MEASURES OF ASSOCIATION

Cohort studies

Case control studies
ASSOCIATION BETWEEN TWO VARIABLES

The following limited definition will be used in this course. Two or more events, characteristics, or other attributes are said to be associated if the probability of occurrence of one of these depends on the occurrence of one or more of the others.

For example, an exposure, $A$, and a disease, $B$ are said to be associated if

$$ P\{B|A\} \neq P\{B|\overline{A}\} $$

or, in terms of incidences:

$$ r\{B|A\} \neq r\{B|\overline{A}\} $$
CHARACTERISTICS OF ASSOCIATIONS RELEVANT TO THIS COURSE

• Statistical Significance of Association

◊ Statistical Tests of Null Hypothesis that there is No Association between Exposure and Disease

\[ H_0: P(B|A) = P(B|\overline{A}) \]

against alternative hypothesis

\[ H_a: P(B|A) \neq P(B|\overline{A}) \]

Statistical tests do not tell us anything about the magnitude of the association!

• Strength Of The Association

◊ Risk Ratio (Relative Risk)

◊ Odds Ratio (Relative Odds)

• Impact of the Association

◊ Attributable Risk Percent
MEASURES OF ASSOCIATION IN A COHORT STUDY

Relative Risk

Odds ratio

Attributable risk percent
Relative Risk—also referred to as Risk Ratio:

Ratio of two incidence rates in a cohort study

In a two by two table:

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease</th>
<th>Non-disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Absent</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

\[
RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}
\]

RR > 1, Risk greater in exposed than unexposed

RR < 1 “Protective effect” Exposure reduces risk of disease

We want to test \( H_0: RR=1 \)
EXAMPLE USING PERSON-TIME INCIDENCE RATES

Format for Data:

\[ E = \text{Exposed} \]
\[ \bar{E} = \text{Not Exposed} \]
\[ N = \text{Number of Cases of Disease} \]
\[ D = \text{Number of Person - Time Units} \]

In an AJE paper (Njolstad et al, AJE 1998; 147:49-58), 6098 men and 5,556 women were followed for an average of 12 years. The following shows their respective incidence rates of Diabetes Mellitus:

\[
\begin{align*}
E = \text{Men} & \quad E = \text{Women} & \quad \text{Total} \\
N & \quad N_E = 87 & \quad N_E = 75 & \quad N_T = 162 \\
D & \quad D_E = 71,537 & \quad D_E = 66,694 & \quad D_T = 138,231 \\
\frac{N}{D} \times 10^{-3} & \quad 1.22 & \quad 1.12 & \quad 1.17
\end{align*}
\]

Rate in “exposed” (males) = \[
\frac{\frac{N_E}{D_E}}{1000} = \frac{87}{71,537} \times 1000 = 1.22 \text{ per thousand}
\]

Rate in “unexposed” (females) = \[
\frac{\frac{N_E}{D_E}}{1000} = \frac{75}{66,694} \times 1000 = 1.12 \text{ per thousand}
\]
Relative Risk=$\frac{1.22}{1.12}=1.08$

Males are 1.08 times more likely than females to contract diabetes over the 12 year period in this study.

Is this RR significantly different than 1.00?
LARGE SAMPLE CHI-SQUARED TEST FOR TESTING EQUALITY OF PERSON-TIME INCIDENCE RATES
(See Rothman, Modern Epidemiology, Formula 11-1)
(Also used in PEPI)

\[ \chi^2 = \left( \frac{N_e - N_t \left( \frac{D_e}{D_t} \right)^2}{N_t \left( \frac{D_e}{D_t} \right) \left( \frac{D_t}{D_t} \right)} \right) \]

\( D_e = \) person yrs in exposed (or among those with factor)
\( D_e = \) person yrs in unexposed (or among those without factor)
\( D_t = \) total person yrs
\( N_e = \) Cases in exposed (or among those with factor)
\( N_t = \) Total number of cases
\[
\frac{162 \times \left( \frac{71,537}{138,231} \times \frac{66,694}{138,231} \right)}{\left( 87 - 162 \left[ \frac{71,537}{138,231} \right] \right)^2} = 0.247
\]

\[
Z = \sqrt{0.247} = 0.497
\]

p = 0.62
Confidence Intervals for Relative Risk (person years)

Asymmetric Sampling Distribution of Relative Risk
Requires taking Confidence Limits of Log of Relative Risk (Breslow)

\[ \frac{N_e}{D_e} \]

Relative Risk: \[ \frac{D_e}{N_e} \]
\[ \frac{N_{\bar{e}}}{D_{\bar{e}}} \]

Confidence Interval for RR:

\[ e^{\ln(\text{RR}) \pm \text{SE}\{\ln(\text{RR})\}} \]

where SE can be approximated by:

\[ \text{SE}\{\ln(\text{RR})\} = \sqrt{\frac{1}{N_e} + \frac{1}{N_{\bar{e}}}} \]

\[ N_e = \text{Cases in exposed} \]
\[ N_{\bar{e}} = \text{Cases in unexposed} \]
For this example:

\[ SE\{\ln(\text{RR})\} = \sqrt{\frac{1}{87} + \frac{1}{75}} = .158 \]

\[ \ln(\text{rr}) = \ln(1.08) = .077 \]

95% CI: \[ e^{.077 \pm 1.96 \times .158} \]

Lower limit = .79
Upper limit = 1.47
RATES2 - Comparison of Person-Time Incidence Rates
Sunday, 1st September 2002.

<table>
<thead>
<tr>
<th>Stratum 1:</th>
<th>Numerator</th>
<th>Denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A:</td>
<td>87</td>
<td>71537</td>
</tr>
<tr>
<td>Category B:</td>
<td>75</td>
<td>66694</td>
</tr>
</tbody>
</table>

**STRATUM 1**  
Rate A = 1.216 per 1000  
Rate B = 1.125 per 1000

Large-sample test:  
Z = 0.497  
Two-tailed P = 0.619

**RATE RATIO** = 1.081  
S.E. of log rate ratio = 0.158

<table>
<thead>
<tr>
<th>C.I.</th>
<th>Cornfield</th>
<th>Breslow</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.84 to 1.40</td>
<td>0.83 to 1.40</td>
</tr>
<tr>
<td>95%</td>
<td>0.80 to 1.47</td>
<td>0.79 to 1.47</td>
</tr>
<tr>
<td>99%</td>
<td>0.72 to 1.62</td>
<td>0.72 to 1.62</td>
</tr>
</tbody>
</table>

**RATE DIFFERENCE** = 0.092 per 1000  
S.E. = 0.184 per 1000

| 90% confidence interval | -0.21 to 0.39 per 1000 |
| 95% confidence interval | -0.27 to 0.45 per 1000 |
| 99% confidence interval | -0.38 to 0.57 per 1000 |

10915.3 person-time units are needed in group B to avoid 1 event.

95% CI = 2211.0 to infinity [down to "number needed to harm": 3716.8]
Example 2

COMPARING TWO RATES FROM COHORT STUDIES (Non-Stratified Analysis, Person Years Method)

(Data from Doll-Hill, 1966 - See PEPI Manual for Complete Reference)

<table>
<thead>
<tr>
<th>CHD Deaths</th>
<th>Smokers</th>
<th>Non-Smokers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_e$=630</td>
<td>$N_e=101$</td>
<td>$N_t=731$</td>
<td></td>
</tr>
<tr>
<td>$D_e=142,247$</td>
<td>$D_e=39,220$</td>
<td>$D_t=181,467$</td>
<td></td>
</tr>
</tbody>
</table>

Incidence:

Smokers: $(630/142,247) \times 100,000 = 442.89$

Non-Smokers $(101/39,220) \times 100,000 = 257.52$

Risk Ratio: $RR = \frac{442.89}{257.52} = 1.720$
\[
\left( \frac{630 - 731 \left[ \frac{142,247}{181,467} \right]^2}{731 \times \left\{ \frac{142,247}{181,467} \times \frac{39,220}{181,467} \right\} } \right) = 26.23
\]

\[ z = \sqrt{26.23} = 5.12 \]

\[ p < .000 \]
Confidence Intervals for Relative Risk

\[
\frac{N_e}{D_e} \quad \frac{N_r}{D_r}
\]

Relative Risk:

\[
\text{Confidence Interval for RR:}
\]

\[e^{\ln(RR) \pm SE\{\ln(RR)\}}\]

where

\[SE\{\ln(RR)\} = \sqrt{\frac{1}{N_e} + \frac{1}{N_r}}\]

For this example:

\[\ln(RR) = 0.5422\]

\[Z = 1.96\]

95% CI = \[e^{0.5422 \pm 1.96 \times (0.1072)}\]

Upper 95% CI for RR = 2.1219
Lower 95% CI for RR = 1.3939
### RATES2 - Comparison of Person-Time Incidence Rates

**Monday, 2nd September 2002.**

<table>
<thead>
<tr>
<th>Stratum 1:</th>
<th>Numerator</th>
<th>Denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A:</td>
<td>630</td>
<td>142247</td>
</tr>
<tr>
<td>Category B:</td>
<td>101</td>
<td>39220</td>
</tr>
</tbody>
</table>

**STRATUM 1**

- Rate A = 442.892 per 100,000
- Rate B = 257.522 per 100,000
- Large-sample test: $Z = 5.121$, Two-tailed $P = 0.000$ [3.04E-07]
- Rate Ratio = 1.720
  - S.E. of log rate ratio = 0.107
  - Cornfield 90% C.I.: 1.44 to 2.05
  - Breslow 90% C.I.: 1.44 to 2.05
  - 539.5 person-time units are needed in group B to avoid 1 event.
  - 95% CI = 405.9 to 803.9

**RATE DIFFERENCE** = 185.370 per 100,000
- 90% confidence interval = 134.20 to 236.54 per 100,000
- 95% confidence interval = 124.39 to 246.35 per 100,000
- 99% confidence interval = 105.23 to 265.51 per 100,000
METHODS FOR INCIDENCE RATE RATIOS (PROPORTIONS)

To test $H_0: RR=1$

Cohort Study of 1 year incidence of Acute Myocardial Infarction, for those with Severe Systolic Hypertension vs Normal Systolic Blood Pressure

<table>
<thead>
<tr>
<th>MI</th>
<th>No MI</th>
<th>Total followed</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>a + c</td>
<td>b + d</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypertension</th>
<th>MI</th>
<th>No MI</th>
<th>Total followed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe</td>
<td>180</td>
<td>9820</td>
<td>10,000</td>
</tr>
<tr>
<td>Normal</td>
<td>30</td>
<td>9970</td>
<td>10,000</td>
</tr>
</tbody>
</table>

$RR$ (Risk ratio or relative risk): $\frac{180}{30} = 6.0$ 

One year risk of MI 6 times higher in those with severe hypertension compared to those with normal blood pressure
To Test Null Hypothesis use Chi Square Test

$$\sum \frac{(O - E)^2}{E}$$

or, for 2x2 tables can use:

$$\frac{(|ad - bc|)^2 \times N}{N_1 N_2 M_1 M_2}$$

or Continuity Correction

$$\frac{\left(\frac{|ad - bc| - N}{2}\right)^2 \times N}{N_1 N_2 M_1 M_2}$$

(or Fisher exact for expected frequencies less than 5)

Results (from PEPI Rates1)

Chi square=108.28, p<.000

95% Confidence Intervals

CI calculated in Log scale (RR asymmetrically distributed)

$$SE(\text{log RR}) = \sqrt{\frac{b}{a(a + b)} + \frac{d}{c(c + d)}}$$

$$95\% \text{ CI (RR)} = e^{\{\text{log } rr \pm [1.96 \times SE(\text{log } rr)]\}}$$
\[ \ln(6.0) = 1.79 \]

\[ \text{SE(log RR)} = \sqrt{\frac{9820}{180(180 + 9820)} + \frac{9970}{30(30 + 9970)}} = 0.197 \]

95% CI: \( e^{[1.79 \pm (1.96 \times 0.197)]} = 4.08, 8.82 \)

**Short cut method for Confidence Interval**

Lower limit (95%) = \( RR \times e^{-[1.96 \times \text{SE(log rr)}]} \)

Upper limit (95%) = \( RR \times e^{[1.96 \times \text{SE(log rr)}]} \)

Lower: 6.0 \( \times \exp\ (-1.96 \times 0.197) = 4.08 \)

Upper: 6.0 \( \times \exp\ (1.96 \times 0.197) = 8.83 \)
RATES1 – Comparison of Rates or Proportions
Sunday, 1st September 2002.

DATA
Stratum 1:                            "X"                       "Not-X"
  Category A:                      180                          9820
  Category B:                       30                          9970

STRATUM 1    Rate A = 1.800 per 100    Rate B = 0.300 per 100
Chi-square                    = 108.280   P = 0.000  [ 2.33E -25 ]
Chi-square (cont. correction) = 106.841   P = 0.000  [ 4.82E -25 ]
Rate ratio = 6.000        S.E. of log rate ratio = 0.197
  90% confidence interval = 4.34 to 8.29
  95% confidence interval = 4.08 to 8.82
  99% confidence interval = 3.61 to 9.96
  [Low-bias estimator of rate ratio in population = 5.807]
Rate difference  = 1.500 per 100    S.E. = 0.144 per 100
  90% confidence interval = 1.254 to 1.746 per 100
  95% confidence interval = 1.208 to 1.792 per 100
  99% confidence interval = 1.120 to 1.880 per 100
66.7 subjects are needed in group B to avoid 1 case.
  95% CI = 55.8 to 82.8
Odds ratio = 6.09
CROSS SECTIONAL STUDIES

Measures of association in cross sectional studies

Prevalence Rate Ratio

\[
\frac{\text{Prevalence Rate (exposed)}}{\text{Prevalence Rate (unexposed)}}
\]

If based on survey of individuals, can use statistical procedures used for cumulative incidence rates

Two bias factors, ratio of disease durations, and ratio of point prevalence estimates complements determine the relation between PRR and the Relative Risk

\[
\text{PRR} = \text{RR} \times \frac{\frac{\text{Dur}_+}{\text{Dur}_-} \times \frac{1 - \text{Pre}v_+}{1 - \text{Pre}v_-}}
\]

\[
= \text{RR} \times \frac{\frac{\text{Dur}_+}{\text{Dur}_-} \times \frac{1 - \text{Pre}v_+}{1 - \text{Pre}v_-}}
\]
ODDS AND ODDS RATIOS

*Odds* = The probability of an event occurring divided by the probability of the event not occurring.

Example: If the incidence of a disease during a particular time period is 0.010 per person, then the odds of a person getting the disease = 0.010/0.990 = 0.0101

*Odds Ratio*: The ratio of two odds

In epidemiology, there are two types of odds ratios that are generally used:

*Disease Odds Ratio*: The odds of getting a disease, B, if an exposure, A, is present divided by the odds of getting disease B if the exposure A is not present.

*Exposure Odds Ratio*: The odds of being exposed to a factor, A, if a disease, B, is present divided by the odds of being exposed to factor A if disease B is not present.
DISEASE ODDS IN TERMS OF INCIDENCES

Odds of getting a disease:

\[ o(B) = \frac{r(B)}{1 - r(B)} \]

Incidence rate / 1 - Incidence rate

Odds of getting a disease if factor A is present:

\[ o(B|A) = \frac{r(B|A)}{1 - r(B|A)} \]

Incidence rate when factor a present/ 1 – incidence rate when factor A is not present

Odds of getting a disease if factor A is absent:

\[ o(B|\bar{A}) = \frac{r(B|\bar{A})}{1 - r(B|\bar{A})} \]

Disease Odds Ratio (ratio of two odds):

\[
\frac{\frac{r(B|A)}{1 - r(B|A)}}{\frac{r(B|\bar{A})}{1 - r(B|\bar{A})}}
\]
DISEASE ODDS RATIO CALCULATED
FROM A FOURFOLD TABLE

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>$\bar{B}$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>$\bar{A}$</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+d</td>
</tr>
</tbody>
</table>

$$o\{B|A\} = \frac{r\{B|A\}}{1-r\{B|A\}} = \frac{a}{1 - \frac{a}{a+b}} = \frac{a}{b}$$

$$o\{B|\bar{A}\} = \frac{r\{B|\bar{A}\}}{1-r\{B|\bar{A}\}} = \frac{c}{1 - \frac{c}{c+d}} = \frac{c}{d}$$

$$OR = \frac{o\{B|A\}}{o\{B|\bar{A}\}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc} = \text{Cross Product Ratio}$$
EXPOSURE ODDS RATIO

• Appropriate for case-control studies

• Sample of Persons with Disease B is Selected

• Sample of Persons not having Disease B is Selected

• Presence of Attribute A is Measured in Both Groups
<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease present</th>
<th>Disease absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>(\bar{A})</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{Exposure Odds Ratio} &= \frac{\text{Disease present}}{\text{Disease absent}} \\
&= \frac{\frac{a}{a+c}}{\frac{b}{b+d}} \\
&= \frac{a \cdot (b+d)}{b \cdot (a+c)} \\
&= \frac{ad}{bc}
\end{align*}
\]

Disease Odds Ratio = Exposure Odds Ratio
ODDS RATIO

Can be used as a measure in itself

Will estimate Relative Risk when Risk of disease is small

Incidence rate in exposed:
\[
\frac{a}{a+b} \approx \frac{a}{b} \quad \text{when } a \text{ is small relative to } b
\]

Incidence rate in unexposed:
\[
\frac{c}{c+d} \approx \frac{c}{d} \quad \text{when } c \text{ is small relative to } b
\]

So, Ratio of Incidence Rates
\[
\frac{a}{a+b} \approx \frac{a}{b} \approx \frac{ad}{bc}
\]
### EXAMPLE

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>Non disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>10</td>
<td>1000</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>5</td>
<td>1000</td>
</tr>
</tbody>
</table>

\[
a=10 \ b=1000, \ c=5 \ d=1000
\]

\[
\text{Relative Risk: } \frac{\frac{10}{1010}}{\frac{5}{1005}} = 1.99
\]

\[
\text{Odds Ratio: } \frac{10 \times 1000}{5 \times 1000} = 2.00
\]
Example when disease is not rare

From Szklo

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>Non-disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>650</td>
<td>1920</td>
<td>2570</td>
</tr>
<tr>
<td>Non-Exposed</td>
<td>170</td>
<td>2240</td>
<td>2410</td>
</tr>
</tbody>
</table>

\[
RR = \frac{650}{170} \cdot \frac{2570}{2410} = 3.59
\]

\[
\text{ODDS RATIO} = \frac{650}{1920} \cdot \frac{170}{2240} = 4.46
\]

Odds ratio will exaggerate magnitude of relative risk, but for rare diseases, difference will be small
RELATIONSHIP BETWEEN ODDS RATIO AND RELATIVE RISK

Where \( q_+ = \) Incidence in exposed
\( q_- = \) Incidence in unexposed

\[
\text{OR} = \frac{q_+}{1-q_+} = \frac{q_+}{1-q_+} \times \frac{1-q_-}{q_-} = \frac{q_+}{q_-} \times \left\{ \frac{1-q_-}{1-q_+} \right\}
\]

\[
\frac{q_+}{q_-} = \text{Relative Risk}
\]

\[
\text{Bias term} = \frac{1-q_-}{1-q_+}
\]

If risk is greater in exposed, bias term will be greater, OR will overestimate RR

If exposure is protective \((q_+ < q_-)\) or, \(\text{RR} < 1\), then OR will overestimate protective effect.

However, if incidence in exposed and unexposed is small, bias term will be close to 1.
INTERPRETATION OF ODDS RATIOS, RISK RATIOS, AND P-VALUES

• P-Values are Measures of the Statistical Significance of the difference between the rates. They tell you how likely the observed difference or a more extreme difference is likely to be observed by chance if the null hypothesis is true.

• P-Values tell you nothing at all about the strength of the association.

• Risk Ratios and Odds Ratios are Measures of the strength of the Association.
### ODDS RATIOS, SIGNIFICANCE TESTING, CONFIDENCE INTERVALS, FOR TWO BY TWO TABLES

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>Non Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Non-Exposed</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
</tr>
</tbody>
</table>

**OR:** \( \frac{ad}{bc} \)

**Confidence Interval:** OR asymmetrical, so take CI of ln(OR)

\[
\text{SE } \{ \ln(\text{OR}) \} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}
\]

**95% CI**{ ln(OR)} = \( \ln(\text{OR}) \pm 1.96 \times \text{SE}[\ln(\text{OR})] \)

**95% CI** (OR) = \( \exp \{ \ln(\text{OR}) \pm 1.96 \times \text{SE}[\ln(\text{OR})] \} \)
Hypothesis testing:  H$_0$: OR=1

Usual Chi square test:  For 2x2 table can use

\[
\frac{(|ad - bc|)^2 \times N}{N_1 N_2 M_1 M_2}
\]

or

Continuity Correction

\[
\frac{\left(|ad - bc| - \frac{N}{2}\right)^2 \times N}{N_1 N_2 M_1 M_2}
\]

$N_1 = a + c$,  
$N_2 = b + d$  
$M_1 = a + b$  
$M_2 = c + d$  
$N = a + b + c + d$
EXAMPLE: CASE-CONTROL STUDY OF RELATIONSHIP BETWEEN OROPHARYNGEAL CANCER AND REGULAR USE OF MOUTHWASH
(Wynder et al, 1983)

<table>
<thead>
<tr>
<th>Oropharyngeal Carcinoma</th>
<th>Regular Use of Mouthwash</th>
<th>Prevalence of Regular Use of Mouthwash (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes (Cases)</td>
<td>259</td>
<td>312</td>
</tr>
<tr>
<td>No (Controls)</td>
<td>205</td>
<td>363</td>
</tr>
<tr>
<td>Total</td>
<td>464</td>
<td>675</td>
</tr>
</tbody>
</table>

\[
\text{OR} = \frac{259 \times 363}{312 \times 205} = 1.47 \\
\text{Ln(OR)} = 0.385
\]

\[
\text{SE \{ln(OR)\}} = \sqrt{\frac{1}{259} + \frac{1}{312} + \frac{1}{205} + \frac{1}{363}} = 0.121
\]

\[
95\% \text{ CI(OR)} = \exp\{0.385 \pm 1.96 \times 0.121\}
\]

Lower Limit: 1.15  Upper Limit: 1.88

\[X^2 = 10.13, \ p = 0.001\]
ATTRIBUTABLE RISK in EXPOSED ($AR_{exp}$)

The incidence of a disease in exposed individuals that can be attributed to the exposure.

$$AR_{exp} = r(A \mid E) - r(A \mid \bar{E})$$

where
A= Disease
E= Exposed
$\bar{E}$ = Not Exposed
r = incidence
INDICATORS RELATED TO ATTRIBUTABLE RISK

1. Attributable Fraction or Percent Among Exposed ($AF_e$) in Szklo ($%AR_{exp}$)

\[
%AR_{exp} = \frac{r(A \mid E) - r(A \mid \bar{E})}{r(A \mid E)}
\]

This represents the proportion of disease in the exposed that is attributable to the exposure.

Note: Dividing both numerator and denominator by $r(A \mid \bar{E})$, we obtain:

\[
%AR_{exp} = \frac{RR - 1}{RR}
\]

($%AR_{exp}$ is sometimes referred to as the clinical attributable risk)

\[
\% POPAR = \frac{r(A) - r(A|\bar{E})}{r(A)}
\]

Note that if $p_e$ = the proportion of the population exposed, then the above expression is algebraically equal to: (see Szklo p.102)

\[
\% POPAR = \frac{p_e(RR - 1)}{1 + p_e(RR - 1)}
\]
Attributable Risk in Case Control Studies can be determined using the Odds Ratio instead of the RR, when the disease is rare

\[
\%AR_{\text{exp}} = \frac{OR - 1}{OR}
\]

For population attributable risk, can use percent exposed in controls to represent exposure in general population, assuming controls are representative of general population.

\[
p_e = \text{proportion exposed in controls}
\]

\[
\%POPAR = \frac{p_e (RR - 1)}{1 + p_e (RR - 1)}
\]
If RR is confounded by other factors, need to use the adjusted RR or OR instead of unadjusted.

Attributable risk has public health implications in terms of prevention. If prevalence of exposure is high, number of cases due to factor could be large even if relative risk is small.

Ex. 50% exposed, RR of 1.5

\[
\% \text{POP AR} = \frac{.50 \times (1.5-1.0)}{.50 \times (1.5-1.0) + 1} = .20
\]

With a small risk of 1.5 and a large proportion of exposed, 20% of cases in population due to exposure!

Alternatively if RR is high but prevalence of exposure is small, population attributable risk will be small

Ex. RR=4.0, prevalence of exposure is 1%

\[
\% \text{POP AR} = \frac{.01 \times (4.0-1)}{.01 \times (4.0-1) + 1} = .029
\]

Only 3% of cases in population due to exposure!
EXAMPLE OF ATTRIBUTABLE RISK

In a large nation, 55% of all women are married (or are cohabitant with) smokers. A large cohort study conducted in this nation estimates that the yearly incidence of lung cancer is $6.2/10^5$ person years in the cohort as a whole, $5.5856/10^5$ person years in women not married or cohabitant with smokers, and $6.7027\times10^5$ person years in women who are married or cohabitant with smokers.

$$RR=6.7027/5.5856 = 1.20$$

$$AR_{exp}=7027/10^5 \text{ py} - 5.5856/10^5 \text{ py} = 1.1171/10^5 \text{ py}$$

$$AR_{exp} = \frac{1.20 - 1.00}{1.20} = 0.167$$

Among female cases, 17% due to husband’s smoking

$$\%POPAR = \frac{0.55 \times (1.20 - 1.00)}{1 + 0.55 \times (1.20 - 1.00)} = 0.0991$$

10% of cases in female population due to husband’s smoking
WHAT ARE CAUSAL ASSOCIATIONS?
(See text by Leon Gordis)

1. Exposure A both necessary and sufficient to result in Disease B.

   A implies B and B implies A

   (Rabies and Bite from an Infected Animal)

2. Exposure A is necessary but not sufficient to result in Disease B.

   B implies A but A does not imply B
   (Infection with mumps virus and clinical mumps)

3. Exposure A is sufficient but not necessary to result in Disease B.

   A implies B but B does not imply A
   (Alpha1 Antitrypsin Deficiency and Emphysema)

4. Exposure A is neither necessary nor sufficient to result in Disease B

   Example:
   B implies A+C or D+E or F+G

   (Site specific cancer can result if one or more of three initiator-promoter combinations are present)
Demonstrating Causation in Observational Studies

“Cigarette Paper” Criteria (Hill)

1. Strength of the Association
2. Dose-Response Relationship
3. Consistency of the Relationship
4. Temporally Correct Direction
5. Biological Plausibility
6. Specificity of Association
7. Cessation of Exposure
8. Consideration of Alternative Explanations
EPA CRITERIA FOR CAUSATION (LIMITED TO IDENTIFICATION OF CARCINOGENS)

Three criteria that must be met before a causal association can be inferred between exposure and cancer in humans.

1. There is no bias that can explain the association.

2. The possibility of confounding has been considered and ruled out as explaining the association.

3. The association is unlikely to be due to chance.