JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Estimating Risk

Sukon Kanchanaraksa, PhD
Johns Hopkins University

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section A

Relative Risk

Incidence of Disease (Attack Rate) = **Absolute Risk**

Attack Rates from Food-Borne Outbreak Exercise

	Attack Rate (%)			
Food	(1) Ate	(2) Not Ate		
Egg salad	83	30		
Macaroni	76	67		
Cottage cheese	71	69		
Tuna salad	78	50		
Ice cream	78	64		
Other	72	50		

Attack Rates from Food-Borne Outbreak Exercise

	Attack Rate (%)		Difference of Attack Rates	
Food	(1) Ate	(2) Not Ate	(1)–(2)	
Egg salad	83	30	53	
Macaroni	76	67	9	
Cottage cheese	71	69	2	
Tuna salad	78	50	28	
Ice cream	78	64	14	
Other	72	50	22	

Attack Rates from Food-Borne Outbreak Exercise

	Attack Rate (%)		Difference of Attack Rates	Ratio of Attack Rates
Food	(1) Ate	(2) Not Ate	(1)–(2)	(1)/(2)
Egg salad	83	30	53	2.77
Macaroni	76	67	9	1.13
Cottage cheese	71	69	2	1.03
Tuna salad	78	50	28	1.56
Ice cream	78	64	14	1.21
Other	72	50	22	1.44

Approaches to the Measurement of Excess Risk

- **Ratio of risks**

$$\frac{\text{Risk in exposed}}{\text{Risk in non-exposed}}$$

- **Differences in risks**

(Risk in exposed) – (Risk in non-exposed)

Relative Risk or Risk Ratio

$$\text{Relative risk (RR)} = \frac{\text{Risk in exposed}}{\text{Risk in non-exposed}}$$

Cohort Study

Then follow to see whether

Calculate and compare

First, identify

	Disease develops	Disease does not develop	Totals	Incidence of disease
Exposed	a	b	a+b	$\frac{a}{a+b}$
Not exposed	c	d	c+d	$\frac{c}{c+d}$

$\frac{a}{a+b}$ = Incidence in exposed $\quad \frac{c}{c+d}$ = Incidence in not exposed

Cohort Study

Then follow to see whether

Calculate and compare

First, identify

	Disease develops	Disease does not develop	Totals	Incidence of disease
Exposed	a	b	a+b	$\frac{a}{a+b}$
Not exposed	c	d	c+d	$\frac{c}{c+d}$

$\frac{a}{a+b}$ = Incidence in exposed

$\frac{c}{c+d}$ = Incidence in not exposed

$$\text{Relative Risk} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Cohort Study

Then follow to see whether calculate

First select

	Develop CHD	Do not develop CHD	Totals	Incidence of disease
Smoke cigarettes	84	2916	3000	$\frac{84}{3000}$
Do not smoke cigarettes	87	4913	5000	$\frac{87}{5000}$

$$\text{Relative Risk} = \frac{\frac{84}{3000}}{\frac{87}{5000}} = \frac{28.0}{17.4} = 1.61$$

Interpreting Relative Risk of a Disease

- If RR = 1
  - Risk in exposed = Risk in non-exposed
  - No association
- If RR > 1
  - Risk in exposed > Risk in non-exposed
  - Positive association; ? causal
- If RR < 1
  - Risk in exposed < Risk in non-exposed
  - Negative association; ? protective

Cross-Tabulation Table (Food-Borne Outbreak Exercise)

Attack Rates of Sore Throat

Tuna Salad \ Egg Salad	Ate	Did not eat
Ate	46/53 (87%)	3/10 (30%)
Did not eat	8/12 (67%)	3/10 (30%)

Cross-Tabulation Table (Food-Borne Outbreak Exercise)

Relative Risk of Sore Throat

Tuna Salad \ Egg Salad	Ate	Did not eat
Ate	2.9	1.0
Did not eat	2.2	1.0

The baseline group for comparison is the no exposure group—i.e., those who did not eat tuna salad and did not eat egg salad

Exposure-Disease Tables Expanded from the Cross-Tabulation Table (Food-Borne Outbreak Exercise)

Both Tuna Salad and Egg Salad

	Sore Throat: Yes	Sore Throat: No	Total
Ate	46	7	53
Did not eat either	3	7	10

RR = (46/53)/(3/10) =2.9

Tuna Salad Only

	Sore Throat: Yes	Sore Throat: No	Total
Ate	3	7	10
Did not eat either	3	7	10

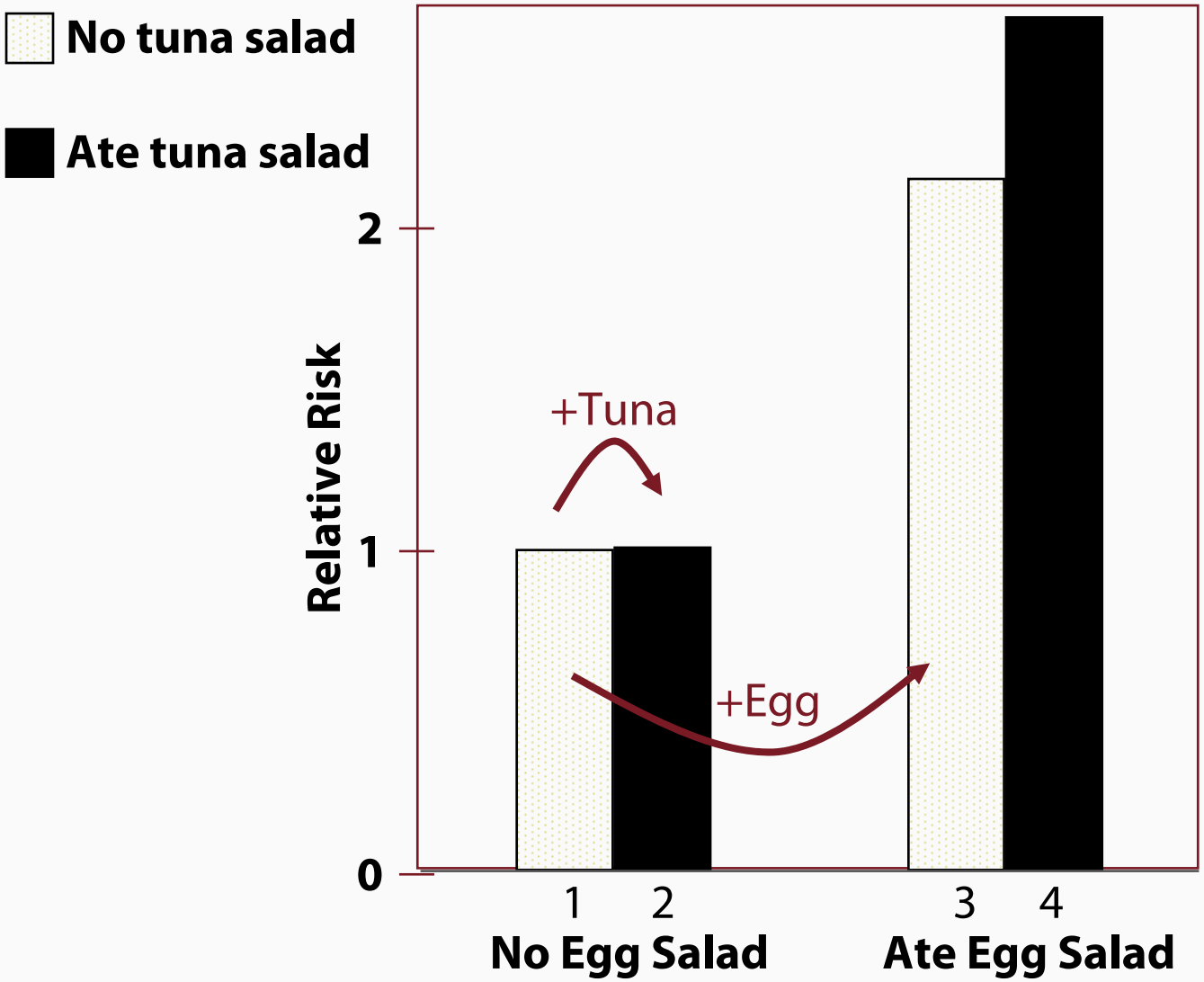
RR = (3/10)/(3/10) =1.0

Egg Salad Only

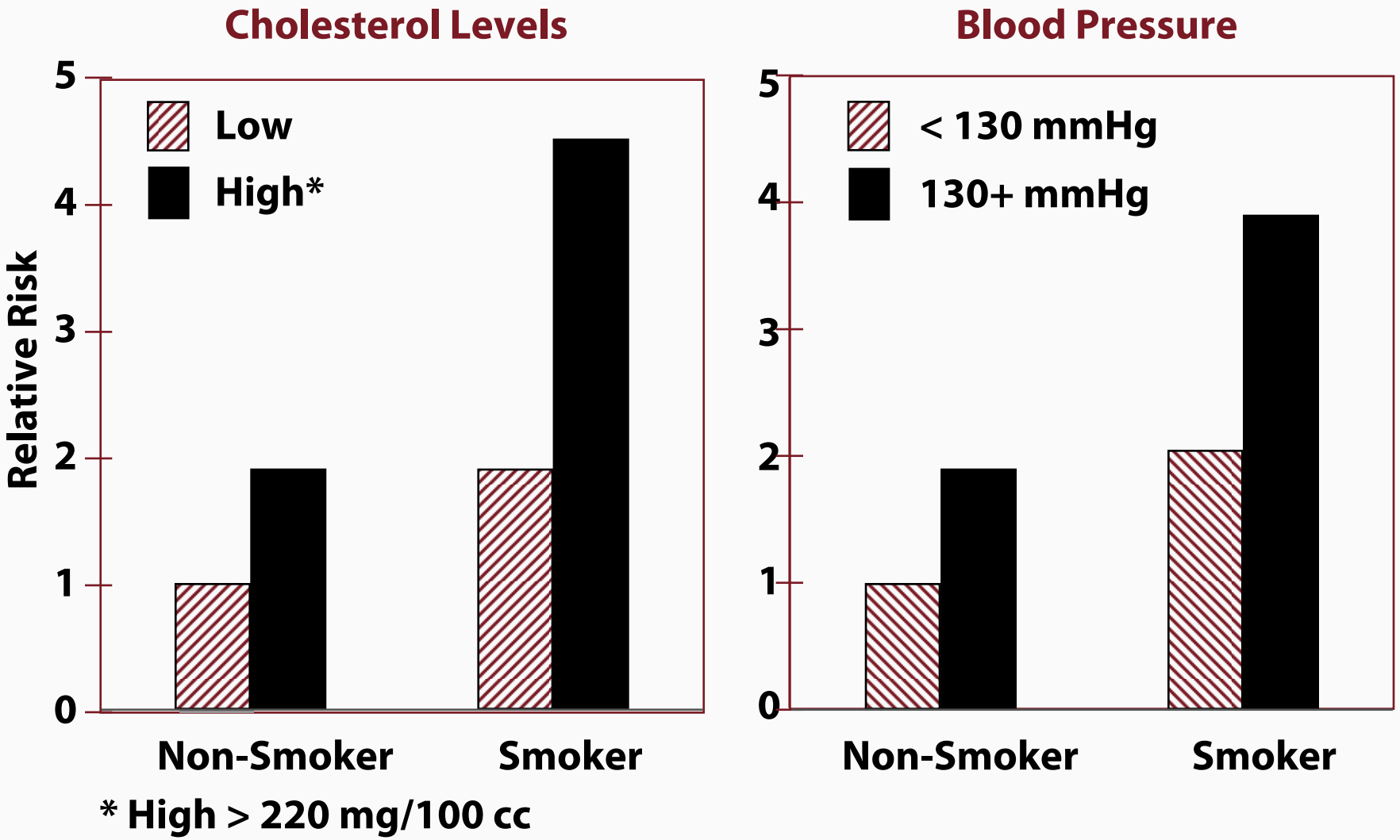
	Sore Throat: Yes	Sore Throat: No	Total
Ate	8	4	12
Did not eat either	3	7	10

RR = (8/12)/(3/10) =2.2

Relative Risk by Food Items



Relative Risk for MI and CHD Death in Men Aged 30–62 in Relation to Cigarette Smoking



Source: Doyle et al, 1964

Relationship between Serum Cholesterol Levels and Risk of Coronary Heart Disease by Age and Sex

Serum Cholesterol mg/dL	Men		Women	
	Aged 30–49	Aged 50–62	Aged 30–49	Aged 50–62
Incidence Rates (per 1,000)				
< 190	38.2	105.7	11.1	155.2
190–219	44.1	187.5	9.1	88.9
220–249	95.0	201.1	24.3	96.3
250+	157.5	267.8	50.4	121.5

Source: Doyle et al, 1964

Incidence Rates and RR of CHD in Relation to Serum Cholesterol Levels by Age and Sex

Serum Cholesterol mg/dL	Men		Women	
	Aged 30–49	Aged 50–62	Aged 30–49	Aged 50–62
Incidence Rates (per 1,000)				
< 190	38.2	105.7	11.1	155.2
190–219	44.1	187.5	9.1	88.9
220–249	95.0	201.1	24.3	96.3
250+	157.5	267.8	50.4	121.5
Relative Risk*				
< 190	1.0	2.8	0.3	4.1
190–219	1.2	4.9	0.2	2.3
220–249	2.5	5.3	0.6	2.5
250+	4.1	7.0	1.3	3.2

* RR of 1.0 set at level for males 30–49 yrs of age with cholesterol level < 190 mg/dL.

Incidence Rates and RR of CHD in Relation to Serum Cholesterol Levels by Age and Sex

Serum Cholesterol mg/dL	Men		Women	
	Aged 30–49	Aged 50–62	Aged 30–49	Aged 50–62
Incidence Rates (per 1,000)				
< 190	38.2	105.7	11.1	155.2
190–219	44.1	187.5	9.1	88.9
220–249	95.0	201.1	24.3	96.3
250+	157.5	267.8	50.4	121.5
Relative Risk*				
< 190	1.0	2.8	0.3	4.1
190–219	1.2	4.9	0.2	2.3
220–249	2.5	5.3	0.6	2.5
250+	4.1	7.0	1.3	3.2

* RR of 1.0 set at level for males 30–49 yrs of age with cholesterol level < 190 mg/dL.

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section B

Odds Ratio

Interpreting Odds

- "Odds" is often known as the ratio of money that may be won versus the amount of money bet
- In statistics, an **odds** of an event is the ratio of:
  - **The probability that the event WILL occur** *to* **the probability that the event will NOT occur**
    - For example, in 100 births, the probability of a delivery being a boy is 51% and being a girl is 49%
    - The odds of a delivery being a boy is **51/49 = 1.04**
- In simpler term, an odds of an event can be calculated as:
  - **Number of events divided by number of non-events**

Calculating Risk in a Cohort Study

	Develop Disease	Do Not Develop Disease
Exposed	a	b
Non-exposed	c	d

The probability that an exposed person develops disease $= \frac{a}{a + b}$

The probability that a non-exposed person develops disease $= \frac{c}{c + d}$

Applying Concept of Odds

- Let's borrow the concept of odds and apply it to disease and non-disease
- So, **the odds of having the disease** is the ratio of the probability that the disease will occur to the probability that the disease will not occur
- Or, **the odds of having the disease** can be calculated as the number of people with the disease divided by the number of people without the disease
- [Note: in the exposure-disease 2x2 table, the **odds of having a disease** in the exposed group is the same as the odds that an exposed person develops the disease]

Calculating Odds in a Cohort Study

	Develop Disease	Do Not Develop Disease
Exposed	a	b
Non-exposed	c	d

The odds that an exposed person develops disease $= \frac{a}{b}$

The odds that a non-exposed person develops disease $= \frac{c}{d}$

Calculating Odds in a Cohort Study

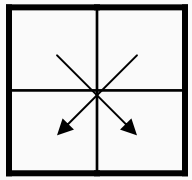
	Develop Disease	Do Not Develop Disease
Exposed	a	b
Non-exposed	c	d

Odds ratio is the ratio of the odds of disease in the exposed to the odds of disease in the non-exposed

$$OR = \frac{\text{odds that an exposed person develops the disease}}{\text{odds that a non-exposed person develops the disease}} = \frac{\frac{a}{b}}{\frac{c}{d}}$$

Disease Odds Ratio in a Cohort Study

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a}{b} \times \frac{d}{c} = \frac{ad}{bc}$$



Calculating Odds Ratio in a Case-Control Study

	Case	Control
History of Exposure	a	b
No History of Exposure	c	d

The odds that a case was exposed $= \frac{a}{c}$

The odds that a control was exposed $= \frac{b}{d}$

Calculating Odds Ratio in a Case-Control Study

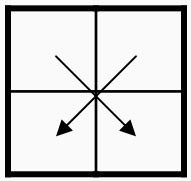
	Case	Control
History of Exposure	a	b
No History of Exposure	c	d

Odds ratio (OR) is the ratio of the odds that a case was exposed to the odds that a control was exposed

$$OR = \frac{\text{odds that a case was exposed}}{\text{odds that a control was exposed}} = \frac{\frac{a}{c}}{\frac{b}{d}}$$

Exposure Odds Ratio in a Case-Control Study

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a}{c} \times \frac{d}{b} = \frac{ad}{bc}$$



Odds Ratio versus Relative Risk

- Odds ratio can be calculated in a cohort study and in a case-control study
  - The exposure odds ratio is equal to the disease odds ratio
- Relative risk can only be calculated in a cohort study

When Is Odds Ratio a Good Estimate of Relative Risk?

- When the "cases" studied are representative of all people with the disease in the population from which the cases were drawn, with regards to history of the exposure
- When the "controls" studied are representative of all people without the disease in the population from which the cases were drawn, with regards to history of exposure
- When the disease being studied is not a frequent one

When Is Odds Ratio a Good Estimate of Relative Risk?

- If the incidence of the disease is low, then:

$$a+b \sim b$$
$$c+d \sim d$$

- Therefore:

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

$$\sim \frac{a/b}{c/d} = \frac{ad}{bc} = OR$$

Comparing OR to RR: Disease Is Infrequent

	Develop Disease	Do not Develop Disease	
Exposed	200	9800	10,000
Non-Exposed	100	9900	10,000

$$\text{Relative Risk} = \frac{200/10,000}{100/10,000} = 2$$

$$\text{Odds Ratio} = \frac{200 \times 9900}{100 \times 9800} = 2.02$$

Comparing OR to RR: Disease Is NOT Infrequent

	Develop Disease	Do not Develop Disease	
Exposed	50	50	100
Non-Exposed	25	75	100

$$\text{Relative Risk} = \frac{50/75}{50/25} = 2$$

$$\text{Odds Ratio} = \frac{50 \times 75}{50 \times 25} = 3$$

Interpreting Odds Ratio of a Disease

- **If OR = 1**
  - Exposure is not related to disease
  - No association; independent
- **If OR > 1**
  - Exposure is positively related to disease
  - Positive association; ? causal
- **If OR < 1**
  - Exposure is negatively related to disease
  - Negative association; ? protective

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section C

Odds Ratio in Unmatched and Matched Case-Control

Unmatched Case-Control Study: Example

CASE	CONTROL
E	N
E	**E**
N	N
E	N
N	**E**
N	N
E	N
E	**E**
E	N
N	N

E = Exposed
N = Not exposed

Assume a study of 10 cases and 10 unmatched controls, with these findings

Unmatched Case-Control Study: Example

CASE	CONTROL
E	N
E	**E**
N	N
E	N
N	**E**
N	N
E	N
E	**E**
E	N
N	N

E = Exposed
N = Not exposed

Thus, 6 of 10 cases were exposed, and 3 of 10 controls were exposed. In a 2x2 table, we have the following:

	Case	Control
Exposed	6	3
Not Exposed	4	7

Unmatched Case-Control Study: Example

CASE	CONTROL
E	N
E	**E**
N	N
E	N
N	**E**
N	N
E	N
E	**E**
E	N
N	N

E = Exposed
N = Not exposed

	Case	Control
Exposed	6	3
Not Exposed	4	7

$$OR = \frac{ad}{bc} = \frac{6 \times 7}{3 \times 4} = 3.5$$

Quick Pause

In a hypothetical 2x2 table with the following rows and columns, is the OR calculated correctly?

	Control	Case
Exposed	8	3
Not Exposed	4	7

$$OR = \frac{ad}{bc} = \frac{8 \times 7}{3 \times 4} = 4.7$$

Quick Pause

	Control	Case
Exposed	8	3
Not Exposed	4	7

Incorrect!

$$OR = \frac{ad}{bc} = \frac{8 \times 7}{3 \times 4} = 4.7$$

Why?

Odds Ratio in a Case-Control Study

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a}{c} \times \frac{d}{b} = \frac{ad}{bc}$$

$$= \frac{(\#\text{ cases exposed}) \times (\#\text{ controls not exposed})}{(\#\text{ cases not exposed}) \times (\#\text{ controls exposed})}$$

The numerator is the product of cases exposed and controls not exposed.

Case-Control Study: Example

	Cases CHD	Controls (without disease)
Smoked cigarettes	112	176
Did not smoke cigarettes	88	224
Total	200	400

% Smoking cigarettes $\frac{112}{200} = 56\%$ $\frac{176}{400} = 44\%$

$$OR = \frac{ad}{bc} = \frac{112 \times 224}{176 \times 88} = 1.62$$

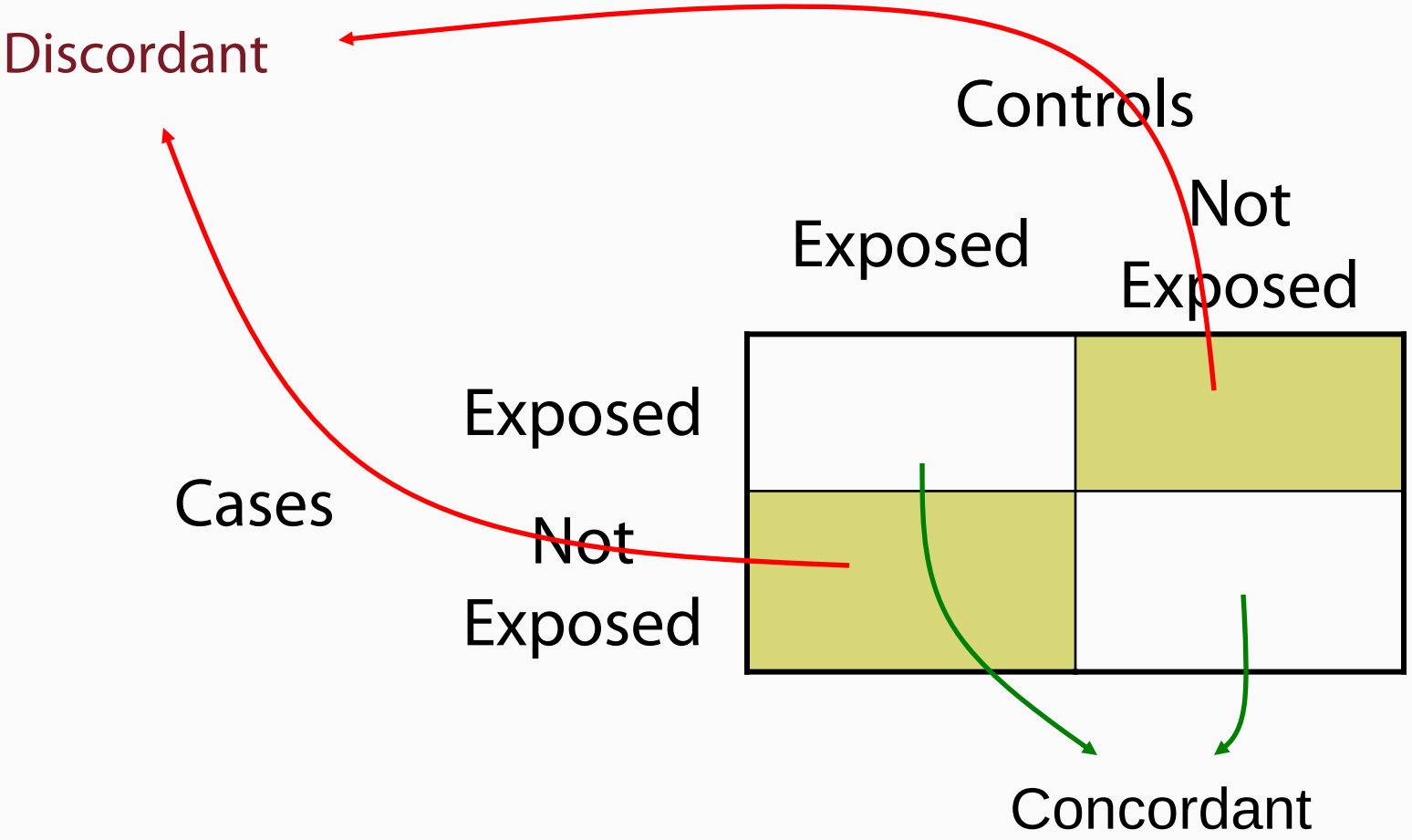
Matched Case-Control Study

- In a **matched case-control study,** one or more controls are selected to match to a case on certain characteristics, such as age, race, and gender
- When one control is matched to a case, the case and the matched control form a **matched pair**

Concordant and Discordant Pairs

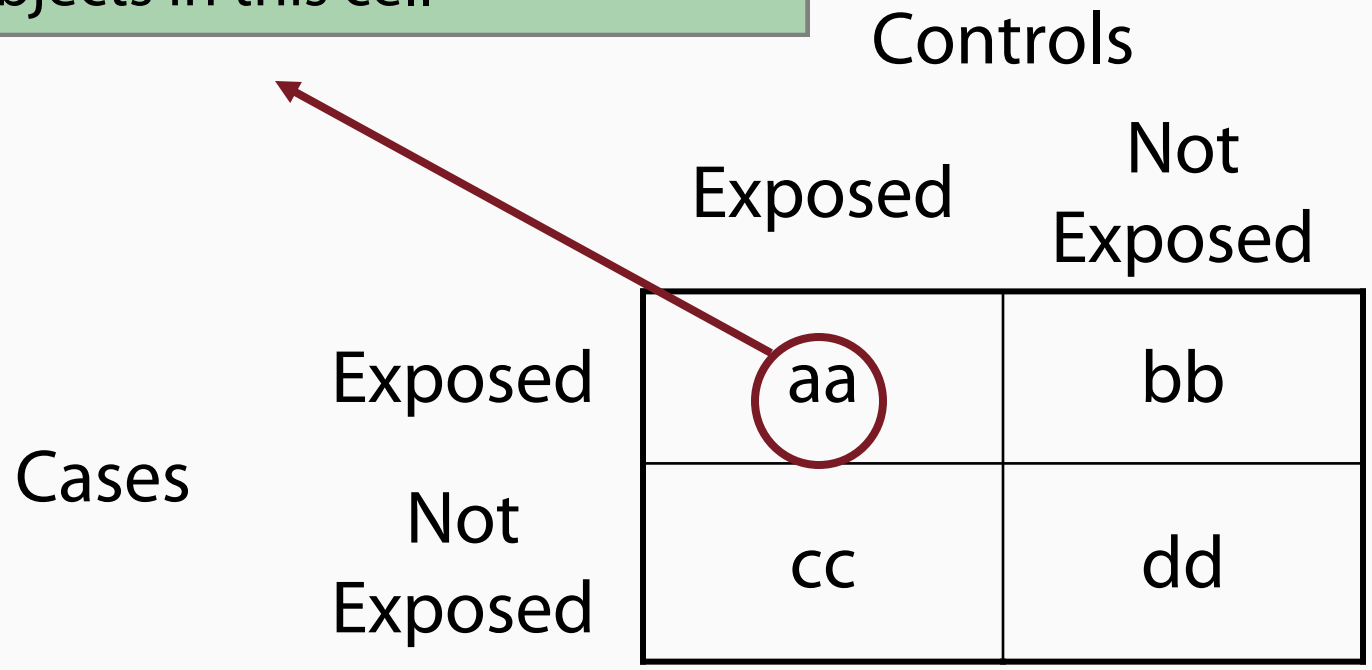
- We can define two types of matched pairs by the similarity or difference of the exposure of the case and control in each pair
- **Concordant pairs** are:
  1. Pairs in which both the case and the control were exposed, and
  2. Pairs in which neither the case nor the control was exposed
- **Discordant pairs** are:
  3. Pairs in which the case was exposed but the control was not, and
  4. Pairs in which the control was exposed and the case was not

2x2 Table in a Matched Case-Control Study



2x2 Table in a Matched Case-Control Study

"aa" = number of matched pairs
2 x aa subjects in this cell



Total number of subjects = 2 x (aa+bb+cc+dd)

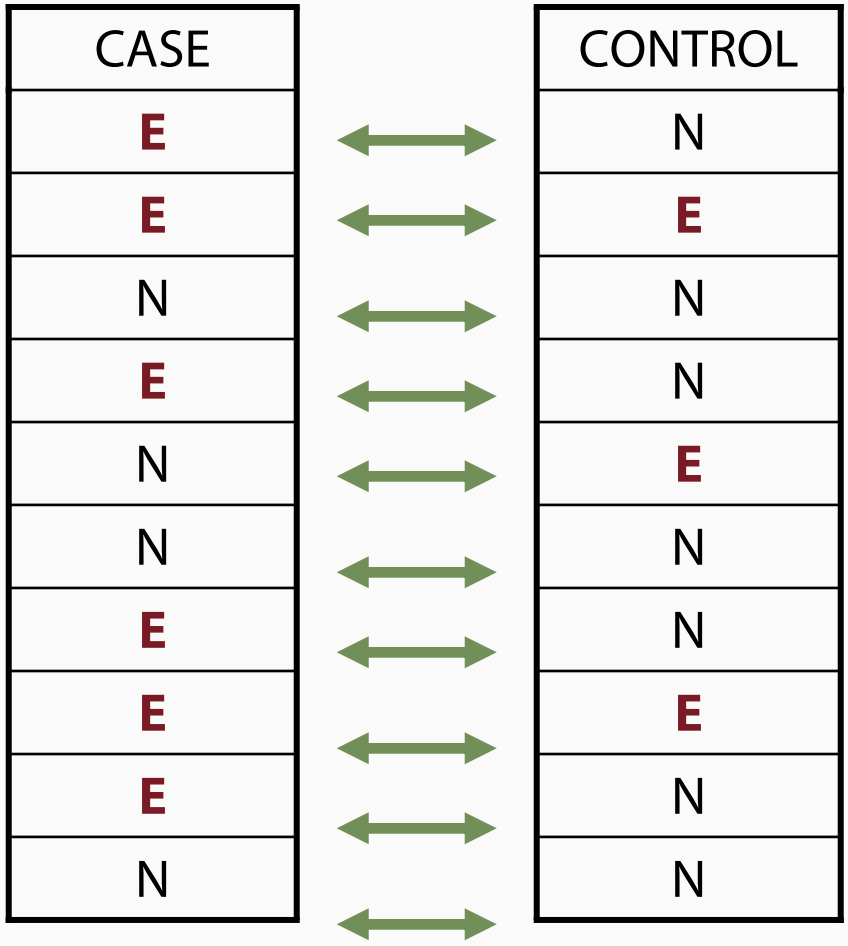
OR from 2x2 Table in a Matched Case-Control Study

$$\text{Odds ratio (matched)} = \frac{bb}{cc}$$

		Controls: Exposed	Controls: Not Exposed
Cases	Exposed	aa	bb
	Not Exposed	cc	dd

Note: bb is not the product of b and b (not b x b); it is the number of pairs

Matched Case-Control Study: Example

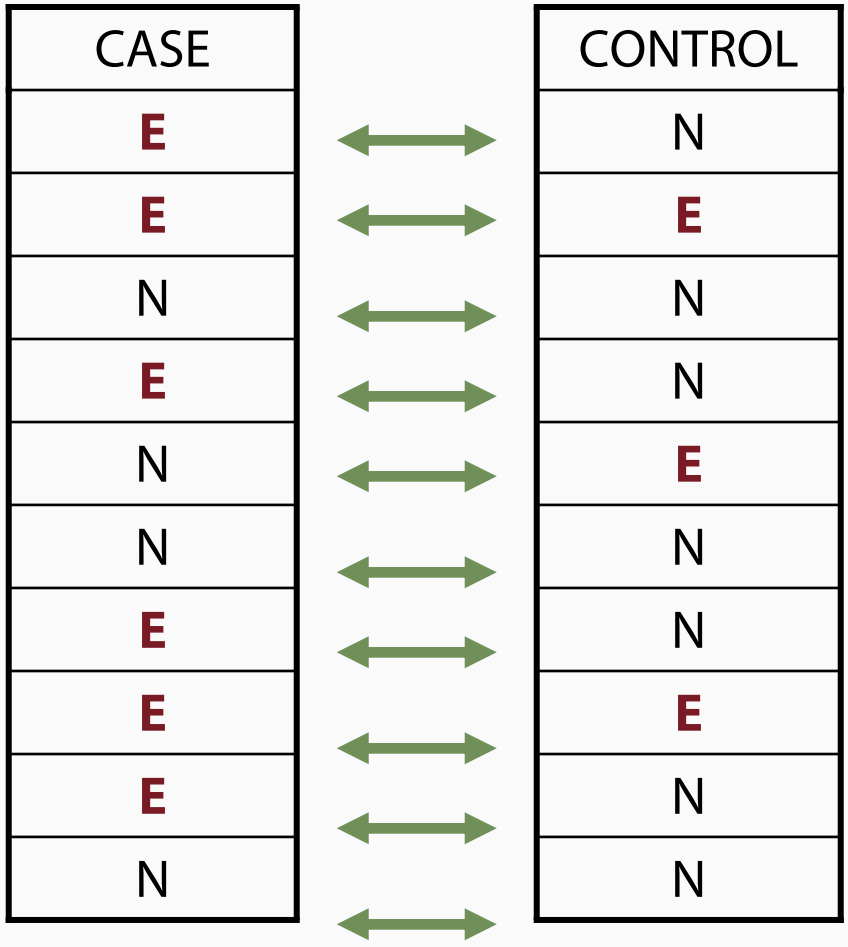


Assume a study of 10 cases and 10 controls in which each control was matched to a case resulting in 10 pairs.

E = Exposed

N = Not exposed

Matched Case-Control Study: Example



E = Exposed
N = Not exposed

Cases

	Controls: Exposed	Controls: Not Exposed
Cases: Exposed	2	4
Cases: Not Exposed	1	3

$$\text{Matched OR} = \frac{4}{1} = 4$$

Review: Matched Case-Control Study

		Controls: Exposed	Controls: Not Exposed
Cases	Exposed	2	4
	Not Exposed	1	3

Q1. How many pairs?

Q2. How many subjects?

Q3. What are the discordant pairs?

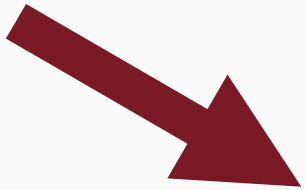
Q4. Which is the "bb" cell?

Q5. What is the "bb" cell?

Review: Unmatching a Matched 2x2 Table

Matched CC

Cases \ Controls	Exposed	Not Exposed
Exposed	2	4
Not Exposed	1	3



Unmatched 2x2

Exposure \ Disease	Yes	No
Exposed		
Not Exposed		

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section D

Attributable Risk

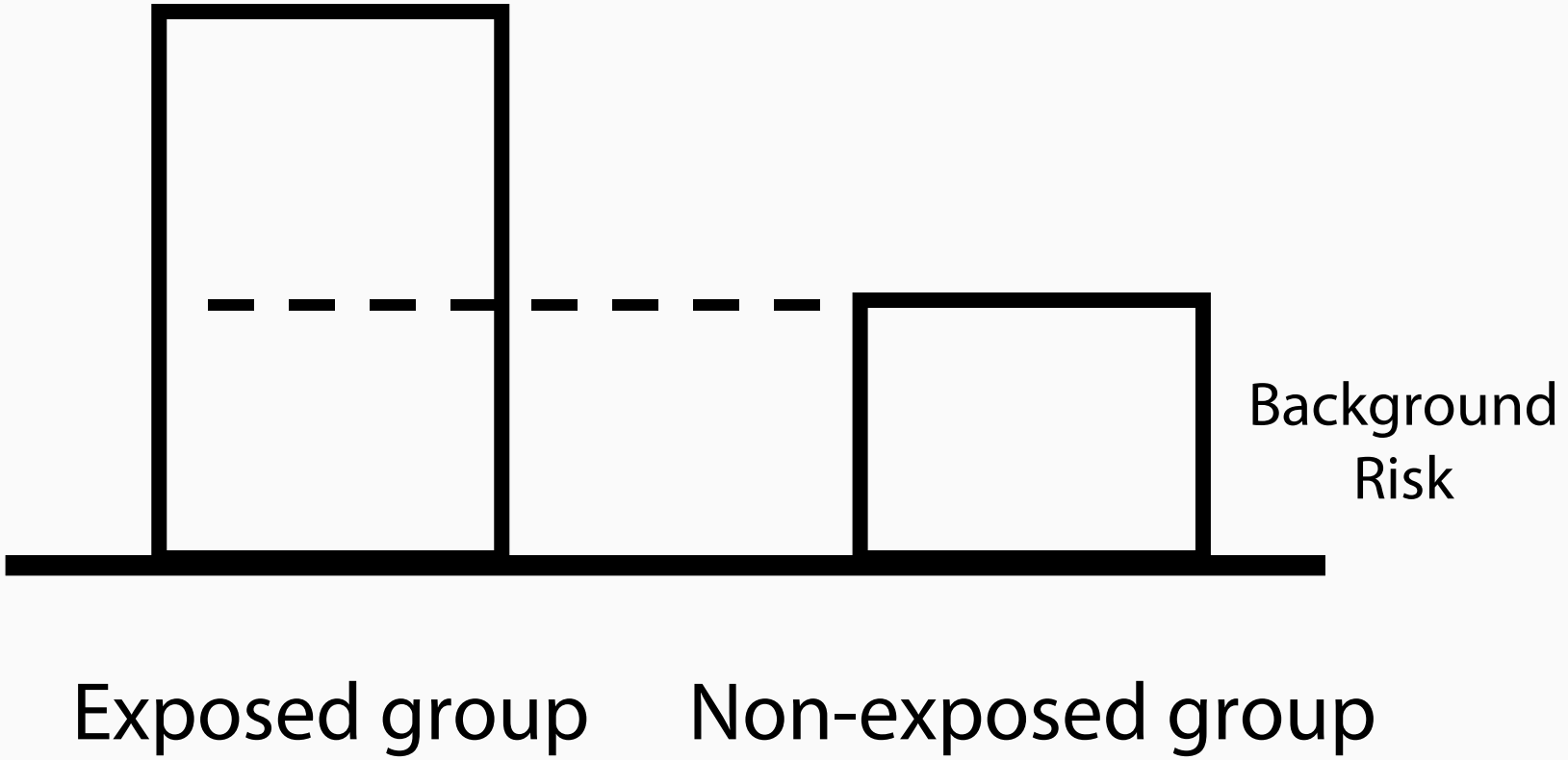
Attributable Risk

- **Attributable risk (AR)** is a measure of excess risk that is attributed to the exposure
- Attributable risk in the exposed group equals the difference between the incidence in the exposed group and the incidence in the non-exposed (baseline) group

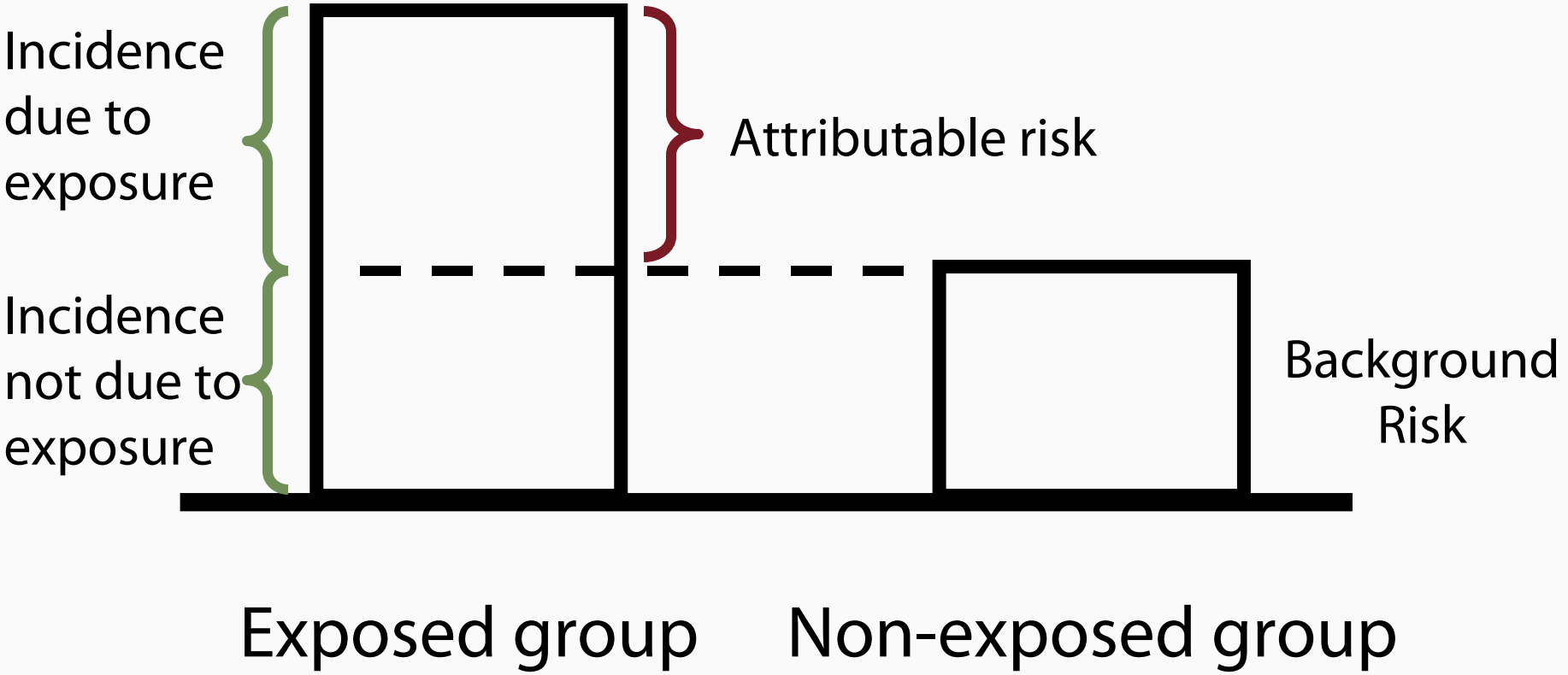
Attack Rates from Food-Borne Outbreak Exercise

	Attack Rate (%)		Difference of Attack Rates	
Food	(1) Ate	(2) Not Ate	(1)–(2)	
Egg salad	83	30	53	
Macaroni	76	67	9	
Cottage cheese	71	69	2	
Tuna salad	78	50	28	
Ice cream	78	64	14	
Other	72	50	22	

Risk in Exposed and Non-Exposed Groups



Risk in Exposed and Non-Exposed Groups



Risk in Exposed and Non-Exposed Groups

1. Incidence attributable to exposure (attributable risk)

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{exposed group}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)$$

Risk in Exposed and Non-Exposed Groups

1. Incidence attributable to exposure (**attributable risk**)

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{exposed group}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)$$

2. Proportion of incidence attributable to exposure (**proportional attributable risk**)

$$= \frac{\left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{exposed group}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)}{\textbf{Incidence in exposed group}}$$

Example: Cohort Study

	Develop CHD	Do not develop CHD	Totals	Incidence of disease
Smoke cigarettes	84	2916	3000	28.0 per 1,000
Do not smoke cigarettes	87	4913	5000	17.4 per 1,000

Attributable Risk in Smokers

1. The incidence in smokers which is attributable to their smoking

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{smokers}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-smokers}\end{array}\right)$$

= 28.0 – 17.4 = 10.6/1,000/year

Proportion Attributable Risk in Smokers

2. The proportion of the total incidence in the smokers which is **attributable** to their smoking

$$= \frac{\left(\text{Incidence in smokers}\right) - \left(\text{Incidence in non-smokers}\right)}{\text{Incidence in smokers}}$$

$$= \frac{28.0 - 17.4}{28.0} = \frac{10.6}{28.0} = 0.379 = 37.9\%$$

Risk in the Total Population

- Population is a mix of exposed and non-exposed groups

Attributable Risk in the Total Population

3. Incidence attributable to exposure

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{total population}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)$$

Attributable Risk in the Total Population

3. Incidence attributable to exposure

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{total population}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)$$

4. **Proportion** of incidence attributable to exposure

$$= \frac{\left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{total population}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)}{\textbf{Incidence in total population}}$$

Attributable Risk in the Total Population

3. Incidence attributable to smoking in the total population

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{total population}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-exposed group}\end{array}\right)$$

Attributable Risk in the Total Population

- If the incidence in the total population is unknown, it can be calculated if we know:
  - Incidence among smokers
  - Incidence among nonsmokers
  - Proportion of the total population that smokes

Attributable Risk in the Total Population

- We know that:
  - The incidence in smokers = 28.0/1,000/year
  - The incidence in nonsmokers = 17.4/1,000/year
- From another source, we learn that:
  - The proportion of smokers in the population is **44%**
- So, we know that:
  - The proportion of nonsmokers in the population is **56%**

Attributable Risk in the Total Population

- Incidence in total population =

(Incidence in smokers)(Percent smokers in population) + (Incidence in non-smokers)(Percent non-smokers in population)

(28.0/1000) (.44) + (17.4/1000) (.56)

= 22.1/1000/year

Attributable Risk in the Total Population

3. Incidence attributable to smoking

$$= \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{total population}\end{array}\right) - \left(\begin{array}{c}\textbf{Incidence in}\\ \textbf{non-smokers}\end{array}\right)$$

(22.1/1000/year) – (17.4/1000/year)

= 4.7/1000/year

Attributable Risk in the Total Population

4. Proportion of incidence attributable to exposure

$$= \frac{\left(\text{Incidence in total population}\right) - \left(\text{Incidence in non-smokers}\right)}{\text{Incidence in total population}}$$

$$\frac{22.1 - 17.4}{22.1}$$

$$= 21.3\%$$

Lung Cancer, CHD Mortality in Male British Physicians

	Age-Adjusted Death Rates/100,000				
	Smokers	Non-Smokers	RR	AR	%AR
Lung cancer	140	10	14.0	130	92%
CHD	669	413	1.6	256	38%
%AR = Proportion attributable risk					

Source: Doll and Peto (1976). BMJ, 2:1525.

Lung Cancer, CHD Mortality in Male British Physicians

	Age-Adjusted Death Rates/100,000				
	Smokers	Non-Smokers	RR	AR	%AR
Lung cancer	140	10	14.0	130	92%
CHD	669	413	1.6	256	38%
%AR = Proportion attributable risk					

Source: Doll and Peto (1976). BMJ, 2:1525.