Epidemiology: Basics of Study Design, 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 and injuries in *groups* of people and the *factors* that influence their distribution.
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

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

Prevalence vs. Incidence rates

➢ Prevalence
   ➢ Proportion of persons in a population who have a particular status (presence of disease or some other health-related condition) at either
     1) A specified point in time
     2) A specified period in time
   ➢ Point vs. period prevalence

Incidence

➢ Incidence
   ➢ The rate of new occurrences of a condition in a population
   ➢ (New occurrences of a condition [=event] during a specified time period / the population during that same specified time period)

Prevalence vs. Incidence rates

➢ Prevalence (continued)
   ➢ Amount of disease prevailing in a population at a given time or within a given period
   ➢ What if we are interested in how quickly new cases are developing in a population?
   ➢ What if the condition of interest lasts a long time (years)?
     E.g., osteoarthritis

Study Designs In Epidemiology...

Basic Considerations; Fundamental Designs

Exposure
OR
Genetic Background
OR
Combination of Both

Disease or Other Outcome

Association

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

Environmental Exposure $\rightarrow$ Disease nonoccurrence

Experimental Study Design

Disease nonoccurrence

Unethical to perform experiments on people if exposure is harmful

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.

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

Clinical observations (case series)

Ecological or Cross-Sectional studies

Case - control studies

Cohort studies

Randomized trials*

*If potential beneficial intervention identified

Correlation between dietary fat intake and breast cancer by country.

Incidence Ratio per 100,000 Women

Per Capita Supply of Fat Calories
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 had higher fat intake

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

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

Stress and TMD
Percent reporting frequent stress:
- Those with TMD symptoms = 56%
- Those without symptoms = 21%

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.

End of Follow-Up

Relative Risk

Risk in exposed
Risk in non-exposed

Analytical Design of a Cohort Study

Relative Risk Calculation for Cohort Study
Interpreting RR of a Disease

If \( RR = 1 \) Risk in exposed equal to risk in unexposed (no association)
If \( RR > 1 \) Risk in exposed greater than risk in unexposed (positive association; possibly causal)
If \( RR < 1 \) Risk in exposed less than risk in unexposed (negative association; possibly protective)

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 periods of time
- Potential bias due to drop outs from study

Drop out in cohort study of oral health of older adults

<table>
<thead>
<tr>
<th>Time</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>907</td>
</tr>
<tr>
<td>3 years</td>
<td>611</td>
</tr>
<tr>
<td>7 years</td>
<td>425</td>
</tr>
</tbody>
</table>

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

Concurrent Cohort Study Begun in 1995

1995

Defined Population
Non-Randomized (Occurs naturally)
Exposed
Develop Disease
Do Not Develop Disease
Not Exposed
Develop Disease
Do Not Develop Disease

2015
Retrospective Cohort Study Begun in 1995

Defined Population
Non-Randomized

Exposed
Develop Disease
Do Not Develop Disease

Not Exposed
Develop Disease
Do Not Develop Disease

1975
1985
1995