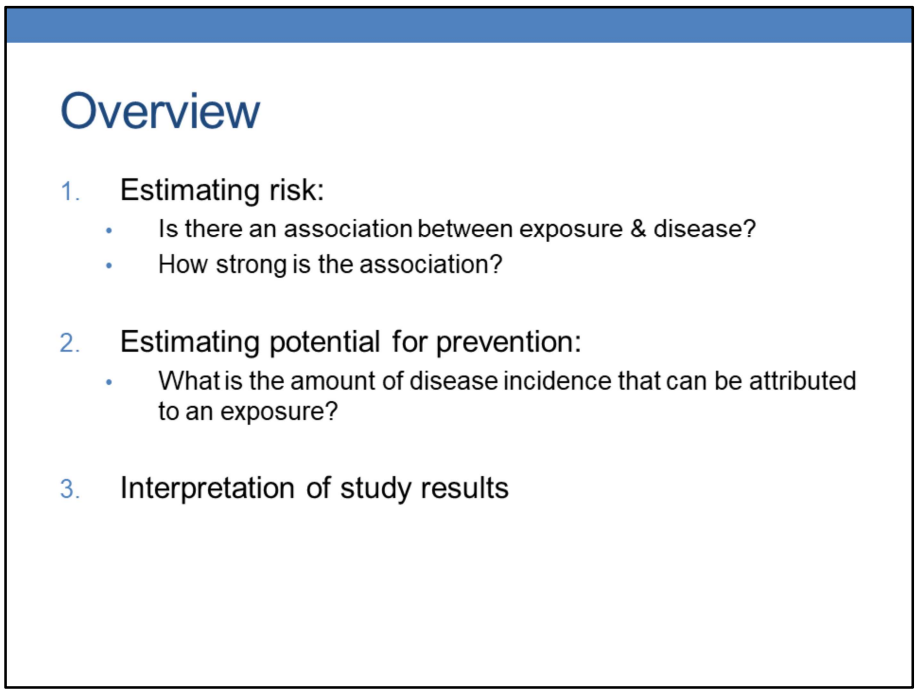
Epidemiologic Measures of Association and Impact Part I

Nasir Mushtaq, PhD, MBBS
Associate Professor

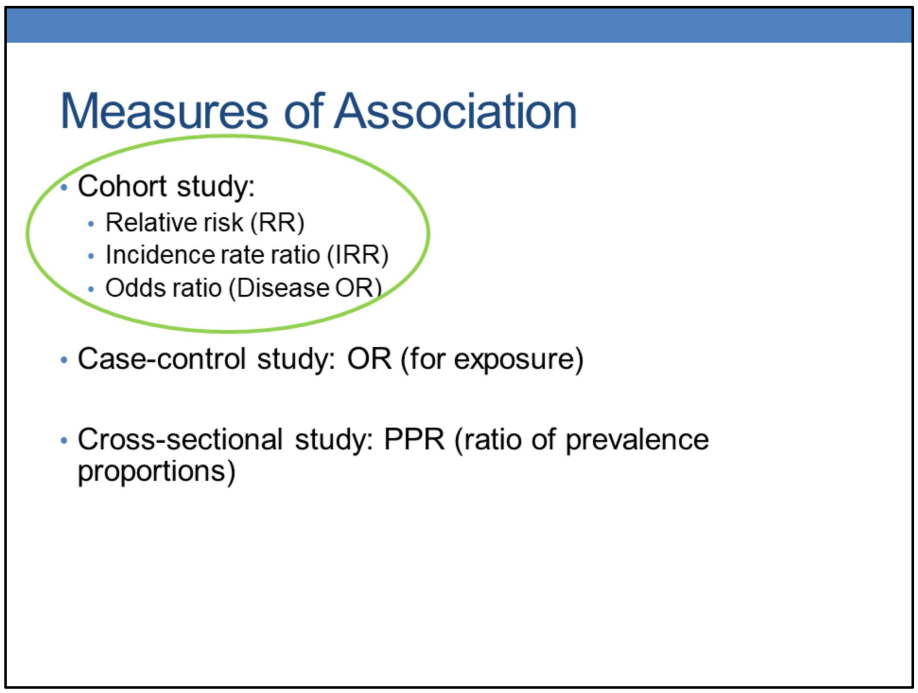
Department of Biostatistics and Epidemiology
Hudson College of Public Health

Department of Family and Community Medicine
OU-TU School of Community Medicine

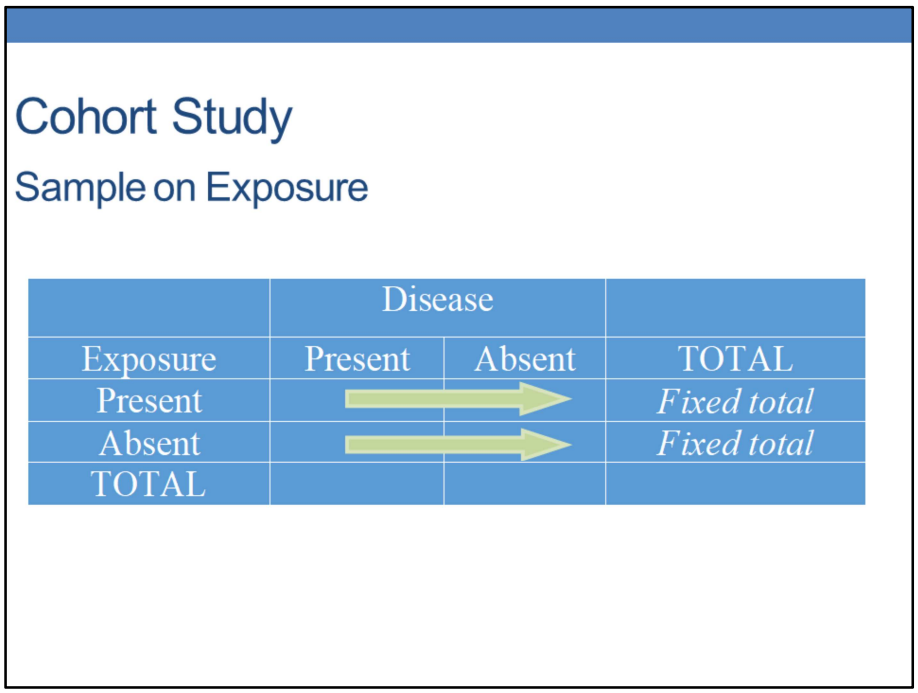
Welcome to this three-part series focused on epidemiologic measures of association and impact.



This three-part series will focus on the estimation of the association between exposures and outcomes. We will learn about epidemiologic measures that can be used to determine if there an association between exposure and disease. Furthermore, we will learn about measures that can be used to summarize the strength of the association between exposure and disease outcome. Next, we will lean about epidemiologic measures that can be used to quantify the potential for prevention. Specifically, we will learn about measures that can be used to quantify the amount of disease incidence that can be attributed to a particular exposure. Throughout this series, we will focus on the interpretation of study results, and in particular, measures of the association between exposure and disease outcome.



In this first module, we will focus on the interpretation of measures of association that can be calculated from cohort studies. We will discuss the calculation of relative risk and incidence rate ratios. While we can calculate the disease odds ratio, this measure is less common. In the second module, we will focus on the interpretation of measures from case-control and cross-sectional studies.



Recall that when conducting a prospective cohort study, we sample groups of individuals based on exposure status, either the exposure is present or the exposure is absent, and then we follow participants forward in time and compare the incidence of disease between the exposed and unexposed groups to determine if exposure is associated with the disease outcome. The number of exposed and unexposed individuals is fixed. As an example, we might enroll 100 smokers and 100 never smokers in a prospective cohort study and compare the cumulative incidence of COPD between the groups.

Measures of Association in Cohort Studies

Relative Risk (RR)

$$RR = \frac{Risk\ in\ exposed}{Risk\ in\ unexposed}$$

$$RR = \frac{P(D|E)}{P(D|NE)}$$

Risk = cumulative incidence

The relative risk is a common measure of association that is calculated using data arising from a prospective cohort study. The relative risk is calculated as the risk of the disease outcome in the exposed group divided by the risk of disease among the unexposed group. In probability notation, the relative risk is the probability of disease among the exposed divided by the probability of disease among the unexposed. Recall that the risk is calculated as the cumulative incidence of disease at a particular point in time.

Relative Risk

	Disease	No Disease	Total
Exposed	a	b	a+b
Not Exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

Risk of disease among **exposed** = a/(a+b) = p(D|E)

In terms of standard notation for a 2x2 table where the columns represent disease status and the rows represent exposure status, the risk of disease among the exposed is calculated as the number of exposed participants with disease divided by the number of exposed participants.

Relative Risk

	Disease	No Disease	Total
Exposed	a	b	a+b
Not Exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

Risk of disease among **unexposed** = c/(c+d) = p(D | NE)

Similarly, the risk of disease among the unexposed is calculated as the number of unexposed participants with disease divided by the number of unexposed participants.

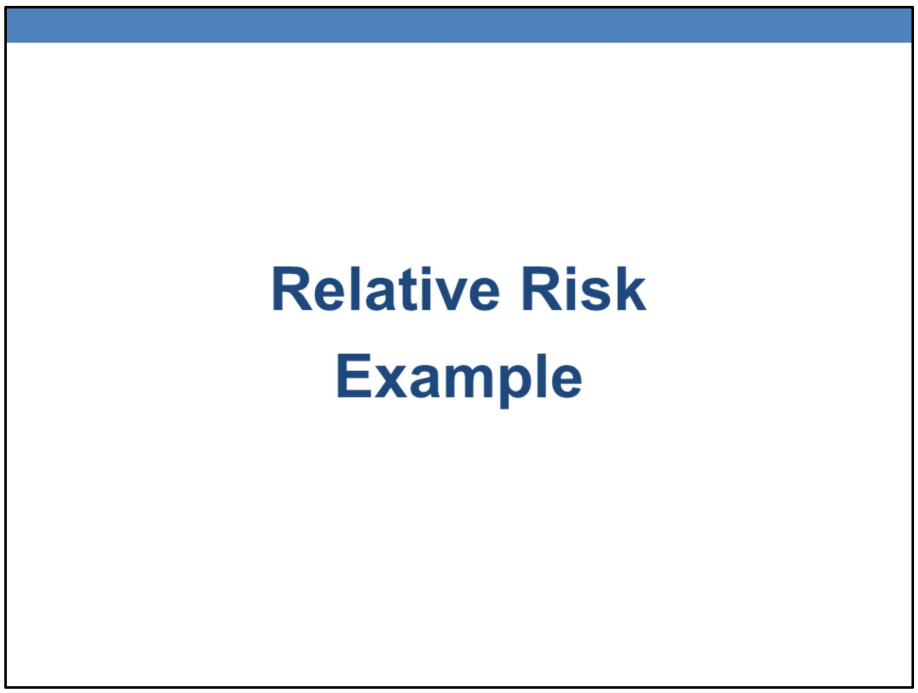
Relative Risk

	Disease	No Disease	Total
Exposed	a	b	a+b
Not Exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

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

$$= \frac{Risk_{exposed}}{Risk_{not\ exposed}} = \frac{P(D|E)}{P(D|NE)}$$

Then, the relative risk is calculated as the ratio of these probabilities, namely, the risk among the exposed divided by the risk among the unexposed.



Now, let's consider a data example and calculate and interpret a relative risk value.

Risk in Smokers

	Lung cancer	No Lung cancer	Total
Smokers	90	710	800
Nonsmokers	10	1190	1200
Totals	100	1900	2000

Risk in smokers = $\frac{a}{(a+b)} = \frac{90}{800} = 0.1125$

i.e. 112.5 / 1000 smokers

Let's consider data from 800 smokers and 1200 non-smokers. Assume that these data come from men aged 60 years or older who are followed over a 20 year period. Among the 800 smokers, 90 developed lung cancer over a 20 year period. This results in a risk of lung cancer among smokers of 0.1125 or 112.5 among 1000 smokers.

Risk in Nonsmokers

	Lung cancer	No Lung cancer	Total
Smokers	90	710	800
Nonsmokers	10	1190	1200
Totals	100	1900	2000

Risk in nonsmokers = $\frac{c}{(c+d)} = \frac{10}{1200} = 0.0083$

i.e. 8.3 / 1000 nonsmokers

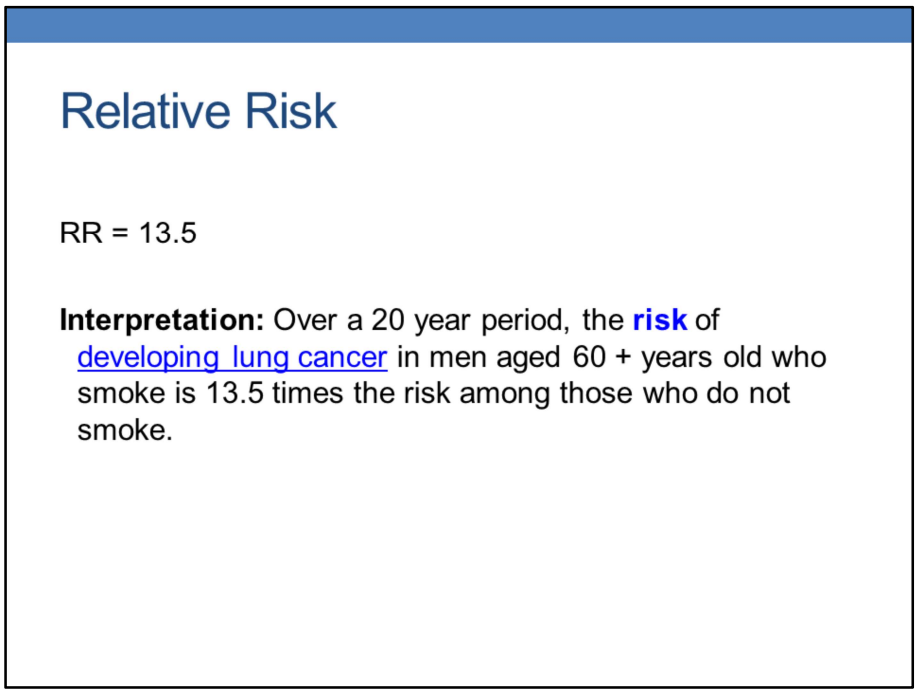
Similarly, among the 1200 non-smokers, 10 developed lung cancer over a 20 year period. This results in a risk of lung cancer among non-smokers of 0.0083 or 8.3 among 1000 non-smokers.

Relative Risk

	Lung cancer	No Lung cancer	Total
Smokers	90	710	800
Nonsmokers	10	1190	1200
Totals	100	1900	2000

$$RR = \frac{Risk\ in\ smokers}{Risk\ in\ nonsmokers} = \frac{0.1125}{0.0083} = 13.5$$

With these risk values, the relative risk, comparing the risk among smokers to the risk among non-smokers, is 13.5



The relative risk value of 13.5 can be interpreted as over a 20 year period, the **risk** of developing lung cancer in men aged 60 + years old who smoke is 13.5 times the risk among those who do not smoke.

Interpretation of RR

- RR = 1 → No association
- RR > 1 → Exposed are at higher risk of outcome than the unexposed
- RR < 1 → Exposed are at lower risk of outcome

When interpreting a relative risk, we compare the relative risk to a value of 1. If the relative risk is 1, this indicates that the risk of disease is equal between the exposed and unexposed and therefore, there is no association between exposure and outcome.

If the value of the relative risk is greater than 1, this indicates that the risk of disease is higher among the exposed compared to the unexposed.

If the value of the relative risk is less than 1, this indicates that the risk of disease is lower among the exposed compared to the unexposed.

Cohort Study
Incidence Rates

	Disease present	Person-time
Exposure		
Present		
Absent		
Total		

In the previous example, we calculated the cumulative risk of a disease outcome after a particular number of years of follow-up. In the smoking example, participants were followed for 20 years for the development of lung cancer.

In some studies, the duration of follow-up will differ among the participants and therefore, we calculate an incidence density measure that reflects not only the number of patients who developed the disease outcome of interest but also reflects the amount of time that the patients are at risk for developing the disease.

Incidence Rate

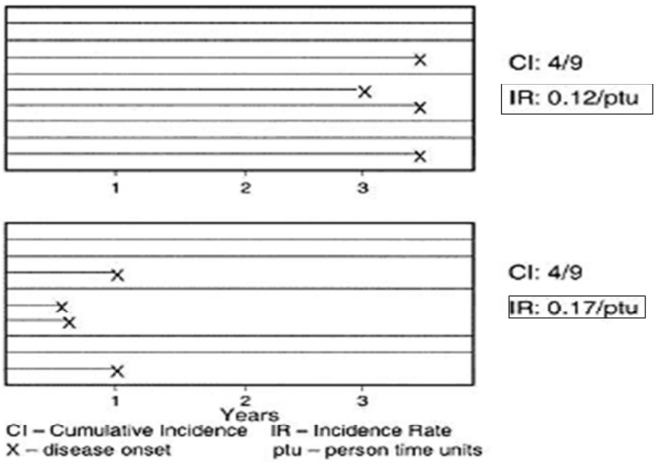
- IR = $\frac{Number\ of\ cases}{Person\ time}$
- IR in exposed = $\frac{Number\ of\ cases\ among\ exposed}{Person\ time\ in\ exposure}$
- IR in nonexposed = $\frac{Number\ of\ cases\ among\ nonexposed}{Person\ time\ in\ nonexposure}$

Recall that we calculate the incidence rate, or incidence density, as the number of cases divided by the person-time at risk.

The incidence rate for the exposed patients is calculated as the number of disease cases among the exposed participants divided by the total person-time at risk among the exposed patients.

The incidence rate for the unexposed patients is calculated as the number of disease cases among the unexposed participants divided by the total person-time at risk among the unexposed patients.

Cumulative Incidence

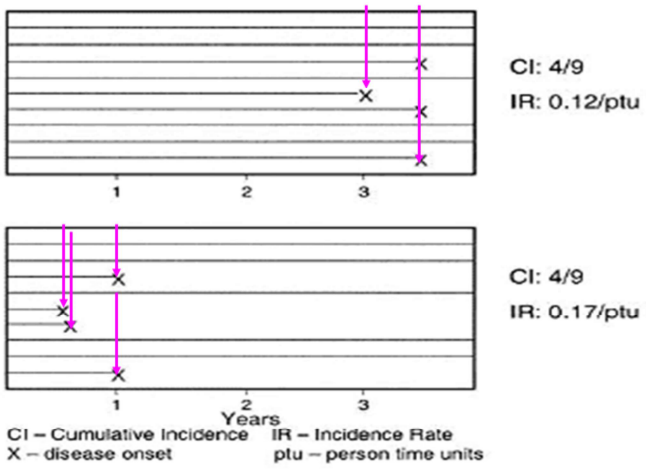


Lautenbach L. et al., Practical Handbook for Healthcare Epidemiologists

Let's consider two different scenarios to better understand the difference between the cumulative incidence and the incidence rate. In each figure, the X symbol indicates the timing of disease onset. The line is drawn to represent the duration of time at risk for each patient prior to the onset of disease.

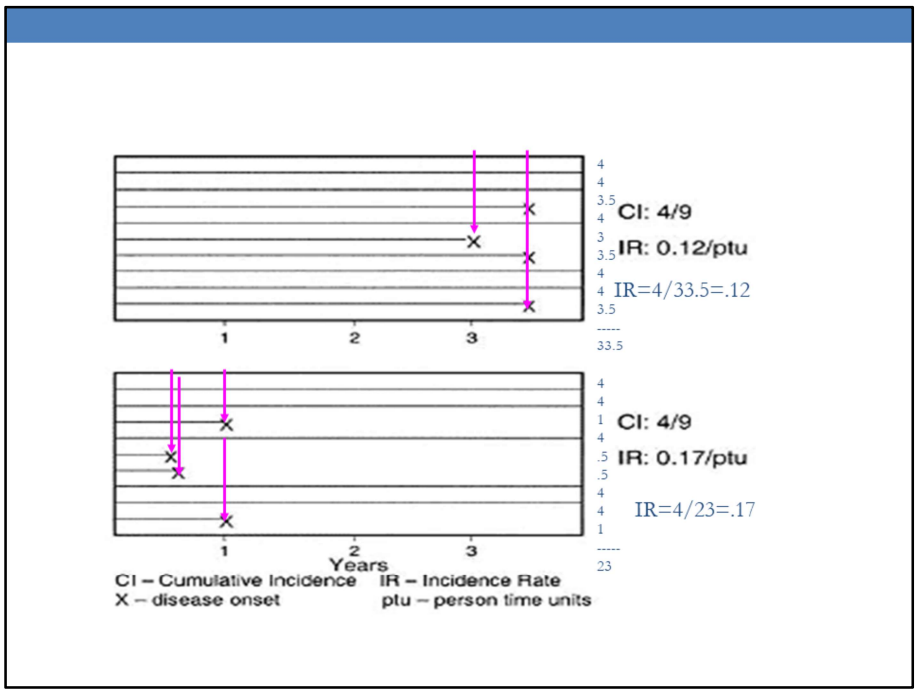
In each example, 4 out of nine patients developed the disease of interest. Therefore, over the 4-year follow-up period, the cumulative incidence is the same, 4 of 9. However, we note that in the first example, the patients developed the disease more slowly than in the second example. When we account for the timing of disease onset, we see that the incidence rate, calculated as the number of diseased patients divided by the total person-years at risk, is higher for the second example than the first. This higher incidence rate indicates that the disease was occurring more quickly, or in other words, that diseased cases were being observed after shorter periods of patient-time at risk.

Cumulative Incidence vs. Incidence Rate



Lautenbach L. et al., Practical Handbook for Healthcare Epidemiologists

Let's compare the timing of disease onset more carefully between the two data scenarios. We first note the timing of the events for the patients in the first and second series of data.



Then, we can calculate the total time at risk among the nine patients in each scenario. We see that the total time at risk in the first scenario is 33.5 person-years compared to a total time at risk of 23 person-years in the second example. Given the smaller denominator of person-time at risk in the second example, the incidence rate will be higher.

Measure of Association in Cohort Studies with Person-Time

Incidence Rate Ratio (IRR)

- $\text{IRR} = \dfrac{\text{Incidence rate in exposed}}{\text{Incidence rate in unexposed}}$

- $\text{IRR} = \dfrac{\text{Number of diseased in exposed} / \text{Person time in exposure}}{\text{Number of diseased in unexposed} / \text{Person time in nonexposure}}$

In settings where it is important to account for the duration of patient follow-up, we can calculate an incidence rate ratio as the ratio of the incidence rate among the exposed patients divided by the incidence rate among the unexposed patients.

Incidence Rate Ratio Example

Now, let's consider a data example for the calculation of an incidence rate ratio.

Incidence Rate in Exposed

	Diseased	Person time
Exposed	6	1410
Nonexposed	2	1505
Total	8	2915

$$\text{IR in Exp} = \frac{6}{1410} = 0.00425$$

i.e., 42.55 per 10,000 person day

To calculate the incidence rate among the exposed participants, we divide the number of patients developing disease, in this case, 6, by the total person-days at risk contributed by the exposed participants. This results in an incidence rate among the exposed participants of 0.00425 or 42.55 per 10,000 person days at risk.

Incidence Rate in Nonexposed

	Diseased	Person time
Exposed	6	1410
Nonexposed	2	1505
Total	8	2915

$$\text{IR in Nonexp} = \frac{2}{1505} = 0.001329$$

i.e., 13.29 per 10,000 person day

To calculate the incidence rate among the unexposed participants, we divide the number of patients developing disease, in this case, 2, by the total person-days at risk contributed by the unexposed participants. This results in an incidence rate among the unexposed participants of 0.001329 or 13.29 per 10,000 person days at risk.

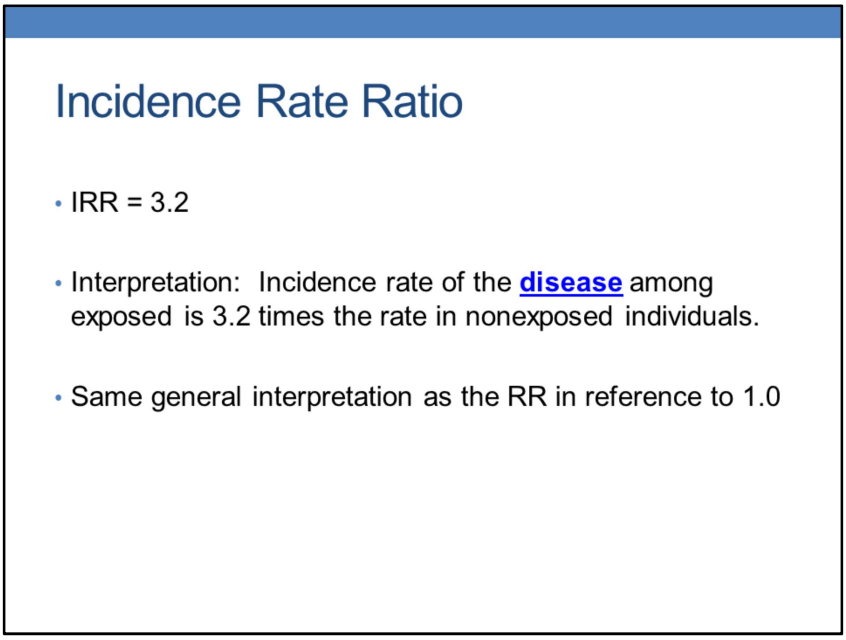
Incidence Rate Ratio

	Diseased	Person time
Exposed	6	1410
Nonexposed	2	1505
Total	8	2915

$$IRR = \frac{\text{IR in exposed}}{\text{IR in nonexposed}}$$

$$= \frac{42.55}{13.29} = 3.2$$

The resulting incidence rate ratio is 42.55 divided by 13.29 that results in a value of 3.2.

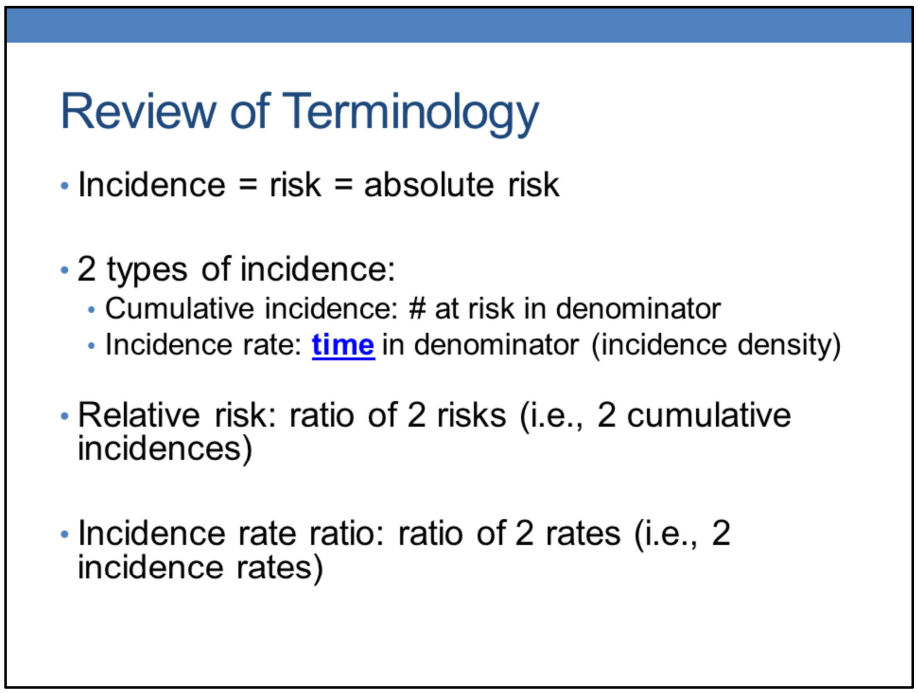


The incidence rate ratio value of 3.2 can be interpreted as the incidence rate of the **disease** among exposed is 3.2 times the rate in nonexposed individuals.

We compare the resulting incidence rate ratio to a value of 1, which indicates that the incidence rate is equal between the exposed and unexposed groups.

If the value of the incidence rate ratio is greater than 1, this indicates that the incidence rate of disease is higher among the exposed compared to the unexposed.

If the value of the incidence rate ratio is less than 1, this indicates that the incidence rate of disease is lower among the exposed compared to the unexposed



Recall that we have studied two different types of incidence calculations, one that is based on the cumulative incidence where the denominator is the number at risk, where we assume that all participants at risk have an equal duration of follow up, and the second that is based on the incidence rate or incidence density, where we use a measure of person-time at risk in the denominator. The incidence rate measure is appropriate when participants have a varying duration of follow-up.

The relative risk is the ratio of two cumulative incidence measures and the incidence rate ratio is the ratio of two incidence rate values. For both ratios, we compare the resulting value to 1 (equal incidence between exposed and unexposed) to aid in interpretation of the association between exposure and outcome.

Summary

- Cohort study design
  - Sample exposed and unexposed
- Measures
  - Relative risk
  - Incidence rate ratio

In this first module, we focused on the interpretation of measures of association that can be calculated from cohort studies, namely the relative risk, based on cumulative incidence, and the incidence rate ratio, that accounts for time at risk.

In the next module, we will discuss measures of association that are commonly used for case-control studies and cross-sectional studies.