Do Not Develop Disease

The Framingham Study - 1948

- Study of cardiovascular disease
- Framingham - 20 miles from Boston
- Population under 30,000
- Selected:
  - Residents between 30 - 62 years of age
  - Free of cardiovascular disease
- Sample size of 5,000

Exposures Being Investigated

- Exposures defined as:
  - Smoking
  - Obesity
  - Elevated blood pressure
  - Elevated cholesterol levels
  - Low levels of physical activity
- Study population examined every two years
  - Passing 60-year follow-up mark!

Biases in Cohort Studies

- Bias in the assessment of the outcome
  - If person assessing disease also knows exposure status
    - May be addressed by blinding
- Information bias
  - Quality of information not comparable between exposed and non-exposed individuals

Biases in Cohort Studies (Cont’d)

- Biases from non-response and losses to follow-up
  - Non-participation / non-response and dropouts can complicate interpretations of findings
- Analytic bias
  - Biases of epidemiologists/biostatisticians analyzing the data

- A key practical drawback (even if you go retro)?
  - For rare outcome(s), need to follow lots of people to get meaningful # of events
- An alternative study offering solution?...
  - The case-control gambit
Odds Ratio: The Measure of a Case-Control Study

<table>
<thead>
<tr>
<th>History of Exposure</th>
<th>Cases (With Disease)</th>
<th>Controls (Without Disease)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No History of Exposure</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Odds ratio = \( \frac{a/c}{b/d} = \frac{ad}{bc} \)

Example: Case-Control Study of CHD and Cigarette Smoking

<table>
<thead>
<tr>
<th></th>
<th>CHD</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke cigarettes</td>
<td>112 (a)</td>
<td>176 (b)</td>
</tr>
<tr>
<td>Do not smoke cigarettes</td>
<td>88 (c)</td>
<td>224 (d)</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>% Smoking cigarettes</td>
<td>56.0</td>
<td>44.0</td>
</tr>
</tbody>
</table>

Odds ratio = \( \frac{a/c}{b/d} = \frac{112/88}{176/224} = 1.62 \)

What does this number mean?
- The odds of smoking are 1.62 times as great for CHD cases as they are for persons without CHD (i.e., controls)

Selection of Cases and Controls
- Incident or prevalent cases
  - New cases versus pre-existing diagnosed cases
  - Better to use new cases (avoid selective survival, etc.)

Matching
- Individual matching
  - E.g., for each 19 year old white female select a control who is 19, white, and female
- Group matching
  - E.g., if 25% of cases are married then select to assure 25% of controls married
Advantages of case-control studies

- ‘Relatively’ quick and inexpensive
- Good for rare disease
- Can examine multiple exposures at same time

Disadvantages of case-control studies

- Information on exposure relies on records or recall – can be inaccurate
- Can be difficult to establish correct temporal relationship

Recall Problems

- Limitations in human recall
  - E.g., 34% of circumcised thought they were not
- Recall bias
  - Person with illness may be more likely to concentrate on (or conjure up) remembering exposure

What could go wrong…???

- Problems that can arise across the study design spectrum…

More Bias

- Selection bias
  - In the choice of cases vs. controls, exp vs. non-exp
- Surveillance bias
  - E.g., Phlebitis and oral contraceptive use; MDs looking for it more closely
- Misclassification bias
  - Differential
  - Non-differential

Reported Causes of Death

“A mother died in infancy”
“Decreased had never been fatally sick”
“Died suddenly, nothing serious”
“Went to bed feeling well, but woke up dead”
“Died suddenly without the aid of a physician”
Information Bias

- Bias in abstracting records
- Recall bias (case-control)
- Interviewer bias
- Bias from surrogate interviews (us. case-control)
- Non-response bias

Summary: Epidemiologic Study Types

- Studies of group characteristics
  - Ecologic
    - Potential ecologic fallacy
- Studies of individual characteristics*
  - Cross-sectional
  - Case-Control
  - Cohort
  - Intervention/Clinical Trial

* [within one group vs. another(s)]

✓ Done!