Outline

- Definitions: What is Epidemiology/ Biostatistics
- Measures of morbidity and mortality data
- Study designs
  - Cross sectional
  - Cohort
  - Case control
  - Randomized controlled studies

Definition of Epidemiology: Old

- In Greek: EPI DEMOS LOGO
- Old definition: Study of epidemics
  Occurrence of a group of cases of illness clearly in excess of normal expectancy at one point in time
**Definition of Epidemiology**: New

- **Newer definition**: Study of distribution and determinants of disease in human populations

**Definition of Statistics**

Statistics is a field of science concerned with:

1. Collection, organization, summarization and analysis of data
2. Drawing of inferences about a body of data when only a part of the data is observed.

Daniel W.

**Definition of Biostatistics**

Biostatistics is the application of statistical methods to biological and health related questions. It is a component required for the completion of an epidemiological study
**Inference:** Drawing conclusions about a population from a sample selected from the population.

**Why do we need biostatistics?**

- **Descriptive:** Collection, organization, summarization and analysis of data.
- **Inferential:** Drawing inferences about a body of data when a part of the data is observed.

**Biostatistics**

**Measures of disease**

- Measures of disease are morbidity and mortality.
- These measures are expressed by:
  - Absolute numbers
  - Rates
  - Ratios
Measures of disease: morbidity

Incident cases
- Number of NEW cases of a disease occurring in the population during a specified time period

Prevalent cases
- Number of cases of a disease that are PRESENT in the population

Type of variables

- Dependent variable: Variable of interest. Known as well as the outcome.

- Independent variable: Predicts or influences the dependent variable. Known as well as the exposure.

Examples:
- Smoking during pregnancy ---------- Low birth weight
- Physiotherapy versus drug ---------- Low back pain
- Physical activity ------------------ Prostate cancer
Cross Sectional Studies

- Called also prevalence studies
- Exposure and disease are assessed simultaneously among individuals in a defined population.
Cross Sectional Studies

Exposure & outcome
Exposure & no outcome
No exposure & outcome
No exposure & no outcome

Sample Population

Exposure | Outcome | Yes | No |
---|---|---|---|
Yes | a | b |
No | c | d |

Prevalence rate of outcome in exposed ($P_E$) = $\frac{a}{a + b}$
Prevalence rate of outcome in non-exposed ($P_Ē$) = $\frac{c}{c + d}$
Prevalence Rate Ratio (PRR) = $\frac{P_E}{P_Ē} = \frac{a}{a + b} \div \frac{c}{c + d}$

Cross Sectional Studies: Example

Association between air pollution and chronic respiratory disease (in a specific plant, at a specific time of the year)

| Air pollution | Chronic Respiratory disease |
|---|---|---|
| Yes | 25 | 111 | 136 |
| No | 12 | 221 | 233 |

Prevalence rate among exposed group = $25/136 = 0.18$
Prevalence rate among non-exposed group = $12/233 = 0.05$
Prevalence rate ratio = $0.18/0.05 = 3.6$
### Study Designs

- Experimental studies
- Observational studies
  - Analytic studies
  - Descriptive studies
- Randomized Controlled trials
- Case control
- Cohort
- Cross sectional

### Cohort Study: Direction of inquiry

**POPULATION**

- Population
- People without the outcome

**Sample of people exposed**

- Outcome
- No Outcome

**Sample of people not exposed**

- Outcome
- Outcome

Start with **EXPOSURE**
Follow up
Check for **OUTCOME**

### Cohort Study: Two by two table

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Risk of outcome in exposed ($R_E$) = $a / a + b$

Risk of outcome in non-exposed ($R_{Ee}$) = $c / c + d$

Relative risk (RR) = \[
\frac{R_E}{R_{Ee}} = \frac{a / (a + b)}{c / (c + d)}\]
**Cohort Study: Relative risk**

Relative Risk provides the best estimate of the strength or magnitude of the exposure-outcome association and is useful in making causal inferences.

- $RR = 1$ indicates no exposure-outcome association: risk in exposed = risk in non-exposed.
- $RR < 1$ indicates negative association (protective): risk in exposed < risk in non-exposed.
- $RR > 1$ indicates positive association (causal): risk in exposed > risk in non-exposed.

**Cohort Study: Example**

<table>
<thead>
<tr>
<th>Coronary heart disease</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>Obesity</td>
<td>5</td>
<td>45</td>
</tr>
</tbody>
</table>

Risk of outcome in exposed ($R_E$) = $12 / 50 = 0.24$

Risk of outcome in non-exposed ($R_{E\bar{E}}$) = $5 / 50 = 0.10$

Relative Risk = $RR = 0.24 / 0.10 = 2.4$

**Study Designs**

Experimental studies

Observational studies

- Analytic studies
- Descriptive studies

Randomized Controlled trials

Case control

Cohort

Cross sectional
Case Control Study

Start with OUTCOME Go backwards Check for EXPOSURE

Case Control Study

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th>Cases</th>
<th>Controls</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 (OR) = \( \frac{ad}{bc} \)

Case Control Study: Example

<table>
<thead>
<tr>
<th></th>
<th>Lung Cancer</th>
<th>No Lung Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smokers</td>
<td>400</td>
<td>200</td>
</tr>
<tr>
<td>Non smokers</td>
<td>600</td>
<td>800</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

Odds Ratio (OR) = \( \frac{400 \times 800}{200 \times 600} = 2.7 \)
Study Designs

- Cross sectional
- Case control
- Experimental studies
- Randomized Controlled trials
- Observational studies
- Analytic studies
- Descriptive studies
- Cohort
- Cross sectional

Randomized Clinical Trials

Assign exposure randomly

**POPULATION**
- Population
  - People without the outcome
  - Sample of people exposed
  - Sample of people not exposed

Assign exposure randomly

ASSIGN EXPOSURE ➔ Follow up ➔ Check for **OUTCOME**

Randomized Clinical Trials

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>c</td>
</tr>
</tbody>
</table>

Risk of outcome in exposed \( R_1 \) = \( \frac{a}{a+b} \)

Risk of outcome in non-exposed \( R_2 \) = \( \frac{c}{c+d} \)

Relative risk (RR) = \( \frac{R_1}{R_2} = \frac{a}{c} \)
Randomized clinical trials: example

<table>
<thead>
<tr>
<th></th>
<th>Smoking Abstinence</th>
<th>No smoking Abstinence</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotine patch</td>
<td>77</td>
<td>68</td>
<td>145</td>
</tr>
<tr>
<td>Placebo patch</td>
<td>24</td>
<td>120</td>
<td>144</td>
</tr>
</tbody>
</table>

Relative Risk (RR) = \( \frac{77}{145} = 3.1 \)

Definition of some basic terms

- **Population**: The largest collection of entities for which we have interest at a particular time.
- **Sample**: A part of a population.
- **Variable**: A characteristic of the objects under observation that takes on different values for different cases, example: age, gender, diastolic blood pressure.

**Definition of some basic terms**

- **Simple random sample**: is when a sample \( n \) is drawn from a population \( N \) in such a way that every possible sample of size \( n \) has the same chance of being selected.
- **Parameter**: descriptive measure computed from data of a population.
- **Statistic**: descriptive measure computed from data of a sample.
Scales of measurements

Categorical variable

- Nominal
- Ordinal
- Dichotomous
- Polytomous

Continuous variable

Categorical variable

I. Nominal scale: Unordered data
   a. Dichotomous: Two categories. Example:
      - Death
   b. Polytomous: More than 2 categories.
      Example:
      - Blood group
      - Race

Categorical variable

II. Ordinal scale: Predetermined order among response classification. (need not assume equal distances between categories).
   Example:
   - Education
   - Social Class
III. Continuous scale: Not restricted to integers, equal distances between categories.
Example:

- Blood pressure
- Age
- Weight
- Cholesterol level

Continuous variable