Measurement in Epidemiology: Frequency, Association, and Impact

Mayfong Mayxay M.D., Ph.D. (Trop Med)

GFMER - WHO - UNFPA - LAO PDR

Training Course in Reproductive Health Research

Vientiane, 12 October 2009
Types of Epidemiologic Measures

1. Measures of disease frequency
2. Measures of association
3. Measures of potential impact
Rationale

1. Careful & accurate measurement of disease occurrence (morbidity & mortality): constitutes fundamental basis of studies

2. Studies are designed to:
   - describe & compare disease trends;
   - identify disease determinants;
   - evaluate public health interventions aimed at controlling health problems
Measures of Disease Frequency

- Measures of disease frequency in **mathematical quantity**
  - Count
  - Proportion (percentage)
  - Rate
  - Ratio
- Measures of disease frequency in **epidemiology**
  - Prevalence
  - Incidence
Importance of Denominator

\[
\frac{a}{b} \quad \text{Numerator}
\]
\[
\text{Denominator}
\]

Example 1:

- 500 cases of dengue fever in Vientiane
- 120 cases of dengue fever in Phongsaly

Which one is more infected?

- Vientiane: \( \frac{200}{800,000} = 0.25/1,000 \)
- Phongsaly: \( \frac{120}{300,000} = 0.4/1,000 \)
Example 2: Dengue outbreak in Savannakhet

Is the situation worse?

Answer: depends on population size, difference in reporting method (more sensible), definition of case !!!
Count, Proportion, Rate, Ratio

• **Counts:**
  - Simplest & most basic measure - absolute number of persons who have disease or characteristic of interest.
  - Useful for health planners & administrators: for allocation of resources (e.g. quantity of ORS needed by diarrheal cases)
  - Count of No. cases of a disease, is used for surveillance of infectious disease for early detection of outbreaks.
Limited values of counts

- Number of persons with characteristic, e.g., cases of dengue hemorrhagic fever, depends on the size of the population at risk of the disease in an area. The bigger this group, the higher is the expected number of cases.

- The duration of observation also affects the frequency of cases; the longer the observation period, the more cases can occur.

→ Count does not contain these elements!
Proportion (percentage, frequency)

- Proportion: + a included in the denominator
  \[
  \frac{a}{a + b}
  \]
  + No measurement unit; > 0 to \leq 1
  + Often expressed as %

Example: From 7,999 females aged 16 – 45 y, 2,496 use modern contraceptive methods. The proportion of those who use modern contraceptive methods = \(\frac{2,496}{7,999} \times 100 = 31.2\%\)
**Rate**

**Definition:** Frequency of events, that occur in a defined time period, divided by the average population of risk.

\[
\text{Rate} = \frac{\text{Numerator}}{\text{Denominator}} \times \text{Constant multiplier}
\]

**Crude death rate** = \[
\frac{\text{Number of deaths} \ (\text{defined place and time period})}{\text{Mid-period population} \ (\text{same place and population})} \times 1000
\]
• **Ratio:** A fraction in which the numerator is not part of the denominator. \( \frac{a}{b} \)

• a and b are two mutually exclusive frequency

• Example:
  - Number of hospital beds per 100,000
  - Male and female dengue infection ratio = 70/35 or 2 males to one female (2 : 1)
Mortality, Fatality, Death

<table>
<thead>
<tr>
<th><strong>Mortality rate:</strong> Death of a particular disease/event in the total population (e.g., maternal mortality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pregnancy – related death (defined place and time period)</td>
</tr>
<tr>
<td>Number of live birth (same place and time period)</td>
</tr>
<tr>
<td>$\times 100,000$</td>
</tr>
<tr>
<td>Maternal mortality rate in Laos in 2000 $\sim \frac{530}{100,000}$</td>
</tr>
</tbody>
</table>

- **Fatality rate:** Mortality among cases of a particular disease
- **Death rate:** Mortality of all diseases among the total population
Commonly Used Rates for Population Study

**Age-specific death rate**

\[
\text{Age-specific death rate} = \frac{\text{Number of death in a particular age group (defined place and time period)}}{\text{Mid-period population (same age group, place and time period)}} \times 1,000
\]

**Cause specific death rate**

\[
\text{Cause specific death rate} = \frac{\text{Number of death due to a particular cause (defined place and time period)}}{\text{Mid-period population (same place and time period)}} \times 1,000
\]
<table>
<thead>
<tr>
<th>Infant mortality rate</th>
<th>Number of death to infants &lt;1 year of age (defined place and time period)</th>
<th>Number of live births (same place and time period)</th>
<th>x 1,000</th>
</tr>
</thead>
</table>

**Infant mortality rate in Laos in 2005 ~ 70 / 1,000**

<table>
<thead>
<tr>
<th>Peri-natal mortality rate</th>
<th>Number of deaths to infants &lt;7 days of age (same place and time period)</th>
<th>Number of live births (same place and time period)</th>
<th>x 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stillbirth (defined place and time period) + Number of stillbirth (same place and time period)</td>
<td>+</td>
<td>Number of live births (same place and time period)</td>
<td></td>
</tr>
</tbody>
</table>
Prevalence

- Number of existing cases of disease
- Proportion of individuals in a population with disease or condition at a specific point of time

Prevalence = \frac{\text{No. of cases observed at time } t}{\text{Total No. of individuals at time } t}
Example of Prevalence

- The prevalence of hypertension (systolic BP ≥ 95 mmHg) on May 1-2, 2009 in Lao men aged 30-69 years in Xienglairkhok village was:

\[
\frac{276 \text{ persons with systolic BP ≥ 95 mmHg}}{1,853 \text{ Lao men aged 30-69 years at the time of survey}} \times 100 = 15\%
\]
Prevalence divided into two types:

* Point prevalence
- 01/01/2009: case No. 2, 4, 5
- 31/12/2009: case No. 6, 7, 10

* Period prevalence between 01/01-31/12/2009:
Case No. 2, 3, 4, 5, 6, 7, 9, 10
Incidence

• Measures of **new cases** of disease that develop in a population during a specified period of time

• 2 types of incidence
  1. Cumulative incidence (incidence)
  2. Incidence rate = incidence density
Cumulative Incidence (CI) = Incidence

No. of individuals who get the disease during a certain period

\[ \text{CI} = \frac{\text{No. of individuals who get the disease}}{\text{No. of individuals in the population at the beginning of the period}} \]

- A proportion
- Has no dimension
- Varies between 0 and 1
Example of Cumulative Incidence

• The population statistic of Lab Lair District in 2001 revealed that there were 5,572 women aged 20-39 years who were sex workers. Based on the record of CHAS, among those women, 45 were HIV + ve during 2002-2005.

• What is the cumulative incidence of HIV + ve among those women during a period of 4 years?

• Cumulative incidence = 45 / 5,572 = 0.008 or 0.8%
Incidence Rate or Density (ID)

No. of new cases that occur during a calendar period

\[ ID = \frac{\text{No. of new cases}}{\text{Amount of pop-time contributed by the observed candidate pop. During that period (time at risk)}} \]

• Not a proportion
• Has dimension (unit of ID is time)
• Varies between 0 to infinity
Example of Incidence Density

• In 2003, the number of new cases of STI was 29 among the men aged 40-44 years in Lab Leu District. The person-years was 41,532 among that group of people.

• What is the incidence density or incidence rate of STI + ve among those people?

• Incidence density = 29 / 41,532 person-years = 0.0007/year
Prevalence vs Incidence

Incidence

Prevalence

Death or cure
Measures of Association

- Statistical relationship between two or more events, characteristics, or other variables
- Statistical relationship between exposure and disease
- Association is not causation!
• **Absolute**
  – Risk difference
    exposed - unexposed

• **Relative**
  – Risk ratios
  – Odds ratios
    exposed / unexposed
• The relative risk of myocardial infarction in men compared with women is: 5

\[
\text{Risk ratio} = \frac{\text{Risk}_{\text{men}}}{\text{Risk}_{\text{women}}} = \frac{5 \text{ cases/1000 PY}}{1 \text{ case/1000 PY}} = 5
\]

• The absolute risk difference between men and women is: 4 cases/1000 PY

\[
5 \text{ cases/1000 PY} - 1 \text{ case/1000 PY} = 4 \text{ cases/1000 PY}
\]
Epidemiologic Measures of Association

- Relative risk
- Odds ratio
- Attributable risk/population attributable risk percent
- Standardized mortality ratios
Cause - Effect Relationship

Cause/risk/exposure $\rightarrow$ Effect/disease/outcome

Population $\rightarrow$ People without disease $\rightarrow$ Exposed

Exposed $\rightarrow$ Disease $\leftrightarrow$ No disease

Not exposed $\rightarrow$ Disease $\leftrightarrow$ No disease
2 x 2 Tables in Epidemiology

Used to summarize frequencies of disease and exposure and used for calculation of association

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>Total</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>c</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>d</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>a + c</td>
</tr>
</tbody>
</table>
## 2 x 2 Tables in Epidemiology

Used to summarize frequencies of disease and exposure and used for calculation of association

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th></th>
<th></th>
<th>Total Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (exposed)</td>
<td>Yes</td>
<td>$a$</td>
<td>$b$</td>
<td>total # exposed</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
<td>total # unexposed</td>
</tr>
<tr>
<td>No (unexposed)</td>
<td>Yes</td>
<td>$c$</td>
<td>$d$</td>
<td>total # with disease</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
<td>total # with no disease</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>Total Population</td>
</tr>
</tbody>
</table>
Relative Risk

- The ratio of the risk of disease in persons exposed compared to the risk in those unexposed
- Often, a measure of association between incidence of disease and exposure of interest

\[
RR = \frac{\text{Incidence rate of disease in exposed}}{\text{Incidence rate of disease in unexposed}}
\]
### Relative Risk

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

Relative Risk = \[
\frac{a}{a + b} \div \frac{c}{c + d}
\]
### Example: Relative Risk

<table>
<thead>
<tr>
<th></th>
<th>Develop CHD</th>
<th>Do Not Develop CHD</th>
<th>Totals</th>
<th>Incidence per 1000/yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smokers</td>
<td>84</td>
<td>2916</td>
<td>3000</td>
<td>28.0</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>87</td>
<td>4913</td>
<td>5000</td>
<td>17.4</td>
</tr>
</tbody>
</table>

Incidence in smokers = $\frac{84}{3000} = 28.0$

Incidence in non-smokers = $\frac{87}{5000} = 17.4$

Relative risk = $\frac{28.0}{17.4} = 1.61$
Interpretation of Relative Risk

• $1 = \text{No association between exposure \& disease}$
  – incidence rates are identical between groups

• $> 1 = \text{Positive association}$
  – exposed group has higher incidence than non-exposed group

• $< 1 = \text{Negative association or protective effect}$
  – non-exposed group has higher incidence
  – example: $0.5 = \text{half as likely to experience disease}$
Interpretation of Relative Risk

- A relative risk of 1.0 or greater indicates an increased risk
- A relative risk less than 1.0 indicates a decreased risk
Odds Ratio

• The ratio of the odds of a condition in the exposed compared with the odds of the condition in the unexposed

• Usually applied to prevalence studies rather than incidence studies

\[
\text{OR} = \frac{\text{odds of disease in exposed}}{\text{odds of disease in unexposed}}
\]
## Odds Ratio

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>$a$</td>
<td>$b$</td>
<td>$a + b$</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>$c$</td>
<td>$d$</td>
<td>$c + d$</td>
</tr>
<tr>
<td>Total</td>
<td>$a + c$</td>
<td>$b + c$</td>
<td></td>
<td>$a + b + c + d$</td>
</tr>
</tbody>
</table>

**Odds Ratio**

$$\frac{a}{b} = \frac{ad}{bc}$$
Based on the Odds Ratio formula, what is the Odds Ratio for each disease status in this famous smoking study?

### Smoking and Carcinoma of the Lung

<table>
<thead>
<tr>
<th>Disease Status</th>
<th>Number of smokers</th>
<th>Number of non-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males Lung cancer</td>
<td>647</td>
<td>2</td>
</tr>
<tr>
<td>Males Controls</td>
<td>622</td>
<td>27</td>
</tr>
<tr>
<td>Females Lung cancer</td>
<td>41</td>
<td>19</td>
</tr>
<tr>
<td>Females Controls</td>
<td>28</td>
<td>32</td>
</tr>
</tbody>
</table>

Measures of Potential Impact

Reflect the expected contribution of a study factor to the frequency of a disease in a particular population. These measures are useful for predicting the efficacy or effectiveness of therapeutic maneuvers and intervention strategies within a specific population, e.g., vaccine.

Essentially, potential impact measures are a combination of frequency and association measures.