BIOSTATISTICS

FINAL

Lecture 4 (Analytic studies 1)

Given by: Dr. Israa Al-Rawashdeh

Done by Doctor:

Abdullah Daradkh

Mahmoud Al-Otoom

MUTAH UNIVERSITY

Analytic studies

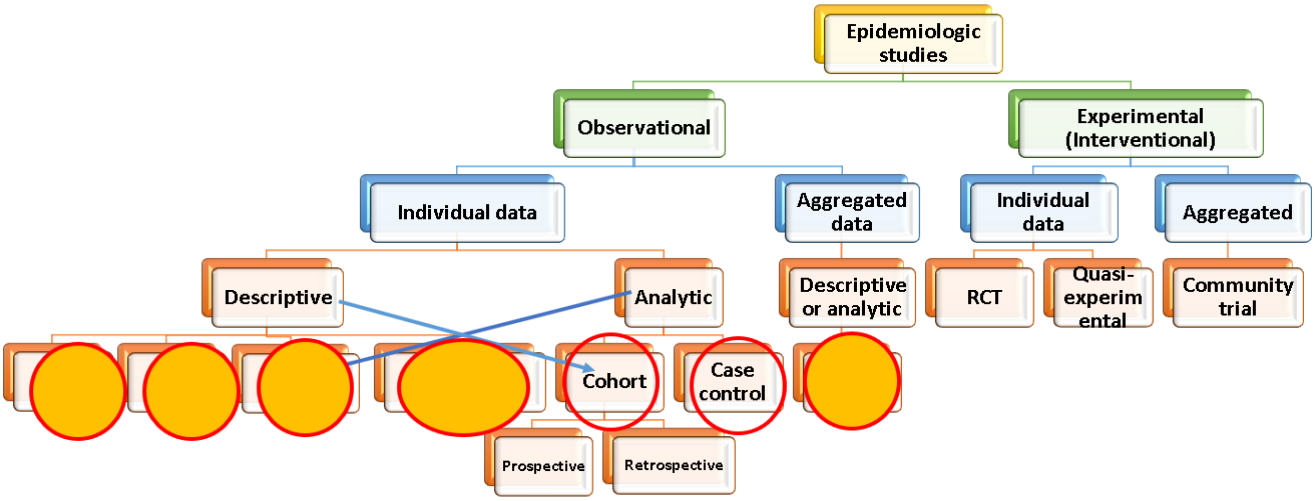
Recap

Last week we have covered:

Descriptive studies; types and characteristics

Cross-sectional studies don't have causal relationship (we can't decide which caused which)

But in analytic studies we can find causal relationship.



Case-Control Studies:

A case-control study is a type of observational (analytic) study in which two existing groups differing in outcome are identified and compared on the basis of some supposed causal attribute.

- **A case-control study is designed to help determine if an exposure is associated with an outcome (i.e., disease or condition of interest).**

In a case-control study patients who have developed a disease are identified and their *past* exposure to suspected aetiological factors is compared with that of controls or referents who do not have the disease.

هذا النوع سيحل المشكلة التي لم تستطع ال cross-sectional studies و longitudinal حلها و هي دراسة العلاقة بين two groups different in outcome
one group will be the cases and the other one will be the control.

- **First, identify the cases (a group known to have the outcome) and the controls (a group known to be free of the outcome). Second, *look back in time* to learn which subjects in each group had the exposure(s), comparing the frequency of the exposure in the case group to the control group.**

- **By definition, a case-control study is always retrospective (حدثت في الماضي) because it starts with an outcome then traces back to investigate exposures. When the subjects are enrolled in their respective groups, the outcome of each subject is already known by the investigator.**
- **Case-control studies are *the most frequently* undertaken analytical epidemiological studies. The most destructive study cross- section**

When to Conduct a Case-Control Study?

- **Appropriate for investigating outbreaks (e.g. a study of Hepatitis A after eating from a Cafeteria)**
- **The outcome of interest is rare (e.g. a study of risk factors for uveal melanoma, or corneal ulcers).**
- **Multiple exposures may be associated with a single outcome → we don't know which exposure developed the disease.**
- **Funding or time is limited**
- **Outcomes with long latent periods (AIDS)**
- **Ideal for preliminary investigation of a suspected risk factor for a common condition; conclusions may be used to justify a more costly and time-consuming longitudinal study later**

FEATURES OF CASE-CONTROL STUDIES

1-DIRECTIONALITY

From Outcome to exposure (backward)

2-TIMING

***Retrospective** for exposure*, **but case-ascertainment can be either retrospective or concurrent.**

ال exposure تكون دائما retrospective

أما ال outcome قد يكون حصل في الماضي أي أن المريض قد أصيب قبل مدة أو قد تكون حدثت حاليا في وقت الدراسة

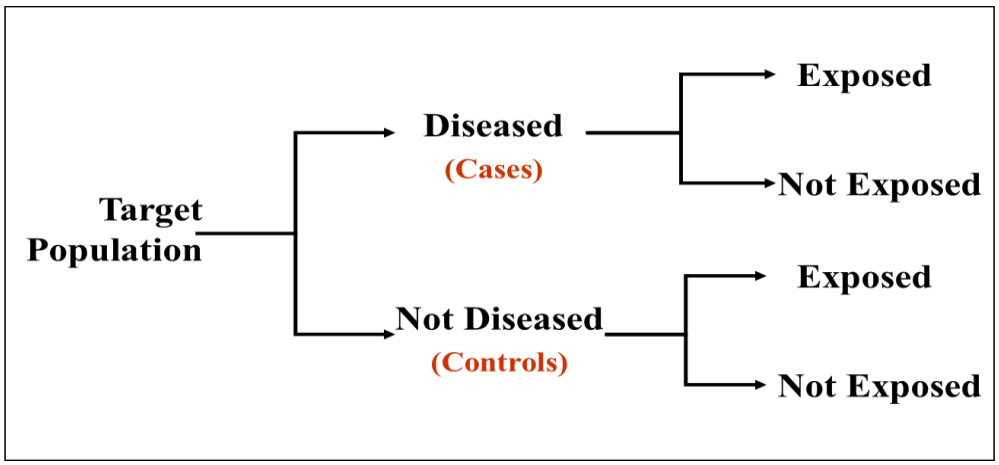
4-SAMPLING

Almost always on outcome, with matching of controls to cases

(first, we identify cases (outcome) then we match them with control)

Case Control Study Design

نحدد ال population الذي سنأخذ منه العينة و من ثَم identify cases(disease) و نقسمهم إلى exposed & not exposed بناءً على ال exposure الذي يقوم الباحث بتحديده و من ثَم نقوم ب matching مع مجموعة أخرى من نفس المجتمع الذي نسميه control (الذين لم يصابوا بالمرض) و نرى إن كانوا exposed o not exposed



Selecting Cases

1-The starting point of most case-control studies is the identification of cases.

2-Select cases after the Diagnostic Criteria And Definition of the disease is clearly established

Case definition

– needs a precise definition

– inclusion and exclusion criteria

– e.g. could be based on

.....clinical features (history, examination) .....clinical measurement

.....laboratory data .....post-mortem findings

3-Source of cases: Cases may be recruited from a hospital, clinic, GP registers or may be population bases. Population based case control studies are generally more expensive and difficult to conduct.

يجب أن تكون دراسة الحالات ممثلة ال cases مثلا اذا كان في المستشفى مئة حالة انفلونزا طيور لن نأخذ المئة ماملين بل سنأخذ عشرين حالة و يجب أن تكون هذه العينة representative لكل الحالات المئة

Selection may be from *incident* or *prevalent* cases:

1-Incident cases are those derived from ongoing ascertainment of cases over time. (newly diagnosed during a defined time period)

2-Prevalent cases are derived from a cross-sectional survey (existing cases).

******Incident cases are preferable to prevalent cases for reducing**

1-recall bias

The patient can't remember what she/he were exposed because of long time

مثل مادة ال asbestos التي كانت تؤدي لحدوث سرطان الرئة فكيف لنا أن نعلم أنها هي المسبب للمرض إن تم التعرض لها قبل وقت طويل و لم لا يكون المسبب مادةً أو شيئاً آخر ؟!

2-(b) Selection bias or (over-representation of certain cases (less severe , more severe))

3- Confusion of the direction of causality (disease influences exposure so unsure of the direction of causality)

→The most desirable way to obtain cases is to include all incident cases in a defined population over a specified period of time

Selecting Controls

Is more difficult than choosing cases (cases based on outcome but control has many conditions):

1-Controls should come from the same population at risk for the disease as the cases (comparable)

2-Controls should be representative of the target population

THREE QUALITIES NEEDED IN CONTROLS (to be comparable)

1-Key concept: Comparability is more important than representativeness in the selection of controls

2-The control should resemble the case in all respects except for the presence of disease

3-The control must be at risk of getting the disease.

A pool of potential controls must be defined.

This pool must mirror the study base of the cases.

i.e. Cases emerge within a study base. Controls should emerge from the same study base, except that they are not cases (matching).

For example, if cases are selected exclusively from hospitalized patients, controls must also be selected from hospitalized patients.

If cases must have gone through a certain ascertainment process (e.g. (e.g. mammogram-detected breast screening), controls must have also. cancer) etc..

يعني ال control تشبه ال cases في كل الخصائص باستثناء أنها غير مصابة بالمرض.

- **يجب أخذ المجموعتين من نفس المكان فإذا أخذت ال cases من المستشفى يجب أخذ ال control من نفس المكان و هكذا.**

****Multiple controls can be used to help add statistical power when cases are excessively difficult to obtain**

****Using more than one control group provides credibility (reliability) to the results**

****More than 3 controls for a case is usually not cost-efficient as well as not effective for statistical considerations**

Measuring exposure status (Ascertainment of exposure)

- Exposure status is measured to assess the presence or level of exposure for each individual for the period of time prior to the onset of the disease or condition under investigation. Various methods can be used to ascertain exposure status. These include:

1-Personal recall, using either a self-administered questionnaire or an interview. recall bias and interviewer bias

--------Some protection may be afforded by blinding interviewers and carefully phrasing interview questions

2-Historical records (more accurate) (e.g. Medical records, Employment records, Pharmacy records) → as medical records its problem is → incomplete information

3-Biological markers of exposure

A gropup of people exposed to BCG vaccine (used against TB) . when we take a blood sample after many years we will find BCG vaccine → so we identify the exposure.
**** problems**
A- not all of exposures has a biological marker

B- its concentration may decrease with time or by a disease .

- ✓ **The procedures used for the collection of exposure data should be the same for cases and controls.**

****Potential confounders** need to be accurately assessed in order to be controlled in the analysis

****Confounding** arises when an exposure and an outcome are both strongly associated with a third variable.

مثال في إحدى الدراسات قاموا بإيجاد علاقة بين شرب الكحول و سرطان الرئة . بعدها بمدة وجدوا أن معظم المصابين بسرطان الرئة هم مدخنين . و بعدها وجدوا أن العلاقة بين سرطان الرئة و التدخين أقوى من علاقته مع شرب الكحول

Smoking was a confounding factor.

Case-Control Study: Analysis Format

Exposure	Cases	Controls
Yes	a	b
No	c	d

Exposure odds ratio (OR) ≈ RR when disease is rare

Odds of being exposed among the cases = a/c
Odds of being exposed among the controls = b/d

Exposure odds ratio = (a/c)/(b/d) = (a*d)/(b*c)

Odds means probability or chance.

A ratio that measures the odds of exposure for cases compared to controls

Odds of exposure = number exposed ÷ number unexposed

OR Numerator: Odds of exposure for cases

OR Denominator: Odds of exposure for controls

Calculating the Odds Ratio

Exposure Status	Disease Status: CHD cases (Cases)	Disease Status: No CHD (Controls)
Smoker	112	176
Non-smoker	88	224
Total	200	400

$$\text{Odds Ratio} = \frac{AD}{BC} = \frac{112 \times 224}{176 \times 88} = 1.62$$

	OR<1	OR=1	OR>1
Odds comparison between cases and controls	Odds of exposure for cases are less than the odds of exposure for controls	Odds of exposure are equal among cases and controls	Odds of exposure for cases are greater than the odds of exposure for controls
Exposure as a risk factor for the disease?	Exposure reduces disease risk (Protective factor)	Particular exposure is not a risk factor	Exposure increases disease risk (Risk factor)

(1 = no association, > 1 = possible association, < 1 = protective effect)

Interpreting the Odds Ratio

The odds of exposure for cases are 1.62 times the odds of exposure for controls.
OR

Those with CHD are 1.62 times more likely to be smokers than those without CHD

OR

Those with CHD are 62% more likely to be smokers than those without CHD

Limitations Of C-C studies:

- **Particularly prone to bias; especially selection, recall and observer (interview) bias**
- **Problems with assessing direction (potential for reverse causality)**
- **Not suitable for rare exposures**
- **Not suitable for studying multiple outcomes for a single exposure**
- **Cannot estimate incidence or prevalence ……. (prevalence = cross sectional study) (incidence = cohort study)**
- **Further limitations if using prevalent cases**

SOME IMPORTANT DISCOVERIES MADE IN CASE CONTROL STUDIES

1950's
- Cigarette smoking and lung cancer

1970's
- Diethyl stilbestrol and vaginal adenocarcinoma
- Post-menopausal estrogens and endometrial cancer

1980's
- Aspirin and Reyes syndrome
- Tampon use and toxic shock syndrome
- AIDS and sexual practices

1990's
- Vaccine effectiveness
- Diet and cancer

Cohort studies(study the exposure)

Cohort: a group of individuals who share a common characteristic

– e.g. age, sex, workers in a factory, etc.. What are other examples of cohorts?

Cohort defined by its exposure to a potential risk factor

– e.g. factory workers exposed or not exposed to a chemical

Cohort members should be free of the outcome under investigation at the start of the study

The outcome of interest could be:

– development of a disease (so the cohort are disease free at the start)

– death (or survival) in a cohort of people with a disease

– other outcomes e.g. admission to hospital

Cohort study is undertaken to support the existence of association between suspected cause and disease

A major limitation of cross-sectional surveys and case-control studies is difficulty in determining if exposure or risk factor preceded the disease or outcome.

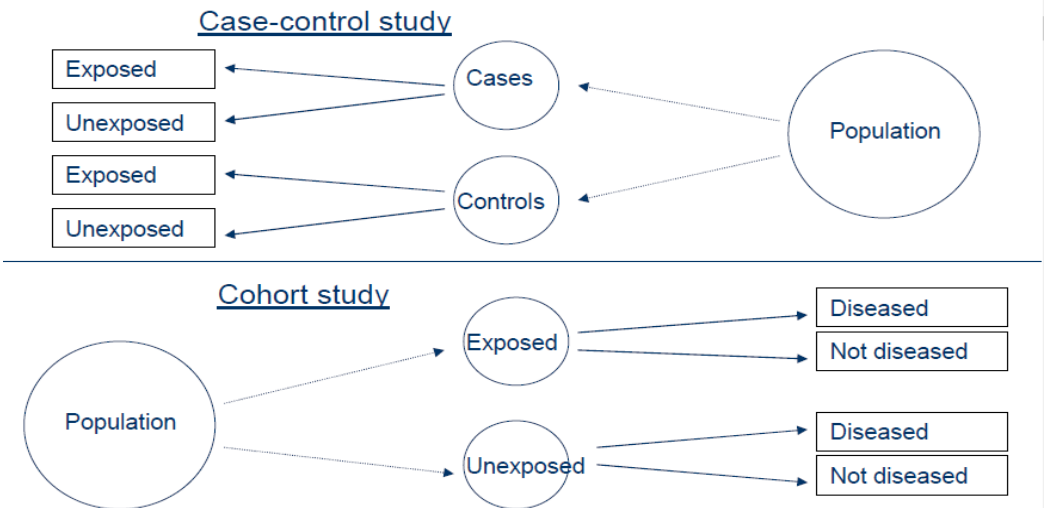
(in cross-sectional surveys and case-control studies we cant know which cased which)

Cohort Study:

Key Point:

Presence or absence of risk factor is determined before outcome occurs.

Case-control vs cohort



The direction in case control studies is from the outcome to the exposure

But in cohort studies is from the exposure to the out come

– incidence studies → the development of disease will be in the future

– longitudinal studies → we follow up the participants

– follow-up studies

– (prospective studies)

May be:

– descriptive

– analytical

Two main types:

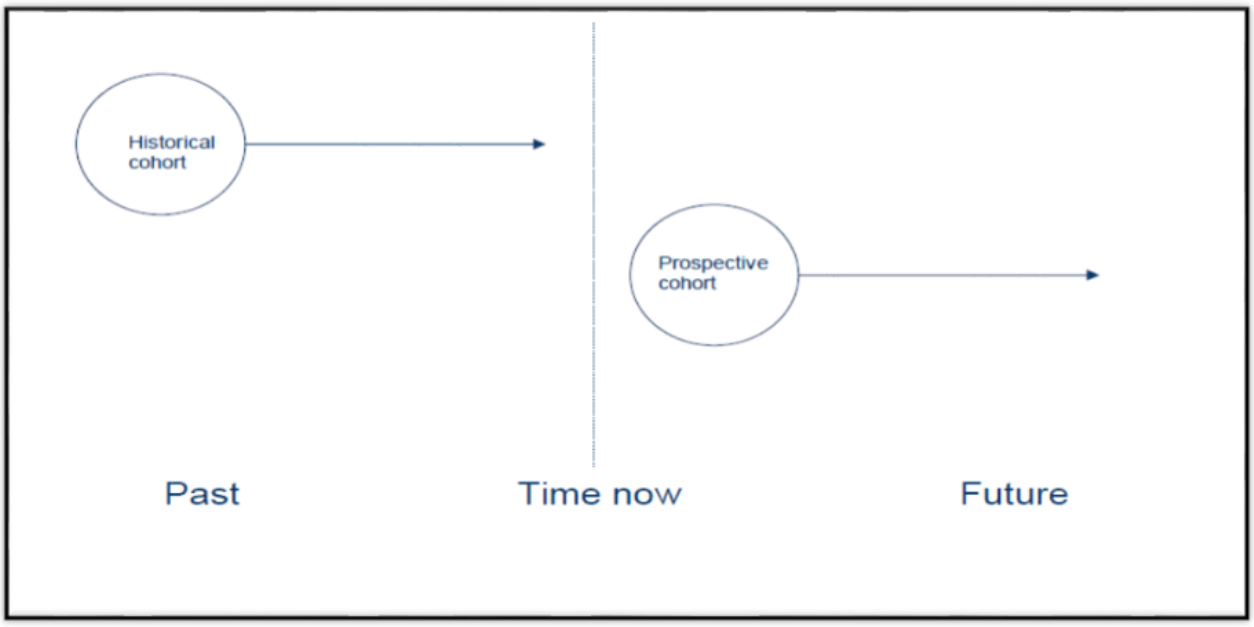
– *Prospective cohort studies*

Start now and follow-up into the future

– *Retrospective (or historical) cohort studies*

Use existing data on exposures and outcomes

Prospective vs retrospective cohort studies



Retrospective cohort VS case-control

Elements of cohort study

A-Selection of study subjects

B-Obtaining data on exposure

C-Selection of comparison group

D-Follow up

E-Analysis

A-Selection of study population

1-Both the cohorts are free of the disease.

2-Both the groups should equally susceptible to disease (both groups are affected by the same risk)

3-Both the groups should be comparable

4-Diagnostic and eligibility criteria for the disease should be defined well in advance.

B-Obtaining data on exposure

Common exposure

Select study population before classifying individuals by exposure status

Sample of general population

Occupational group (workforce or occupational group) e.g. study of Jordanian doctors investigating the adverse effects of smoking

Rare exposure

Select on basis of exposure, Person having exposure to some physical, chemical or biological agent

e.g. factory workers exposed to a chemical

C-Selection of comparison group, Unexposed group?

- **Internal comparison**
  - **Only one cohort involved in study**
  - **Sub classified and internal comparison done**
  - **E.g. Workers in the same factory who are not exposed (or less exposed)**
- **External comparison**
  - **More than one cohort in the study for the purpose of comparison**
  - **e.g. Workers in other factories**
  - **e.g. Cohort of radiologist compared with ophthalmologists**
- **Comparison with general population rates**
  - **If no comparison group is available we can compare the rates of study cohort with general population. → WITH SIMILAR CHARACTERISTICS**
  - **e.g. Cancer rate of uranium miners with cancer in general population (be aware of Healthy Worker Effect)**

Exposure assessment

Data may be obtained by:

– Questionnaire / interview , Medical examination , Blood samples and other tests, Medical records

Exposure may vary over time

e.g. smoking status→ close follow up لذلك نقوم بالتأكد من صحة معلومات المشارك كل مدة

– may need to reassess at regular intervals

– use more complex statistical methods for analysis

Prospective cohort study

– exposure data collected before outcomes

Retrospective cohort study

– relies on pre-existing exposure data

– potential issues with accuracy and consistency

By obtaining the data of exposure we can classify cohorts as

1. Exposed and non exposed and
2. By degree exposure we can sub classify cohorts

D-Follow-up and outcomes

Follow up is the most critical part of the study

May need long follow-up period

– Cost implications

– Occupational cohorts may be easier to follow-up

Retrospective cohort studies

– follow-up period already occurred and avoids some of the costs

Loss to follow-up a potentially serious problem

– especially if related to developing outcome

– and there are differences between exposure groups

Outcome data may be collected by

– interview / questionnaire / Periodic medical examination

– contacting cohort members or family or doctor

– relying on routine sources (e.g. death certification)

Outcomes should be collected without knowledge of exposure status

– otherwise potential for observer bias

Observer bias → the person who makes the study shouldn't know the people who were exposured & who weren't exposured

→Some loss to follow up is inevitable due to death, change of address, migration, change of occupation.

→ Loss to follow-up is one of the draw-back of the cohort study. (Attrition)

E-Analysis

**Analysis of a cohort study uses either the risk or the rate ratio of disease in the exposed cohort compared with the rate or risk in the unexposed cohort.

***Risks and rates* can be further manipulated to provide additional information on the effects of the exposure of interest, such as risk ratios, rate ratios, attributable risks (risk or rate differences) and attributable risk percent.

Risk and Rates

Risk is defined as the number of new cases divided by the total population-at-risk at the beginning of the follow-up period.

$$Risk = \frac{\#new\ cases}{total\ \#\ of\ individuals\ at\ risk}$$

A rate is the number of new cases of a health outcome divided by the total person-time-at-risk for the population.

Person-time is calculated by the sum total of time all individuals remain in the study without developing the outcome of interest (the total amount of time that the study members are at risk of developing the outcome of interest).

Person-time can be measured in days, months, or years, depending on the unit of time that is relevant to the study. A rate measures the rapidity of health outcome occurrence in the population.

$$Rate = \frac{\#new\ cases}{total\ person-time\ at\ risk}$$

Two-by-two tables are generally used to organize the data from a study :

	Disease	No disease	total
Exposed	a	b	a+b
Non-exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

Risk ratio

- The risk ratio is defined as the risk in the exposed cohort divided by the risk in the unexposed cohort (the comparison group).
- A risk ratio may vary from zero to infinity

$$\text{Risk ratio} = \frac{\text{Risk (cumulative incidence) in the exposed group}}{\text{Risk (cumulative incidence) in the unexposed group}}$$

Incidence rate

Incidence among exposed = R1= a/a+b

Incidence among non-exposed = R2= c/c+d

Relative Risk

$$\text{Risk ratio} = \frac{\text{Risk (cumulative incidence) in the exposed group}}{\text{Risk (cumulative incidence) in the unexposed group}} = \frac{a/a+b}{c/c+d} = \frac{R1}{R2}$$

Example

Suppose researchers conduct a cohort study and gather the following data on the effects of air pollution exposure on respiratory illness among factory workers

	Disease	No disease	total
Exposed	60	140	200
Non-exposed	25	175	200
Total	85	315	400

In this study, the risk in the exposed group is 60/200, or 0.30 cases per person (30 cases per 100 people), and the risk in the unexposed group is 25/200, or 0.125 cases per person (13 cases per 100 people).

Therefore, the RR is 0.30/0.125 = 2.4

A risk ratio of 2.4 means: that the exposed group has *2.4 times* the risk of developing respiratory illness as the unexposed group.

Rate ratio

The rate ratio is defined as rate of health outcome occurrence in the exposed group divided by the rate of health outcome occurrence in the unexposed group or less exposed (comparison group)

$$\bullet \text{ Rate ratio} = \frac{\text{Incidence rate in the exposed group}}{\text{Incidence rate in the unexposed group}}$$

Example

	Disease	No disease	Person-year at risk
Exposed	60	140	175
Non-exposed	25	175	188
Total	85	315	363

The rate in the exposed cohort is 60/175 person- years= 0.34 cases/person-year.

The rate in the unexposed cohort is 25/188 person-years= 0.13 cases/ person-year.

The rate ratio in this study is 0.34/0.13= 2.6.

This rate ratio reveals that respiratory illness among workers exposed to air pollution is developing at 2.6 times the rate that respiratory illness is developing among workers not exposed to air pollution.

The following table may be applied to both risk and rate ratios.: •

Risk ratio or	Exposure
<1	Exposure is protective
=1	Exposure is neither preventive nor
>1	Exposure is harmful

Attributable Risk (Rate difference)

In order to find the absolute effect of an exposure a health outcome the attributable rate (AR), or rate difference, must be computed.

The term attributable risk (AR) is same as rate difference (RD).

The attributable rate is the excess rate among the exposed population attributed to exposure.

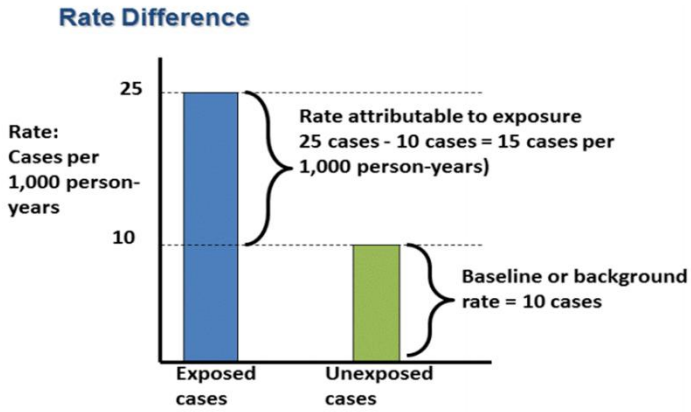
It is defined as the rate in the exposed minus the rate in the unexposed.

The rate difference can also be reported as a percent

Estimation of Risk

$$AR = \frac{\text{Incidence ratr of disease among exposed} - \text{incidence rate of disease among non exposed}}{}$$

$$OR \quad AR\% = \frac{RR-1}{RR} * 100$$



Smoking	Lung cancer		Total
	YES	NO	
YES	70	6930	7000
NO	3	2997	3000
	73	9927	10000

- **Incidence of lung cancer among smokers**
  - **70/7000 = 10 per 1000**
- **Incidence of lung cancer among non-smokers**
  - **3/3000 = 1 per thousand**
- **RR = 10 / 1 = 10**
- **(lung cancer is 10 times more common among smokers than non smokers)**
- **AR = 10 – 1 / 10 X 100**
  - **= 90 %**

Find out RR and AR for above data

Analysis (risk / odds ratio example)