Introduction to Epidemiology

Charles DiMaggio PA-C, MPH, PhD
Goals

- Define Epidemiology
- Discuss Concepts in Infectious Disease Epidemiology
- Introduce Biostatistical Concepts
- Introduce Study Designs
Defining Epidemiology

- Epidemiology is What Epidemiologists Do
- Epidemiology is Reasoned Argument
- Epidemiologists Count Things
- Epidemiologists are Clinicians to a Community
Defining Epidemiology

Epidemiology is the study of the distribution and determinants of disease in human populations.
History

Nebuchadnezzar and Daniel

James Lind – Scurvy
- HMS Salisbury 1740 – 1744

John Snow – Cholera
- 19th Century London
- Laid the foundation for modern epidemiology
## The Epidemiology of Cholera

<table>
<thead>
<tr>
<th>Water Company</th>
<th>Number of Houses</th>
<th>Deaths from Cholera</th>
<th>Deaths per 10,000 Houses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southwark and Vauxhall</td>
<td>40,046</td>
<td>1263</td>
<td>315</td>
</tr>
<tr>
<td>Lambeth</td>
<td>26,107</td>
<td>98</td>
<td>37</td>
</tr>
<tr>
<td>Rest of London</td>
<td>256,423</td>
<td>1422</td>
<td>59</td>
</tr>
</tbody>
</table>
Austin Bradford Hill

“The highest returns can be reaped by imagination in combination with a logical and critical mind, a spice of ingenuity coupled with an eye for the simple and humdrum, and a width of vision and pursuit of facts that is allied with an attention to detail that is almost nauseating”

“Nature makes the experiments, and we watch and understand them if we can”
Concepts in Infectious Disease
Epidemiology

- Glossary of Disease Transmission
- The Epidemiologic Triad
- Chain of Transmission
Glossary of Disease Transmission

- Epidemic
- Outbreak
- Cluster
- Endemic
- Pandemic
Epidemiologic Triad

- **Agent**
  - Nutritive, Chemical, Physical, Infectious
- **Host (Person)**
  - Inborn, Acquired, Behavioral
  - Susceptible, Immune, Infected
- **Environment**
Disease Transmission

- **Time (Epidemic Curve)**

- **Chain of Transmission**
  - Source
  - Portal of Exit
  - Mode
    - direct vs. indirect (airborne, vector, vehicle, droplet nuclei, fomites)
  - Portal of Entry
Introduction to Biostatistical Concepts

- Measures of Disease Frequency
- Measures of Effect
- Statistical Significance
Measures of Disease Frequency

- Ratios (Odds)
- Proportions (Risk)
- Rates
  - Incidence Rate
  - Cumulative Incidence
- Relative Rates
- Difference Between a Rate and a Risk
Measures of Effect

The Four-Fold Table

Relative Effects = Ratios

Absolute Effects = Differences

Attributable Proportion
# Difference vs. Ratio

Mortality Rates per 100,000 person-years form lung cancer and coronary artery disease for smokers and non-smokers

<table>
<thead>
<tr>
<th></th>
<th>Smokers</th>
<th>Non-Smokers</th>
<th>Odds Ratio</th>
<th>Risk Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Cancer</td>
<td>48.3</td>
<td>4.5</td>
<td>10.8</td>
<td>43.8</td>
</tr>
<tr>
<td>CAD</td>
<td>294.7</td>
<td>169.5</td>
<td>1.7</td>
<td>125.13</td>
</tr>
</tbody>
</table>
The Importance of Significance

- Chance and Probability
- P-values
- Confidence Intervals

1 is the loneliest number
Study Designs

- Overview
- Ecologic Studies
- Cross-sectional Studies
- Randomized Clinical Trials
- Case-Control Studies
Overview of Study Designs

- Ceteris Paribus
- Experiments and Quasi-Experiments
- Observational Studies
  - Descriptive Studies
    - Ecologic, Cross-sectional
  - Analytic Studies
    - Randomized Clinical Trials, Cohort, Case Control
Ecologic Studies

- The Ecologic Fallacy
- Durkheim
- Robinson
- Lung Cancer and Pollution
Cross-Sectional Studies

Directionality

Incidence-Prevalence Bias

<table>
<thead>
<tr>
<th></th>
<th>1-2 yrs</th>
<th>3-7 yrs</th>
<th>&gt;7 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>% who have ever received AFDC</td>
<td>30%</td>
<td>40%</td>
<td>30%</td>
</tr>
<tr>
<td>% receiving AFDC at particular time</td>
<td>7%</td>
<td>28%</td>
<td>65%</td>
</tr>
</tbody>
</table>
Randomized Clinical Trial

- Selection of Study Population
  - Reference, experimental, actual
- Random Allocation
  - Placebo
- Blinding
  - Double, triple
- Weaknesses
Cohort Study

- Definition and Conduct
- Types
- Strengths
- Weaknesses
- Analysis

- RR = (A/A+B) / (C/C+D)
Case-Control Studies

- Definition and Conduct
- Strengths
- Weaknesses
  - Temporality
  - Control Group
  - Recall Bias
- Analysis: The Odds Ratio
The Odds Ratio

- Prospective vs. Retrospective Approach
- Need for a New Measure of Effect
- Exposure Odds Ratio
- Disease Odds Ratio
- Rare Disease Assumption
- \( OR = \frac{ad}{bc} \)
Threats to Validity

Confounding
- Causally related to the disease
- Associated with the exposure
- Not a result of study design

Bias
- Systematic error
- Design, conduct or analysis
Confounding

- **Definition**

- **Examples**

\[
\text{OR} = \frac{ad}{bc} = \frac{(90)(60)}{(60)(90)} = 2.25
\]

<table>
<thead>
<tr>
<th>Myocardial Infarction</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee Drinking</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>90</td>
<td>60</td>
<td>150</td>
</tr>
<tr>
<td>No</td>
<td>60</td>
<td>90</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>150</td>
<td>300</td>
</tr>
</tbody>
</table>
# Confounding

## Control

### Design, Analysis

<table>
<thead>
<tr>
<th>Smokers</th>
<th>Non-Smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>MI</td>
</tr>
<tr>
<td>No MI</td>
<td>No MI</td>
</tr>
<tr>
<td>Coffee</td>
<td>Coffee</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

**Totals**

| 100 | 50  | 50  | 100 |

**OR =**

| 1.0  | 1.0  |
Bias

- Recall Bias
- Diagnosis Bias
- Hawthorne Effect
- Selection Biases
Selection Biases

- Detection (Surveillance) Bias
- Incidence Prevalence (Survivor) Bias
- Loss to Follow up (Non Response Bias)
- Health Worker Effect
- Volunteer Bias
- Berkson’s Bias
Berkson’s Bias

Truth
- No relationship between vag bleed and endometrial CA
- OR = 1

Bias
- Probability of admission varies by diagnosis
- OR = 4.3

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Endometrial</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaginal Bleeding</td>
<td>Yes</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Endometrial</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaginal Bleeding</td>
<td>Yes</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>163</td>
</tr>
</tbody>
</table>
Standardization of Rates

Age-Specific Mortality Rates (per 1000): Sweden and Panama

<table>
<thead>
<tr>
<th>Age</th>
<th>Sweden</th>
<th>Panama</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-29</td>
<td>1.1</td>
<td>5.3</td>
</tr>
<tr>
<td>30-59</td>
<td>3.6</td>
<td>5.2</td>
</tr>
<tr>
<td>&gt;60</td>
<td>45.7</td>
<td>41.6</td>
</tr>
</tbody>
</table>

Next, we choose some standard age distribution, let’s say:

<table>
<thead>
<tr>
<th>Age</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-29</td>
<td>3,145,000</td>
</tr>
<tr>
<td>30-59</td>
<td>3,057,000</td>
</tr>
<tr>
<td>&gt;60</td>
<td>1,294,000</td>
</tr>
</tbody>
</table>

Multiply the age-specific mortality rate by the standard population rates:

<table>
<thead>
<tr>
<th>Age</th>
<th>Sweden</th>
<th>Panama</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-29</td>
<td>1.1 x 3,145,000 = 3,459.5</td>
<td>5.3 x 3,145,000 = 16,668.5</td>
</tr>
<tr>
<td>30-59</td>
<td>3.6 x 3,057,000 = 11,005.2</td>
<td>5.2 x 3,057,000 = 15,896.4</td>
</tr>
<tr>
<td>&gt;60</td>
<td>45.7 x 1,294,000 = 59,135.8</td>
<td>41.6 x 1,294,000 = 53,838.4</td>
</tr>
</tbody>
</table>

Divide by the total standard populations to get the standardized mortality rates:

Sweden = 73,599.5 / 7,496,000 x 1000 = 9.8 per 1000

Panama = 86,403.3 / 7,496,000 x 1000 = 11.5 per 1000