University of Pittsburgh

آغا خان یونیورسٹی
THE AGA KHAN UNIVERSITY

South Asian Cardiovascular Research Methodology Workshop

Basic Epidemiology

Measures of Association

Thomas Songer, PhD

Epidemiologic Reasoning

1. **Suspicion that a factor (exposure) may influence occurrence of disease**

-
-
  - **Do subpopulations have higher or lower rates?**
  - **Are disease rates increased in the presence of certain factors?**
-
-

Epidemiologic Reasoning

2. Formulation of specific hypotheses
   - Based on suspicions concerning influence of a particular factor on disease occurrence

3. Conduct analytic studies
   - Hypotheses are tested to determine if statistical associations between factors (exposures) and disease exist
   - Study population is assembled from individuals with disease or outcome of interest and an appropriate comparison group

Epidemiologic Reasoning

4. Assess validity of association

   -
   - Is the association valid?
   - Are there alternative explanations for the association?
     - Chance
     - Bias
     - Confounding

Epidemiologic Reasoning

5. Make a judgement of whether a cause-effect relation between factor (exposure) and disease exists

- What is the magnitude of the association?
- Are the findings consistent with previous studies (or conflicting)?
- Are the findings biologically credible?
- Can underlying biological mechanisms that support the association be identified?

Epidemiologic Measures

- Measures of disease frequency - measures disease risk or burden in a population

–

–

Epidemiologic Measures

- **Measures of association**
  - **frequency relative to other factors**
  - **Indications of how more or less likely one is to develop disease as compared to another**

Epidemiologic Measures of Association

- **Absolute**
  - **Risk difference**

    exposed - unexposed
- **Relative**
  - **Risk ratios**
  - **Odds ratios**

    exposed / unexposed

Epidemiologic Measures of Association

- **in men compared with women is : 5**

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

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

5 cases/1000 PY - 1 case/1000 PY = 4 cases/1000 PY

Epidemiologic Association

- **Statistical relationship between two or more events, characteristics, or other variables**

⇩

- **Statistical relationship between exposure and disease**

- **Association is not causation!**

Risk Factor

- A factor (exposure) found to be associated with a health condition
- an attribute or exposure that increases the probability of occurrence of disease
  - behaviour
  - genetic
  - environmental
  - social

-- time
-- person
-- place

Epidemiologic Measures of Association

-
-
-

attributable risk percent

- Standardized mortality ratios

2 x 2 Tables in Epidemiology

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

	Disease		
Exposure	Yes	No	Total
Yes	*a*	*b*	*a* + *b*
No	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *c*	*a* + *b* + *c* + *d*

2 x 2 Tables: Contents of Cells

a = number of individuals who are exposed and have the disease

b = number who are exposed and do not have the disease

c = number who are not exposed and have the disease

d = number who are both non-exposed and non-diseased

2 x 2 Tables in Epidemiology

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

	Disease		
Exposure	Yes	No	Total
Yes (exposed)	*a*	*b*	*total # exposed*
No (unexposed)	*c*	*d*	*total # unexposed*
Total	*total # with disease*	*total # with no disease*	*Total Population*

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**

$$RR = \frac{\text{Incidence rate of disease in exposed}}{\text{Incidence rate of disease in unexposed}}$$

Exposure	Disease: Yes	Disease: No	Total
Yes	*a*	*b*	*a* + *b*
No	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *c*	*a* + *b* + *c* + *d*

$$\text{Relative Risk} = \frac{\mathbf{a / (a + b)}}{\mathbf{c / (c + d)}}$$

Relative Risk

Incidence in smokers = 84/3000 = 28.0
Incidence in non-smokers = 87/5000 = 17.4
Relative risk = 28.0/17.4 = 1.61

Interpretation of Relative Risk

- **1 = No association between exposure and disease**
  - **incidence rates are identical between groups**
- **> 1 = Positive association**
  - **exposed group has higher incidence than non-exposed group**
- **< 1 = Negative association or protective effect**
  - **non-exposed group has higher incidence**
  - **example: .5 = half as likely to experience disease**

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

At times, epidemiologists will choose to express disease frequency in terms of odds

What are odds?

Measures of Disease Association

The chance of something happening can be expressed as a risk and/or as an odds:

Risk = the chances of something happening / the chances of *all* things happening

Odds = the chances of something happening / the chances of it *not* happening

Example: If I choose a student randomly from this class, how likely is it that I will choose you?

Risk (probability) = 1/9 = .111

Odds = 1/8 = .125

Measures of Disease Association

Example: Among 100 people at baseline, 20 develop influenza over a year.

The risk is 1 in 5 (i.e. 20 among 100) = .2 The odds is 1 to 4 (i.e. 20 compared to 80) = .25

Odds

- What are odds?
- Let p = the probability of an event
- 1-p = the probability that the event does not occur
- Odds of the event = p/1-p
  - If the probability of an event is 0.7, the the odds of winning are 0.7/0.3 = 2.33

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

$$OR = \frac{\text{odds of disease in exposed}}{\text{odds of disease in unexposed}}$$

Exposure	Disease: Yes	Disease: No	Total
Yes	*a*	*b*	*a* + *b*
No	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *c*	*a* + *b* + *c* + *d*

$$\text{Odds Ratio} = \frac{\mathbf{[a / (a + b)] / [1 - (a/(a+b))]}}{\mathbf{[c / (c + d)] / [1 - (c/(c+d))]}}$$

Odds Ratio

Exposure	Disease: Yes	Disease: No	Total
Yes	*a*	*b*	*a* + *b*
No	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *c*	*a* + *b* + *c* + *d*

$$\mathbf{Odds\ Ratio} = \frac{\mathbf{[\ a / b\ ]}}{\mathbf{[\ c / d\ ]}} = \frac{\mathbf{[\ ad\ ]}}{\mathbf{[\ bc\ ]}}$$

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

Disease Status	# of smokers	# of nonsmokers	
Males Lung cancer	647	2	
Males Controls	622	27	
Females Lung cancer	41	19	
Females Controls	28	32	

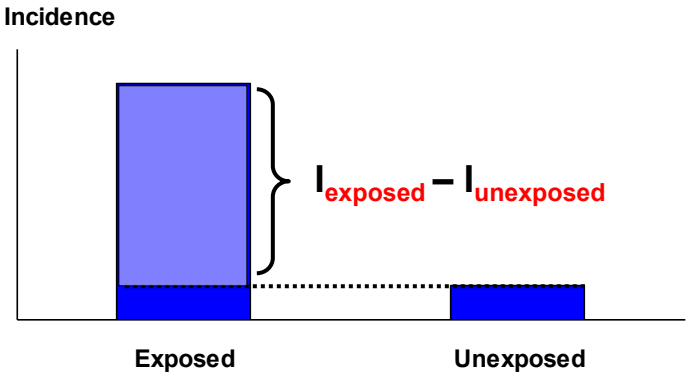
Doll R. Bradford, Hill A. Smoking and carcinoma of the lung: preliminary report. British Medical Journal 1950, 2: 739-748.

Difference Measures

- **Attributable risk**
  - # of cases among the exposed that could be eliminated if the exposure were removed
  - = Incidence in exposed - Incidence in unexposed
- **Population attributable risk percent**
  - Proportion of disease in the study population that could be eliminated if exposure were removed

$$= \frac{\text{Incidence in total population - Incidence in unexposed}}{\text{incidence in total population}}$$

Attributable Risk



I = Incidence

Attributable Risk

- Rate of disease in the population that can be directly attributed to the exposure
- equals incidence rate in exposed minus incidence rate in the unexposed

$$= \quad A / (A + B) \quad - \quad C / (C + D)$$

AR: Fast driving and Automobile Deaths

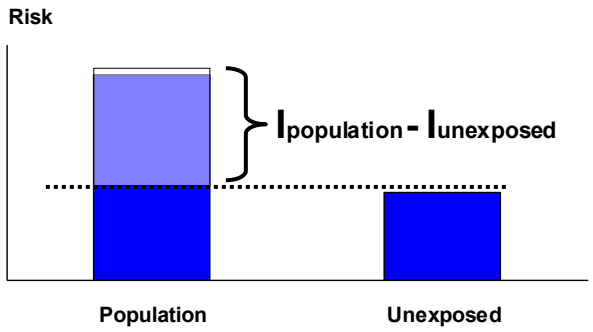
	Dead	Not dead		Risk	RD
Fast	100	1900	2000	0.05	
					0.04
Slow	80	7920	8000	0.01	
	180	9820	10000		

Population Attributable Risk (PAR)

- Excess risk of disease in total population attributable to exposure
- Reduction in risk which would be achieved if population entirely unexposed
- Helps determining which exposures relevant to public health in community

$$PAR = I_{population} - I_{unexposed}$$

Population Attributable Risk



Population Attributable Risk Percent (PAR%)

- PAR expressed as a percentage of total risk in population

$$PAR\% = \frac{I_{population} - I_{unexposed}}{I_{population}} \times 100$$

PAR: Fast driving

	Dead	Not dead		Risk
Fast	100	1900	2000	0.050
Slow	80	7920	8000	0.010
	180	9820	10000	0.018

$$PAR = 0.018 - 0.010 = 0.008$$

$$PAR\% = \frac{0.018 - 0.010}{0.018} \times 100 = 44\%$$

Conclude

- **44% of driving-related deaths in population were presumably due to fast driving**